Shape Analysis of 3D Human Faces and Applications

Boulbaba Ben Amor

Université Lille1 - Sciences et Technologies

Mémoire présenté pour l'obtention de

L'Habilitation à Diriger des Recherches

Shape Analysis of 3D Human Faces and Applications

Par

Boulbaba Ben Amor

December 2, 2014

Committee

M.	Rama Chellappa	Minta Martin Professor of Engineering, CAR, University of Maryland, College Park, USA	(Rapporteur)
Mme.	Maja Pantic	Professor, i·BUG, Imperial College London, UK	(Rapporteur)
M.	Atilla Baskurt	Professor, LIRIS (UMR CNRS 5205), INSA Lyon, France	(Rapporteur)
Mme.	Bernadette Dorizzi	Professor, Intermedia, Institut Mines-Télécom, France	(Examinateur)
M.	Alberto Del Bimbo	Professor, MICC, University of Florence, Italy	(Examinateur)
Mme.	Sophie Tison	Professor, LIFL (CNRS 8022), Lille1 University, France	(Examinateur)
M.	Mohamed Daoudi	Professor, LIFL (CNRS 8022), Institut Mines-Télécom, France	(Garant d'HDR)

Acknowledgements

Thanks for reading. . .

Contents

Part I

Synthesis of Activities

Curriculum Vitae

Boulbaba BEN AMOR, 36 years, married, 1 child.
Associate Professor (Maître de conférences) with Institut Mines-Télécom/Télécom Lille,
Member of Computer Science Laboratory of Lille (LIFL UMR CNRS/Lille1 8022),
Cité scientifique - Rue Guglielmo Marconi, 59653 Villeneuve d'Ascq Cedex BP 20145 - FRANCE.
Tél : + (33) 3 20 43 64 18
Fax : + (33) 3 20 33 55 99
E-mail : benamor@telecom-lille.fr
Web : http://www.telecom-lille.fr/people/benamor/

Education

2003-2006	**Ph.D. in Computer Science,** with **highest honor**,
	École Centrale de Lyon – France.
	Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS UMR 5205 CNRS).
2002-2003	**Master Degree in Computer Science**, Images & Systems,
	École Centrale de Lyon – France.
	Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS UMR 5205 CNRS).
1999-2002	**Engineer Degree in Computer Science,**.
	National Engineering School of Sfax (ENIS) – Tunisia.
1997-1999	**Two-year classes to prepare for admission to Engineer Schools,**
	Preparatory Institute for Engineering Studies of Sfax (IPEIS) – Tunisia.
1997	**Baccalauréat** in Mathematics – Tunisia.

Positions

2007-present	**Associate Professor (Maître de conférences), Institut Mines-Télécom**,
	Department of Computer Science & Networks, Télécom Lille.
	Laboratoire d'Informatique Fondamentale de Lille (LIFL UMR CNRS/Lille1 8022).
2014	**Visiting Professor (January-August), Florida State University – USA**,
	Supported conjointly by Institut Mines-Télécom and Florida State University.
	Statistical Shape Analysis and Modeling Group (SSAMG), Department of Statistics, FSU.
2006-2007	**Assistant Professor (A.T.E.R), Ecole Centrale de Lyon – France**,
	Department of Mathematics & Computer Science, École Centrale de Lyon.
2005-2006	**Temporary Lecturer, EPSI de Lyon – France**,
	École Privé des Sciences Informatique, Lyon.

Research Grants

2010 - 2013 — ANR[1] **blanc international 3D Face Analyzer**, Partners: **École Centrale de Lyon/LIRIS, Lille1/LIFL, Baihang University/IRIP, North China University of Technology (NCUT)** . *Leader of work-package, co-supervision of two post-docs.*

2008 - 2011 — ANR Face Analysis and Recognition using 3D (FAR-3D), Partners: **Lille1/LIFL, École Centrale de Lyon/LIRIS, Eurecom, THALES R&T** . *Leader of work-package, co-supervision of one Ph.D. and one Master projects, development of the 3D Face Authentication demonstration.*

2008 - 2013 — CPER-CIA (Contrat Plan État Région – Campus Intelligence Ambiante) contract funded by the Région Nord-Pas de Calais. *Leader of the 3D acquisition platform.*

2008 - 2011 — CMCU French-Tunisian project, Partners: **Lille1/LIFL, École Centrale de Lyon/LIRIS, ENSI/CRISTAL, ENIS/REGIM** . *Supervision of one Ph.D. student.*

2005 - 2007 — Techno-vision IV[2] **(Identification par l'Iris et le Visage via la Vidéo)** co-funded by Ministère de la Recherche and Ministère de la Défense, Partners: **GET - INT, GET - ENST, IMEDIA - INRIA, Institut EURECOM, Let It Wave, LIRIS - École Centrale de Lyon, IBISC - Université d'Evry, THALES (TRT), URATEK** . *Design of acquisition settings and evaluation protocols of 3D face recognition algorithms.*

Students Supervision

Ph.D. thesis (5)

2008 - 2011 — Hassen Drira, *Statistical Computing on Manifolds for 3D Face Analysis and Recognition*, defense on July, 2011, **University of Lille1**, *ANR funding, co-supervision with Prof. Mohamed Daoudi – **rate 50%**.*

2009 - 2012 — Ahmed Maalej, *3D Facial Expressions Recognition using Shape Analysis and Machine Learning*, defense on May 2012, **University of Lille1**, *Funded by Institut Mines-Télécom/Région Nord-Pas de Calais, co-supervision with Prof. Mohamed Daoudi – **rate 50%**.*

2009 - 2012 — Lahoucine Ballihi, *3D Facial Biometrics by learning Geometric Features*, defense on May 2012, **University of Lille1 & University Mohammed V-Agdal (Rabat, Morocco)**, *Funded by STIC Morocco program, co-supervision with Prof. Mohamed Daoudi & Pr. Driss Aboutajdine (University Mohammed V-Agdal)– **rate 40%**.*

2011 - present — Baiqiang Xia, *Learning 3D Geometric Features for Soft Biometrics Recognition*, defense planned for November 2014, **University of Lille1**, *Funded by the China Scholarship Council (CSC), co-supervision with Prof. Mohamed Daoudi – **rate 70%**.*

2012 - present — Taleb Alashkar, *3D Face Dynamic Sequences Analysis: Application to Face and Facial Expression Recognition*, defense planned for October 2015, **University of Lille1**, *Funded by the Future & Rupture program supported by Institut Mines-Télécom, co-supervision with Prof. Mohamed Daoudi & Prof. Stefano Berretti (University of Firenze, Italy) – **rate 70%**.*

---

[1] www.agence-nationale-recherche.fr/en/project-based-funding-to-advance-french-research/

Post-docs (2)

2012 - 2013 — **Lahoucine Ballihi**, **Post-doc,** *University of Lille1*, *Funded by the ANR, co-supervision with Prof. Mohamed Daoudi* – **rate 50%**.

2011 - 2012 — **Hassen Drira**, **Post-doc,** *University of Lille1*, *Funded by the ANR, co-supervision with Prof. Mohamed Daoudi* – **rate 50%**.

Masters (2)

2010 - 2011 — **Rim Slama**, **Master,** *3D Face Recognition under Occlusions*, defense on July 2011, **Ecole Nationale des Sciences de l'Informatique, Tunisie**, *co-supervision with Prof. Mohamed Daoudi* – **rate 50%**.

2012 — **Marion Detourmignies**, **final year study project at Telecom Lille,** *Expression Analysis from Depth-consumer 3D Sensors*, defense on September 2012, **Télécom Lille**, *co-supervision with Prof. Mohamed Daoudi & Prof. Ludovic Macaire (LAGIS Lab.)*– **rate 50%**.

Collaborations

Statistical Shape Analysis and Modeling Group (SSAMG), Florida State University, USA, I visited the FSU for 8 months (December 2013 - August 2014) as a **Visitor Scholar** – My stay was supported conjointly by Institut Mines-Télécom and Florida State University. Collaboration with *Prof. Anuj Srivastava* giving rise to 6 common journal publications and several international conferences.

Media Integration and Communication Center (MICC), University of Firenze, Italy, collaboration with *Prof. Stefano Berretti*, resulted in 3 journal publications. Co-supervision of the thesis of Taleb Alashkar supported by the Future & Rupture program of Institut Mines-Télécom (2013-2015). Taleb Alashkar received financial help of DAI N11 from University of Lille1 for a 2-months staying at Media Integration and Communication Center (MICC), University of Firenze.

Laboratory of Intelligent Recognition and Image Processing (IRIP), Beihang University, China, collaboration with *Prof. Yunhong Wang* and *Dr. Di Huang*, since 2010. Collaboration part of the ANR-NSFC 3D Face Analyzer project. This collaboration have resulted in 2 conference papers. Baiquiang Xia, Ph.D. student at University of Lille1, spent 6 weeks with the IRIP Lab., in 2013.

Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS), Ecole Centrale de Lyon, France, collaboration with *Prof. Liming Chen* and *Dr. Mohsen Ardabilian*. We have participated together to the ANR FAR-3D and ANR blanc international 3D Face Analyzer projects. We co-wrote the first chapter of the book *3D Face Modeling and Recognition* [O1], published in 2013. I was invited to the Ph.D. committee of Wael Ben Soltana, the defense organized in December 2012.

Invited talks and Seminars

CAR, Center for Automation Research, University of Maryland College Park, USA, July 2014.

SSAMG Group, Florida State University, USA, January 2014.

IRIP Lab., Baihang University, China, April 2013 & July 2011.

VIVA Lab., University of Ottawa, Canada, July 2012.

Bell Labs, Belgium, March 2013.

Workshop « 3D Human Modeling », IMAGINA event, Monaco, February 2011[2]

Workshop « Essor du Nomadisme : une sécurité suffisante ? », International forum on Cybercriminality, Lille, January 2010.[3]

Services to the community

Journal Review Committees

Reviewer for several journals as – *IEEE Transactions on Image Processing – IEEE Transactions on Multimedia – IEEE Transactions on Information Forensics & Security – Image and Vision Computing – Pattern Recognition Letters – Computers & Graphics – EURASIP Journal on Image and Video Processing – International Journal of Pattern Recognition and Artificial Intelligence – IET Computer Vision.*

Conferences Review Committees

Reviewer for several international conferences and workshops as *IAPR International Conference on Pattern Recognition (ICPR) – IEEE International Conference on Image Processing (ICIP) – IEEE/RSJ International Conference on Intelligent Robots and Systems – EUSIPCO – SITIS – CORESA – Eurographics Workshop on 3D Object Retrieval (2010) – FG 3D Face Biometrics Workshop (FG'13) – and others.*

Conference Program Committees

Member of technical committees of the following conferences and workshops *3D Face Biometrics Workshop (in conjunction with the Automatic Face and Gesture Recognition conference, 2013)– 3D Object Retrieval Workshop (in conjunction with ACM Multimedia 2010) – International Conference on Signal Image Technology & Internet Based Systems, SITIS – International Conference on Intelligent Machines, ICIM – CORESA (COmpression et REprésentation des Signaux Audiovisuels) – TAIMA (Ateliers de Travail sur le Traitement et l'Analyse de l'Information).*

Conferences Organizing Committees

Member of the organization committee of *15th edition of CORESA (COmpression et REprésentation des Signaux Audiovisuels), Lille, 24-25 May 2012.*

Member of the organization committee of the international conference *Shape Modeling International (SMI), which will take place in Lille in 2015.*

---

[2]http://www.imagina.mc/2011/content/Industry-speakers/industry_speakersUK.php
[3]https://cybercriminalite.wordpress.com/

Participation to Ph.D. committees

Member of the thesis committee of Wael Ben Soltana, École Centrale de Lyon, December 2012.

Member of the thesis committee (midterm) of François Lamare, Télécom SudParis, September 2014.

Awards

Best Paper Award[4] received for our paper titled – Can 3D Shape of the Face Reveal Your Age? – in the international conference on Computer Vision Theory and Applications (VISAPP) 2014, with *Baiqiang Xia (PhD student), Mohamed Daoudi and Hassen Drira.*

Affiliation to Scientific Societies

Member of IEEE (Institute of Electrical and Electronics Engineers).

Member of GDR-ISIS[5], Image et Vision theme.

Member of AFRIF[6] (Association Française de Reconnaissance et d'Interprétation des Formes).

Publications

Articles in international peer-reviewed journals (7)

[**J1**] **Boulbaba Ben Amor**, Hassen Drira, Stefano Berretti, Mohamed Daoudi, Anuj Srivastava, *4D Facial Expression Recognition by Learning Geometric Deformations*. Accepted for publication as regular paper in the IEEE Tran. on Cybernetics, 2014 (published online). **(Impact Factor: 3.236)**.

[**J2**] Hassen Drira, **Boulbaba Ben Amor**, Anuj Srivastava, Mohamed Daoudi, Rim Slama, *3D Face Recognition under Expressions, Occlusions, and Pose Variations*. IEEE Trans. Pattern Anal. Mach. Intell. 35(9): 2270-2283, 2013. **(Impact Factor: 4.795)**.

[**J3**] Lahoucine Ballihi, **Boulbaba Ben Amor**, Mohamed Daoudi, Anuj Srivastava, Driss Aboutajdine, *Boosting 3-D-Geometric Features for Efficient Face Recognition and Gender Classification*. IEEE Transactions on Information Forensics and Security 7(6): 1766-1779, 2012. **(Impact Factor: 1.895)**.

[**J4**] Ahmed Maalej, **Boulbaba Ben Amor**, Mohamed Daoudi, Anuj Srivastava, Stefano Berretti, *Shape analysis of local facial patches for 3D facial expression recognition*. Pattern Recognition 44(8): 1581-1589, 2011.**(Impact Factor: 2.632)**.

[**J5**] Stefano Berretti, **Boulbaba Ben Amor**, Mohamed Daoudi, Alberto Del Bimbo, 3*D facial expression recognition using SIFT descriptors of automatically detected keypoints*. The Visual Computer 27(11): 1021-1036, 2011. **(Impact Factor: 0.91)**.

---

[4]`visapp.visigrapp.org/PreviousAwards.aspx`
[5]`gdr-isis.fr/`
[6]`www.afrif.asso.fr/`

[**J6**] <u>**Boulbaba Ben Amor**</u>, Hassen Drira, Lahoucine Ballihi, Anuj Srivastava, Mohamed Daoudi, *An experimental illustration of 3D facial shape analysis under facial expressions.* Annals of telecommunications 64(5-6): 369-379, 2009. (**Impact Factor: 0.56**).

[**J1-s**] Baiqiang Xia, <u>**Boulbaba Ben Amor**</u>, Mohamed Daoudi, *Combining Face Averageness and Symmetry for 3D-based Gender Classification.* <u>**Accepted for publication**</u> in Pattern Recognition, 2014 (**Impact Factor: 2.584**).

Articles under review, or in progress (3)

[**J2-s**] <u>**Boulbaba Ben Amor**</u>, Jingyong Su, Anuj Srivastava, *Action Recognition Using Rate-Invariant Analysis of Skeletal Shape Trajectories.* <u>**Major Revisions**</u> in IEEE Trans. Pattern Anal. Mach. Intell., 2014.

[**J3-s**] Baiqiang Xia, <u>**Boulbaba Ben Amor**</u>, Mohamed Daoudi, *Joint Gender, Ethnicity and Age Estimation from 3D Faces - An Experimental Illustration of their Mutual Correlations.* **Submitted** to International Journal of Computer Vision, 2014.

[**J4-s**] <u>**Boulbaba Ben Amor**</u> and Anuj Srivastava, *Sequential Test of Skeletal Shape Trajectories for Early Action Recognition.* (**In preparation**).

Book Chapters (4)

[**O1**] Contributor in three chapters of the book*"3D Face Modeling, Analysis and Recognition"*[7], Editors Mohamed Daoudi, Anuj Srivastava, Remco Veltkamp, ISBN: 978-0-470-66641-8, Wiley, August 2013.

- Chapiter I, *3D Face Modeling*, <u>**Boulbaba Ben Amor**</u>, Mohsen Ardabilian and Liming Chen.

- Chapiter III, *3D Face Surface Analysis and Recognition Based on Facial Curves*, Hassen Drira, Stefano Berretti, <u>**Boulbaba Ben Amor**</u>, Mohamed Daoudi, Anuj Srivastava, Alberto del Bimbo and Pietro Pala.

- Chapter V, *Applications*, Stefano Berretti, <u>**Boulbaba Ben Amor**</u>, Hassen Drira, Mohamed Daoudi, Anuj Srivastava, Alberto del Bimbo and Pietro Pala.
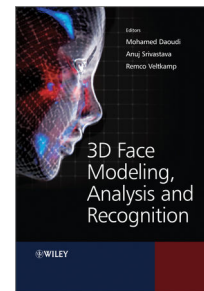
[**O2**] Contributor in the chapter *Elastic Shape Analysis of Functions, Curves, and Trajectories* of the book *"Riemannian Computing in Computer Vision"*[8], Editors Pavan Turaga, and Anuj Srivastava (**In preparation**).

Articles in selective international peer-reviewed conferences and workshops (24)

[**C1**] Baiqiang Xia, <u>**Boulbaba Ben Amor**</u>, Mohamed Daoudi, Hassen Drira: *Can 3D Shape of the Face Reveal your Age ?* International Conference on Computer Vision Theory and Applications (VISAPP 2014), **Best Paper Award**[9].

---

[7]http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470666412.html
[8]http://www.public.asu.edu/~pturaga/RCCV/Welcome.html
[9]http://www.visapp.visigrapp.org/PreviousAwards.aspx

[C2] Baiqiang Xia, **<u>Boulbaba Ben Amor</u>**, Hassen Drira, Mohamed Daoudi, Lahoucine Ballihi: *Gender and 3D facial symmetry: What's the relationship?* 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG 2013).

[C3] Baiqiang Xia, **<u>Boulbaba Ben Amor</u>**, Huang Di, Daoudi Mohamed, Wang Yunhong, Drira Hassen: *Enhancing Gender Classification by Combining 3D and 2D Face Modalities*, 21th European Signal Processing Conference (EUSIPCO) (2013).

[C4] Mohamed Daoudi, Hassen Drira, **<u>Boulbaba Ben Amor</u>**, Stefano Berretti: *A dynamic geometry-based approach for 4D facial expressions recognition.* EUVIP 2013: 280-284.

[C5] Hassen Drira, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Stefano Berretti: *A Dense Deformation Field for Facial Expression Analysis in Dynamic Sequences of 3D Scans.* HBU 2013: 148-159.

[C6] Lahoucine Ballihi, Anuj Srivastava, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Driss Aboutajdine: *Which 3D geometric facial features give up your identity?* ICB 2012: 119-124.

[C7] Hassen Drira, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Anuj Srivastava, Stefano Berretti: *3D dynamic expression recognition based on a novel Deformation Vector Field and Random Forest.* ICPR 2012: 1104-1107.

[C8] Wael Ben Soltana, Mohsen Ardabilian, Pierre Lemaire, Di Huang, Przemyslaw Szeptycki, Liming Chen, Nesli Erdogmus, Lionel Daniel, Jean-Luc Dugelay, **<u>Boulbaba Ben Amor</u>**, Hassen Drira, Mohamed Daoudi, Joseph Colineau: *3D face recognition: A robust multi-matcher approach to data degradations.* ICB 2012: 103-110.

[C9] Remco C. Veltkamp, Stefan van Jole, Hassen Drira, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Huibin Li, Liming Chen, Peter Claes, Dirk Smeets, Jeroen Hermans, Dirk Vandermeulen, Paul Suetens: SHREC '11 Track: *3D Face Models Retrieval.* 3DOR 2011: 89-95.

[C10] Lahoucine Ballihi, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Anuj Srivastava, Driss Aboutajdine: *Selecting 3D Curves on the Nasal Surface using AdaBoost for Person Authentication.* 3DOR 2011: 101-104.

[C11] Stefano Berretti, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Alberto Del Bimbo: *Person Independent 3D Facial Expression Recognition by a Selected Ensemble of SIFT Descriptors.* 3DOR 2010: 47-54.

[C12] Hassen Drira, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Anuj Srivastava: *Pose and Expression-Invariant 3D Face Recognition using Elastic Radial Curves.* BMVC 2010: 1-11.

[C13] Stefano Berretti, Alberto Del Bimbo, Pietro Pala, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi: *A Set of Selected SIFT Features for 3D Facial Expression Recognition.* ICPR 2010: 4125-4128.

[C14] Ahmed Maalej, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Anuj Srivastava, Stefano Berretti: *Local 3D Shape Analysis for Facial Expression Recognition.* ICPR 2010: 4129-4132.

[C15] Hassen Drira, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Anuj Srivastava: *Nasal Region Contribution in 3D Face Biometrics Using Shape Analysis Framework.* ICB 2009: 357-366.

[C16] Hassen Drira, **<u>Boulbaba Ben Amor</u>**, Anuj Srivastava, Mohamed Daoudi: *A Riemannian analysis of 3D nose shapes for partial human biometrics.* ICCV 2009: 2050-2057.

[C17] Salwa Said, **<u>Boulbaba Ben Amor</u>**, Mourad Zaied, Chokri Ben Amar, Mohamed Daoudi: *Fast and efficient 3D face recognition using wavelet networks.* ICIP 2009: 4153-4156.

[C18] Joseph Colineau, Johan D'Hose, **<u>Boulbaba Ben Amor</u>**, Mohsen Ardabilian, Liming Chen, Bernadette Dorizzi: *3D Face Recognition Evaluation on Expressive Faces Using the IV2 Database.* ACIVS 2008: 1050-1061.

[C19] **<u>Boulbaba Ben Amor</u>**, Mohsen Ardabilian, Liming Chen: *Toward a region-based 3D face recognition approach.* ICME 2008: 101-104, 2008.

[C20] **<u>Boulbaba Ben Amor</u>**, Mohsen Ardabilian, Liming Chen: *An Improved 3D Human Face Reconstruction Approach Based on Cubic Splines Models.* 3DPVT 2006: 279-286.

[C21] **<u>Boulbaba Ben Amor</u>**, Mohsen Ardabilian, Liming Chen: *New Experiments on ICP-Based 3D Face Recognition and Authentication.* ICPR (3) 2006: 1195-1199.

[C22] **<u>Boulbaba Ben Amor</u>**, Mohsen Ardabilian, Liming Chen: *Enhancing 3D Face Recognition By Mimics Segmentation.* ISDA (3) 2006: 150-155, 2005.

[C23] **<u>Boulbaba Ben Amor</u>**, Mohsen Ardabilian, Liming Chen: *3D Face Modeling Based on Structured-Light Assisted Stereo Sensor.* ICIAP 2005: 842-849.

[C24] **<u>Boulbaba Ben Amor</u>**, Mohsen Ardabilian, Liming Chen: *Efficient and low-cost 2.5D and 3D face photography for recognition.* SITIS 2005: 36-42.

National reviewed journals and conferences (7)

[J7] Lahoucine Ballihi, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Anuj Srivastava, Driss Aboutajdine, *Sélection de caractéristiques géométriques faciales 3D par boosting pour la reconnaissance des visages*, Traitement du signal (Journal), volume 29, 3-5 - 2012 - pp.383-407.

[C25] Lahoucine Ballihi, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Anuj Srivastava, Driss Aboutajdine, *Quelles caractéristiques géométriques faciales 3D donnent votre identité ?*, 18e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle (RFIA), Lyon, 2012.

[C26] Hassen Drira, Rim Slama, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Anuj Srivastava: *Une nouvelle approche de reconnaissance de visages 3D partiellement occultés*, 18e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle (RFIA), Lyon, 2012.

[C27] **<u>Boulbaba Ben Amor</u>**, Drira Hassen, Daoudi Mohamed, Ardabilian Mohsen, Ben Soltana Wael, Chen Liming, Lemaire Pierre, Erdogmus Nesli, Dugelay Jean Luc, Colineau Joseph: *Fusion d'Experts pour une Biométrie Faciale 3D Robuste aux Déformations*, 18e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle (RFIA), Lyon, 2012.

[C28] Ahmed Maalej, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi: *Analyse locale de la forme 3D pour la reconnaissance d'expressions faciales*, 13ème édition d'ORASIS, journées francophones des jeunes chercheurs en vision par ordinateur, Juin 2011.

[C29] Lahoucine Ballihi, **<u>Boulbaba Ben Amor</u>**, Mohamed Daoudi, Anuj Srivastava, D. Aboutajdine: *Sélection de courbes de la surface nasale pour l'authentification en utilisant Adaboost*, 17e congrès francophone RFIA, Caen, 2010.

[C30] **<u>Boulbaba Ben Amor</u>**, Mohsen Ardabilian, Liming Chen: *Acquisition 3D de visages par vision active : une solution de bout en bout*, 2e conférence COmpression et REprésentation des Signaux Audiovisuels 2005, Renne, 2005.

Software (1)

[**S1**] Mohsen Ardabilian, Liming Chen, Karima Ouji, **<u>Boulbaba Ben Amor</u>**, *Procédé d'acquisition et de modélisation 3D sans contact et avec ajout de texture*, École Centrale de Lyon, Janvier 2010.

Ph.D. Thesis (1)

[**T1**] **<u>Boulbaba Ben Amor</u>**, *Contributions to 3D face Modeling and Recognition*, Ph.D. Dissertation, co-supervised by *Prof. Liming Chen* and *Dr. Mohsen Ardabilian*, Ecole Centrale de Lyon, N 2006-39, December 2006.

Teaching Activities

TELECOM Lille (2007 – Present)

Télécom Lille is a French public engineering school (grande école) founded by Université de Lille1 and Institut Mines-Télécom, in 1991. Open to holders of a French Baccalaureate, two-year degree students (roughly equivalent to a Bachelor's degree) and to professionals, Télécom Lille has broadened access to engineering degrees. Since my arrival in September 2007, I have designed new classes related to my research topics that are proposed to the last year students. In particular, I have targeted two topics: *Introduction to Biometrics* and *3D Acquisition, Modeling and Analysis*. Since 2011, I'm the responsible and the coordinator (with another colleague in the same department) of the *Systems & Networks Security* module, proposed to the last year students.

2011 - present: Coordinator of the **Systems & Networks Security module** at Télécom Lille,
This module is proposed to the last year students. It provides a rich set of security topics, ranging from computer system security to network security and security of emerging ICT applications, and it is composed of lectures, conferences, and hands-on experiments, given by both guest lecturers from industry and researchers from academia, offering technical and scientific perspectives on dealing with security threats and attacks. The students are expected to gain insightful understanding on those significant security issues and the latest security technologies, as well as learning to solve practical security problems in creative and efficient ways. The module contains fundamental topics (Cryptography, Risk assessment, Dependability, Network Protocols, Intrusion Detection Techniques, etc.) and a second part which changes continuously according to the hot new topics in security (Cyber-criminality, Trust and Privacy, etc.).

2011 - present: **Introduction to Biometrics**,
Lectures and a practical project, proposed within the Systems & Networks Security module at Télécom Lille. My research activities in the field of face recognition pushed me to introduce this class. It is designed to first highlight the need of emerging biometrics and related applications. Then, it studied more deeply some of the technologies such as Iris, Fingerprint and Face following a pattern recognition methodology (preprocessing, feature extraction, pattern matching, datasets, evaluations, etc.). Usually, a seminar from a company (Morpho[10], etc.) is proposed to give the industrial point-of-view, the systems constraints and the new applications.

2008 - present: **3D Acquisition and Analysis Techniques**,
Lectures, practices and demonstrations, proposed to the Multimedia module at Télécom Lille. I have

---

[10]www.morpho.com/?lang=en

introduce this class to allow our students discover multiple technologies of 3D imaging (stereo-vision, structured-light, laser rangefinder, time-of-flight, etc.). Furthermore, I have introduced practices on 3D mesh processing and analysis using the Visualization Toolkit (VTK) library[11]. With these practices, the students discover tools for 3D mesh processing and scientific visualization. They work with a class library entirely designed in an Object-Oriented fashion.

2007 - present: **Algorithms and C/C++ programming**,
Lectures, practices and project, proposed to the 1st year students at Télécom Lille. First, with the Algorithms teaching, the students learn how to design solutions for a given problem and how to write an algorithm. Then, a translation to the C/C++ language which is built and executed, allow to touch on the whole process.

2007 - present: **Data Structure *(E-learning)***,
Distant lectures, proposed within the apprenticeship curriculum at Télécom Lille. Here, the students are apprentice and spend half of their time in a company and the other half at school for practices and exams. I note that Télécom Lille has developed a strong expertise in e-learning, since 2000.

2007 - 2012: **Multimedia Indexing and Retrieval**,
Java development project, proposed to the last year students in the Multimedia module. With this project, the students design and implement a graphical user interface and a set of image descriptors for image retrieval. The project ends with an evaluation procedure using the appropriate criteria Recall/Precision, F1-score, etc. This project is proposed with the lectures on Multimedia Retrieval given by Prof. Mohamed Daoudi.

École Centrale de Lyon (2004 − 2007)

École centrale de Lyon, founded in 1857, is one of the oldest post-bachelor graduate schools (grande école) in France. During my Ph.D. graduation, I got the chance to work with the professors at École Centrale de Lyon and to be involved in some classes,

2004 - 2007: **Algorithms and Data Structure**,
Practices proposed to the 1st year students at École Centrale de Lyon (with Prof. Liming Chen).

2004 - 2007: **Oriented Object Programming**,
Practices proposed to the 2nd year students at École Centrale de Lyon (with Prof. Christian Vial).

2006 - 2007: **Information System Architecture (Java language)**,
Practices proposed to the 3rd year students at École Centrale de Lyon (with Prof. Christian Vial).

EPSI Lyon (2005 − 2006)

EPSI (École Privée des Sciences Informatiques) is a French private school founded in 1961.

2005 - 2006: **Software Engineering**,
Lectures, practices and projects proposed to the engineer students at EPSI-Lyon. It was a nice experience at this private school. In fact, I was the responsible of the Software Engineering module to study modeling methodologies like UML (Unified Modeling Language) or MERISE. Several projects have been proposed to the students to practice and design applications.

---

[11]http://www.vtk.org/

Part II

Research Activities

Notations

TABLE 1: List of symbols and their definitions used in this *habilitation*.

Symbol	Definition /Explanation
X	$n \times 3$ matrix of landmarks coordinates in $\mathbb{R}^3$.
$\|.\|_F$	the matrix Frobenius norm.
$\mathcal{C}$	preshape space.
$\langle X, Y \rangle$	Riemannian metric on $\mathcal{C}$
$\mathcal{S}$	shape space.
$T_X(\mathcal{C})$	tangent space to $\mathcal{C}$ at $X \in \mathcal{C}$.
$[X]$	equivalence class of X (orbit of X).
$d_{\mathcal{C}}$	distance on $\mathcal{C}$, $d_c(X, Y) = cos^{-1}(\langle X, Y \rangle)$.
$exp_X(V)$	exponential map function ($exp : T_X(\mathcal{C}) \to \mathcal{C}$).
$exp_X^{-1}(Y)$	inverse exponential map function ($exp^{-1} : \mathcal{C} \to T_X(\mathcal{C})$).
$SO(3)$	rotation group in $\mathbb{R}^3$.
$\gamma(t)$	a parametrized geodesic path on Kendall's shape space.
$L(\gamma)$	the length of γ.
I	the unit interval $[0, 1]$.
$\mathbb{L}^2$	(also $\mathbb{L}^2(I, \mathbb{R}^3)$) refer to the infinite-dimensional function space of $\mathbb{R}^3$.
β	a parametrized continuous curve in $\mathbb{R}^3$.
q	the SRV function, $q(t) = \dot{\beta}(t)/\sqrt{\|\dot{\beta}(t)\|}$.
$|.|$	the standard Euclidean norm in $\mathbb{R}^3$.
$\langle f_1, f_2 \rangle$	the the standard $\mathbb{L}^2$ metric $\int_0^1 f_1(s) f_2(s) ds$
$\|.\|$	the standard $\mathbb{L}^2$ norm.
$\psi(t)$	a parametrized geodesic path on the shape space of continuous curves.
$L(\psi)$	the length of ψ.
$\dot{\psi}(t)$	derivative of ψ with respect to t.
Γ	group of orientation-preserving re-parametrizations in $[0, 1]$.
$[q]$	equivalence class of SRVF under rotation and re-parametrization.
$\mathcal{C}, \mathcal{S}$	preshape space and shape space, respectively.
$\mathcal{F}$	a facial surface represented as indexed collection of radial curves $\{\beta^{(\alpha)}\}_{\alpha \in [0, 2\pi]}$.
$\mathcal{M}$	the manifold of facial surfaces $\mathcal{M} = \mathcal{S}^{[0, 2\pi]}$.
$d_{\mathcal{C}}, d_{\mathcal{S}}, d_{\mathcal{M}}$	metrics on $\mathcal{C}$, $\mathcal{S}$ and $\mathcal{M}$, respectively.
Ψ	a geodesic path between elements of $\mathcal{M}$.
$\psi^{(\alpha)}$	a geodesic path between curves of index α.
$\bar{\mathcal{F}}$	a sample mean on $\mathcal{M}$ of a set of facial shapes.
$\mathbb{S}^1$	the unit circle.
r_l	a landmark of index l on the face.
λ	a variable for the value of the distance from the landmark r_l.
c_λ	a closed curve extracted around an arbitrary landmark r.
$d_{\mathcal{S}[0, \lambda_0]}$	metric on $\mathcal{S}^{[0, \lambda_0]}$ the manifold of facial patches.

Chapter 1

Extended Summary

Over the last decade, I have directed my research towards the topic of 3D shape analysis and recognition with a particular focus on studying facial surfaces and their dynamics (deformations). I have targeted fundamental applications in pattern recognition – face recognition, facial expression classification and soft-biometrics (gender, age, etc.) estimation – based on three-dimensional data. With the emergence of modern shape theory and related approaches based on differential geometry, conducted mainly in 2D domain, I have been attracted by the elegant theory and relevant geometric and statistical tools that it offers. In particular, viewing shapes as elements of finite- or infinite-dimensional manifolds, the definition of Riemannian structures (or metrics) on these manifolds, and computing statistics on them (sample mean of shapes, sample covariance, etc.). These tools are suitable and computationally efficient to be applied to pattern recognition problems, as shown in the literature of 2D domain. Throughout this *habilitation*, my goal was to develop shape analysis frameworks for 3D faces and dynamic faces (4D faces), with the specific interrelated goals of (1) matching facial surfaces (i.e. dense registration of the shapes modulo rigid and non-rigid transformations); (2) comparing 3D shapes (i.e. define shape-preserving metrics); (3) measuring/quantifying the deformations between them; and (4) averaging a population of shapes to get representative shape, for instance. As I shall describe later on, the developed methodology is shown to be promising from both empirical and theoretical perspectives and illustrate how several applications can profit from the underlying mathematical framework.

Historically, I was among the first researchers focusing on 3D face modeling and recognition (Ben Amor et al., 2006a) by investigating the (rigid) registration approaches, such as the well-known *Iterative Closest Point* (ICP) algorithm and develop variants to perform region-oriented ICP to handle the facial expression variations. It is important to mention that despite the slight overlapping with my early research conducted within my Ph.D. project[1], the methodology that I adopted later presents in someway a scientific jump in several fronts. On the theoretical side, viewing the 3D facial shapes as elements of Riemannian manifolds and the use of tools from **Differential geometry** to interpolate between them and to quantify their

---

[1]Refer to my earlier publications (Ben Amor et al., 2006a), (Ben Amor et al., 2006b), and (Ben Amor et al., 2008) for more details.

divergence, was my significant methodological evolution. Furthermore, modeling the shape variations by action of groups on shape representations, as previously formulated by D.G. KENDALL and ULF GRENANDER which leads to the **Group theory**, was my second methodological trend. This approach resulted in a registration-comparison (leaded jointly) solution of 3D facial shapes. With the use of appropriate mathematical representations and elastic shape analysis models, an efficient approach to register facial surfaces in presence of the variations cited above, has been proposed. The contributions of this *habilitation* are following:

- ↪ A mathematical representation of 3D faces through multiple curves suitable for shape analysis;

- ↪ An effective elastic model for 3D face registration and analysis leaded jointly;

- ↪ An efficient way to deform one surface into another which provides interpolation tools on the underlying manifolds;

- ↪ Robust face recognition solutions under expression, occlusion and pose variations, as shown through the extensive set of experiments conducted on publicly available Benchmarks;

- ↪ Relevant geometric tools to capture and learn deformations across 3D faces,

- ↪ Algorithms for facial expression recognition from static and dynamic 3D data;

- ↪ Morphology-inspired geometric features for soft-biometrics estimation.

In summary, this *habilitation* provides comprehensive frameworks for shape analysis of faces with applications to facial biometrics and facial expression analysis. Different pieces of the proposed computational framework, our contributions and the application perspectives are raised in the following items:

3D Faces, Elements of Infinite-dimensional Riemannian Manifolds – In *Pattern Recognition*, shapes are considered as one of the main cues for scene understanding, in addition to colors, textures and motions. Shapes are usually represented as sparse landmarks or continuous contours which delimit the external boundary of the objects. However, with the advancements of 3D scanning technologies, like laser rangefinder, structured-light and time-of-flight cameras, it is possible to capture depth-maps of the objects in the scene, which is a more complete description of the external boundary of shapes. These 3D images (depth-maps) could be often converted to 3D point clouds, 3D surfaces (meshes) or textured surfaces. Hence, there was an outstanding need for developing suitable tools and computational strategies for shape interpretation under many variablities of two kinds – **(1) shape-preserving transformations** such as the rigid transformations (scaling, translations and rotations), and **(2) shape-changing transformations** as the non-rigid deformations, which affect the shape. Because most of the geometric features (landmark configurations, curves, surfaces, deformations) do not belong to Euclidean spaces but rather to curved manifolds, a new methodology has emerged over the three last decades for studying shapes and model their variabilities. The emerging *school* of modern shape theory, promoted

by researchers as D.G. KANDELL and U. GRENANDER, consider the space of shapes representations as a geometrical object akin to a Riemannian manifold on which it is possible to resort conventional tools from Differential Geometry. Thus, shapes (or their mathematical representations) live on spaces on which one should impose Riemannian metrics required to quantify shape similarity/difference, as a first goal. There has been an increasing interest in using Riemannian frameworks for shape analysis of objects. One reason for its increasing popularity is the breadth of the tools that it offers. On one hand it allows to remove all shape-preserving transformations from the representation space, using the notion of quotient spaces and, on the other, it provides geometric tools and algorithms for computing statistics (sample means, covariances, etc.) of shapes useful in statistical modeling, for example. Our first contribution relay on this point, after a pre-processing step, trough a new mathematical representation, we map the facial surfaces on an infinite dimensional Riemannian manifold. Then, we derive algorithms to quantify shapes divergence and compute statistics on the underlying shape space.

According to several studies (including my earlier work (Ben Amor et al., 2006a)), the most important variability that we seek to filter out is the facial deformation. The scaling, the location and the rotations could be removed by introducing pre-processing steps and using previous solutions (Kendall, 1984). To tackle this issue, we have proposed to represent the facial surfaces as collections of radial curves emanating from the nose tip (as a reference point). This impose a first level of parameterization of the facial surfaces. The second level is to adopt an elastic model to pairwise register and compare the 3D curves through the faces. Recall that in practice the 3D faces subject of the study are presented as 3D meshes devoid of order (or parameterization) on them. Thus, representing facial surfaces as indexed collections of parametrized curves is a crucial step in our shape analysis methodology. Some previous works (including ours) have proposed similar representations through sets of curves (Samir et al., 2006, 2009)(Ben Amor et al., 2009, Drira et al., 2009a)(Berretti et al., 2011), however, the radial curves representation is more suitable to handle facial deformations (as we will discuss in Chapter 2, Section 2.4.1). The next step was to propose a mathematical representation of 3D faces which account for the stretching, shrinking and bending of the facial surfaces. We have adopted the Square Root Velocity Framework, proposed in (Joshi et al., 2007, Srivastava et al., 2011) to compare facial surfaces through their parametrized radial curves. Based on this representation, shapes (of 3D faces) are viewed as elements of an infinite-dimensional curved manifold. The shape space is a quotient space of a pre-shape space under the action of the groups of transformations. Tools from differential geometry are used to define a Riemannian metric on the shape space and compute geodesics (most efficient way to deform on shape to another) under the elastic metric.

Shape Analysis of Static Faces – Due to the increasing importance of shape analysis of objects in different applications, including 3D faces, a variety of mathematical representations and techniques have been suggested in the literature. The difficulty in analyzing shapes of objects comes from the fact that: (1) Shape representations, metrics, and models should be invariant to certain transformations. For instance, rigid motions and re-parameterization of facial surfaces do not change their shapes, and any shape analysis of faces should be invariant to these transformations. Several tools have been proposed in our framework including (1) interpolating between 3D facial shapes using geodesics on manifolds;

(2) measuring distances between shapes (length of geodesics); (3) averaging a set of shapes to define sample means and statistical models on their tangent spaces. These ideas have been applied to 3D face recognition under challenging variations such as expressions, occlusions, and pose variations (Drira et al., 2010a, 2013a). Here, facial surfaces are parametrized as collections of curves are elements of an infinite-dimensional manifold as described earlier. This methodology have been also successfully applied to expression analysis from 3D static faces (Maalej et al., 2010, 2011). Here, shape analysis of local facial patches (local surfaces around pre-defined landmarks) is used to capture and learn deformation patterns for each class of expressions. Taking another direction, we have proposed in (Berretti et al., 2011) to use SIFT feature descriptors of depth images around keypoints to perform 3D facial expression recognition.

Shape Analysis of Dynamic (4D) Faces – With the advancement in 3D imaging technologies, it is possible nowadays to capture dynamic flows of objects for the purposes of merging them or making analysis of their dynamics (motions). To reach the latter goal, an accurate dense registration of meshes (frames) across time is important. One basic idea to capture facial deformations across 3D video sequences is to track mesh vertices densely along successive 3D frames. Specifically, in comparing shapes of faces, it is important that similar biological parts are registered to each other across different faces. Furthermore, it is important to use techniques that allow a joint registration and comparisons of surfaces in a comprehensive framework, rather than in two separate steps. These two issues – invariance and registration – are naturally handled using Riemannian methods where one can choose metrics that are invariant to certain transformations and form quotient spaces (termed shape spaces) by forming equivalence classes of objects that have the same shape. The elastic Riemannian metric that we used provides a nice physical interpretation of measuring deformations between facial curves using a combination of stretching and bending. We have introduced in (Drira et al., 2012, Ben Amor et al., 2014a) a novel method to capture densely and faithfully the facial deformations, grounding on Riemannian geometry, called *Dense Scalar Fields (DSFs)*. The elastic deformations are accurately captured by the DSF features and used for expression classification. The main motivation of using a Riemannian approach is to perform registration that matches corresponding anatomical features, and obtain deformation fields that are physically interpretable.

Learning Geometric Features for Classification – In this part of our work, we utilize ideas from two growing but disparate fields in computer vision – shape analysis using tools from differential geometry and Machine Learning techniques – to learn the geometric features for classification tasks. For example, in (Ballihi et al., 2012a,b) we have used the Adaboost algorithm to select and highlight salient geometrical facial features that contribute most in 3D facial biometrics. Firstly, a large set of geometric features (curves) are extracted using level sets (circular curves) and streamlines (radial curves) of the Euclidean distance functions of the facial surface; together they approximate facial surfaces with arbitrarily high accuracy. Then, we used the shape analysis techniques in order to measure shapes difference and capture their deformations that are fed up to the Boosting step. A second idea was to derive from the geometric features Euclidean representations, then apply Machine Learning algorithms to perform face classification. For example, the Dense Scalar Fields are the magnitude of the shooting vectors (elements of different

tangent spaces), which bring their initial representation to an Euclidean vector space. Thus, applying traditional classification algorithms (Random Forest, Hidden Markov Models, Support Vector Machines, etc.) for face classification become possible. These ideas have been tested in facial expression recognition (Maalej et al., 2011, Ben Amor et al., 2014a) and more recently in soft-biometrics estimation (Xia et al., 2013a, 2014a), such as gender classification and age estimation.

Manuscript Organization

After this executive summary of the *habilitation* contributions, Chapter 2 describes scientific challenges of 3D face analysis and provide a brief review of modern shape theory with an essential background of shape analysis of landmark-based representations and continuous parametrized curves. It also details our methodology for shape analysis of 3D faces and shows our practical and unified computing framework. Chapter 3 targets our first application, 3D face recognition under pose, expression and occlusion variations. In Chapter 4 we explore the role of shapes in facial expression analysis and recognition, from static or dynamic data. We present in Chapter 5 our third application related to soft-biometrics estimation using Morphology-inspired geometric descriptions. Finally Chapter 6 presents our current research and some perspectives for future directions.

Chapter 2

Shape Analysis on Riemannian Manifolds

The goal of this chapter is to provide a brief literature review and essential theoretical aspects required to introduce our contributions in 3D shape analysis. We will start by reviewing some pioneering work from the literature related to modern shape theory ranging from landmarks (a dataset of registered anchor points), in particular the Kendall's shape analysis methodology, to landmark-free geometrical representations, including 3D continuous curves. Then, we will introduce the geometrical shape analysis framework grounding on Riemannian geometry proposed in our research for 3D shape analysis. In particular, we will emphasize on two key ideas – **(1)** the mathematical representation of the studied shapes, and **(2)** the geometric tools defined on the space of these representations for the purposes to measure distances between shapes, interpolate between them, quantify their shape difference, estimate an average shape as well as estimate the shape variability within a class of shapes. One important difficulty to lead shape analysis is that the study of the representations involves non-linear spaces (i.e. not vector spaces), thus tools commonly used to process the shapes are not valid. For that reason, shape analysis often implies *Differential Geometry* to study curved spaces formed by shape representations and to develop appropriate tools to resolve the tasks above-cited. Another important fundamental aspect which has emerged is to consider the variabilities induced by the transformations such as rotations, translation, scaling, and deformations (diffeomorphisms) as actions of certain groups on the space of representations. Thus, the study of the groups and their actions which involves knowledge in *Group theory*, is required to conduct transformation(s)-invariant shape analysis. Later, through this *habilitation*, we will discuss the benefits of the geometric shape analysis methodology and will provide some illustrations involving 3D face analysis applications.

Recall that the aim of *shape analysis* is to develop mathematical descriptions of shapes, so that shapes of objects can also be measured and quantitatively compared. In particular, goals of shape analysis include:

↪ **Quantifying similarity (or difference) between shapes** – The space of shape descriptors is equipped with a sort of distance function d. This function measures the difference between shapes, enabling us to make statements such as shape x is more similar to y than to z (formally, $d(x, y) \leq d(x, z)$).

↪ **Finding the most efficient way to deform one shape into another** – By viewing shapes as elements of a given shape space, it is often possible to define paths (curves) on that shape space connecting two arbitrary shapes. Minimal curves (in term of length) provide the most efficient way to deform one shape into another and can be seen as an interpolation between these shapes. This needs first algorithms to produce dense correspondence between the shapes that seek to optimally register them.

↪ **Defining a mean shape to represent a shape class** – Given a set of shapes, points of the shape space, one seek here to define a sample representative mean shape of that class. As shape spaces are curved spaces, we need to use appropriate approaches from differential geometry involving tangent spaces to design algorithms to estimate a sample mean.

↪ **Shape classification** – Suppose we have a database of previously-observed shapes from each of several different classes. Given a test shape, we would like to determine which of these classes the test shape most likely belongs to. In classification tasks, it is often useful to define probability densities of shapes given a mean shape and a set of training samples.

Thanks to tools from Differential Geometry, achieving the goals above-mentioned become possible. In the following sections, I shall review essential concepts and related methods from state-of-the-art which offered elegant theories and computational algorithms for shape analysis. From this point-of-view, my research is indebted to the seminal work of D.G. KENDALL (Kendall, 1984) and his colleagues (Dryden and Mardia, 1998) on statistical shape analysis of landmark-based representations, the work of U. GRENANDER on shape theory formulation (Grenander, 1993), and the recent work of A. SRIVASTAVA and his group in elastic shape analysis of continuous curves in $\mathbb{R}^n$, in particular (Klassen et al., 2004), (Joshi et al., 2007), and (Srivastava et al., 2011). I should also cite the work conducted in M. DAOUDI's group in the field of 3D face recognition using iso-level curves (Samir et al., 2006, 2009), before I joined his group and get involved in the research on that topic. Next section will provide a brief literature review of 3D face analysis as a particular challenging case of 3D shape analysis.

2.1 3D Faces, Particular 3D Shapes

The exploitation of the 3D shape of the face rather than its appearance with definition of innovative algorithms for 3D face matching has been a growing field of research in recent years. In addition to face recognition, many applications have been targeted as facial expressions analysis and Action Units

detection, and soft-biometrics estimation (gender, ethnicity and age). The up-to-date list of the solutions proposed can be derived from the survey of (Bowyer et al., 2006a) and the literature reviews of (Drira et al., 2013a)(Ben Amor et al., 2014a)(Xia et al., 2014b). A collection of 3D face data sets have also been made available, thus providing the opportunity to have fair comparison between different solutions. The Face Recognition Grand Challenge (FRGC) initiative, directed by NIST, provides common data sets to be used as a reference for training (FRGCv1) and evaluation (FRGCv2) ; a 3D face recognition contest was launched in 2005, with the final results published in 2006 (Phillips et al., 2005). Since then, several 3D face recognition approaches have been developed. In the same direction, some research groups are interested on performing facial expression recognition from 3D still images (Wang et al., 2006a)(Berretti et al., 2011)(Fang et al., 2011)(Gong et al., 2009a)(Le et al., 2011)(Mpiperis et al., 2008a)(Tang and Huang, 2008a) and very recently from 4D face sequences (Sun et al., 2010a) and (Sandbach et al., 2012a). Thanks to the publicly available databases such as BU-3DFE and Bosphorus for static analysis and BU-4DFE for dynamic analysis, many research groups have investigated the problem with the assumption that the 3D facial geometry could be more informative (in terms of deformations) that planar images, thus their use can achieve better performances as demonstrated in (Sandbach et al., 2012a).
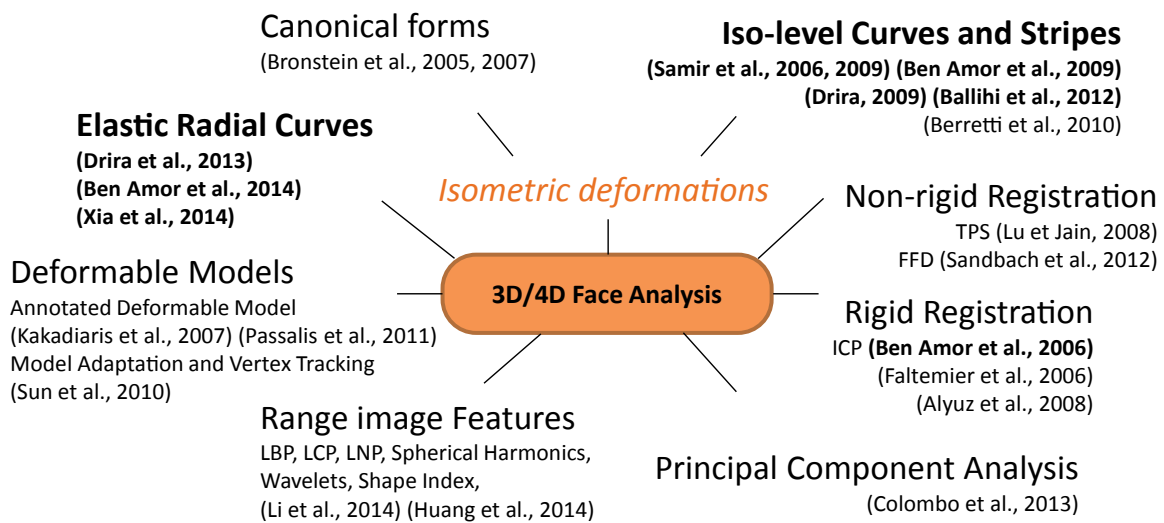

FIGURE 2.1: Literature review of 3D face analysis approaches for recognition, expression recognition and soft-biometrics estimation.

Show in figure 2.10 is an overview of related approaches used in different applications, in particular face and facial expression recognition. The fundamental ingredient of these approaches (performing on static or on dynamic 3D data) is the facial matching (or correspondence). For example in face recognition it is important to match accurately different anatomical parts of the face before using a defined metric to derive a dissimilarity measure between the face (Bronstein et al., 2005a, 2007a). This step is also very important to achieve dynamic analysis (Sun et al., 2010a) (Sandbach et al., 2012b). With the definition of an appropriate metric, this is the most important step in 3D facial shape analysis because of the multiple variabilities exhibited by the facial surface,

Pose variations (missing data) Facial expression variations Internal and External occlusions
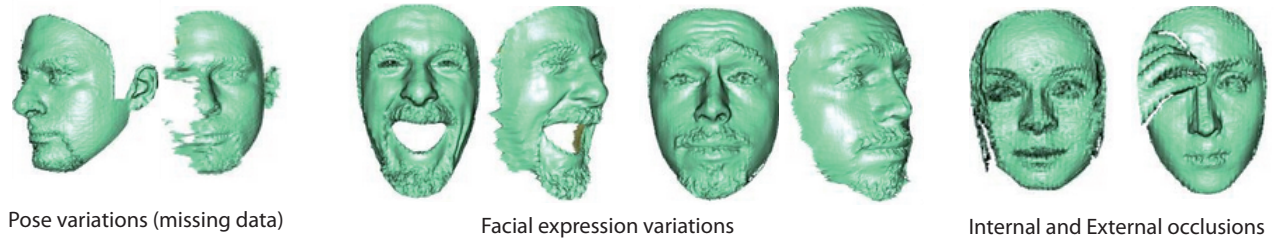
FIGURE 2.2: Different challenges of 3D face analysis: Pose variations, deformations caused by the expressions, occlusions, missing data, etc. (these examples are taken from the Bosphorus dataset).

↪ **Facial Expressions** induce non-rigid (often elastic) transformations to the facial surfaces (center of Figure 2.2). One seek here to use metrics suitable to compare facial surfaces modulo such transformations. An assumption which states an approximation of the facial deformations to be isometric under the intrinsic surface metric has been first proposed in (Bronstein et al., 2005b)(Bronstein et al., 2007a) and considered by other groups (Samir et al., 2009)(Ben Amor et al., 2009) using iso-level curves and (Berretti et al., 2010a) based on iso-stripes, to handle the deformation problem. Another direction consists to use *Deformable Face Models* as illustrated in (Kakadiaris et al., 2007) (annotated deformable model) and (Sun et al., 2010a) (Model Adaptation and Vertex Tracking) to perform facial registration. The use of existing solution for non-rigid registration such as the Thin-Plate Splines (TPS) in (Lu et al., 2006a) and the Free-Form Deformation (FFD) in (Sandbach et al., 2012b) algorithms is explored to perform an accurate registration.

↪ **External Occlusions** result in external parts which will over the facial mesh (the right part of Figure 2.2). One seek here to first remove the external occlusions which results in missing data, then apply one of the following strategies, (1) tolerate the missing data using local approaches (Mian et al., 2008)(Huang et al., 2011) or region-based approaches (Faltemier et al., 2008), or (2) complete the missing data using generative statistical models (such as Gappy-PCA applied on range images, as proposed in (Colombo et al., 2009)) then perform the analysis.

↪ **Head Pose Variations** induce self-occlusions which results in missing data on the facial surface and luck of matching between the faces (left part of Figure 2.2). A rough registration step is necessary to perform face comparison (Ben Amor et al., 2006b)(Lu et al., 2006a).

Accordingly, several works have investigated the problem of face analysis using static and dynamic 3D data with fundamental problems of invariance and registration. From the description above, the proposed approaches tackle one of the challenging problems. Abundant works have been proposed for non-rigid registrations of facial surfaces, and others have been oriented to handle external occlusions or significant pose variations. In this *habilitation*, we propose a novel methodology which can handle the above-cited challenges under a unified Riemannian framework. As we will illustrate later, the – invariance and registration – problems are solved jointly (in the same step) using an elastic model suitable to handle the deformations. This framework allows also for formal statistical inferences, such as the estimation of

missing facial parts using Principal Component Analysis on tangent spaces and computing average shapes. These tools are useful in data completion and hierarchical clustering for efficient recognition. Facial deformations are captured using the shooting directions (elements of tangent spaces of the underlying manifold) computed between faces and serves to classify the facial expressions. Lastly, Morphology-inspired perspectives are extracted using the same Riemannian framework to perform soft-biometrics estimation and study the correlation between them. Although the ideas of using tools from Differential Geometry and Machine Learning are disjoint, and rarely used together, we have successfully used them to learn *geometric* features for *classification.*

2.2 Literature Review on *Modern Shape Theory*

Shapes of natural or man-made objects extracted from imaged scenes are important cues used in detection, recognition, retrieval, clustering and classification. It points out the external form of someone or something as produced by their outline and is usually viewed as a set of landmarks or continuous boundary. A formal and intuitive definition of shape has been introduced in (Kendall, 1984) then in (Dryden and Mardia, 1998) and considered later by many researchers in this field:

***Shape** is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object (definition taken as is from (Dryden and Mardia, 1998)).*

This topic is not new[1], that is, the biologist D'Arcy W. Thompson (Thompson, 1917) have proposed theories in geometric species warping, which have been presented in his book *On Growth and Form.* Thompson explored the degree to which differences in the forms of related animals could be described by means of relatively simple mathematical transformations (Thompson, 1917). Thompson essentially applied the idea of morphing to biological structures, whereby he plotted the contour landmarks of various like-species and generated deformed grids based on the results.

The seminal work of Kendall (Kendall, 1984) (and later of (Dryden and Mardia, 1998)) and Bookstein (Bookstein, 1986) resulted in elegant and comprehensive statistical shape analysis theory, that influenced the modern theory of shapes and inspired many researchers with the introduction of methods and techniques derived from differential geometry. Herein, shapes are represented by **sets of ordering landmark points** and their statistical variability imposes to deal with a set of Euclidean shape-preserving transformations such us **scaling**, **translation**, and **rotation**. Starting with this space, Kendall methodically and rigorously proceed to remove variability due to translation, rotation, and (optionally) scaling to arrive at their space of shape representations, which is the space of orbits under the action of the rotation group. He also equipped this orbit space with a Riemannian metric, making possible to quantify

---

[1]This review cannot claim to provide an exhaustive account of the subject. Related theory and references listed in this section mostly rely on the author's view of the field. We encourage the reader to refer to the nice and comprehensive survey written recently by L. Younes in (Younes, 2012) for a more complete picture.

shape divergences (geodesic distance) and to provide geodesics between shapes. The notion of *geodesic distance* is a basic tool for statistical shape analysis. Later, in (Dryden and Mardia, 1998) the authors have discussed planar Procrustes analysis to highlight the main components of shape analysis. They have defined tangent-space probability models to Kendall's shape manifolds.

From a similar point-of-view, Grenander's shape theory formulation (Grenander et al., 1991) viewed **continuous shapes** as points on an infinite-dimensional, differentiable manifold. The variations between shapes are modeled by actions of *Lie groups* on this manifold. Low-dimensional groups, such as rotation, translation and scaling, change the object instances keeping the shape unchanged, while high-dimensional groups (diffeomorphisms) smoothly change the object shapes (Younes, 1998). This theory proposed to view the set of shape representations (the shape space) as quotient of the pre-shape space, obtained by modding out shape-preserving transformations. The idea is to assume that there is a template object, which represents the reference shape. Then, the variability of the shape is analyzed through the deformations of this reference shape towards the actual observations: a shape difference is encoded by the transformation that deforms one to reach the other. In (Grenander et al., 1991), the authors used Markov models to represent the boundaries of non-rigid objects. They demonstrated how these models can be used to detect objects in noisy images. By using deformable templates (Grenander, 1993) with shape deformations modeled as action of diffeomorphisms, (Basri et al., 1998) described how to measure similarity between objects in terms of **the amount of stretching and bending** necessary to turn the shape of one object into the shape of another one. Work toward elastic shape analysis (using **elastic models**) has been carried out by L. YOUNES in (Younes, 1998) followed by (Michor and Mumford, 2003) and (Mio et al., 2007), where some shape descriptors and metrics were derived. The key idea in elastic analysis is that the mapping is nonlinear, i.e. points that are matched together are at unequal distances from their origins. Such a matching can be considered as an elastic matching, as one curve has to (locally) stretch, compress and bend to match the other.

Modern vision problems require a unifying framework for the statistical study of shapes and the development of computational algorithms. This demand has spawned numerous studies of shapes in recent years. For example, (Klassen et al., 2004) introduced a representation for planar curves parametrized by arc-length in which each curve is represented by its angle function $\theta(t)$, defined as the elevation angle of the tangent vector of the curve at t (values are chosen so that θ is continuous). This representation is invariant to translation and reparametrization, and can be made rotation-invariant by vertically shifting each angle function so that it has an average value of π. However, since all curves are required to be parametrized by arc-length, it is not possible to reparametrize curves to improve the registration between them. They applied this framework to statistical modeling and analysis using large collections of shapes (Srivastava et al., 2005). Other parametric shape analysis frameworks do allow the curves to be reparametrized; these are referred to elastic shape analysis frameworks, since they model deformations as combinations of bending and stretching. Mio et al. (Mio et al., 2007) used a parametric curve representation along with an **elastic metric** obtained as a weighted combination of stretching and bending energies. A general discussion of several classes of metrics on the space of smooth planar curves modulo

reparametrization that can be introduced for this purpose can be found in (Michor and Mumford, 2007). More recently, (Joshi et al., 2007, Srivastava et al., 2011) presented a special representation of curves, called the Square-Root Velocity Function (SRVF), under which a specific elastic metric becomes an $\mathbb{L}^2$ metric and simplifies the shape analysis. The later family of approaches will be discussed in detail in Section 2.4. Before coming to them, I shall present in the Section 2.3 the Kendall shape analysis approach related to registered landmark representations.

When the state-of-the-art on shape analysis of landmarks configurations and curves (1D-functions, 2D-curves (planar), ..) is now well established, there is nowadays a new trend to shape analysis of surfaces. The main difference with continuous curves is that when the arc-length parametrization is a natural choice to impose an "order" on curves, surfaces are completely devoid of order. So, one should first impose a parametrization then move to the shape analysis task. It is well-known that the main difficulty in 3D shape analysis is the registration of vertices across the 3D meshes before their comparisons. In addition, the analysis methods should achieve desired invariance (which is closely related to the chosen metric) to transformations such as rigid motions (rotations, translations and global scaling) and to reparameterizations (deformations). Some recent work studied shape spaces of parametrized surfaces and the invariance to rigid transformations and change of parametrization (Bauer and Michor, 2011) and (Kurtek et al., 2012).

Because of the difficulty of the over-mentioned problems, our focus was directed to facial surfaces and all the variabilities than are able to exhibit. We can cite the head pose variations which results in rigid transformations (mainly rotations) and missing data due to the self occlusions. Also, the facial expressions which non-rigidly deform the surface and some times change its topology (when the mouth is open for instance). When moving our faces in front of the 3D camera, it introduces translations, scaling and spatial resolution changes. All these changes makes the problem difficult and conventional tools are no more able to handle all the challenges in one step. For example, the well-known Iterative Closest Point algorithm (Besl and McKay, 1992) works only for rigid surfaces thus do not tolerate deformations.

Our idea is to represent the facial surfaces by collections of curves. The facial curves have a fixed ordering and only one-dimensional diffeomorphisms are needed for their registration. One can view our methodology as a natural transit from curves to 3D surfaces, with applications to 3D face analysis. In the following sections, we shall give a brief review of Kendall's approach for shape analysis, then move to continuous curves. We will close the chapter by presenting our methodology for shape analysis of 3D faces and the derived tools to register facial surfaces, compare and average them and quantify accurately their divergence. These tools will be applied in the next chapters for face recognition (Chapter 3), expression recognition (Chapter 4) and soft-biometrics estimation (Chapter 5).

2.3 Kendall's Shape Space and Procrustes Analysis

Kendall's Shape Analysis approach provides an elegant and comprehensive framework applied to discrete landmarks representation and provides tools derived from differential geometry which can be used to come to the shape analysis aims cited above. This framework is useful in many applications involving registered landmarks configurations, in medical images, animal anatomy understanding, etc. Let $X \in \mathbb{R}^{n \times 3}$ represents a discrete points set or a configuration of n landmarks in $\mathbb{R}^3$ (here, we will restrict ideas to the landmarks in $\mathbb{R}^3$, extension to $\mathbb{R}^k$ is straightforward). Starting with this space, ([Kendall, 1984](#)) methodically and rigorously proceed to remove variability due to translation, rotation, and (optionally) scaling to arrive at their space of shape descriptors, which is the space of orbits under the action of the rotation group. They also equip this orbit space with a Riemannian metric, making it possible to compute distances and geodesics between shapes. More formally, they are interested in analyzing shapes of the set of configurations X, i.e analysis should be invariant to rotations, translations, and global scaling. The setup for this shape analysis is described next. To remove translation, a possible common way is to force X to satisfy: $\sum_{i=1}^{n} X_{i,j} = 0$, for $j = 1, 2, 3$, via translation. Similarly, to remove the scale we assume that $\|X\|_F = \sqrt{\sum_{i,j} X_{i,j}^2} = 1$. Let $\mathcal{C}_0$ be the set of all such centered configurations of n landmarks in $\mathbb{R}^3$, i.e. $\mathcal{C}_0 = \{X \in \mathbb{R}^{n \times 3} | \sum_{i=1}^{n} X_{i,j} = 0 \text{ for } j = 1, 2, 3\}$. $\mathcal{C}_0$ is a $3(n-1)$ dimensional vector space and can be identified by $\mathbb{R}^{3(n-1)}$. Since in some applications it is required to filter out the scale variability, the *pre-shape* space can be defined to be:

$$\mathcal{C} = \{X \in \mathcal{C}_0 | \|X\|_F = 1\}.$$

$\mathcal{C}$ is a unit sphere in $\mathbb{R}^{3(n-1)}$ and, thus, is $(3n-4)$ dimensional. The tangent space at any pre-shape X is given by:

$$T_X(\mathcal{C}) = \{V \in \mathcal{C}_0 | \text{trace}(V^T X) = 0\} .$$

While the translation and scaling are removed from the representation, one still need to account for the rotation variability. For any $X \in \mathcal{C}$, we define an equivalence class: $[X] = \{XO | O \in SO(3)\}$ that represents all rotations of a configuration X. The set of all such equivalence classes, $\mathcal{S} = \{[X] | X \in \mathcal{C}\} = \mathcal{C}/SO(3)$, is called the *shape space* of configurations having n landmarks. The tangent space at any shape $[X]$ is defined to be:

$$T_{[X]}(\mathcal{S}) = \{V \in \mathcal{C}_0 | \text{trace}(V^T X) = 0, \ \text{trace}(V^T X S) = 0\} , \tag{2.1}$$

where S is any 3×3 skew-symmetric matrix. The first condition makes V tangent to $\mathcal{C}$ and the second makes V perpendicular to the rotation orbit and, together they force V to be tangent to the shape space $\mathcal{S}$.
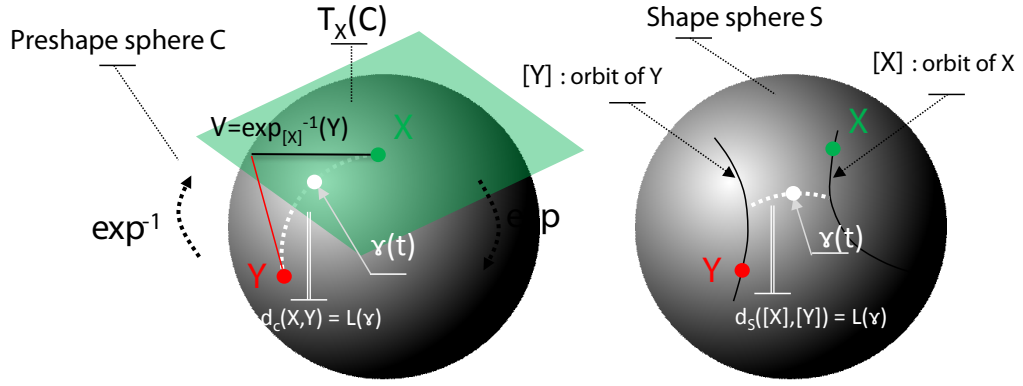
FIGURE 2.3: A pictorial of the spherical structure of the manifold $\mathcal{C}$, on the left panel, X (green) and Y (red) are elements of $\mathcal{C}$ the geodesic γ connecting them (white path), $T_X(\mathcal{C})$ (green) is the velocity space of $\mathcal{C}$ attached to X, and the shooting vector V obtained by $\exp_X^{-1}(Y)$. The right panel shows the Shape Sphere with two elements $[X]$ and $[Y]$ on it as well as the geodesic path γ (white path) and the distance between $[X]$ and $[Y]$ on $\mathcal{S}$ defined as the length of γ, $L(\gamma)$.

In view of the spherical structure of $\mathcal{C}$, the expressions for the exponential map and its inverse are well known, and can be easily adapted to $\mathcal{S}$. Shown in Fig. 2.3 are the geometry of the *pre-shape* (left) and *shape* (right) spaces with notations that will be detailed by the following items:

1. **Exponential Map**: The exponential map is given by: for any $V \in T_{[X]}(\mathcal{S})$,

$$\exp_{[X]}(V) = \left[\cos(\theta)X + \frac{\sin(\theta)}{\theta}V \right], \tag{2.2}$$

$$\theta = \sqrt{\langle V, V \rangle} = \sqrt{\text{trace}(VV^T)} \ .$$

2. **Inverse Exponential**: The inverse exponential map is given by:

$$\exp_{[X]}^{-1}([Y]) = \frac{\theta}{\sin(\theta)}(YO^* - \cos(\theta)X), \tag{2.3}$$

$$\theta = \cos^{-1}(\langle X, YO^* \rangle) = \cos^{-1}(\text{trace}(X(YO^*)^T)) \ .$$

Here O^* is the optimal rotation of Y that aligns it with X: $O^* = \text{argmin}_{O \in SO(3)} \|X - YO\|_F^2$.

3. **Geodesic Path**: Also, assuming standard Riemannian metric on $\mathcal{S}$, the geodesic between any two configurations $[X], [Y] \in \mathcal{S}$ is given by $[\gamma(\tau)]$ where

$$\gamma(\tau) = \frac{1}{\sin(\theta)}(\sin((1-\tau)\theta)X + \sin(\tau\theta)YO^*) \tag{2.4}$$

$$\theta = \cos^{-1}(\langle X, YO^* \rangle) \ , \tag{2.5}$$

(a) Source   (b) Target

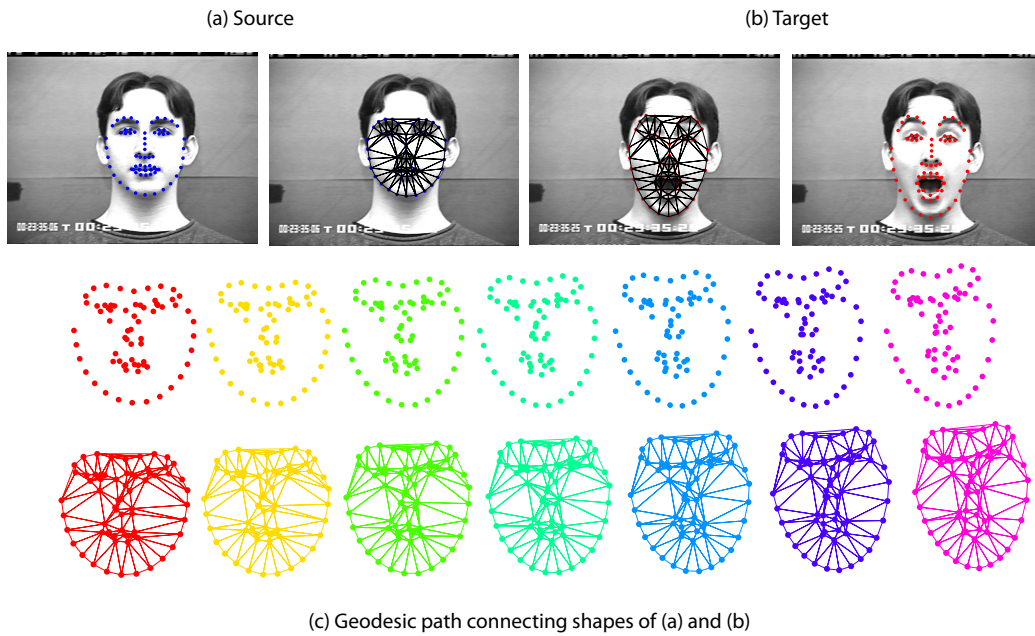(c) Geodesic path connecting shapes of (a) and (b)

FIGURE 2.4: An illustration of face landmarks interpolation under two different facial expressions (a) Neutral face with extracted 2D landmarks and (b) Surprised face with extracted landmarks. Part (c) gives the interpolation on the shape space between shapes from (a) and (b) called also geodesic connecting these shapes.

where O^* is the optimal rotation as stated in the previous item. To illustrate this idea, let's consider facial images given in (a) and (b) of figure 2.4 and the landmarks extracted on them. One can apply the interpolation formula given by Eq. 2.4 to these configurations and get the geodesic $\gamma(\tau)$ at discrete τ, given in panel (c).

4. **Procrustes Metric**: This θ given in Eq. 2.4 is also the geodesic distance separating $[X]$ and $[Y]$ (length of the shortest path connecting $[X]$ and $[Y]$) in the shape space $\mathcal{S}$, i.e.

$$d_s([X], [Y]) = \theta = \cos^{-1}(\langle X, YO^* \rangle) .$$

5. **Parallel Translation:** For shapes $[X]$ and $[Y]$, and a tangent vector $V \in T_{[X]}(\mathcal{S})$, the parallel transport of V to $[Y]$, along a geodesic connecting $[X]$ and $[Y]$, is given by:

$$V_{[X] \to [Y]} = V - \frac{2 \langle V, YO^* \rangle}{\|X + YO^*\|_F^2} (X + YO^*) . \tag{2.6}$$

6. **Mean and Median Shape Estimation**: Another important tool in shape analysis of landmarks configurations is computing their statistical summaries. Given a set of landmarks, one would like to compute their statistical mean as a template for that shape. In case of noisy observations and in the presence of outliers, one often uses a median instead of the mean to minimize the influence of outliers. Therefore, we need tools to compute mean and median of sets of shapes under the chosen

shape metric. For a set of given shapes $[X_1], \ldots, [X_k]$, the two quantities of interest are defined to be:

$$\text{Mean} \quad \hat{\mu} \quad = \quad \underset{[X] \in \mathcal{S}}{\operatorname{argmin}} \sum_{i=1}^{k} d_s([X], [X_i])^2,$$

$$\text{Median} \quad \hat{m} \quad = \quad \underset{[X] \in \mathcal{S}}{\operatorname{argmin}} \sum_{i=1}^{k} d_s([X], [X_i]) \ .$$

The formulas for computing the inverse exponential map (exp^{-1}) and the exponential map (exp) are given by Eqns. 2.3 and 2.2, respectively.

As stated earlier, the spherical structure of the manifold of interest keeps the methodology quite simple and computationally efficient. For example, the great circles on the n-sphere (unit sphere in $\mathbb{R}^{n+1}$) are geodesics of the n-sphere. A great circle is given by the intersection of the sphere with a 2-plane that pass through the origin in the Euclidean space $\mathbb{R}^{n+1}$ and the ending points of the geodesic, which could be computed explicitly using Eq.2.4. Same comment for the remaining geometric tools as the Exponential map (exp) and Inverse Exponential map (exp^{-1}) given by Eq. 2.2 and Eq. 2.3, respectively. This elegant framework and the suite of geometric tools that it can provide can be applied in many applications where registered landmarks configurations are available.

Shape spaces of landmarks are still an active research topics and are used in large number of applications. Recently, based on Kendall's approach, we have proposed in **[J1-s]** (Major Revisions in PAMI) a set of geometric tools for shape analysis of skeletal data estimated from the depth information measured by depth sensors such as the Kinect. The proposed framework studies the problem of classifying actions of human subjects using depth movies generated by the Kinect or other depth sensors. Representing human body as dynamical skeletons, we study the evolution of their (skeletons') shapes as trajectories on Kendall's shape manifold. More details bout this work and some on-going research are given in the last chapter (Chapter 6) of this *habilitation* as one of our future research directions. The shape theory of Kendall and his school (Kendall, 1984)(Dryden and Mardia, 1998) meets many of the requirements of shape analysis, but the use of landmarks is a drawback. Additionally, while the simplicity of the shape representation is very attractive from a computational standpoint, it often leads to unsatisfactory interpolations. More recent shape analysis approaches have been introduced based on Kendall's achievements but oriented to continuous parametrized curves which will be the subject of the next section.

2.4 Shape Space of Parametrized Curves and Elastic Analysis

Several applications of shape analysis required modeling the objects of interest as continuous boundaries, planar curves from 2D imaging and 3D surfaces from 3D imaging, for instance, the contours extracted from imaged scenes of humans (in this case we talk about silhouettes) or animals or any other general

object. In medical image analysis, boundaries of anatomical parts can be approximated by surfaces in $\mathbb{R}^3$. More recently, in face analysis one can use rangefinder laser scanner to capture the external boundary of the frontal part of the human head, the face, and use it in tasks of surgery simulation, realistic face animation or recognition.

Among the last decades, there has been a considerable effort in shape analysis of continuous objects, in particular continuous curves. Consequently, an abundant literature on shape analysis of curves represented mathematically to be elements of infinite-dimensional Riemannian manifolds. (Ghorbel, 1992) proposed to use Fourier descriptors and moment features for global shape representation. This approach has been used in (Ghorbel et al., 1996, Daoudi et al., 1999) for motion analysis and in (Ghorbel, 1998) for image coding. An harmonic analysis (Fourier transform for separating the shape description and the transformations) allows them to derive the notions of shape, shape space and the invariant feature space as well as a metric to compare them. (Younes, 1999) defined shape space of planar curves and derive a Riemannian approach for interpolation and comparison. (Michor and Mumford, 2006) have studied different choices of Riemannian metrics for comparing shapes of closed planar curves. More recently, (Mio et al., 2007) have proposed a family of elastic metrics able that account for the stretching and bending to deform one shape into another. In particular the Fisher-Rao metric (Mio et al., 2007) allows to keep the distance between shapes unchanged when changing their parametrization. (Joshi et al., 2007) have introduced an interesting mathematical representation of curve shapes called SRVF (Square Root Velocity Function). Taking this representation, the elastic metric for comparing shapes of curves becomes the simple $\mathbb{L}^2$-metric, as shown in (Srivastava et al., 2011). This point is very important as it simplifies the analysis of curves, under the elastic metric, to the standard functional analysis. As many contributions introduced in this *habilitation* build on the **Square Root Velocity (SRV)** framework, I shall briefly recall ideas and properties of such representation in shape analysis of curves. Then, in section 2.5, we will describe our 3D shape analysis methodology applied to facial surfaces.

2.4.1 The Square Root Velocity (SRV) Representation

Let $\beta : I \to \mathbb{R}^3$, represent a parameterized smooth curve, where $I = [0, 1]$. Note that the set of all such curves is the Hilbert space $\mathbb{L}^2(I, \mathbb{R}^3)$ or simply $\mathbb{L}^2$. Given a function β, let $\|\beta\| = \sqrt{\int_I |\beta(t)|^2 dt}$ its $\mathbb{L}^2$ norm[2]. If β_1 and β_2 are two arbitrary elements of $\mathbb{L}^2$, then we can compute the distance between them given by $\|\beta_1 - \beta_2\|$. However, the distance value can vary with the parametrization function, so that $\|\beta_1 - \beta_2\| \neq \|\beta_1 \circ \gamma - \beta_2 \circ \gamma\|$ ($\gamma : I \to I$ is a re-parameterization function and let Γ be the set of all such functions), which has deep repercussions on shape analysis. To handle this issue, in stead of using β, (Joshi et al., 2007) proposed to represent it mathematically using the *Square Root Velocity Function (SRVF)* given by,

$$q(t) = \dot{\beta}(t)/\sqrt{|\dot{\beta}(t)|}, \tag{2.7}$$

---

[2]From now, we will use $|\cdot|$ to denote the standard Euclidean norm in $\mathbb{R}^3$ and $\|\cdot\|$ to denote the $\mathbb{L}^2$-norm.

q is a special function of β that simplifies computations under the elastic metric introduced in (Mio et al., 2007), as we will show later. More precisely, as shown in (Srivastava et al., 2011), the elastic metric for comparing shapes of curves becomes the simple $\mathbb{L}^2$-metric under the SRVF representation. Furthermore, under the $\mathbb{L}^2$-metric, the group of re-parametrization acts by isometries on the manifold of q functions, which is not the case for the original curve β, so that we have $\|q_1 - q_2\| = \|q_1 \circ \gamma - q_2 \circ \gamma\|$. Also note that if we are given $q \in \mathbb{L}^2(I, \mathbb{R}^3)$, there exists a curve β (unique up to a translation) such that the given q is the SRV function of that curve. This curve can be obtained using the equation: $\beta(t) = \int_0^t q(s)|q(s)|ds$.

2.4.2   *Pre-shape Space* of SRV Functions

Consider the space $\mathcal{C} = \{q : I \to \mathbb{R}^3 | \ \|q\| = 1\} \subset \mathbb{L}^2(I, \mathbb{R}^3)$, be the pre-shape space of such representations. $\mathcal{C}$ is a space of shape descriptors which are invariant with respect to translations and scaling, that is, if two curves differ only by a scaling or a translation, then both of these curves will have the same representative in $\mathcal{C}$. Because the q function involves the first derivative ($\dot{\beta}$), it is translation-invariant, but still varies under scaling, rotations, and reparametrizations. It is made scaling invariant by first scaling all the q functions to unit length ($\|q\| = \sqrt{\int_I |q(t)|^2 dt} = \sqrt{\int_I |\dot{\beta}(t)| dt} = L(\beta) = 1$). With the $\mathbb{L}^2$ metric on its tangent spaces, $\mathcal{C}$ becomes a Riemannian manifold. However, $\mathcal{C}$ is called a pre-shape space because curves that differ only by a rotation or a reparametrization will still have different representatives in $\mathcal{C}$. Now, that we have identified the pre-shape space, $\mathcal{C}$ sub-manifold of $\mathbb{L}^2$, the tangent space at a point $q \in \mathcal{C}$ is the set of all vectors which are orthogonal to q, $T_q(\mathcal{C}) = \{v \in \mathbb{L}^2(I, \mathbb{R}^3) | \langle q, v \rangle = 0\}$. Each tangent space is equipped with the usual $\mathbb{L}^2$ inner product, and the geodesics on $\mathcal{C}$ are the great circle paths (where a great circle path is the intersection of $\mathcal{C}$ with a two-dimensional subspace of $\mathbb{L}^2$). The geodesic path between any two points $p, q \in \mathcal{C}$ is given analytically by, $\psi : [0, 1] \to \mathcal{C}$, where

$$\psi(t) = \frac{1}{\sin(\theta)} \left(\sin((1 - t)\theta)p + \sin(\theta t)q \right), \tag{2.8}$$

and the geodesic length (the geodesic distance) is given by,

$$d_\mathcal{C}(p, q) = L(\psi) = \theta = cos^{-1}(\langle p, q \rangle). \tag{2.9}$$

Figure 2.5 illustrates the spherical structure of the preshape space $\mathcal{C}$ of the SRV functions. Given two elements q_1 and q_2 the geodesic connecting them is given explicitly by the parametrization of the minor-arc of the great circle passing throw the center of the hyper-sphere, q_1 and q_2. The shooting vector (or the initial velocity vector) v at q_1 pointing towards q_2 along the geodesic ψ connecting q_1 and q_2 is given by $\dot{\psi(t)}_{|t=0}$, the first derivative of ψ with respect to the parameter t taken at $t = 0$.
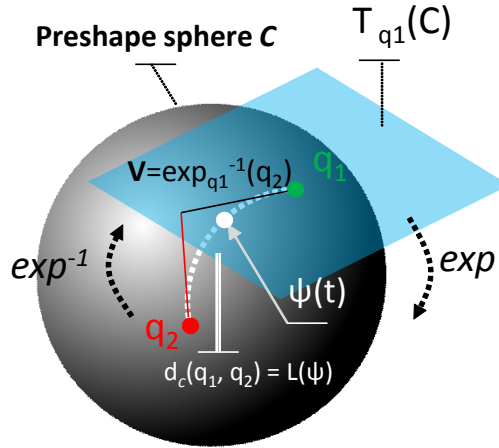
FIGURE 2.5: Geometry of the pre-shape space $\mathcal{C}$, q_1 and q_2 are elements of $\mathcal{C}$, $\psi(t)$ denote the geodesic connecting q_1 and q_2 and $T_{q_1}(\mathcal{C})$ is the tangent space to $\mathcal{C}$ on q_1.

The exponential map $(exp : T_{q_1}(\mathcal{C}) \to \mathcal{C})$ and its inverse $(exp^{-1} : \mathcal{C} \to T_{q_1}(\mathcal{C}))$ are also given with explicit formulas (similarly to the Kendall's shape space of landmarks configurations, presented in Section 2.3) and could be easily extended to the shape space that will be presented in the next section.

2.4.3 *Shape Space* as Quotient Space and Elastic Metric

To study the *shapes* of parametrized curves in a rotation and reparametrization-invariant fashion, one should identify all rotations and re-parameterizations of a curve as an equivalence class which is,

$$[q] = \text{closure}\{\sqrt{\dot{\gamma}(t)}Oq(\gamma(t))|O \in SO(3), \quad \gamma \in \Gamma\}. \tag{2.10}$$

Here Γ denote the group of orientation-preserving diffeomorphisms of I to itself (representing reparametrizations), and $SO(3)$ denote the group of orthogonal 3×3 real matrices with determinant $+1$ (representing rotations in $\mathbb{R}^3$). The set of such equivalence classes, denoted by $\mathcal{S} \doteq \{[q]|q \in \mathcal{C}\}$ is the *shape space* of SRV functions defined in $\mathbb{R}^3$. $\mathcal{S}$ is a metric space with the metric inherited from the larger space $\mathcal{C}$. To obtain geodesics and geodesic distances between elements of $\mathcal{S}$, one needs to solve the optimization problem:

$$(O^*, \gamma^*) = \underset{(O,\gamma) \in SO(3) \times \Gamma}{\text{argmin}} d_{\mathcal{C}}(q_1, \sqrt{\dot{\gamma}}O(q_2 \circ \gamma)). \tag{2.11}$$

While for a fixed rotation $O \in SO(3)$, the optimization over Γ is done using the Dynamic Programming (DP) algorithm (Bellman, 1957), for a fixed $\gamma \in \Gamma$, the optimization over $SO(3)$ is performed using SVD (Singular Value Decomposition). By iterating between these two steps, we can reach a solution for the

joint optimization problem. Let $q_2^*(t) = \sqrt{\dot{\gamma}^*(t)}O^*q_2(\gamma^*(t)))$ be the optimal element of $[q_2]$, associated with the optimal rotation O^* and the best re-parameterization γ^* of the second curve, then the geodesic distance between $[q_1]$ and $[q_2]$ in $\mathcal{S}$ is given by (Eqn. 2.12):

$$d_{\mathcal{S}}([q_1], [q_2]) \doteq d_{\mathcal{C}}(q_1, q_2^*) \tag{2.12}$$

The geodesic path connecting $[q_1]$ and $[q_2]$ elements of $\mathcal{S}$ is given by (Eqn. 2.8), with a difference that q_2^* replace q_2. Invariance to rotation and reparametrization is achieved by modding out the actions of $SO(3)$ and the group Γ of orientation-preserving diffeomorphisms. Both of these group actions are isometric with respect to the $\mathbb{L}^2$ metric, which gives rise to a distance function on the space of orbits. Let us recall how the actions of the groups Γ and $SO(3)$ (and $SO(3) \times \Gamma$) act on the preshape space. Details and proofs are given in (Srivastava et al., 2011) and (Robinson, 2012), we briefly recall some of them here for convenience.

- **Action of $SO(3)$ on $\mathcal{C}$** – the rotation group $SO(3)$ acts from the left on $\mathcal{C}$ by multiplication. This action is isometric; for any $q_1, q_2 \in \mathcal{C}$ and any $O \in SO(3)$, we have

$$
\begin{aligned}
\langle Oq_1, Oq_2 \rangle &= \int_I \langle Oq_1(t), Oq_2(t) \rangle \, dt \tag{2.13} \\
&= \int_I \langle q_1(t), q_2(t) \rangle \, dt \tag{2.14} \\
&= \langle q_1, q_2 \rangle
\end{aligned}
$$

The second equality comes from the fact that $SO(3)$ acts by isometry on $\mathbb{R}^3$. The following equality is trivial: $\|Oq_1 - Oq_2\| = \|q_1 - q_2\|$.

- **Action of Γ on $\mathcal{C}$** – The reparametrization group Γ acts on $\mathcal{C}$ from the right as follows: for $\gamma \in \Gamma$ and $q \in C$,

$$(q, \gamma) = \sqrt{\dot{\gamma}}(q \circ \gamma) \tag{2.15}$$

The action is defined in this way so that acting on an SRVF is equivalent to reparametrizing the corresponding curve. Now, to see that Γ acts by isometries, let $q_1, q_2 \in \mathcal{C}$, and let $\gamma \in \Gamma$. Then

$$\langle (q_1, \gamma), (q_2, \gamma) \rangle = \int_I \left\langle \sqrt{\dot\gamma} q_1(\gamma(t)), \sqrt{\dot\gamma} q_2(\gamma(t) \right\rangle dt \qquad (2.16)$$

$$= \int_I \langle q_1(\gamma(t)), q_2(\gamma(t) \rangle \sqrt{\dot\gamma} \, dt \qquad (2.17)$$

$$= \int_I \langle q_1(t), q_2(t) \rangle \, dt$$

$$= \langle q_1, q_2 \rangle$$

Here also, we can derive the following equality $\|(q_1, \gamma) - (q_2, \gamma)| = \|q_1 - q_2\|$. Another useful fact is that these two actions commute: reparametrizing and then rotating gives the same result as rotating and then reparametrizing. Let $q \in \mathcal{C}$, $O \in SO(3)$, and $\gamma \in \Gamma$, then,

$$O(q, \gamma) = O(\sqrt{\dot\gamma}(q \circ \gamma)) = \sqrt{\dot\gamma}((Oq \circ \gamma)) = (Oq, \gamma) \qquad (2.18)$$

If q_1 and q_2 are SRVFs of two curves β_1 and β_2, respectively, then it is easy to show that under the $\mathbb{L}^2$ norm, $\|q_1 - q_2\| = \|\sqrt{\dot\gamma}(q_1 \circ \gamma) - \sqrt{\dot\gamma}(q_2 \circ \gamma)\|$, for all $\gamma \in \Gamma$, while $\|\beta_1 - \beta_2\| \neq \|(\beta_1 \circ \gamma) - (\beta_2 \circ \gamma)\|$, in general. This is one important reason why SRVF q is a suitable representation of curves than β for shape analysis.
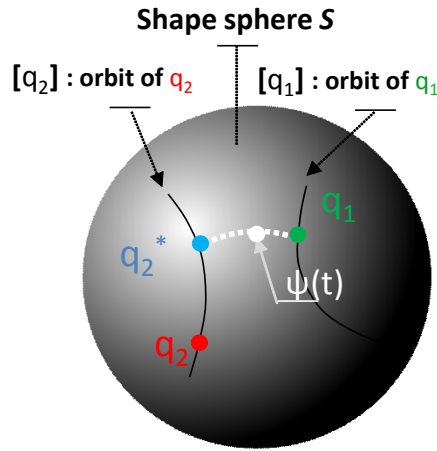

FIGURE 2.6: Shape space $\mathcal{S} = \mathcal{C}/SO(3) \times \Gamma$ with $[q_1]$ and $[q_2]$ elements of $\mathcal{S}$ (orbits of shapes on $\mathcal{C}$). $q_2^* = \sqrt{\dot{\gamma^*}} O^* q_2 \circ \gamma^*$ where $O^* \in SO(3)$, $\gamma^* \in \Gamma$ are the optimal rotation and the best reparametrization, respectively. $\psi(t)$ denote a geodesic connecting $[q_1]$ and $[q_2]$.

- **The Shape Space $\mathcal{S}$** – The feature space of shape representations will be a sort of quotient of $\mathcal{C}$ under the actions of $SO(3)$ and Γ. We then define the shape space as the set of all such closed-up orbits (as defined previously in Eqn. 2.10) by $\mathcal{S} = \{[q] | q \in \mathcal{C}\} = \mathcal{C}/SO(3) \times \Gamma$. Shown in Figure 2.6 a pictorial illustrating the space space $\mathcal{S}$, two orbits $[q_1]$ and $[q_2]$, a geodesic connecting them ψ.

- **Metric on the Shape Space** $\mathcal{S}$ – To define a distance between two closed-up orbits $[q_1]$, $[q_2] \in \mathcal{S}$, one simply take the smallest distance between any pair of representatives. Formally, let $[q_1]$, $[q_2] \in \mathcal{S}$, the metric $d_\mathcal{S}$ is given by Eqn. 2.19:

$$
\begin{aligned}
d_\mathcal{S}([q_1],[q_2]) &= \underset{v\in[q_1],w\in[q_2]}{\arginf} d_\mathcal{C}(v,w) \\
&= \underset{O_1,O_2\in SO(3),\gamma_1,\gamma_2\in\Gamma}{\arginf} d_\mathcal{C}(O_1(q_1,\gamma_1),O_1(q_2,\gamma_2)) \\
&= \underset{O\in SO(3),\gamma\in\Gamma}{\arginf} d_\mathcal{C}(q_1,O(q_2,\gamma))
\end{aligned}
\tag{2.19}
$$

The last equality follows from the fact that both group actions are isometric. Recall that our elastic matching problem is that of finding a pair $(\hat{\gamma},\hat{O})$ subject to,

$$
(\hat{\gamma},\hat{O}) = \underset{\gamma\in\Gamma,O\in SO(3)}{\arginf} \int_I ||q_1(t) - O\sqrt{\dot{\gamma}(t)}q_2(\gamma(t))||^2 dt
\tag{2.20}
$$

- **Optimal Matching** – Find an optimal re-parametrization between q functions, when scaling, translation and rotations are filtered out, can be viewed as an optimal matching between the corresponding parametrized curves β throw their SRV functions. In other worlds, the problem of matching β_1 and β_2 is turned out to find an optimal re-parametrization between $[q_1]$ and $[q_2]$, respectively their SRVF which will realize the minimum distance between the orbits.

(a) Source   (b) Registration result   (c) Target

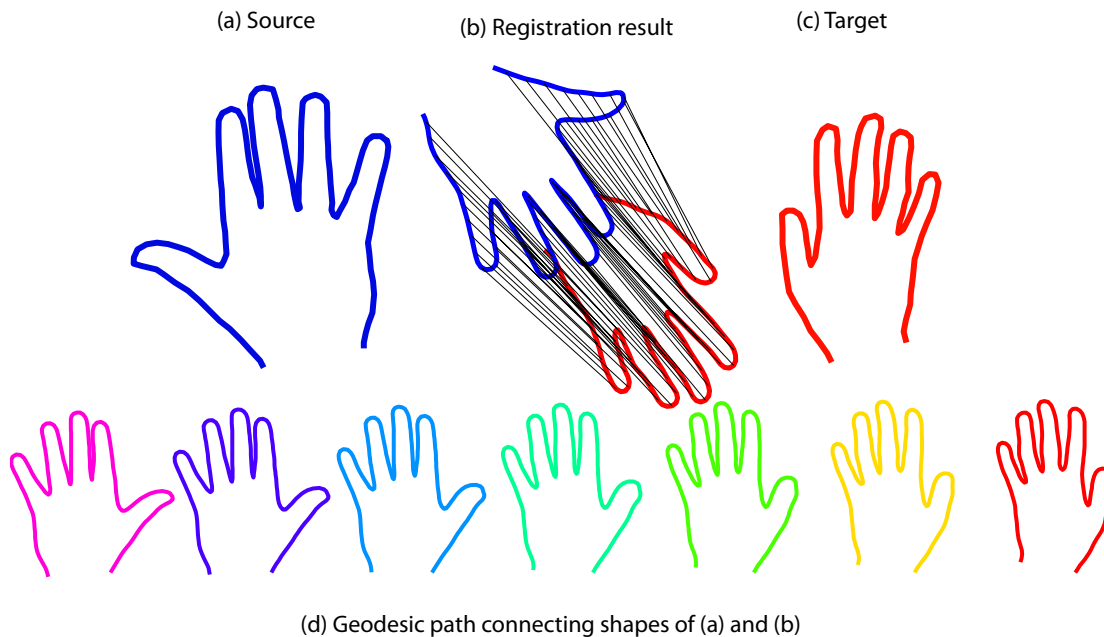(d) Geodesic path connecting shapes of (a) and (b)

FIGURE 2.7: An example of optimal matching of two parametrized curves of hands under deformations. One needs a combination of bending and stretching to match anatomical parts as the tips of the fingers and the valleys between them.

To illustrate the key idea behind the theoretical framework, presented above, for shape analysis of parametrized curves (in $\mathbb{R}^2$), we consider in figure 2.7, two hands boundaries ((a) and (c)) extracted from images. Panel (b) illustrates a dense registration of their shapes, and panel (d) draws a geodesic path connecting them. One should note that this result is important as it allows to (jointly) compute the optimal registration and the distance between shapes.

As stated previously, shape analysis of continuous curves is now rich and one can develop techniques which achieve invariance to the standard transformations (rigid and re-parametrization). In many other applications involving 3D imaging technologies, an outstanding need is observable for developing suitable tools and computational strategies for 3D shape interpretation. Nowadays, this need is growing in computer vision due to the huge advancements in cost-effective 3D sensors. In stead of curves, the output data is three-dimensional and are represented as depth-maps, point clouds or 3D meshes. In the following section, I shall describe our contributions to 3D face shape analysis through a multiple curves approximation. Thus our contributions could be viewed as a transit step to 3D shape analysis which provide a comprehensive framework to compute geodesics and distances, achieve accurate registration between facial surfaces, develop statistical models for 3D face shape completion, etc.

2.5 Shape Analysis of Facial Surfaces

Understanding human faces by computers is still an active research area and find its application in many applications ranging from face detection to emotion interpretation. Classical approaches are based on the face appearance extracted from 2D still images (and videos). As 2D images are the result of the camera projection of the scene, the 3D geometry of the face is lost. Few works emphasize on the role of 2D face geometry based on the Active Appearance Model or the Active Shape Models. With the advancement in 3D imaging systems (mainly optical), there has been an increasing effort (since 2005-2006) to 3D faces. This has been marked by the release of the Face Recognition Grand Challenge dataset (phi) accompanied with a well-defined evaluation protocol, and the publication of first results using this dataset and other in-house datasets (Heseltine et al., 2004)(Lu et al., 2004)(Bronstein et al., 2005c)(Ben Amor et al., 2006a)(Samir et al., 2006)(Berretti et al., 2006)(Bowyer et al., 2006a) [3]. The aim of the research that we present here is to introduce relevant tools and computational algorithms for the purpose of *shape analysis* of facial surfaces (both static, 3D, and dynamic, 4D) with applications to face recognition, facial expression recognition and soft-biometrics estimation. In this section, I shall discuss the main challenges of facial shape analysis and introduce many facets of our contributions to handle these challenges. The details related to each of the above mentioned applications will be discussed in the next chapters. To do so, I shall refer to the mathematical background presented in the previous sections to describe the proposed Riemannian framework for 3D face analysis and its relevant geometric tools used in our target applications.

---

[3]We refer the reader to the next three chapters for a more complete state-of-the-art in each application

2.5.1 Motivation for Elastic Radial Curves

The changes in facial expressions affect different regions of a facial surface differently. For example, during a smile, the top half of the face is relatively unchanged while the lip area changes a lot, and when a person is surprised the effect is often the opposite. If chosen appropriately, curves have the potential to capture regional shapes and that is why their role becomes important. The locality of shapes represented by facial curves is an important reason for their selection. The next question is: Which facial curves are suitable for recognizing people? Curves on a surface can, in general, be defined either as the level curves of a function or as the streamlines of a gradient field. Ideally, one would like curves that maximally separate inter-class variability from the intra-class variability (typically due to expression changes). The past usage of the level curves (Samir et al., 2006, 2009) (of the surface distance function) has the limitation that each curve goes through different facial regions and that makes it difficult to isolate local variability.

FIGURE 2.8: A smile (see middle) changes the shapes of the curves in the lower part of a the face while the act of surprise changes shapes of curves in the upper part of the face.

In contrast, the radial curves with the nose tip as origin, that we propose here, have a tremendous potential. This is because: **(i)** the nose is in many ways the focal point of a face. It is relatively easy and efficient to detect the nose tip (compared to other facial parts) and to extract radial curves, with nose tip as the center, in a completely automated fashion. It is much more difficult to automatically extract other types of curves, e.g. those used by sketch artists (cheek contours, forehead profiles, eye boundaries, etc); **(ii)** Different radial curves pass through different regions and, hence, can be associated with different facial expressions. For instance, differences in the shapes of radial curves in the upper-half of the face can be loosely attributed to the inter-class variability while those for curves passing through the lips and cheeks can largely be due to changes in expressions. This is illustrated in Fig. 2.8 which shows a neutral face (left), a smiling face (middle), and a surprised face (right). The main difference in the middle face, relative to the left face, lies in the lower part of the face, while for the right face the main differences lie in the top half; **(iii)** Radial curves have a more universal applicability. The curves used in the past have worked well for some specific tasks, e.g., lip contours in detecting certain expressions, but they have not been as efficient for some other tasks, such as face recognition. In contrast, radial curves capture the full geometry and are applicable to a variety of applications, including facial expression recognition; **(iv)** In the case of the missing parts and partial occlusion, at least some part of every radial curve is usually

available. It is rare to miss a full radial curve. In contrast, it is more common to miss an eye due to occlusion by glasses, the forehead due to hair, or parts of cheeks due to a bad angle for laser reflection. This issue is important in handling the missing data via reconstruction, as shall be described in chapter 3; **(v)** Natural face deformations are largely (although not exactly) symmetric and, to a limited extent, are radial around the nose. Based on these arguments, we present here a novel geometrical representation of facial surfaces using radial curves that start from the nose tip.
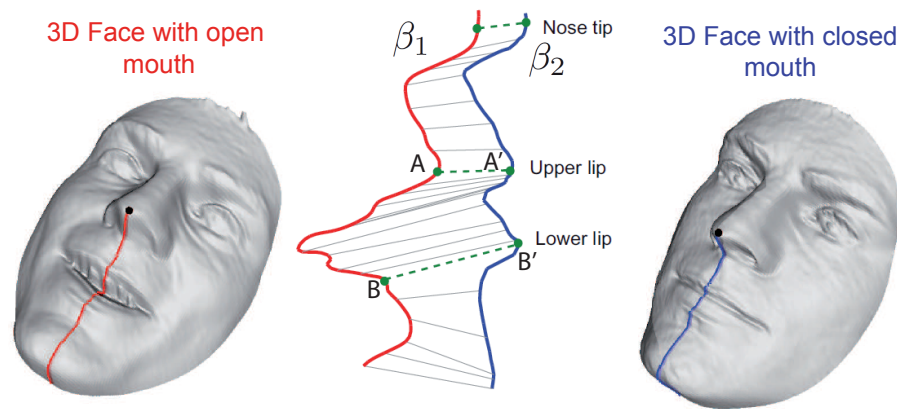

FIGURE 2.9: An example of matching radial curves extracted from two 3D faces of the same person: a curve with an open mouth (on the left) and a curve with a closed mouth (on the right). One needs a combination of stretching and shrinking to match similar points (upper lips, lower lips, etc)

Consider the two parameterized curves shown in Fig. 2.9; call them β_1 and β_2. Our task is to automatically match points on these radial curves associated with two different facial expressions. The expression on the left has the mouth open whereas the expression on the right has the mouth closed. In order to compare their shapes, we need to register points across those curves. One would like the correspondence to be such that geometric features match across the curves as well as possible. In other words, the lips should match the lips and the chin should match the chin. Clearly, if we force an arc-length parameterization and match points that are at the same distance from the starting point, then the resulting matching will not be optimal. The points A and B on β_1 will not match the points A' and B' on β_2 as they are not placed at the same distances along the curves. For curves, the problem of optimal registration is actually the same as that of optimal re-parameterization. This means that we need to find a re-parameterization function $\gamma(t)$ such that the point $\beta_1(t)$ is registered with the point $\beta_2(\gamma(t))$, for all t. The question is how to find an optimal γ for an arbitrary β_1 and β_2? Keep in mind that the space of all such γ is infinite dimensional because it is a space of functions. As described in section 2.4, this registration is accomplished by solving an optimizing problem using the dynamic programming algorithm, but with an objective function that is developed from a Riemannian metric. The chosen metric, termed an **elastic metric**, has a special property that the same re-parameterization of two curves does not change the distance between them. This, in turn, enables us to fix the parameterization of one curve arbitrarily and to optimize over the parameterization of the other. This optimization leads to a proper distance

(geodesic distance) and an optimal deformation (geodesic) between the shapes of curves. In other words, it results in their elastic comparisons (refer to section 2.4 for details).

2.5.2 Mathematical Representation of Facial Shape

A facial surface S will be represented by a collection of radial curves that are defined as follows. Although S is a triangulated mesh, we start the discussion by assuming that it is a continuous surface. Let $\{\beta^{(\alpha)}\}^{\alpha \in [0,2\pi]}$ denote the collection of radial curves on S which makes angles α with a reference radial curve. The reference curve on S is chosen to be the vertical curve after the face has been rotated to the upright position ($\alpha = 0$). Figure 2.10 illustrates the 3D facial surface approximation by collections of radial curves and their pairwise correspondence.
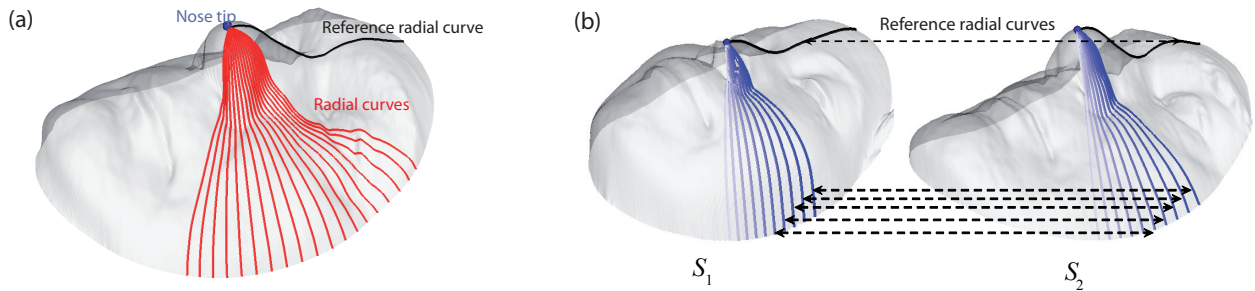

FIGURE 2.10: (a) 3D face approximation by a collection of radial curves; (b) pairwise correspondence of the facial curves.

We will use the elastic shape analysis of parametrized curves described in section 2.4 because it is particularly appropriate in our shape analysis of facial surfaces. This is because (1) such analysis uses the Square-Root Velocity Function (Joshi et al., 2007) representation which allows us to compare local facial shapes in presence of elastic deformations, (2) this method uses a representation under which the elastic metric (Mio et al., 2007) reduces to the standard $\mathbb{L}^2$ metric and thus simplifies the analysis, (3) under this metric the Riemannian distance between curves is invariant to the re-parametrization. To analyze the shape of $S = \{\beta^{(\alpha)}\}^{[0,2\pi]}$, we shall represent each curves $\beta^{(\alpha)}$ mathematically using the square-root representation (presented in section 2.4) such as $q^{(\alpha)}(t) \doteq \dot{\beta}^{(\alpha)}(t)/\sqrt{|\dot{\beta}^{(\alpha)}(t)|}$). Here t is a parameter $\in I$ and $|\cdot|$ is the standard Euclidean norm in $\mathbb{R}^3$. Following the over-mentioned shape representation $\mathcal{F} = \{q^{(\alpha)}\}^{[0,2\pi]}$, facial shapes could be viewed as elements of the Riemannian manifold $\mathcal{M} \doteq \mathcal{S}^{[0,2\pi]}$. Hence, all the geometric tools defined on $\mathcal{S}$ are straightforward in $\mathcal{M}$, as we will describe later.

2.5.3 Joint Shape Registration and Comparison

- **Dense Shape Registration** – The difficulty in analyzing shapes of objects comes from the fact that: (1) Shape representations, metrics, and models should be invariant to certain transformations

that are termed *shape preserving.* For instance, rigid motions and re-parameterizations of facial surfaces do not change their shapes, and any shape analysis of faces should be invariant to these transformations. (2) Registration of points across objects is an important ingredient in shape analysis. Specifically, in comparing shapes of faces, it makes sense that similar biological parts are registered to each other across different faces. Furthermore, it is important to use techniques that allow a joint registration and comparisons of surfaces in a comprehensive framework, rather than in two separate steps. These two issues— invariance and registration—are naturally handled using Riemannian methods where one can choose metrics that are invariant to certain transformations and form quotient spaces (termed shape spaces) by forming equivalence classes of objects that have the same shape. The elastic Riemannian metric used in this work provides a nice physical interpretation of measuring deformations between facial surfaces, through the radial curves representation, using a combination of stretching and bending. Figure 2.11 gives four examples of facial curves matching under facial expression variations using the elastic model.
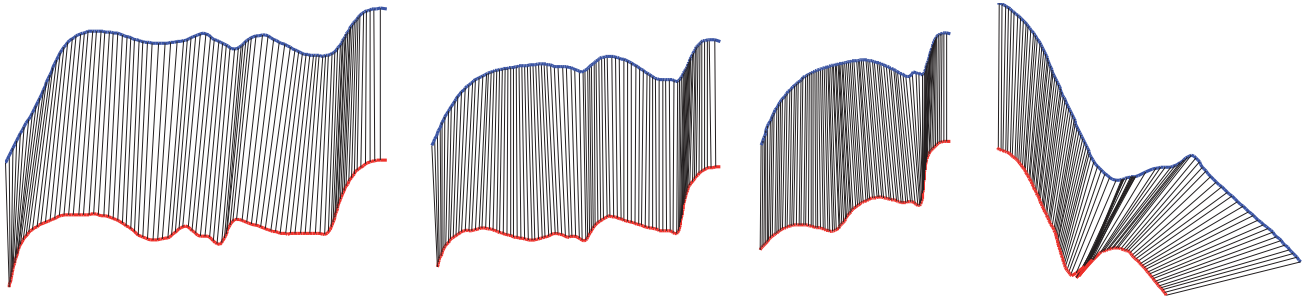

FIGURE 2.11: Four examples of facial curves matching results using the elastic Riemannian method.

Figure 2.12 provides an example of facial shapes registration through the radial curves representation. For better visibility, four different resolutions of the registration result are shown.

- **Metric on the Manifold** $\mathcal{M}$ – To define a distance between two facial shapes $\mathcal{F}_1$, $\mathcal{F}_2 \in \mathcal{M}$, one simply integrate over all α the distance between pairwise-corresponded SRVFs of the curves on the faces $\mathcal{F}_1$ and $\mathcal{F}_2$, given by :

$$d_{\mathcal{M}}(\mathcal{F}_1, \mathcal{F}_2) = \int_0^{2\pi} d_{\mathcal{S}}([q_1^{(\alpha)}], [q_2^{(\alpha)}]) \ d\alpha \tag{2.21}$$

where $\mathcal{F}_i = \{[q_i^{(\alpha)}]^{\alpha \in [0,2\pi]}\}$, the set of SRV functions on the face when varying the parameter α.

2.5.4 Geodesics and Shapes Interpolation

Since we have deformations (geodesic paths) between corresponding curves, we can combine these deformations to obtain deformations between full facial surfaces. In fact, these full deformations can be shown to be formal geodesic paths between faces, when represented as elements of $\mathcal{M} \doteq \mathcal{S}^{[0,2\pi]}$. According to the *Theorem* provided in (Samir et al., 2009) (page 94) adapted to our case, if we are given a path in $\mathcal{M}$
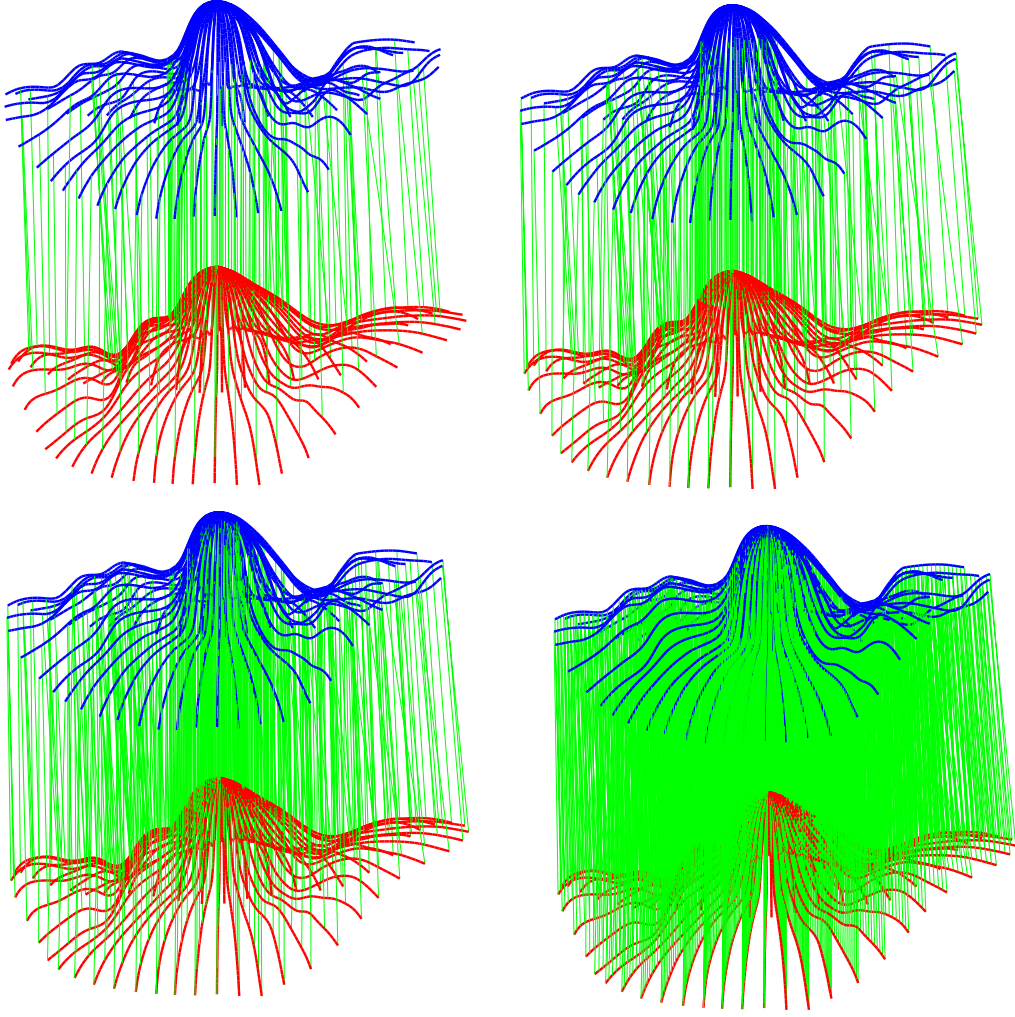
FIGURE 2.12: Results of the optimal matching (dense registrations) of two facial surfaces of same subject with different expressions viewed at different resolutions for better visibility.

represented as $\Psi : [0, 2\pi] \times I \to \mathcal{M}$. For each $\alpha \in [0, 2\pi]$, define $\psi^{(\alpha)} : I \to \mathcal{S}$ by $\psi^{(\alpha)} = \Psi(\alpha, t)$. Then Ψ is a geodesic in $\mathcal{M}$ if $\forall \alpha \in [0, 2\pi]$, $\psi^{(\alpha)}$ is a geodesic in $\mathcal{S}$. Thus, the geodesic path between any two points $\mathcal{F}_1, \mathcal{F}_2 \in \mathcal{M}$ is given by, $\Psi : [0, 1] \to \mathcal{M}$, where

$$\psi^{(\alpha)}(\tau) = \frac{1}{\sin(\theta^{(\alpha)})} \left(\sin((1 - \tau)\theta^{(\alpha)}) q_1^{(\alpha)} + \sin(\theta\tau) q_2^{(\alpha)*} \right) \tag{2.22}$$

and the length of the geodesic is $\theta^{(\alpha)} = d_{\mathcal{C}}(q_1^{(\alpha)}, q_2^{(\alpha)*}) = cos^{-1}(\langle q_1^{(\alpha)}, q_2^{(\alpha)*} \rangle)$. These geodesics provide a tangible benefit, beyond the current algorithms that provide some kind of a similarity score for analyzing faces. In addition to their interpretation as optimal deformations under the chosen metric, the geodesics can also be used for computing the mean shape and measuring the shape covariance of a set of faces, as illustrated later. Figure 2.13 illustrates an example of geodesic path between faces through their radial curves representations viewed under different view angles for better visibility.

The figure 2.14 illustrates the benefit of use of the elastic model (in the *shape space*) over the non-elastic
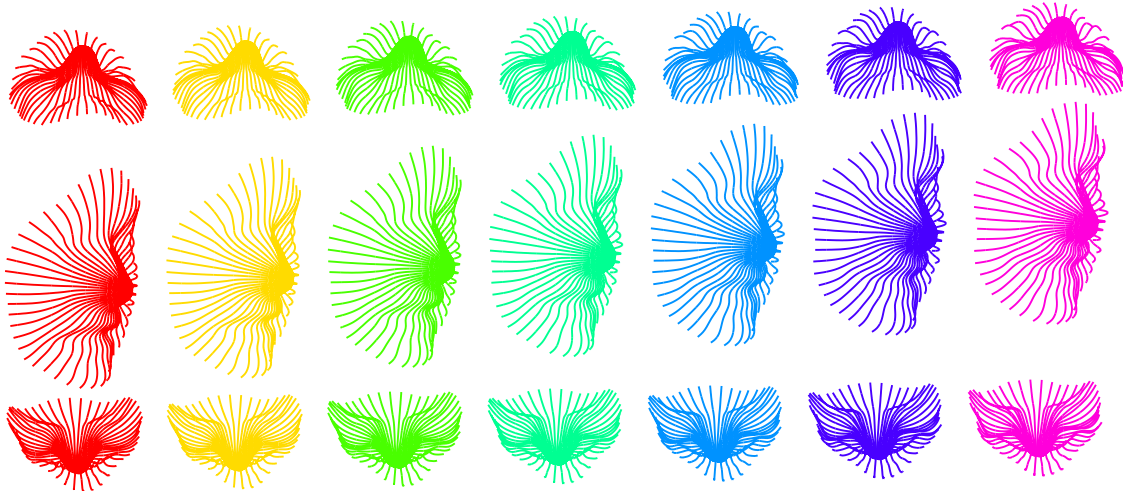
FIGURE 2.13: An example of geodesic path connecting shapes of 3D faces through the radial curves representation. The figure shows three different views of the same geodesic.

model (in the *pre-shape space*) and an Euclidean model (using the ICP algorithm). While, the top panel (a) shows the shape interpolation for an intra-class example (expressive and neutral faces of the same person), the second panel shows an inter-class example (faces of different persons). In both examples, it is clear to note the most natural deformations are exhibited by the intermediate faces (equally spaces points from the geodesic) of the first path performed in the shape space, using the elastic model. The remaining paths in the pre-shape space and in the Euclidean space presents less-natural deformations. Due to the accurate dense correspondence achieved by the elastic model, the first path showed in the panel (b) of this figure represents the most efficient way to deform the source face to the target face when keeping the anatomical parts of the faces in the intermediate faces. The application of these ideas will be shown in Chapter 3 dedicated to 3D face recognition, in which a definition of a metric to compare faces, is a crucial step in designing face recognition algorithms.

2.5.5   Sample Mean of 3D Faces

As mentioned above, an important advantage of our Riemannian approach over many past work on 3D face analysis is its ability to compute summary statistics of a set of faces. For example, one can use the notion of Karcher mean (Karcher, 1977) to define an average face that can serve as a representative face of a group of faces. To calculate a Karcher mean of facial surfaces $\{\mathcal{F}^1, ..., \mathcal{F}^k\}$ in $\mathcal{M}$, we define an objective function:

$$\mathcal{V} : \mathcal{M} \to \mathbb{R}, \mathcal{V}(\mathcal{F}) = \sum_{i=1}^{k} d_{\mathcal{M}}(\mathcal{F}^i, \mathcal{F})^2$$

The Karcher mean is then defined by:

$$\overline{\mathcal{F}} = \arg \min_{\mathcal{F} \in \mathcal{M}} \mathcal{V}(\mathcal{F}).$$

(a) Intra-class geodesic paths
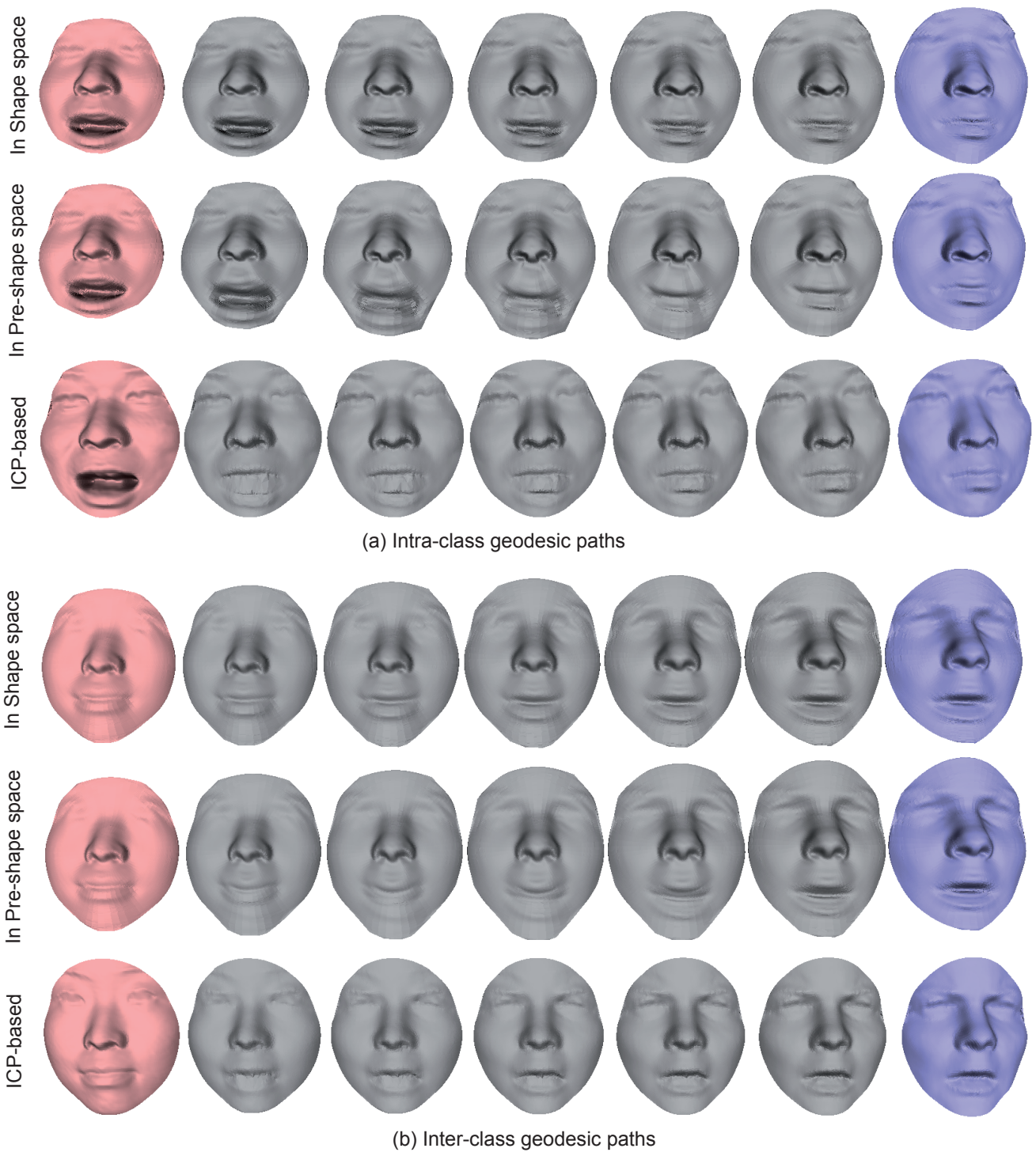

(b) Inter-class geodesic paths

FIGURE 2.14: (a) Intra-class deformations achieved from top-to-bottom in the shape space (using the elastic model); in the preshape space (using an non-elastic model); and in $\mathbb{R}^3$ using the iterative closest point algorithm (Besl and McKay, 1992) to align the faces and to establish a dense correspondence using the closest-point criteria. (b) Inter-class deformations using the three over-mentioned methods.

The algorithm for computing Karcher mean is a standard one, see e.g. (Dryden and Mardia, 1998) and successfully used in (Drira et al., 2009b) to compute the mean of 3D nasal shapes.

(a) Intra-class Karcher Mean
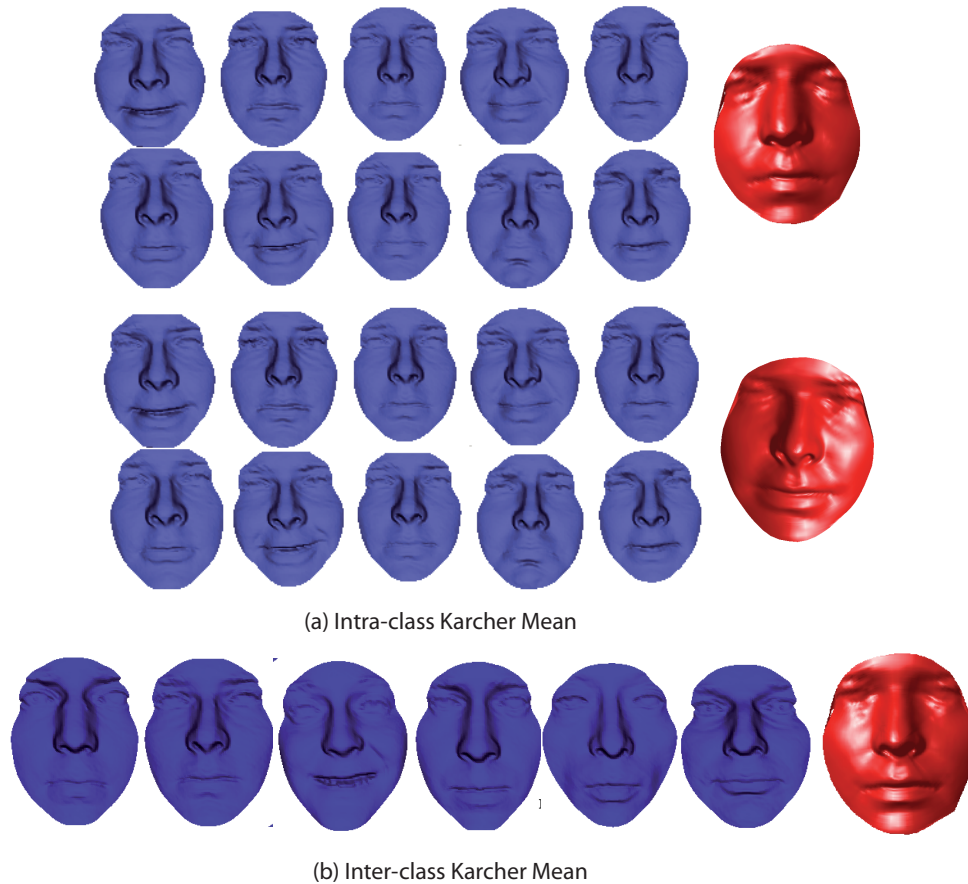
(b) Inter-class Karcher Mean

FIGURE 2.15: Three examples of Karcher means (in red), the first two computes the mean shape of faces within the same class, the third example shows the result for a set shapes comping from different classes.

This minimizer may not be unique and, in practice, one can pick any one of those solutions as the mean face. This mean has a nice geometrical interpretation: $\overline{\mathcal{F}}$ is an element of $\mathcal{M}$ that has the smallest total (squared) deformation from all given facial surfaces $\{\mathcal{F}^1, ..., \mathcal{F}^k\}$. Three examples of Karcher sample means for sets faces are given in figure 2.15. On the left the set of faces to be averaged and on the right (in red) the mean sample shape. When the two first examples illustrate the means of faces taken from a single class of person but with expression variations, the third example average a set of shapes from different classes of different age, different gender and different ethnic group.

We shall use this tool in Chapter 3 in 3D face recognition in two tasks (1) Hierarchical clustering of the gallery for efficient face recognition and (2) design a statistical model for partially-obscured face completion.

2.5.6 Optimal Deformations (Dense Scalar Fields)

In some applications, it is also interesting to capture accurately the deformations exhibited by the 3D faces in order to analyze locally the deformations or to track the facial motions. In order to capture and model the deformations of the face induced by facial movements, we propose to use the same facial

representation by radial curves and to use tools from differential geometry to effectively capture the deformations between two arbitrary faces or along a video of dynamic faces. To this end, we propose to exploit the notion of **shooting vector** along a geodesic to capture the facial deformations between facial shapes, end points of a geodesic on $\mathcal{M}$, and derive our **Dense Scalar Fields** (Drira et al., 2012, Ben Amor et al., 2014a).
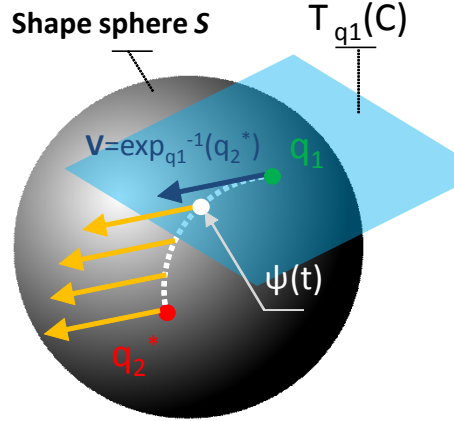

FIGURE 2.16: The tangent vector field (in yellow) along a geodesic ψ and the shooting vector V (arrow in black) element of $T_{q_1}(\mathcal{C})$ at the source point of the same geodesic ψ one $[q_1]$ pointing towards $[q_2]$ (or q_1 pointing towards q_2^*).

More formally, for each SRV representation $q^{(\alpha)}$ of a given original curve $\beta^{(\alpha)}$, the tangent vector field on this geodesic is then written as $\frac{d\psi^{(\alpha)}}{d\tau} : [0, 1] \rightarrow T_{\psi^{(\alpha)}}(\mathcal{S})$, and is obtained by the Eqn. 2.23:

$$\frac{d\psi^{(\alpha)}}{d\tau} = \frac{-\theta}{\sin(\theta)} \left(\cos((1-\tau)\theta)q_1^{(\alpha)} - \cos(\theta\tau)q_2^{(\alpha)*} \right). \tag{2.23}$$

The tangent vector field along a geodesic path ψ connecting q_1 to q_2 is illustrated in Figure 2.16 by yellow arrows. Knowing that on geodesics, the covariant derivative of its tangent vector field is equal to 0, $\frac{d\psi^{(\alpha)}}{d\tau}$ is parallel along the geodesic $\psi^{(\alpha)}$ and one can represent it with $\frac{d\psi^{(\alpha)}}{d\tau}|_{\tau=0}$ without any loss of information. Accordingly, Eqn. (2.23) becomes at the source point q_1 ($\tau = 0$), Eqn. (2.24):

$$\frac{d\psi^{(\alpha)}}{d\tau}|_{\tau=0} = \frac{\theta^{(\alpha)}}{\sin(\theta^{(\alpha)})} \left(q_2^{(\alpha)*} - \cos(\theta^{(\alpha)})q_1^{(\alpha)} \right) \ (\theta^{(\alpha)} \neq 0). \tag{2.24}$$

Figure 2.16 illustrates the idea to map the two radial curves on the Hyper-sphere $\mathcal{C}$ in the Hilbert space through their SRVFs q_1 and q_2, and shows the geodesic path connecting these two points ψ. The tangent vectors of this geodesic path represent a vector field whose covariant derivative is zero. According to this, $\frac{d\psi^*}{d\tau}|_{\tau=0}$ becomes sufficient to represent this vector field, with the remaining vectors obtained by parallel transport of $\frac{d\psi^*}{d\tau}|_{\tau=0}$ along the geodesic ψ^*. In this way, the vector $\frac{d\psi^*}{d\tau}|_{\tau=0}$ represents the dense deformation field between two given SRVF q_α^1 and q_α^2. Based on the above representation, we define

the **Dense Scalar Fields** capable to capture deformations between two 3D faces through their radial multiple radial curves representation and using the elastic model for shape analysis.
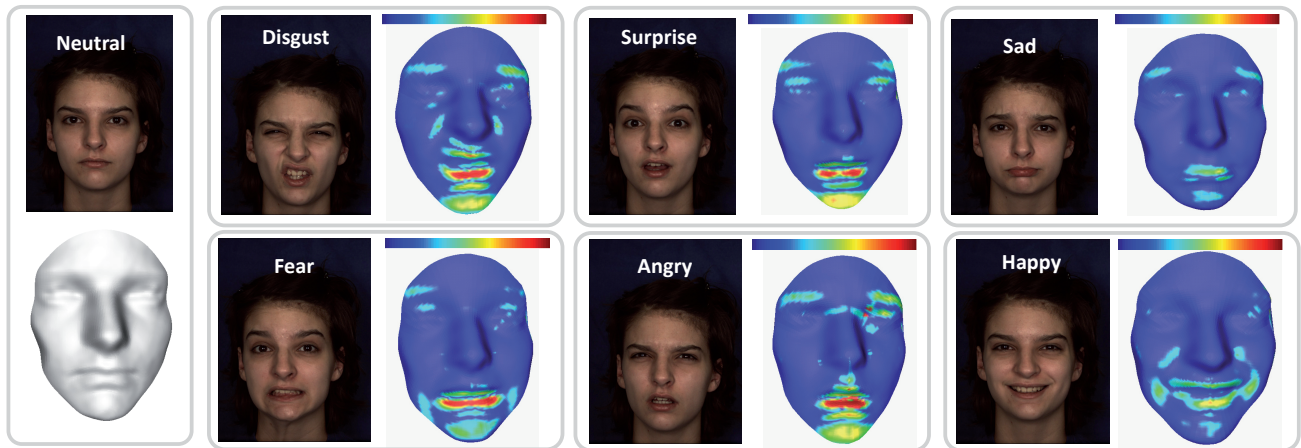

FIGURE 2.17: Deformation Scalar Fields computed between a neutral face of a given subject and the apex frames of the sequences of the six prototypical expressions of the same subject. The neutral scan is shown on the left. Corresponding texture images are also illustrated with each DSFs colormap.

A graphical interpretation of this mathematical representation is given in Figure 2.17. This figure provides the Dense Scalar Fields computed between a neutral face and expressive faces of the same person exhibiting the six universal facial expressions. The amount of deformation between them can be appreciated in this figure, as each facial expression shows different pattern of deformations. This representation will be used in Chapter 4 for facial expression classification from dynamic flows of 3D faces.

2.6 Conclusion

In this chapter, we have provided essential theoretical aspects and related literature to our research. We have started by the Kandall's shape analysis framework and we have presented Differential geometry and Group theory tools to remove shape-preserving transformations (scaling, location, rotation and re-parameterization). Although, the Kendall's approach is applicable only on shapes with registered landmarks representation, it provides a comprehensive framework. When moving to continuous (non-registered) shapes, an additional challenge appears, which is the re-parametrisation. Recent methods, inspired from the seminal Granender's shape theory, view the set of shape descriptors (the shape space) as quotient of the pre-shape space, obtained by modding out shape-preserving transformations. Thus, viewing smooth changes in shapes as action of diffeomorphisms (or the group of) is the key idea to to measure similarity between shapes in terms of the amount of stretching and bending necessary to turn one shape into another. This latter idea presents the foundations of our research on shape analysis and comparison of 3D facial shapes. Find jointly an accurate facial surfaces registration trough representation

by collections of elastic radial curves and return a distance between shapes in terms of the amount of stretching and bending is one of main contributions of our research.

Chapter 3

3D Face Recognition

In this chapter[1] I shall describe our contributions to 3D shape analysis with application to a fundamental problem in computer vision, **Face Recognition**. In fact, the head pose variations and the presence of external occlusions which lead to unregistered and obscured data, as well as the intra-class variations due to facial expressions, cause enormous variability, making the problem highly challenging. Throughout this part of the *habilitation*, I shall describe the application of our comprehensive and unified computational Riemannian framework (presented in Chapter 2) and its ability to cover many facets of 3D shape analysis where the robustness to the aforementioned variations is our central goal. We propose a mathematical representation of 3D facial shapes through collections of curves and use elastic shape analysis to develop a Riemannian framework for analyzing their full shapes. Thus, shapes of 3D faces are viewed as points on some Riemannian manifolds, and one should define appropriate metrics for shape interpolation, distance measurement, and so on. I have benefited to conduct this research from many exemplary efforts and development of elegant and mature methodologies of shape analysis of parametrized curves developed so far[2]. From this point of view, my research is indebted to some of these seminal previous works (cf. (Dryden and Mardia, 1998) or the Kendall's school, (Klassen et al., 2004), (Srivastava et al., 2011), and (Samir et al., 2006, 2009)).

I shall emphasize, in particular, on how our facial surfaces representation, along with the defined elastic Riemannian metric, seems natural for performing, jointly, accurate dense registration (or matching) and comparison of 3D facial surfaces. Both the robustness (invariance) to variations in shape comparison and registration are naturally handled using Riemannian methods where the proposed metric is invariant to certain transformations and form quotient spaces (termed shape spaces) by forming equivalence classes of faces that share the same shape under shape-preserving transformations. We have adopted an elastic Riemannian metric which provides a nice physical interpretation of measuring deformations between facial curves using a combination of **stretching and bending** (Drira et al., 2013a, 2010b). This latter property

---

[1]The content of this chapter is based on our published papers (Ballihi et al., 2012b), (Drira et al., 2013a) and (Drira et al., 2009b).

[2]Refer to the previous chapter for an historical review of this school of shape analysis.

makes the registration of facial surfaces accurate by matching corresponding anatomical features on the facial surfaces. We have targeted the 3D face recognition problem as first application of the theoretical and computational framework (Drira et al., 2013a, 2010b), for which interpolation between 3D shapes through accurate registration, averaging a set of 3D shapes and completion of partially obscured shape are possible. It is shown to be promising from both – empirical and theoretical – perspectives. In terms of the empirical evaluation, our results match or improve the state-of-the-art methods on several prominent databases, each posing a different type of challenge. From a theoretical perspective, this framework allows for formal statistical inferences, such as the estimation of missing facial parts using PCA on tangent spaces and average shapes computation. Because of the importance of efficiency when designing face recognition solutions, we have also introduced two solutions, (1) a feature selection step using Adaboost algorithm, to select and highlight salient geometrical facial features that contribute most in 3D face recognition (Ballihi et al., 2012a,b)and (2) an hierarchical organization, a tree-like Gallery organization of 3D faces. This latter imply comparisons performed only at the nodes of the tree. It has been explored on 3D facial shapes (Drira et al., 2010c) and 3D nasal shapes (Drira et al., 2009b) for partial biometrics. To construct such a shape tree one need to be able to cluster similar shapes.

3.1 Motivation and Related Work

Due to the natural, non-intrusive, and high throughput nature of face data acquisition, automatic face recognition has many benefits when compared to other biometrics. Accordingly, automated face recognition has received a growing attention within the computer vision community over the past three decades. Among different modalities available for face imaging, 3D scanning has a major advantage over 2D color imaging in that nuisance variables, such as illumination and small pose changes, have a relatively smaller influence on the observations. However, 3D scans often suffer from the problem of missing parts due to self occlusions or external occlusions, or some imperfections in the scanning technology. Additionally, variations in face scans due to changes in facial expressions can also degrade face recognition performance. In order to be useful in real-world applications, a 3D face recognition approach should be able to handle these challenges, i.e., it should recognize people despite large facial expressions, occlusions and large pose variations.

Some examples of face scans highlighting these issues are illustrated in Fig. 3.1. We note that most recent research on 3D face analysis has been directed towards tackling changes in facial expressions while only a relatively modest effort has been spent on handling occlusions and missing parts. Although a few approaches and corresponding results dealing with missing parts have been presented, none, to our knowledge, has been applied systematically to a full real database containing scans with missing parts. In this chapter, we adopt the Riemannian framework introduced in Chapter 2, in the process dealing with large expressions, occlusions and missing parts. Additionally, we provide some basic tools for statistical

FIGURE 3.1: Challenges of 3D face recognition: expressions, obscured data, pose variations (self-occlusion) and external occlusions.

shape analysis of facial surfaces. These tools help us to compute a typical or average shape and measure the intra-class variability of shapes, and will even lead to face atlases in the future.

The task of recognizing 3D face scans has been approached in many ways, leading to varying levels of successes. We refer the reader to one of many extensive surveys on the topic, e.g. see (Bowyer et al., 2006b). Below we summarize a smaller subset that is more relevant to our work. **Deformable template-based approaches** – There have been several approaches in recent years that rely on deforming facial surfaces from one into another, under some chosen criteria, and use quantifications of these deformations as metrics for face recognition. Among these, the ones using non-linear deformations facilitate the local stretching, compression, and bending of surfaces to match each other and are referred to as **elastic** methods. For instance, Kakadiaris et al. (Kakadiaris et al., 2007) utilize an *annotated face model* to study geometrical variability across faces. The annotated face model is deformed elastically to fit each face, thus matching different anatomical areas such as the nose, eyes and mouth. In (Passalis et al., 2011), Passalis et al. use automatic landmarking to estimate the pose and to detect occluded areas. The facial symmetry is used to overcome the challenges of missing data here. Similar approaches, but using manually annotated models, are presented in (ter Haar and Velkamp, 2010, Lu and Jain, 2008). For example, (Lu and Jain, 2008) uses manual landmarks to develop a thin-plate-spline based matching of facial surfaces. A strong limitation of these approaches is that the extraction of fiducial landmarks needed during learning is either manual or semi-automated, except in (Kakadiaris et al., 2007) where it is fully automated. **Local regions and features approaches** – Another common framework, especially for handling expression variability, is based on matching only parts or regions rather than matching full faces. Lee et al. (Lee et al., 2005) use ratios of distances and angles between eight fiducial points, followed by a SVM classifier. Similarly, Gupta et al. (Gupta et al., 2007) use Euclidean/geodesic distances between anthropometric fiducial points, in conjunction with linear classifiers. As stated earlier, the problem of automated detection of fiducial points is non-trivial and hinders automation of these methods. Gordon (Gordan, 1992) argues that curvature descriptors have the potential for higher accuracy in describing surface features and are better suited to describe the properties of faces in areas such as the cheeks, forehead, and chin. These descriptors are also invariant to viewing angles. Li et al. (Li et al., 2009) design a feature pooling and ranking scheme in order to collect various types of low-level geometric features, such as curvatures, and rank them according to their sensitivity to facial expressions. Along

similar lines, Wang et al. (Wang et al., 2010) use a signed shape-difference map between two aligned 3D faces as an intermediate representation for shape comparison. McKeon and Russ (McKeon and Russ, 2010) use a region ensemble approach that is based on Fisherfaces, i.e., face representations are learned using Fisher's discriminant analysis. In (Huang et al., 2010), Huang et al. use a multi-scale Local Binary Pattern (LBP) for a 3D face jointly with shape index. Similarly, Moorthy et al. (Moorthy et al., 2010) use Gabor features around automatically detected fiducial points. To avoid passing over deformable parts of faces encompassing discriminative information, Faltemier et al. (Faltemier et al., 2008) use 38 face regions that densely cover the face, and fuse scores and decisions after performing ICP on each region. A similar idea is proposed in (Spreeuwers, 2011) that uses PCA-LDA for feature extraction, treating the likelihood ratio as a matching score and using the majority voting for face identification. Queirolo et al. (Queirolo et al., 2010) use Surface Inter-penetration Measure (SIM) as a similarity measure to match two face images. The authentication score is obtained by combining the SIM values corresponding to the matching of four different face regions: circular and elliptical areas around the nose, forehead, and the entire face region. In (Alyuz et al., 2008), the authors use Average Region Models (ARMs) locally to handle the challenges of missing data and expression-related deformations. They manually divide the facial area into several meaningful components and the registration of faces is carried out by separate dense alignments to the corresponding ARMs. A strong limitation of this approach is the need for manual segmentation of a face into parts that can then be analyzed separately. **Surface distance based approaches** – There are several papers that utilize distances between points on facial surfaces to define features that are eventually used in recognition. (Some papers call it *geodesic distance* but, in order to distinguish it from our later use of *geodesics* on shape spaces of curves and surfaces, we shall call it *surface distance*.) These papers assume that surface distances are relatively invariant to small changes in facial expressions and, therefore, help generate features that are robust to facial expressions. Bronstein et al. (Bronstein et al., 2005a) provide a limited experimental illustration of this invariance by comparing changes in surface distances with the Euclidean distances between corresponding points on a canonical face surface. To handle the open mouth problem, they first detect and remove the lip region, and then compute the surface distance in presence of a hole corresponding to the removed part (Bronstein et al., 2007b). The assumption of preservation of surface distances under facial expressions motivates several authors to define distance-based features for facial recognition. Samir et al. (Samir et al., 2009) use the level curves of the surface distance function (from the tip of the nose) as features for face recognition. Since an open mouth affects the shape of some level curves, this method is not able to handle the problem of missing data due to occlusion or pose variations. A similar polar parametrization of the facial surface is proposed in (Mpiperis et al., 2008b) where the authors study local geometric attributes under this parameterization. To deal with the open mouth problem, they modify the parametrization by disconnecting the top and bottom lips. The main limitation of this approach is the need for detecting the lips, as proposed in (Bronstein et al., 2007b). Berretti et al. (Berretti et al., 2010b) use surface distances to define facial stripes which, in turn, is used as nodes in a graph-based recognition algorithm.

The main limitation of these approaches, apart from the issues resulting from open mouths, is that they
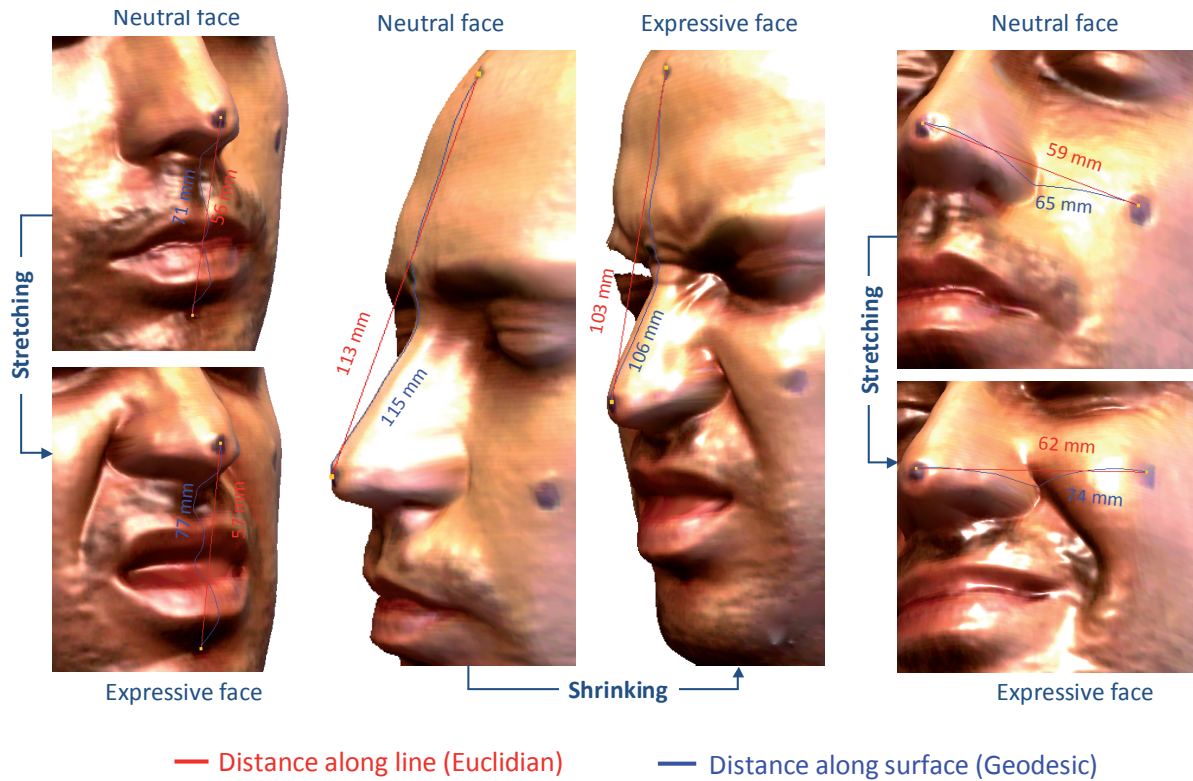
FIGURE 3.2: Significant changes in both Euclidean and surface distances under large facial expressions.

assume that surface distances between facial points are preserved within face classes. This is not valid in the case of large expressions. Actually, face expressions result from the stretching or the shrinking of underlying muscles and, consequently, the facial skin is deformed in a non-isometric manner. In other words, facial surfaces are also stretched or compressed locally, beyond a simple bending of parts. In order to demonstrate this assertion, we placed four markers on a face and tracked the changes in the surface and Euclidean (straight line) distances between the markers under large expressions. Fig. 3.2 shows some facial expressions leading to a significant shrinking or stretching of the skin surface and, thus, causing both Euclidean and surface distances between these points to change. In one case these distances decrease (from 113 mm to 103 mm for the Euclidean distance, and from 115 mm to 106 mm for the surface distance) while in the other two cases they increase. This clearly shows that large expressions can cause stretching and shrinking of facial surfaces, i.e., the facial deformation is elastic in nature. Hence, the assumption of an isometric deformation of the shape of the face is not strictly valid, especially for large expressions. This also motivates the use of elastic shape analysis in 3D face recognition.

3.2 Data Pre-processing and Representation

To analyze facial surfaces (or their shapes), one should first preprocess them and provide the way to represent them as multiple curves. This will be the subject of this section wich will introduce the pipeline of 3D scans pre-processing to denoise the data and to extract the informative part of the face (the facial

surface) and the procedure to extract the radial curves, our approximation (or parametrization) of the 3D facial meshes.

3.2.1 3D Face pre-processing

Since the raw data contains a number of imperfections, such as holes, spikes, and include some undesired parts, such as clothes, neck, ears and hair, the data pre-processing step is very important and non-trivial.
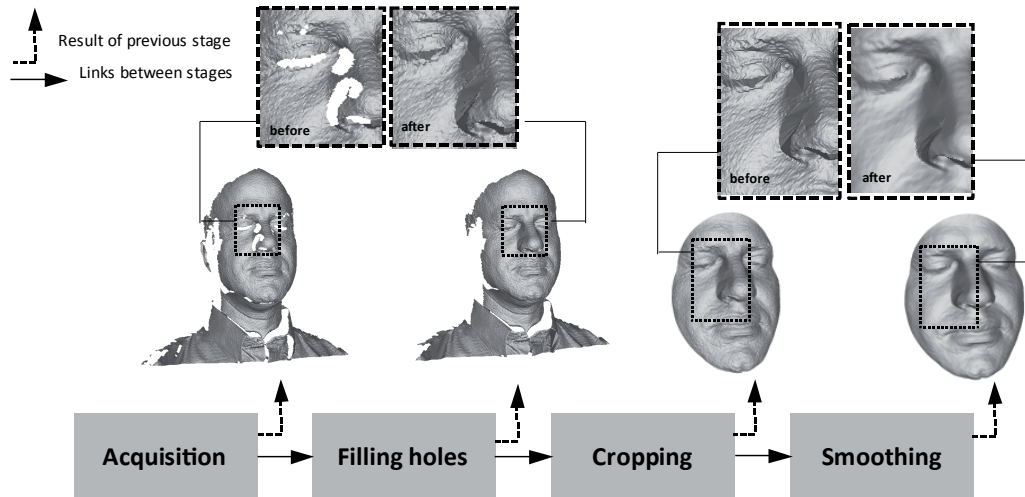

FIGURE 3.3: The different steps of preprocessing: acquisition, filling holes, cropping and smoothing

As illustrated in Fig. 3.3, this step includes the following items:

↪ *The hole-filling filter* identifies and fills holes in input meshes. The holes are created either because of the absorption of laser in dark areas, such as eyebrows and mustaches, or self-occlusion or open mouths. They are identified in the input mesh by locating boundary edges, linking them together into loops, and then triangulating the resulting loops.

↪ *A cropping filter* cuts and returns parts of the mesh inside an Euclidean sphere of constant radius centered at the nose tip, in order to discard as much hair as possible. The nose tip is automatically detected for frontal scans and manually annotated for scans with occlusions and large pose variation.

↪ *A smoothing filter* reduces high frequency components (spikes) in the mesh, improves the shapes of cells, and evenly distributes the vertices on a facial mesh.

The output of the pre-processing step will be a 3D facial mask with the nose tip coordinates, detected automatically. Holes of limited size are filled and the noise filtered out, in addition, only the informative part of the face is kept for the analyze. We have used existing filters from the open source Vizualisation toolkit library[3] to design the outlined pipeline.

---

[3] http://www.vtk.org/

3.2.2   3D Face Representation by Elastic Radial Curves

Each facial surface is represented by an indexed collection of radial curves that are defined and extracted as follows. Let S be a facial surface obtained as an output of the preprocessing step. The reference curve on S is chosen to be the vertical curve after the face has been rotated to the upright position. Then, a radial curve $\beta^{(\alpha)}$ is obtained by slicing the facial surface by a plane $P^{(\alpha)}$ that has the nose tip as its origin and makes an angle α with the plane containing the reference curve. That is, the intersection of $P^{(\alpha)}$ with S gives the radial curve $\beta^{(\alpha)}$. We repeat this step to extract radial curves from S at equally-separated angles, resulting in a set of curves that are indexed by the angle α. Fig. 3.4 shows an example of this process. If needed, we can approximately reconstruct S from these radial curves according to $S \approx \cup_\alpha \beta^{(\alpha)} = \cup_\alpha \{S \cap P^{(\alpha)}\}$. Using these curves, we will demonstrate that the elastic framework is well suited to modeling of deformations associated with changes in facial expressions and for handling missing data.
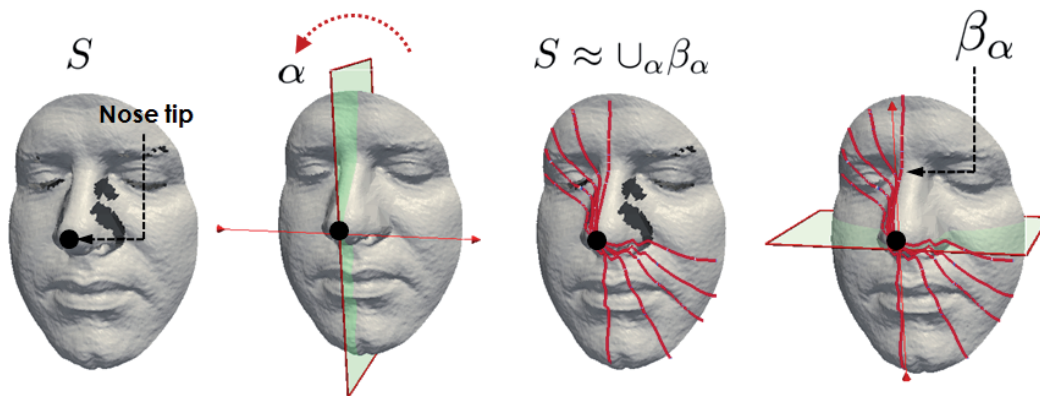

FIGURE 3.4: Extraction of radial curves: images in the middle illustrate the intersection between the face surface and planes to form two radial curves. The collection of radial curves is illustrated in the rightmost image.

In this application (Face Recognition), the probe face is first rigidly aligned to the gallery face using the ICP algorithm. In this step, it is useful but not critical to accurately find the nose tip on the probe face. As long as there is a sufficient number of distinct regions available on the probe face, this alignment can be performed. Next, after the alignment, the radial curves on the probe model are extracted using the plane P_α passing through the nose tip of the **gallery** model at an angle α with the vertical. This is an important point in that only the nose tip of the gallery and a good alignment between gallery-probe is needed to extract good quality curves.

3.3   Tools on the Manifold of 3D Faces

In this section we will use the framework described in Chapter 2, in particular the tools designed to interpolate between shapes (geodesics) and the length of these paths, under the defined elastic metric.

Recall that the central goal is given two facial surfaces how to quantify their difference under deformations (caused by expressions), missing data, occlusions and pose variations. Recall also that throw the representation by elastic radial curves, we view the shapes of 3D faces as element of an infinte-dimentional manifold termed $\mathcal{M}$ equipped with a Riemannian elastic metric ($d_\mathcal{M}$). We have demonstrated the benefits of the Riemannian approach which allows jointly accurate dense correspondence between 3D surfaces (when counting for the elastic deformations that exhibit the expressive faces). These properties are suitable in face recognition as stated in Section 3.1 to handle the facial deformations.

3.3.1 Shape Interpolation – *Geodesics* on the Shape Space

Given two 3D shapes, since we have deformations (geodesic paths) between their pairwise radial curves, we can spatially interpolate these deformations to obtain deformations between full facial surfaces. In fact, these full deformations can be shown to be formal geodesic paths between faces, when represented as elements of $\mathcal{M} = \mathcal{S}^{[0,2\pi]}$. Shown in Fig. 3.5 are examples of some geodesic paths between source and target faces. It illustrates geodesic paths between faces of different subjects, and are termed *inter-class geodesics*. One can note that the shape located at the center of each path correspond to the average shape between the source ans the target. As illustrated, it has an intermediate-smooth morphology between the shapes of the ending points of each geodesic. In other worlds, when considering two facial shapes of different age, different gender or different ethnicity on could generate with our method a typical average face between them, in a mathematical way!

FIGURE 3.5: Three examples of inter-class geodesic paths. In each row, the first and the last 3D shapes represent respectively the source and the target faces, the intermediate shapes are equally-spaces points from the geodesic gerenated by our algorithm.

Shown in Fig. 3.6 are examples geodesic paths between faces of the same person conveying different expressions, and are termed *intra-class geodesics*. In the same way, we can note that the central shape

on each of these paths represent the middle intermediate deformation from the source and the target shapes.
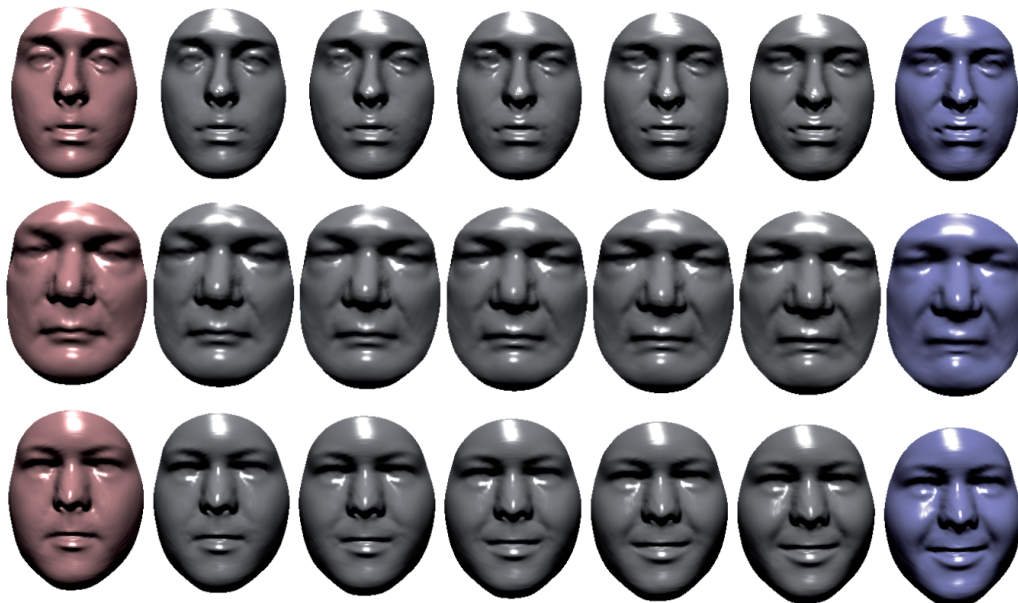

FIGURE 3.6: Three examples of intra-class geodesic paths. In each row, the first and the last 3D shapes represent respectively the source and the target faces, the intermediate shapes are equally-spaces points from the geodesic gerenated by our algorithm.

These geodesics provide a tangible benefit, beyond the current algorithms that provide some kind of a similarity score for analyzing faces. These geodesic paths illustrate the most efficient way to deform one shape into another. The accurate dense correspondence, one of the essential ingredients of shape analysis, plays an important role to get these natural-like deformations (or morphing). In fact, matching accurately different anatomical parts of the faces yields with accurate deformations, hence the importance of use elastic models in shape analysis of facial surfaces. In addition to their interpretation as optimal deformations under the chosen metric, the geodesics can also be used for computing the mean shape and measuring the shape covariance of a set of faces, as illustrated later.

3.3.2 Expression-Robust 3D Face Recognition

We will show in this section the benefits of using our **elastic model** in comparing facial surfaces exhibiting different facial expressions. The goal here is the robustness to the deformations introduced by the expressions. We should note here that in case of open mouth which often results in a hole, the preprocessing algorithm presented in section 3.2.1 will fill the hole by locating the edges and interpolating between them. An other important step is to discrete the shape representation both by indexed collections of radial curves and finite-points representations of each curve, starting from the nose tip. To this end, facial surface S is represented by an indexed collection of radial curves, indexed by the n uniform angles $\mathcal{A} = \{0, \frac{2\pi}{n}, \frac{4\pi}{n}, \ldots, 2\pi\frac{(n-1)}{n}\}$. Thus, the shape of a facial surface can been represented

as an element of the set $\mathcal{S}^n \approx \mathcal{M}$. The indexing provides a correspondence between curves across faces. For example, the curve at an angle α on a probe face is compared with the curve at the same angle on a gallery face. Thus, the distance between two facial surfaces is $d_{\mathcal{M}} : \mathcal{S}^n \times \mathcal{S}^n \to \mathbb{R}_{\geq 0}$, given by $d_{\mathcal{M}}(\mathcal{F}^1, \mathcal{F}^2) \doteq \frac{1}{n} \sum_{\alpha \in \mathcal{A}} d_{\mathcal{S}}([q_1^{(\alpha)}], [q_2^{(\alpha)}])$. Here, $[q_i^{(\alpha)}]$ denotes the SRVF of the radial curve $\beta_i^{(\alpha)}$ on the ith facial surface, $i \in 1, 2$.

To explore experimentally the ability of our elastic model to be robust the facial exressions, we conducted our experiments on the FRGC2.0 dataset (Face Recognition Grand Challenge)(Phillips et al., 2005). The dataset contains 4007 3D scans of 466 subjects with near-frontal pose. About 40% of the scans are expressive. In our experiment we will consider the manual clustering of the 3D scans into three categories: neutral expression, small expression, and large expression as illustrated in Figure 3.7.

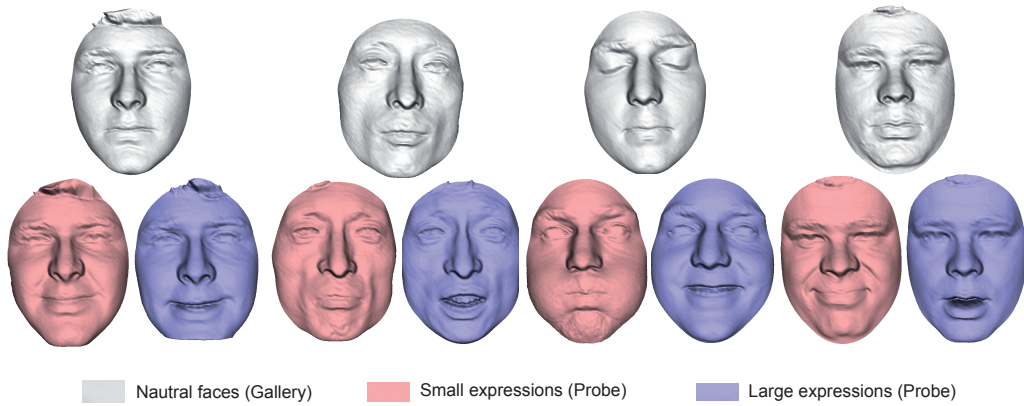Nautral faces (Gallery)   Small expressions (Probe)   Large expressions (Probe)

FIGURE 3.7: Examples of 3D preprocessed scans from the FRGC2.0 dataset. The Gallery faces are almost neutral as shown by the top row. The set of probe faces includes small expressions (in red) and large expressions (in blue) as shown in the second row.

As shown in Figure 3.7, the facial surfaces output of the pre-processing step are of good quality. Now, we will report experimental results both under the identification and verification scenarios, using the commonly used evaluation criteria in biometrics.

I. Identification Scenario: The gallery consists of the first scans for each subject in the database, and the remaining scans make up the probe faces. This dataset was automatically preprocessed as described in the Section 3.2.1. The left panel of Figure 3.8 shows the Cumulative Matching Curves (CMCs) of our method under this protocol for the three cases: neutral vs. neutral, neutral vs. non-neutral and neutral vs. all. Note that our method results in 97.7% rank-1 recognition rate in the case of neutral vs. all. In the difficult scenario of neutral vs. expressions, the rank-1 recognition rate is 96.8%, which represents a high performance, while in the simpler case of neutral vs. neutral the rate is 99.2%. As shown in these CMC curves the recognition rates reaches quickly 100% (at rank 5).

A comparison of recognition performance of our method with several state-of-the-art results is presented in Table 3.1. This time, in order to keep the comparisons fair, we kept all the 466 scans in the gallery. Notice that our method achieved a 97% rank-1 recognition which is close to the highest published results

TABLE 3.1: Comparison of rank-1 scores on the FRGCv2 dataset with the state-of-the-art results.

Method	Rank-1 Recognition Rate (%)
(Spreeuwers, 2011)	99%
(Wang et al., 2010)	98.3%
(ter Haar and Velkamp, 2010)	97%
(Berretti et al., 2010b)	94.1%
(Queirolo et al., 2010)	98.4%
(Faltemier et al., 2008)	97.2%
(Kakadiaris et al., 2007)	97%
(Drira et al., 2013a)	**97%**

on this dataset (Faltemier et al., 2008, Spreeuwers, 2011, Queirolo et al., 2010). Since the scans in FRGCv2 are all frontal, the ability of region-based algorithms, such as (Faltemier et al., 2008, Queirolo et al., 2010), to deal with the missing parts is not tested in this dataset. For that end, one would need a systematic evaluation on a dataset with the missing data issues, e.g. the GavabDB. The best recognition score on FRGCv2 is reported by Spreeuwers (Spreeuwers, 2011) which uses an intrinsic coordinate system based on the vertical symmetry plane through the nose. The missing data due to pose variation and occlusion challenges will be a challenge there as well.
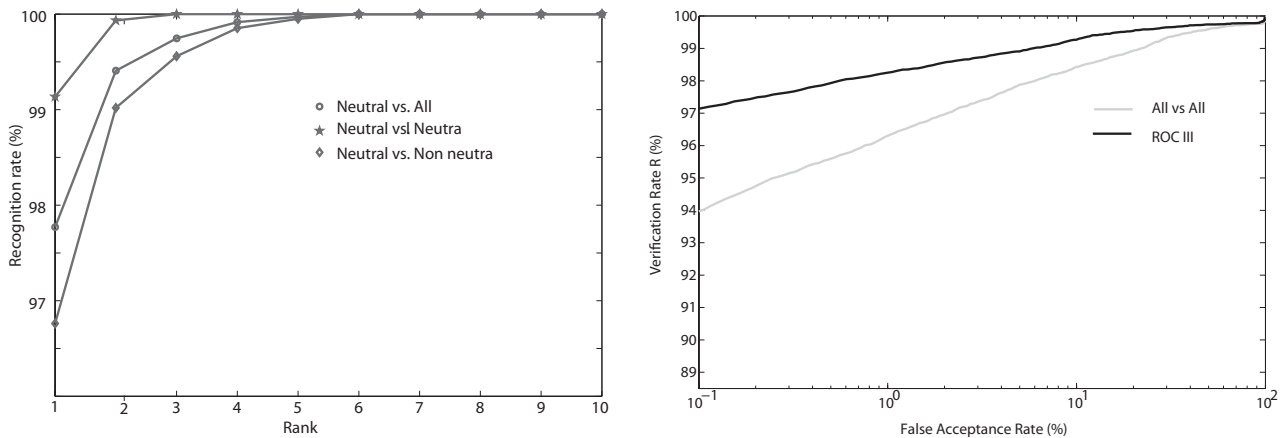

FIGURE 3.8: Left: Cumulative Matching Curves for the Neutral vs. Neutral, Neural vs. Small expressions and Neutral vs. Large expressions on FRGC2.0. Right: ROC curves for the All vs. All and ROC III evaluation protocols.

I. Verification Scenario: In order to evaluate the performance of the proposed approach in the verification scenario, the Receiver Operating Characteristic (ROC) curves for the ROC III mask of FRGCv2 and "all-versus-all" are plotted in the right part of Fig. 3.8. For comparison, Table 3.2 shows the verification results at false acceptance rate (FAR) of 0.1 percent for several methods. For the standard protocol testings, the ROC III mask of FRGC v2, we obtain the verification rates of around 97%, which is comparable to the best published results. In the all-versus-all experiment, our method provides 93.96% VR

at 0.1% FAR, which is among the best rates in the table (Wang et al., 2010, Spreeuwers, 2011, Queirolo et al., 2010).

TABLE 3.2: Comparison of verification rates at FAR=0.1% on the FRGCv2 dataset with state-of-the-art results (the ROC III mask and the All vs. All scenario).

Method	ROC III	All vs. All
(Kakadiaris et al., 2007)	97%	–
(Faltemier et al., 2008)	94.8%	93.2%
(Berretti et al., 2010b)	–	81.2%
(Queirolo et al., 2010)	96.6%	96.5%
(Spreeuwers, 2011)	94.6%	94.6%
(Wang et al., 2010)	98.4%	98.1%
(Drira et al., 2013a)	**97.1%**	**93.9%**

Note that these approaches are applied to FRGCv2 only. Since scans in FRGCv2 are mostly frontal and have high quality, many methods are able to provide good performance. It is, thus, important to evaluate a method in other situations where the data quality is not as good. In the next two sections, we will consider those situations with the GavabDB involving the pose variation and the Bosphorus dataset involving the occlusion challenge.

3.3.3 3D Face Recognition under Large Pose Variations

In situations involving non-frontal 3D scans, some curves may be partially hidden due to self occlusion. The use of these curves in face recognition can severely degrade the recognition performance and, therefore, they should be identified and discarded.

I. Curve Quality Filter: We introduce here a quality filter that uses the continuity and the length of a curve to detect such curves. To pass the quality filter, a curve should be one continuous piece and have a certain minimum length, say of, 70mm. The discontinuity or the shortness of a curve results either from missing data or large noise. We show two examples of this idea in Fig. 3.9 where we display the original scans, the extracted curves, and then the action of the quality filter on these curves. Once the quality filter is applied and the high-quality curves retained, we can perform face recognition procedure using only the remaining curves. That is, the comparison is based only on curves that have passed the quality filter. Let β denotes a facial curve, we define the boolean function *quality*: ($quality(\beta) = 1$) if β passes the quality filter and ($quality(\beta) = 0$) otherwise. Recall that during the pre-processing step, there is a provision for filling holes. Sometimes the missing parts are too large to be faithfully filled using linear interpolation. For this reason, we need the quality filter that will isolate and remove curves associated with those parts.
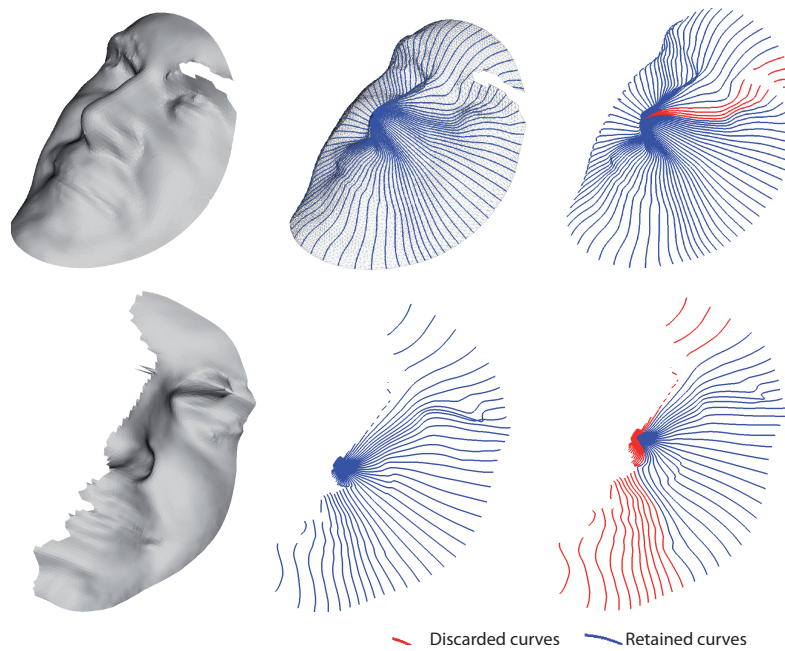
FIGURE 3.9: Curve quality filter: examples of detection of broken and short curves (in red) and complete curves (in blue). The first example show missing data due to bad reflection of the laser by the eyebrow. The second example shows a profile scan with only half of the data available.

II. Experiments on the GavabDB dataset: Since GavabDB (Moreno and Sanchez, 2004) has many noisy 3D face scans under large facial expressions, we will use that database to help evaluate our framework. This database consists of the Minolta Vi-700 laser range scans from 61 subjects – 45 male and 16 female – all of them Caucasian. Each subject was scanned nine times from different angles and under different facial expressions (six with the neutral expression and three with non-neutral expressions). The neutral scans include several frontal scans – one scan while looking up (+35 degree), one scan while looking down (-35 degree), one scan from the right side (+90 degree), and one from the left side (-90 degree). The non-neutral scans include cases of a smile, a laugh, and an arbitrary expression chosen freely by the subject. Figure 3.10 shows the nine scans of one subject from the GavabDB.

FIGURE 3.10: Scans of one subject of the GavabDB dataset under different variations.

One of the two frontal scans with the neutral expression for each person is taken as a gallery model, and the remaining are used as probes. Table 3.3 compares the results of our method with the previously published results following the same protocol. As noted, our approach provides the highest recognition rate for faces with non-neutral expressions (94.54%). This robustness comes from the use of radial, elastic

TABLE 3.3: Recognition results comparison of the different methods on the GavabDB.

Method	(N)	(E)	(N)+(E)	(LD)	(LU)	(RS)	(LS)
(Li et al., 2009)	96.6%	93.3%	94.6%	–	–	–	–
(Moreno et al., 2005)	90.1%	77.9%	–	–	–	–	–
(Mahoor and Abdel-Mottaleb, 2009)	–	72%	78%	85.3%	88.6%	–	–
(ter Haar and Velkamp, 2010)	–	–	–	–	–	–	–
(Mousavi et al., 2008)	–	–	91%	–	–	–	–
(**Drira et al., 2013a**)	**100%**	**94.5%**	**95.9%**	**100%**	**98.3%**	**70.4%**	**86.8%**

curves since: (1) each curve represents a feature that characterizes local geometry and, (2) the elastic matching is able to establish a correspondence with the correct alignment of anatomical facial features across curves. Table 3.3 provides an exhaustive summary of results obtained using GavabDB; our method outperforms the majority of other approaches in terms of the recognition rate. Note that there is no prior result in the literature on 3D face recognition using sideway-scans from this database. Although our method works well on common faces with a range of pose variations within 35 degrees, it can potentially fail when a large part of the nose is missing, as it can cause an incorrect alignment between the probe and the gallery. This situation occurs if the face is partially occluded by external objects such as glasses, hair, etc. To solve this problem, we first restore the data missing due to occlusion.

3.3.4   3D Face Recognition under External occlusions

In this section we target the problem of occlusions by external parts like glasses, hands, hair, etc. We will consider two components, (1) external occlusion detection and removal and (2) missing data recovery.

I. Occlusion Removal: The first problem we encounter in externally-occluded faces is the detection of the external object parts. We accomplish this by comparing the given scan with a template scan, where a template scan is developed using an average of training scans that are complete, frontal and have neutral expressions. The basic matching procedure between a template and a given scan is recursive ICP, which is implemented as follows. In each iteration, we match the current face scan with the template using ICP and remove those points on the scan that are more than a certain threshold away from the corresponding points on the template. This threshold has been determined using experimentation and is fixed for all faces. In each iteration, additional points that are considered extraneous are incrementally removed and the alignment (with the template) based on the remaining points is further refined.

Fig. 3.11 shows an example of this implementation. From left to right, each face shows an increasing alignment of the test face with the template, with the aligned parts shown in magenta, and also an increasing set of points labeled as extraneous, drawn in pink. The final result, the original scan minus the extraneous parts, is shown in green at the end. In the case of faces with external occlusion, we first
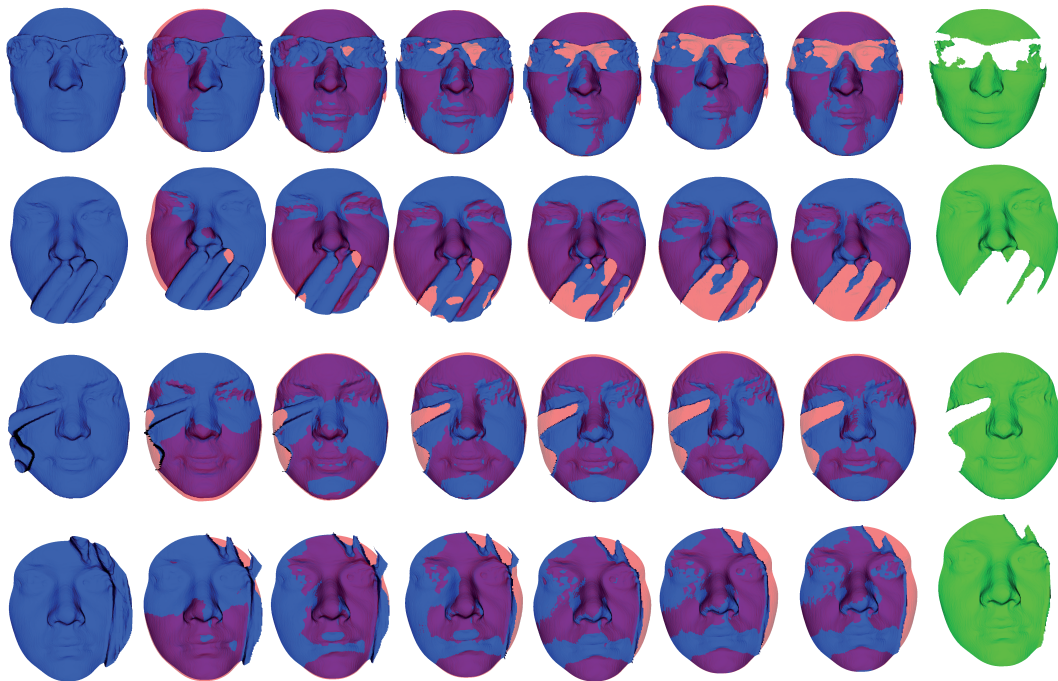
FIGURE 3.11: Gradual removal of occluding parts in a face scan using Recursive-ICP.

restore them and then apply the recognition procedure. That is, we detect and remove the occluded part, and recover the missing part resulting in a full face that can be compared with a gallery face using the metric $d_{\mathcal{S}}$. The recovery is performed using the tangent PCA analysis and Gaussian models, as described in Section 3.4. In order to evaluate our approach, we perform this automatic procedure on the Bosphorus database (Alyuz et al., 2008).

II. Completion of Partially-Obscured Curves: Earlier we have introduced a filtering step that finds and removes curves with missing parts. Although this step is effective in handling some missing parts, it may not be sufficient when parts of a face are missing due to external occlusions, such as glasses and hair. In the case of external occlusions, the majority of radial curves could have hidden parts that should be predicted before using these curves. This problem is more challenging than self-occlusion because, in addition to the missing parts, we can also have parts of the occluding object(s) in the scan. In a non-cooperative situation, where the acquisition is unconstrained, there is a high probability for this kind of occlusion to occur. Once we detect points that belong to the face and points that belong to the occluding object, we first remove the occluding object and use a statistical model in the shape space of radial curves to complete the broken curves. This replaces the parts of face that have been occluded using information from the visible part and the training data.

The core of this problem, in our representation of facial surfaces by curves, is to take a partial facial curve and predict its completion. The sources of information available for this prediction are: (1) the current (partially observed) curve and (2) several (complete) training curves at the same angle that are extracted from full faces. The basic idea is to develop a sparse model for the curve from the training curves and use that to complete the observed curve. To keep the model simple, we use the PCA of the training data, in

an appropriate vector space, to form an orthogonal basis representing training shapes. Then, this basis is used to estimate the coefficients of the observed curve and the coefficients help reconstruct the full curve. Since the shape space of curve $\mathcal{S}$ is a nonlinear space, we use the tangent space $T_\mu(\mathcal{S})$, where μ is the mean of the training shapes, to perform PCA. Let α denote the angular index of the observed curve, and let q_α^1, $q_\alpha^2, \ldots, q_\alpha^k$ be the SRVFs of the curves taken from the training faces at that angle. As described earlier, we can compute the sample Karcher mean of their shapes $\{[q_\alpha^i] \in \mathcal{S}\}$, denoted by μ_α. Then, using the geometry of $\mathcal{S}$ we can map these training shapes in the tangent space using the inverse exponential map. We obtain $v_{i,\alpha} = \exp_{\mu_\alpha}^{-1}(q_\alpha^i)$, where

$$\exp_{q_1}^{-1}(q_2) = \frac{\theta}{\sin(\theta)}(q_2^* - \cos(\theta)q_1), \quad \theta = \cos^{-1}(\langle q_1, q_2^* \rangle) \ ,$$

and where q_2^* is the optimal rotation and re-parameterization of q_2 to be aligned with q_1, as discussed earlier. A PCA of the tangent vectors $\{v_i\}$ leads to the principal basis vectors $u_{1,\alpha}, u_{2,\alpha}, ..., u_{J,\alpha}$, where J represents the number of significant basis elements.

Now returning to the problem of completing a partially-occluded curve, let us assume that this curve is observed for parameter value t in $[0, \tau] \subset [0, 1]$. In other words, the SRVF of this curve $q(t)$ is known for $t \in [0, \tau]$ and unknown for $t > \tau$. Then, we can estimate the coefficients of q under the chosen basis according to $c_{j,\alpha} = \langle q, u_{j,\alpha} \rangle \approx \int_0^\tau \langle q(t), u_{j,\alpha}(t) \rangle \, dt$, and estimate the SRVF of the full curve according to

$$\hat{q}_\alpha(t) = \sum_{j=1}^{J} c_{j,\alpha} u_{j,\alpha}(t) \ , \quad t \in [0, 1] \ .$$

We present three examples of this procedure in Fig. 3.12, with each face corrupted by an external occlusion as shown in column (a). The detection and removal of occluded parts is performed as described in the previous section, and the result of that step is shown in column (b). Finally, the curves passing through the missing parts are restored and shown in (c).

In order to evaluate this reconstruction step, we have compared the restored surface (shown in the top row of Fig. 3.12) with the complete neutral face of that class, as shown in Fig. 3.13. The small values of both absolute deviation and signed deviation, between the restored face and the corresponding face in the gallery, demonstrate the success of the restoration process.

III. Experiments on Bosphorus Dataset: The Bosphorus database is suitable for this evaluation as it contains scans of 60 men and 45 women, 105 subjects in total, in various poses, expressions and in the presence of external occlusions (eyeglasses, hand, hair). The majority of the subjects are aged between 25 and 35. The number of total face scans is 4652; at least 54 scans each are available for most of the subjects, while there are only 31 scans each for 34 of them. The interesting part is that for each subject there are four scans with occluded parts. These occlusions refer to (i) mouth occlusion by hand, (ii) eyeglasses, (iii) occlusion of the face with hair, and (iv) occlusion of the left eye and forehead regions by

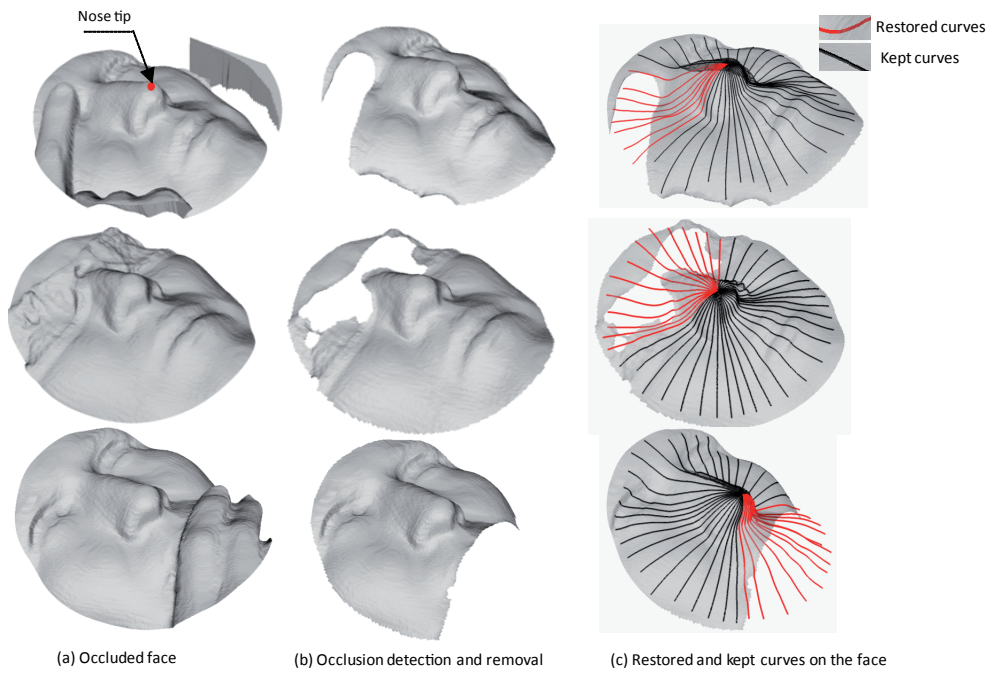(a) Occluded face   (b) Occlusion detection and removal   (c) Restored and kept curves on the face

FIGURE 3.12: (a) Faces with external occlusion, (b) faces after the detection and removal of occluding parts and (c) the estimation of the occluded parts using a statistical model on the shape spaces of curves.
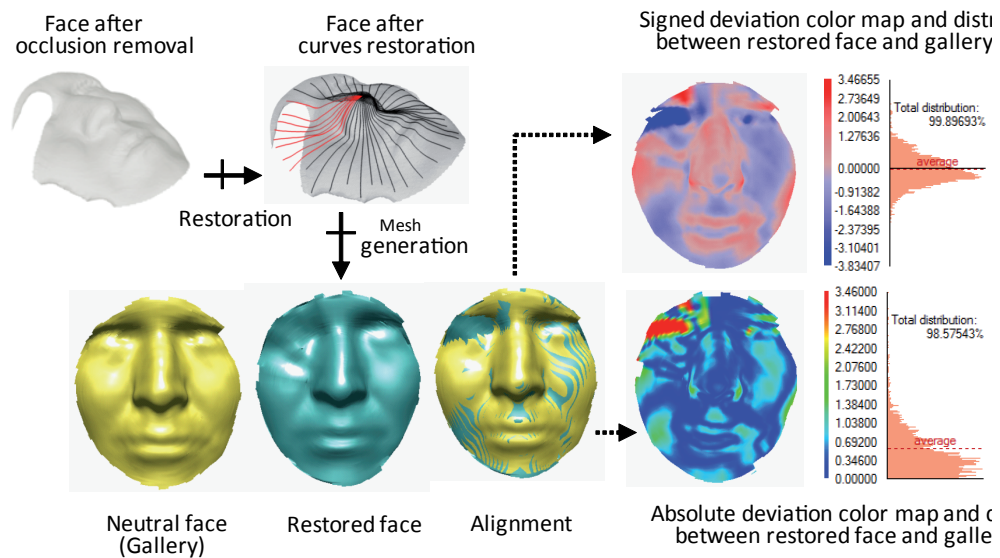

FIGURE 3.13: Illustration of a face with missing data (after occlusion removal) and its restoration. The deviation between the restored face and the corresponding neutral face is also illustrated.

hands. Fig. 3.14 shows sample images from the Bosphorus 3D database illustrating a full scan on the left and the remaining scans with typical occlusions.

We pursued the same evaluation protocol used in the previously published papers: a neutral scan for each person is taken to form a gallery dataset of size 105 and the probe set contains 381 scans that have occlusions. The training is performed using other sessions so that the training and test data are disjoint.
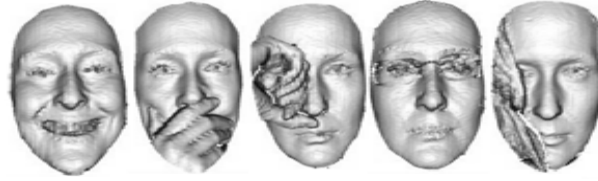
FIGURE 3.14: Examples of faces from the Bosphorus database. The un-occluded face on the left and the different types of occlusions are illustrated.

The rank-1 recognition rate is reported in Fig. 3.15 for different approaches depending upon the type of occlusion. As these results show the process of restoring occluded parts significantly increases the accuracy of recognition. The rank-1 recognition rate is 78.63% when we remove the occluded parts and apply the recognition algorithm using the remaining parts, as described in Section 2.4. However, if we perform restoration, the recognition rate is improved to 87.06%. Clearly, this improvement in performance is due to the estimation of missing parts on curves. These parts, that include important shape data, were not considered by the algorithm described earlier. Even if the part added with restoration introduces some error, it still allows us to use the shapes of the partially observed curves. Furthermore, during restoration, the shape of the partially observed curve is conserved as much as possible.
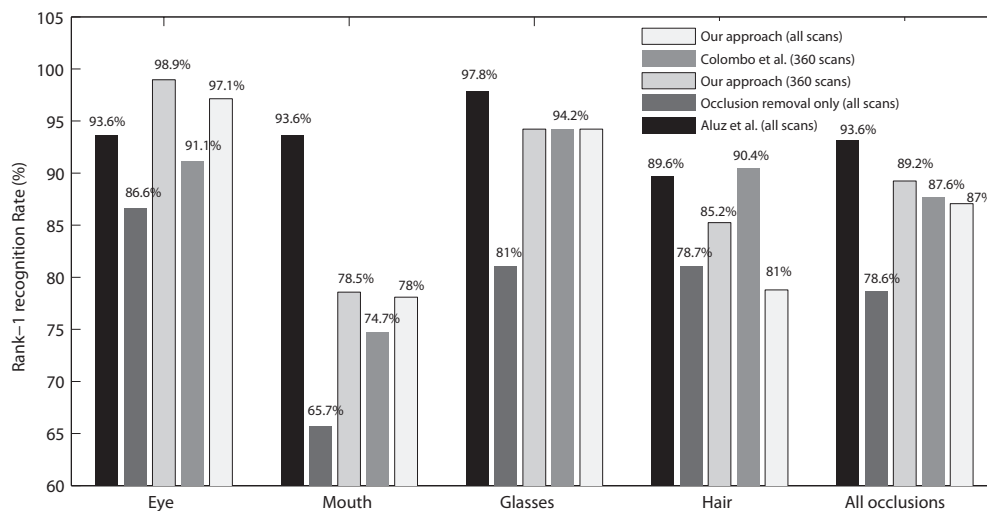

FIGURE 3.15: Recognition results on the Bosphorus database and comparison with state-of-the-art approaches.

Examples of 3D faces recognized by our approach are shown in Fig. 3.12, along with different steps of the algorithm. The faces in the two bottom rows are examples of incorrectly recognized faces by our algorithm without restoration (as described earlier), but after the restoration step, they are correctly recognized. (Alyuz et al., 2008) reported a 93.69% rank-1 recognition rate overall for this database using the same protocol that we have described above. While this reported performance is very good, their processing has some manual components. Actually, the authors partition the face manually and fuse the scores for matching different parts of the face together. In order to compare with (Colombo et al., 2011),

we reduce the probe dataset to 360 by discarding bad quality scans as (Colombo et al., 2011) did. Our method outperforms their approach with an overall performance of 89.25%, although individually our performance is worse in the case of occlusion by hair. It is difficult, in this case, to completely overcome face occlusion. Therefore, during the restoration step, our algorithm tries to keep majority of parts. This leads to a deformation in the shape of curves and, hence, affects the recognition accuracy.

3.4   Towards Efficient Face Recognition Solution

In this section, we explore two complementary ideas to make the recognition scenario more efficient. It is well stated that, compared to 2D-face recognition approaches, the 3D-based approaches required more computational time for several reasons. The most important one is the data representation itself, where the data in 2D are images (ordered pixels), in the second case the data is 3D meshes without any order. We target here the one-to-many matching scenario (identification scenario). To handle these issues, we propose first (1) a tree-like organization of the facial surfaces in the Gallery, and (2) introduce a boosting step to highlight and select the most relevant curves on the face. Thus, a gain on time will happen when parsing the branches of the tree instead of an exhaustive comparison, and use only the relevant curves to compare the probe face to the gallery face.

3.4.1   Hierarchical Gallery Organization

One of the main goals for studying shapes of faces is to conduct biometric identification where query is often compared to a set of gallery shapes. This comparison can be made more efficient if we can organize the gallery elements in form of a hierarchical database, i.e. a tree, where the comparisons are performed only at the nodes of that tree. To construct such a shape tree we need to be able to cluster similar shapes, and that is the problem we study next.

I. Clustering of 3D Faces: Consider the problem of clustering n faces (in $\mathcal{M}$) into k clusters. A general approach is to form clusters in such a way that they minimize total "within-cluster" variance. Let a configuration $\mathcal{F}$ consists of clusters denoted by $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k$, and let μ_is be the mean shapes in $\mathcal{F}_i$s and n_is be the sizes of $\mathcal{F}_i$s. There are several cost functions used for clustering, e.g the sum of traces of covariances within clusters. However, the computation of means μ_is of large shape clusters, and therefore their variances, is computationally expensive, especially when they are updated at every iteration. As a solution, one often uses a variation, called *pairwise clustering* (Hofmann and Buhmann, 1998), where the variance of a cluster is replaced by a scaled sum of distances (squared) between its elements:

$$Q(\mathcal{F}) = \sum_{i=1}^{k} \frac{2}{n_i} \left(\sum_{\mathcal{F}^a \in \mathcal{F}_i} \sum_{b<a, \mathcal{F}^b \in \mathcal{F}_i} d_{\mathcal{M}}(\mathcal{F}^a, \mathcal{F}^b)^2 \right). \tag{3.1}$$

We seek configurations that minimize Q, i.e., $\mathcal{F}^* = \arg\min Q(\mathcal{F})$. Notice that the metric used is the arithmetic mean $d_{\mathcal{M}}$. We will minimize the clustering cost using a Markov chain search process on the configuration space. The basic idea is to start with a configuration of k clusters and reduce Q by re-arranging shapes among the clusters. The re-arrangement is performed in a stochastic fashion using two kinds of moves. These moves are performed with probability proportional to the negative exponential of the Q-value of the resulting configuration. The two types of moves are following. (1) **Move a shape**: Here we select a shape randomly and re-assign it to another cluster. Let $Q_j^{(i)}$ be the clustering cost when a shape N_j is re-assigned to the cluster C_i keeping all other clusters fixed. If N_j is not a singleton, i.e. not the only element in its cluster, then the transfer of N_j to cluster C_i is performed with probability: $P_M(j, i; T) = \frac{\exp(-Q_j^{(i)}/T)}{\sum_{i=1}^{k} \exp(-Q_j^{(i)}/T)}$ $i = 1, 2, \ldots, k$. Here T plays a role similar to temperature in simulated annealing. If N_j is a singleton, then moving it is not allowed in order to fix the number of clusters at k. (2) **Swap two shapes**: Here we select two shapes randomly from two different clusters and swap them. Let $Q^{(1)}$ and $Q^{(2)}$ be the Q-values of the original configuration (before swapping) and the new configuration (after swapping), respectively. Then, swapping is performed with probability: $P_S(T) = \frac{\exp(-Q^{(2)}/T)}{\sum_{i=1}^{2} \exp(-Q^{(i)}/T)}$. In order to seek global optimization, we have adopted a simulated annealing approach. Although simulated annealing and the random nature of the search help in avoiding local minima, the convergence to a global minimum is difficult to establish.

It is important to note that once the pairwise distances are computed, they are not computed again in the iterations. Secondly, unlike k-mean clustering, the mean shapes are never calculated in this clustering. The algorithms for computing Karcher mean and clustering can be applied repeatedly for organizing a large database of human faces into a hierarchy that allows efficient searches during identification process. As an illustration of this idea, we consider the 466 face scans corresponding of distinct subjects of FRGC2.0 dataset. These shapes form the bottom layer of the hierarchy, called level D in Figure 3.16. Then, we compute Karcher mean shapes (representative shapes) for each person to obtain shapes at level C. These shapes are further clustered together and a Karcher mean is computed for each cluster. These mean shapes form the level B of the hierarchy. Repeating this idea a few times, we reach the top of the tree with only one shape. We obtain so the final tree shown in Figure 3.16. If we follow a path from top to bottom of the tree, we see the shapes getting more particularized to groups and then to individuals as illustrated in Figure 3.17.

II. Hierarchical Shape Identification: Once the tree is formed, one can use this representation of data to conduct biometric search in order to reduce time computation. Specially in identification scenario, which needs a comparison of the query shape to the whole gallery dataset. In view of this structure, a natural way is to start at the top, compare the query with the shapes at each level, and proceed down the branch that leads to the closest shape. At any level of the tree, there is a number, say h, of possible shapes and our goal is to find the shape that matches the query best.

III. Experiments on FRGC2.0: We have tested this idea on the FRGC2.0 dataset where the Gallery subset is organized as describes in I and the query scans are matched against the faces in a single branch
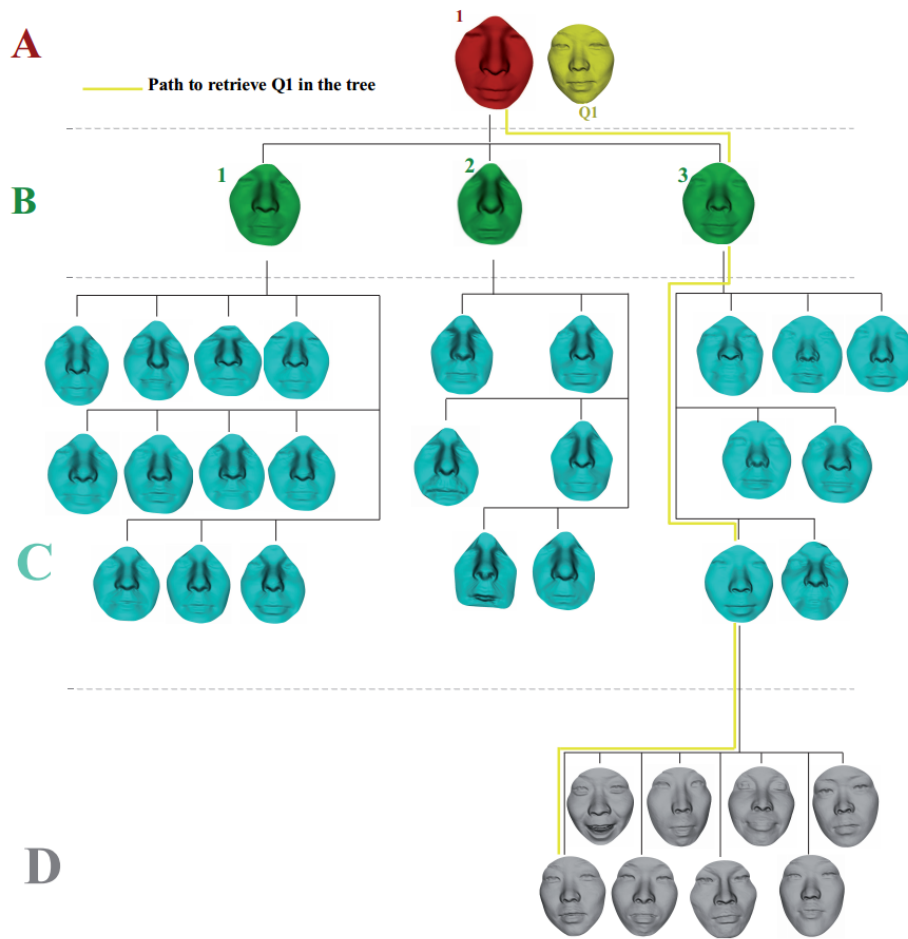
FIGURE 3.16: The tree resulting on hierarchical clustering on the 466 earliest scans of FRGC2.0 dataset (the Gallery).

FIGURE 3.17: Paths from top to bottom in the tree show increasing shape resolutions

of the tree, according to the retrieval procedure described in II. The rank on recognition rate is 92.4% compared to 97% without clustering (by performing exhaustive search along the gallery). The time concuming is reduced from 52 minutes to 3 minutes in average which is 15 times less.

3.4.2 Boosting Geometric Facial Features

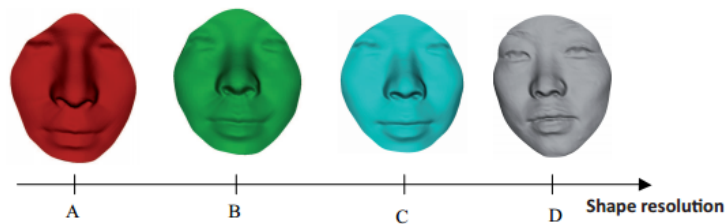In this part, we combine ideas from *shape analysis* using tools from differential geometry and *feature selection* derived from Machine Learning to select and highlight salient 3D geometrical facial features. Notice that we introduce in addition to the radial curves presented above, iso-level curves extracted on the face with a reference point the nose tip and the Euclidean distance as the set function, as shown if Figure 3.19. Our previous studies on iso-level curves on facial surfaces as complete biometrics (Ben Amor et al., 2009) an on the nose (Drira et al., 2009a,b) as partial biometrics, demonstrated their relevance to conduct shape statistics. However, their use assume that surface distances are relatively robust to changes in facial expressions and, therefore, help generate features that are robust to facial expressions. The assumption of preservation of surface distances under facial expressions is no more true when large expressions are conveyed which result in stretching and shrinking of the facial surface. In addition, since an open mouth affects the shape of some level curves, this method is not able to handle the problem of missing data.
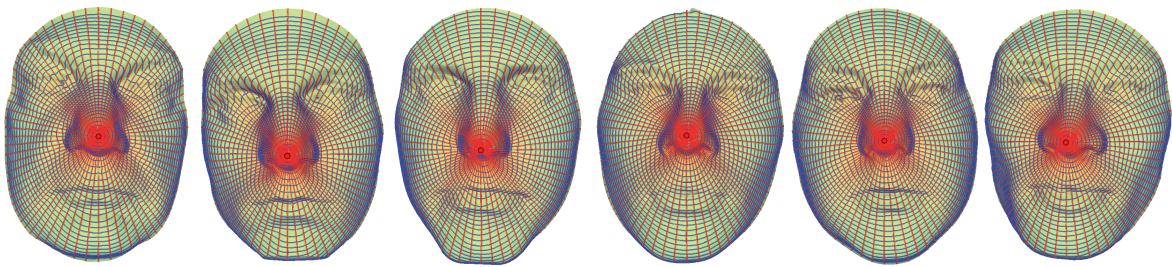

FIGURE 3.18: Radial and iso-level curves extracted on some faces fron FRGC2.0 dataset.

After preprocessing the 3D scans, we represent obtained facial surfaces by finite indexed collections of circular and radial curves. The comparison of pairwise curves, extracted from faces, is based on shape analysis of parameterized curves using differential geometry tools (as described in Chapter 2). The extracted features are trained as weak classifiers and the most discriminative features are selected optimally by adaptive boosting. Accordingly, it consists on the following steps:

- The **Off-line training step**, learns the most salient circular and radial curves from the sets of extracted ones, in a supervised fashion. In face recognition, construct feature vectors by comparing pairwise curves extracted from facial surfaces. Next, feed these examples, together with labels indicating if they are inter-class or not. Thus, the adaptive boosting selects and learns alliteratively the weak classifiers and adding them to a final strong classifier, with suitable weights. As a result of this step, we keep the T-earliest selected features (with highest weights) for the testing step.

- The **On-line test step**, performs classification of a given test face. In the recognition problem, a probe face is compared to the gallery faces using only individual scores computed based on selected features using the $d_{\mathcal{S}}$ metric.

Here, we introduce the feature selection step to highlight (or localize) the most stable and most discriminatory curves. To this end, we use the well-known machine learning algorithm AdaBoost introduced by Freund and Schapire in (Freund and Schapire, 1995). Recall that, boosting is based on iterative selection of weak classifiers by using a distribution of training samples. At each iteration, the best weak classifier is provided and weighted by the quality of its classification. In practice, the individual level curves and radial curves are used as weak classifiers. After M iterations, the most relevant T ($T < M$) facial curves are returned by the algorithm. The set of selected curves returned by Adaboost is shown in Fig. 3.19. The first row of the figure shows the locations of the subset of curves selected by Adaboost on different sessions of the same person with different expressions, whereas, the second row gives curves locations on different subjects. We notice that the boosting algorithm selects the level curves located on the nasal region, which is stable under expressions and radial curves in the way that it avoid two parts, the first one is the lower part of the face since its shape is affected by expressions, particularly when the mouth is open. The second area corresponds to the eye/eyebrow regions. Shapes of radial curves passing throw these regions change when conveying expressions. In contrast, the most stable area cover the nasal/forehead/cheeks regions. As expected and studied in our previous papers (Drira et al., 2009a,b), the use of level curves on the nose is relevant to perform expression-robust partial face analysis.
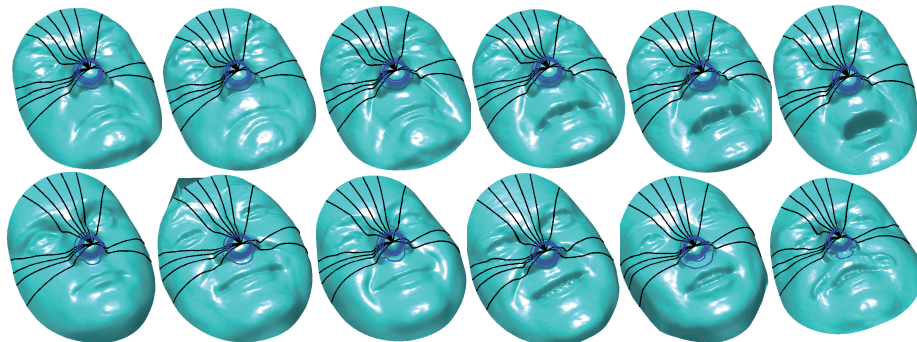

FIGURE 3.19: The most discriminating radial and circular curves selected by Boosting for face recognition, given on different faces.

This approach has been trained and tested on the FRGC2.0 dataset and reported 98.02% as rank-one recognition rate using only a subset of of radial and iso-level curves.

TABLE 3.4: Rank-1/Computation cost (in sec) for different configurations.

Performance	All curves		Selected curves	
	Rank-1 RR	*Time(s)*	*Rank-1 RR*	*Time(s)*
Radial curves	88.65%	1.6	89.04%	0.48
Level curves	66.51%	1.04	85.65%	0.20
Fusion	91.81%	2.64	**98.02%**	**0.68**

Table 3.4 reports the rank-one recognition rates of the radial and the level curves taken individually as well as their fusion. Also, it reports the recognition rates achieved using only the salient features. We notice that the fusion of selected radial and level features achieved the highest performance (98.02%), we can see also an important gain in the computational time (0.68 seconds) to match two facial surfaces.

These ideas (a) hierarchical clustering of the gallery and (b) boosting the geometric curves are presented disjoint. We expect that their use, in the same system, can improve the computational time in a one-to-many matching scenario and keep high recognition performance.

3.5   Conclusion

In this chapter we have demonstrated the application of our Riemannian approach in statistical shape analysis of facial surfaces with the final goal of face recognition. Several problems have been tackled using the unified framework – (1) face comparisons under shape-preserving transformations, (2) shape matching and comparison of 3D shapes robust to deformations caused by the expressions, (3) averaging facial shapes to represent classes/clusters and its use in organizing hierarchically the Gallery for efficient one-to-many matching (4) partially-obscured data completion then shape analysis using statistical models to recover missing data, and (5) Combine ideas from differential geometry and feature selection to select salient curves for more efficient recognition. We have presented results on the mostly used 3D Face datasets (FRGC2.0, Gavab and Bosphorus) designed to develop and test solutions working under variations of facial expression, pose variations and occlusions between gallery and probe scans.

Chapter 4

3D Facial Expression Recognition

In this chapter[1] we introduce two major contributions in the field of facial expression analysis and recognition from static and dynamic 3D data. In particular, we show how to capture facial movements using geometric shape analysis tools introduced in Chapter 2.

Hence, instead of the deformation-robust shape analysis, developed in Chapter 3 oriented to 3D face recognition, we focus here on two relevant questions — How effectively capture the facial deformations resulted from the facial expressions? and how establish facial deformation patterns for successful classification? — To answer these questions, we adopt two approaches : **(1)** A *Landmark-dependent* approach is first introduced in (Maalej et al., 2010, 2011) which uses 3D static scans. Herein, we propose a new facial representation using **Local Facial Patches** extracted around the landmarks, suitable for expression analysis; then **(2)** A *Landmark-free* approach (Drira et al., 2012, Ben Amor et al., 2014a) operating on dynamic data is developed based on a **Global Facial Representation** of 4D data. In both approaches, the issue of shape correspondence is a key point to achieve accurate measurements of the deformations. These deformations which characterize the facial expressions, such as happy or surprise, are executed by facial muscles and they are more prominent in 3D (geometry) than 2D (appearance or texture), which motivates this choice to this shift in the very recent literature. In both approaches that we have introduced, adapted Riemannian shape analysis frameworks are applied to derive statistical analysis of facial shapes. The classification task is resolved using conventional *Machine Learning* techniques, as Random Forest (RF), Support Vector Machine (SVM), Adaboost and Hidden Markov Model (HMM).

After introducing the problem, sections 4.2 and 4.2 describe our contributions in facial expression from static and dynamic data, respectively.

---

[1]The content of this chapter is related to our published papers in the field of facial expression recognition, (Maalej et al., 2011) and (Ben Amor et al., 2014a).

4.1 Introduction

Over the last two decades, automatic recognition of facial expressions emerged as a field of active research, with applications in several different areas, such as HMI (Human-Machine Interaction), psychology, computer graphics, transport security (by detecting driver fatigue, for example), and so on. The importance of facial expressions was first realized and investigated by psychologists, among others. In a seminal work by Mehrabian et al. (Mehrabian and Wiener, 1967) the relative importance of verbal and nonverbal messages in communicating feelings and attitude is described. In particular, they provided evidence that face-to-face communication is governed by the 7%-38%-55% rule, that balances the relevance of verbal, vocal and visual elements, respectively, in communications. Despite this rigid quantification has since been refuted in later studies, it still provides an indication that the words and tone of the voice form only a part of human communication. The non-verbal elements related to the body language (e.g., gestures, posture, facial expressions) also play an important role. Starting from a different point of view, Ekman (Ekman, 1972) conducted the first systematic studies on facial expressions in the late 70s. Ekman also showed that facial expressions can be coded through the movement of face points as described by a set of *action units* (Ekman and Friesen, 1977). Through his experiments, it is demonstrated that there are six *prototypical* facial expressions, representing *anger*, *disgust*, *fear*, *happiness*, *sadness* and *surprise*, plus the *neutral* one that are universally recognized and remain consistent across different ethnicities and cultures. The presence of these prototypical facial expressions is now widely accepted for scientific analysis. These results, in turn, inspired many researchers to analyze facial expressions in video data, by tracking facial features and measuring the amount of facial movements in video frames (Zeng et al., 2009). This body of work demonstrates a collective knowledge that facial expressions are highly dynamical processes, and looking at sequences of face instances can help to improve the recognition performance. We further emphasize that, rather than being just a static or dynamic 2D image analysis, it is more natural to analyze expressions as spatio-temporal deformations of 3D faces, caused by the actions of facial muscles. In this approach, the facial expressions can be studied comprehensively by analyzing temporal dynamics of 3D face scans (3D plus time is often regarded as 4D data). From this perspective the relative immunity of 3D scans to lighting conditions and pose variations give support for the use of 3D and 4D data. Motivated by these considerations, there has been a progressive shift from 2D to 3D in performing facial shape analysis for recognition and expression recognition. In particular, this latter research subject is gaining momentum thanks to the recent availability of public 3D datasets, like the *Binghamton University* 3D Facial Expression database (BU-3DFE) (Yin et al., 2006a), and the *Bosphorus* 3D Face Database (Savran et al., 2008). At the same time, advances in 3D imaging technology (Di3D, 2006, 3DMD, 2010) have permitted collections of large datasets that include temporal sequences of 3D scans (i.e., 4D datasets), such as the *Binghamton University* 4D Facial Expression database (BU-4DFE) (Yin et al., 2008), the 4D dataset constructed at *University of Central Lancashire* (Hi4D-ADSIP) (Matuszewski et al., 2011, 2012), and the dynamic 3D FACS dataset (D3DFACS) for facial expression research (Cosker et al., 2011), which also includes fully coded FACS. This trend has been strengthened further by the

introduction of inexpensive acquisition devices, such as the consumer 3D cameras like Kinect or Asus (Kinect, 2010, Point grey, 2010, Asus, 2010), that provide fast albeit low-resolution streams of 3D data to a large number of users, thus opening new opportunities and challenges in 3D face recognition and facial expression recognition (Berretti et al., 2012, Li et al., 2013).

4.2 Expression Classification from *Static* (3D) Faces

Facial expression recognition has been extensively studied over the past decades especially in 2D domain (e.g., images and videos) resulting in a valuable enhancement. Existing approaches that address facial expression recognition can be divided into three categories: (1) *static* vs. *dynamic*; (2) *global* vs. *local*; (3) *2D* vs. *3D*. Most of the approaches are based on feature extraction/detection as a mean to represent and understand facial expressions. Pantic and Rothkrantz (Pantic and Rothkrantz, 2000) and Samal and Iyengar (Samal and Iyengar, 1992) presented a survey where they explored and compared different approaches that were proposed, since the mid 1970s, for facial expression analysis from either static facial images or image sequences. Whitehill and Omlin (Whitehill and Omlin, 2006) investigated on the Local versus Global segmentation for facial expression recognition. In particular, their study is based on the classification of action units (AUs), defined in the well-known Facial Action Coding System (FACS) manual by Ekman and Friesen (Ekman and Friesen, 1978), and designating the elementary muscle movements involved in the bio-mechanical of facial expressions. They reported, in their study on face images, that the local expression analysis showed no consistent improvement in recognition accuracy compared to the global analysis.

4.2.1 Related Work

As for 3D facial expression recognition, the first work related to this issue was presented by Wang et al. (Wang et al., 2006b). They proposed a novel geometric feature based facial expression descriptor, derived from an estimation of primitive surface feature distribution. A labeling scheme was associated with their extracted features, and they constructed samples that have been used to train and test several classifiers. They reported that the highest average recognition rate they obtained was 83%. They evaluated their approach not only on frontal-view facial expressions of the BU-3DFE database, but they also tested its robustness to non-frontal views. A second work was reported by Soyel and Demirel (Soyel and Demirel, 2007) on the same database. They extracted six characteristic distances between eleven facial landmarks, and using Neural Network architecture that analysis the calculated distances, they classified the BU-3DFE facial scans into 7 facial expressions including neutral expression. The average recognition rate they achieved was 91.3%. Mpiperis et al. (Mpiperis et al., 2008b) proposed a joint 3D face and facial expression recognition using bilinear model. They fitted both formulations, using symmetric and asymmetric bilinear models to encode both identity and expression. They reported an average recognition rate of 90.5%. They also reported that the facial expressions of disgust and surprise

were well identified with an accuracy of 100%. Tang and Huang (Tang and Huang, 2008b) proposed an automatic feature selection computed from the normalized Euclidean distances between two picked landmarks from 83 possible ones. Using regularized multi-class AdaBoost classification algorithm.

In this first approach, we further investigate the problem of 3D identity-independent facial expression recognition. The main contributions of our approach are the following: (1) We propose a new process for representing and extracting patches on the facial surface scan that cover multiple regions of the face; (2) We apply a framework to derive 3D shape analysis to quantify similarity measure between corresponding patches on different 3D facial scans. Thus, we combine a local geometric-based shape analysis approach of 3D faces and several machine learning techniques to perform such classification.

4.2.2   Shape Analysis of Local Patches

Most of the earlier work in 3D shape analysis use shape descriptors such as curvature, crest lines, shape index (e.g., ridge, saddle, rut, dome, etc.). These descriptors are defined based on the geometric and topological properties of the 3D object, and are used as features to simplify the representation and thus the comparison for 3D shape matching and recognition tasks. Despite their rigorous definition, such features are computed based on numerical approximation that involves second derivatives and can be sensitive to noisy data. In case of 3D facial range models, the facial surface labeling is a critical step to describe the facial behavior or expression, and a robust facial surface representation is needed. In Samir et al. (Samir et al., 2006) the authors proposed to represent facial surfaces by an indexed collections of 3D closed curves on faces. These curves are level curves of a surface distance function (i.e., geodesic distance) defined to be the length of the shortest path between a fixed reference point (taken to be the nose tip) and a point of the extracted curve along the facial surface. This being motivated by the robustness of the geodesic distance to facial expressions and rigid motions. Using this approach they were able to compare 3D shapes by comparing facial curves rather than comparing corresponding shape descriptors.

In this approach, we intend to further investigate on local shapes of the facial surface. We are especially interested in capturing deformations of local facial regions caused by facial expressions. Using a different solution, we compute iso-level curves using the Euclidean distance function (which is sensitive to deformations and thus can better capture differences related to variant expressions). To this end, we choose to consider N reference points (landmarks) $\{r_l\}_{1 \leq l \leq N}$ (Figure 4.1 (a)) and associated sets of level curves $\{c_\lambda^l\}_{1 \leq \lambda \leq \lambda_0}$ (Figure 4.1 (b)). These curves are extracted over the patches centered at these points. Here λ stands for the value of the distance function between the reference point r_l and the point belonging to the curve c_λ^l, and λ_0 stands for the maximum value taken by λ. Accompanying each facial model there are 83 manually picked landmarks, these landmarks are practically similar to the MPEG-4 feature points and are selected based on the facial anatomy structure. Given these points the feature region on the face can be easily determined and extracted. We were interested in a subset of 68 landmarks laying
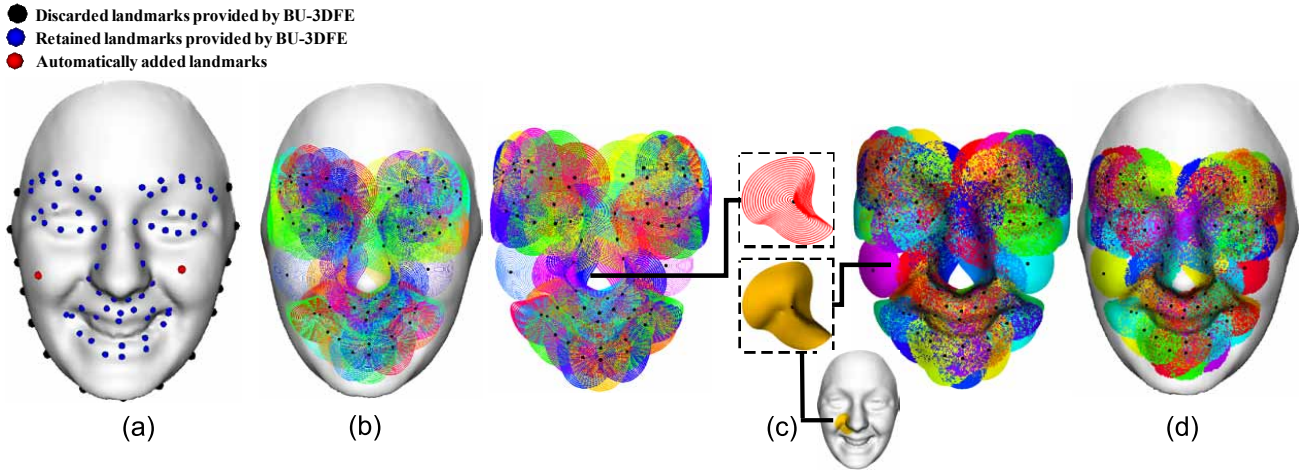
FIGURE 4.1: (a) 3D annotated facial shape model (70 landmarks); (b) 3D closed curves extracted around the landmarks; (c) 3D curve-based patches composed of 20 level curves with a size fixed by a radius $\lambda_0 = 20mm$; (d) Extracted patches on the face.

within the face area, discarding those marked on the face border. Contrary to the MPEG-4 feature points specification that annotates the cheeks center and bone, in BU-3DFE there were no landmarks associated with the cheek regions. Thus, we add two extra landmarks at both cheeks, obtained by extracting the middle point along the geodesic path between the mouth corner and the outside eye corner. We propose to represent each facial scan by a number of patches centered on the considered points. Let r_l be the reference point and P_l a given patch centered on this point and localized on the facial surface denoted by S. Each patch will be represented by an indexed collection of level curves. To extract these curves, we use the Euclidean distance function $\|r_l - p\|$ to characterize the length between r_l and any point p on S. Indeed, unlike the geodesic distance, the Euclidean distance is sensitive to deformations, and besides, it permits to derive curve extraction in a fast and simple way. Using this function we defined the curves as level sets of:

$$\|r_l - .\| : c_\lambda^l = \{p \in S \mid \|r_l - p\| = \lambda\} \subset S, \ \ \lambda \in [0, \lambda_0]. \tag{4.1}$$

Each c_λ^l is a closed curve, consisting of a collection of points situated at an equal distance λ from r_l. The Fig. 4.1 resumes the scheme of patches extraction.

Once the patches are extracted, we aim at studying their shape and design a similarity measure between corresponding ones on different scans under different expressions. This is motivated by the common belief that people smile, or convey any other expression, the same way, or more appropriately certain regions taking part in a specific expression undergo practically the same dynamical deformation process. We expect that certain corresponding patches associated with the same given expression will be deformed in a similar way, while those associated with two different expressions will deform differently. The following sections describe the shape analysis of closed curves in $\mathbb{R}^3$, initially introduced in (Joshi et al., 2007), and its extension to analyze shape of local patches on facial surfaces. We notice here that the shape analysis approach applied for curves here is different to the one described in Chapter 2, because of the

closure constraint. For this reason, we shall describe in details in the following sections the shape analysis approach for closed curves and the extension to facial patches.

4.2.2.1 Shape Analysis of 3D Closed Curves

We start by considering a closed curve β in $\mathbb{R}^3$. While there are several ways to analyze shapes of closed curves, an elastic analysis of the parametrized curves is particularly appropriate in 3D curves analysis. This is because (1) such analysis uses a square-root velocity function representation which allows us to compare local facial shapes in presence of elastic deformations, (2) this method uses a square-root representation under which the elastic metric reduces to the standard $\mathbb{L}^2$ metric and thus simplifies the analysis, (3) under this metric the Riemannian distance between curves is invariant to the re-parametrization. To analyze the shape of β, we shall represent it mathematically using a square-root representation of β (as described in Chapter 2) as follows ; for an interval $I = [0, 1]$, let $\beta : I \longrightarrow \mathbb{R}^3$ be a curve and define $q : I \longrightarrow \mathbb{R}^3$ to be its square-root velocity function (SRVF), given by:

$$q(t) \doteq \dot{\beta}(t)/\sqrt{\|\dot{\beta}(t)\|} \ . \tag{4.2}$$

Here t is a parameter $\in I$ and $\|.\|$ is the Euclidean norm in $\mathbb{R}^3$. We note that $q(t)$ is a special function that captures the shape of β and is particularly convenient for shape analysis, as we describe next. The classical elastic metric for comparing shapes of curves becomes the $\mathbb{L}^2$-metric under the SRVF representation (Srivastava et al., 2011). This point is very important as it simplifies the calculus of elastic metric to the well-known calculus of functional analysis under the $\mathbb{L}^2$-metric. Also, the squared $\mathbb{L}^2$-norm of q, given by: $\|q\|^2 = \int_{\mathbb{S}^1} < q(t), q(t) > dt = \int_{\mathbb{S}^1} \|\dot{\beta}(t)\| dt$, which is the length of β. In order to restrict our shape analysis to closed curves, we define the set: $\mathcal{C} = \{q : \mathbb{S}^1 \longrightarrow \mathbb{R}^3 | \int_{\mathbb{S}^1} q(t)\|q(t)\|dt = 0\} \subset \mathbb{L}^2(\mathbb{S}^1, \mathbb{R}^3)$. Notice that the elements of $\mathcal{C}$ are allowed to have different lengths. Due to a non-linear (closure) constraint on its elements, $\mathcal{C}$ is a non-linear manifold. We can make it a Riemannian manifold by using the metric: for any $u, v \in T_q(\mathcal{C})$, we define:

$$\langle u, v \rangle = \int_{\mathbb{S}^1} \langle u(t), v(t) \rangle \, dt \ . \tag{4.3}$$

So far we have described a set of closed curves and have endowed it with a Riemannian structure. Next we consider the issue of representing the *shapes* of these curves. It is easy to see that several elements of $\mathcal{C}$ can represent curves with the same shape. For example, if we rotate a curve in $\mathbb{R}^3$, we get a different SRVF but its shape remains unchanged. Another similar situation arises when a curve is re-parametrized; a re-parameterization changes the SRVF of curve but not its shape. In order to handle this variability, we define orbits of the rotation group $SO(3)$ and the re-parameterization group Γ as the equivalence classes in $\mathcal{C}$. Here, Γ is the set of all orientation-preserving diffeomorphisms of $\mathbb{S}^1$ (to itself) and the elements of Γ are viewed as re-parameterization functions. For example, for a curve $\beta : \mathbb{S}^1 \to \mathbb{R}^3$ and a function

$\gamma : \mathbb{S}^1 \rightarrow \mathbb{S}^1$, $\gamma \in \Gamma$, the curve $\beta \circ \gamma$ is a re-parameterization of β. The corresponding SRVF changes according to $q(t) \mapsto \sqrt{\dot{\gamma}(t)}q(\gamma(t))$. We set the elements of the orbit:

$$[q] = \{\sqrt{\dot{\gamma}(t)}Oq(\gamma(t))|O \in SO(3), \quad \gamma \in \Gamma\} , \qquad (4.4)$$

to be equivalent from the perspective of shape analysis. The set of such equivalence classes, denoted by $\mathcal{S} \doteq \mathcal{C}/(SO(3) \times \Gamma)$ is called the *shape space* of closed curves in $\mathbb{R}^3$. $\mathcal{S}$ inherits a Riemannian metric from the larger space $\mathcal{C}$ due to the quotient structure.

The main ingredient in comparing and analysing shapes of curves is the construction of a geodesic between any two elements of $\mathcal{S}$, under the Riemannian metric given in Eq.(4.3). Given any two curves β_1 and β_2, represented by their SRVFs q_1 and q_2, we want to compute a geodesic path between the orbits $[q_1]$ and $[q_2]$ in the shape space $\mathcal{S}$. This task is accomplished using a *path-straightening approach* which was introduced in (Klassen and Srivastava, 2006). The basic idea here is to connect the two points $[q_1]$ and $[q_2]$ by an arbitrary initial path α and to iteratively update this path using the negative gradient of an energy function $E[\alpha] = \frac{1}{2}\int_s \langle \dot{\alpha}(s), \dot{\alpha}(s) \rangle ds$. The interesting part is that the gradient of E has been derived analytically and can be used directly for updating α. As shown in (Klassen and Srivastava, 2006), the critical points of E are actually geodesic paths in $\mathcal{S}$. Thus, this gradient-based update leads to a critical point of E which, in turn, is a geodesic path between the given points. In the remainder of this section, we will use the notation $d_{\mathcal{S}}(\beta_1, \beta_2)$ to denote the length of the geodesic in the *shape space* $\mathcal{S}$ between the orbits q_1 and q_2, to reduce the notation.

4.2.2.2   Shape Analysis of 3D Facial Patches

Now, we extend ideas developed in the previous section from analyzing shapes of curves to the shapes of patches. As mentioned earlier, we are going to represent a number of l patches of a facial surface S with an indexed collection of the level curves of the $\|r_l - .\|$ function (Euclidean distance from the reference point r_l). That is, $P_l \leftrightarrow \{c_\lambda^l, \lambda \in [0, \lambda_0]\}$, where c_λ^l is the level set associated with $\|r_l - .\| = \lambda$. Through this relation, each patch has been represented as an element of the set $\mathcal{S}^{[0,\lambda_0]}$. In our framework, the shapes of any two patches are compared by comparing their corresponding level curves. Given any two patches P_1 and P_2, and their level curves $\{c_\lambda^1, \lambda \in [0, \lambda_0]\}$ and $\{c_\lambda^2, \lambda \in [0, \lambda_0]\}$, respectively, our idea is to compare the patches curves c_λ^1 and c_λ^2, and to accumulate these differences over all λ. More formally, we define a distance $d_{\mathcal{S}^{[0,\lambda_0]}}$ given by:

$$d_{\mathcal{S}^{[0,\lambda_0]}}(P_1, P_2) = \int_0^L d_{\mathcal{S}}(c_\lambda^1, c_\lambda^2)d\lambda . \qquad (4.5)$$

In addition to the distance $d_{\mathcal{S}^{[0,\lambda_0]}}(P_1, P_2)$, which is useful in biometry and other classification experiments, we also have a geodesic path in $\mathcal{S}^{[0,\lambda_0]}$ between the two points represented by P_1 and P_2. This
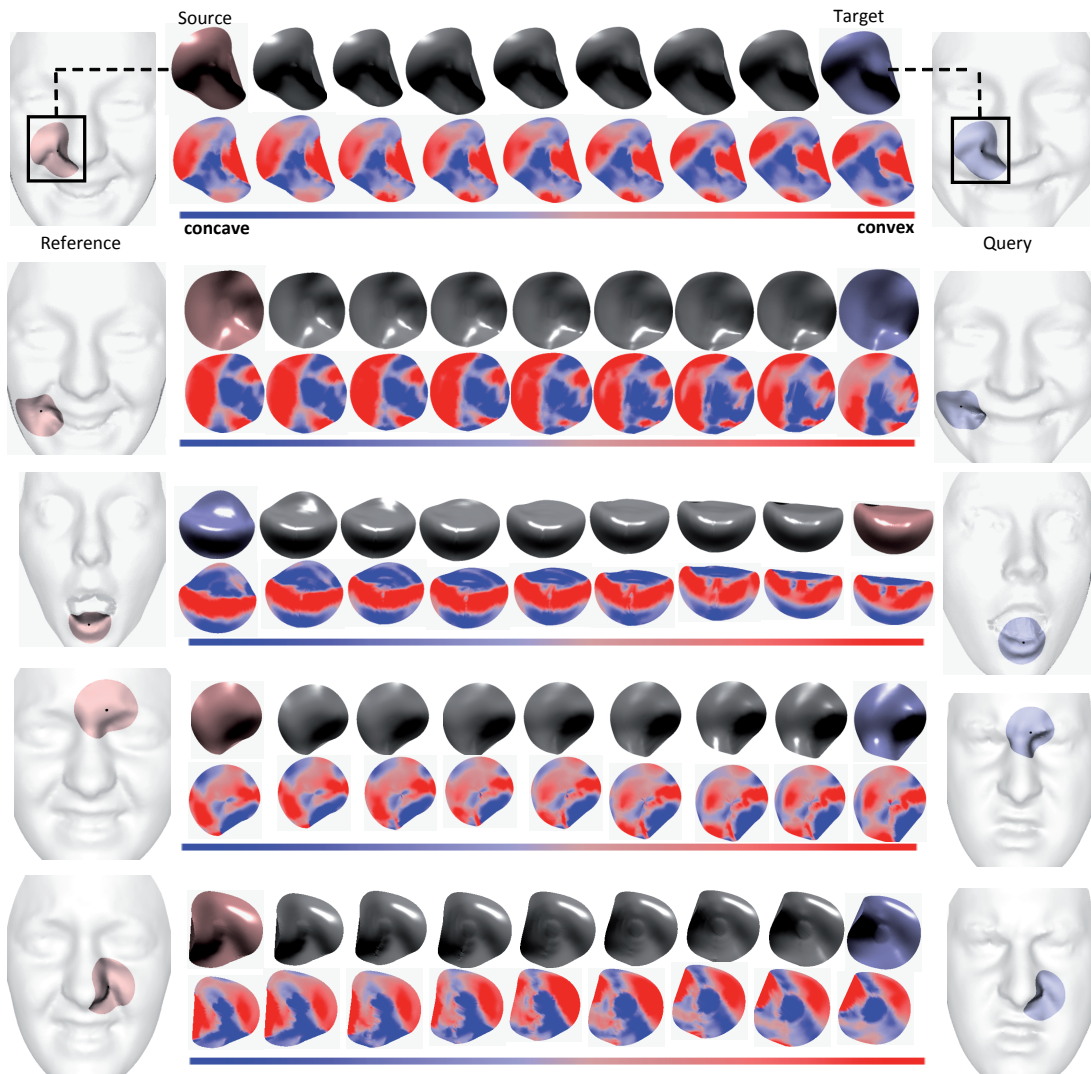
FIGURE 4.2: Examples of intra-class (same expression) geodesic paths with shape and mean curvature mapping between corresponding patches.

geodesic corresponds to the optimal elastic deformations of facial curves and, thus, facial surfaces from one to another. Figure 4.2 shows some examples of geodesic paths that are computed between corresponding patches associated with shape models sharing the same expression, and termed *intra-class geodesics*. In the first column we illustrate the source, which represents scan models of the same subject, but under different expressions. The third column represents the targets as scan models of different subjects. As for the middle column, it shows the geodesic paths. In each row we have both the shape and the mean curvature mapping representations of the patches along the geodesic path from the source to the target. The mean curvature representation is added to identify concave/convex areas on the source and target patches and equally-spaced steps of geodesics. This figure shows that certain patches, belonging to the same class of expression, are deformed in a similar way. In contrast, Figure 4.3 shows geodesic paths between patches of different facial expressions. These geodesics are termed *inter-class geodesics*. Unlike the intra-class geodesics shown in Figure 4.2, these patches deform in a different way.
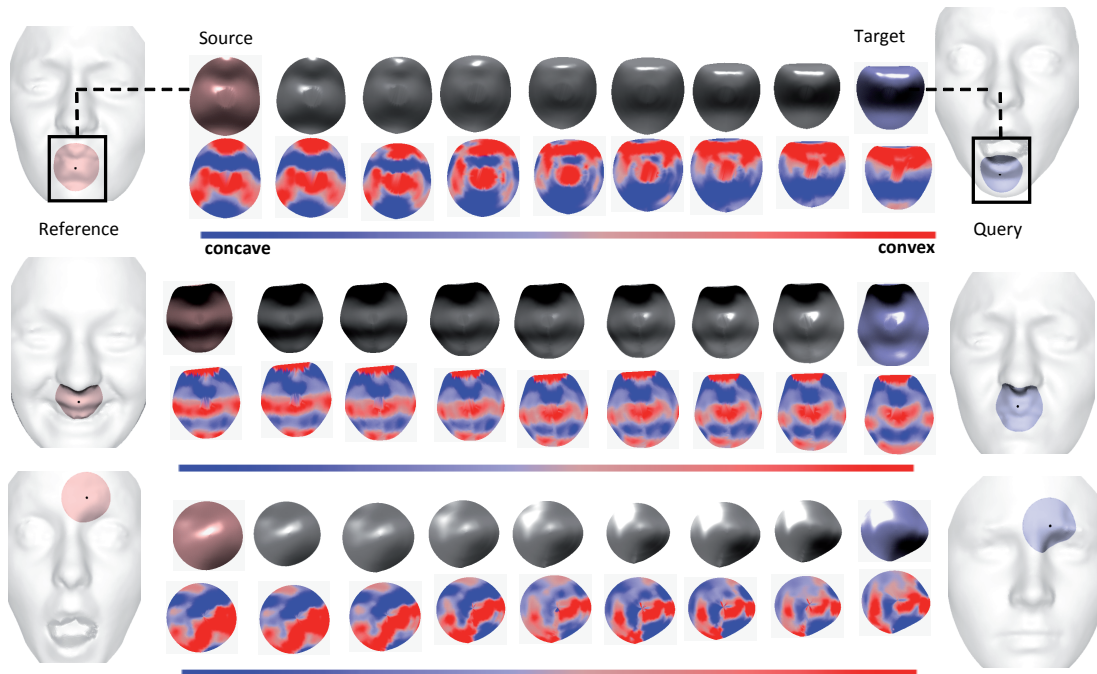
FIGURE 4.3: Examples of inter-class (different expressions) geodesic paths between source and target patches.

4.2.3 Facial Expression Classification

In order to classify expressions, we build a feature vector for each facial scan. Given a candidate facial scan of a person j, facial patches are extracted around facial landmarks. For a facial patch P_j^i, a set of level curves $\{c_\lambda\}_j^i$ are extracted centered on the ith landmark. Similarly, a patch P_{ref}^i is extracted in correspondence to landmarks of a reference scans *ref*. The length of the geodesic path between each level curve and its corresponding curve on the reference scan are computed using a Riemannian framework for shape analysis of 3D curves (see Sections 4.2.2.1 and 4.2.2.2). The shortest path between two patches at landmark i, one in a candidate scan and the other in the reference scan, is defined as the sum of the distances between all pairs of corresponding curves in the two patches as indicated in Eq. (4.5). The feature vector is then formed by the distances computed on all the patches and its dimension is equal to the number of used landmarks $N = 70$ (i.e., 68 landmarks are used out of the 83 provided by BU-3DFED and the two additional cheek points). The ith element of this vector represents the length of the geodesic path that separates the relative patch to the corresponding one on the reference face scan. All feature vectors computed on the overall dataset will be labeled and used as input data to machine learning algorithms such as Multi-boosting and SVM, where Multi-boosting is an extension of the successful Adaboost technique for forming decision committees.

4.2.4 Experimental Results

To investigate facial expression recognition, we have applied our proposed approach on a dataset that is appropriate for this task. In this Section, we describe the experiments, obtained results and comparisons with related work.

I. Database Description and Experimental Setting: BU-3DFE is one of the very few publicly available databases of annotated 3D facial expressions, collected by Yin et al. (Yin et al., 2006b) at Binghamton University. It was designed for research on 3D human face and facial expression and to develop a general understanding of the human behavior. Thus the BU-3DFE database is beneficial for several fields and applications dealing with human computer interaction, security, communication, psychology, etc. There are a total of 100 subjects in the database, 56 females and 44 males. A neutral scan was captured for each subject, then they were asked to perform six expressions namely: Happiness (HA), Anger (AN), Fear (FE), Disgust (DI), Sad (SA) and Surprise (SU). The expressions vary according to four levels of intensity (low, middle, high and highest or 01-04). Thus, there are 25 3D facial expression models per subject in the database. A set of 83 manually annotated facial landmarks is associated to each model. These landmarks are used to define the regions of the face that undergo to specific deformations due to single muscles movements when conveying facial expression (Ekman and Friesen, 1978). In Figure 4.4, we illustrate examples of the six universal facial expressions 3D models including the highest intensity level.

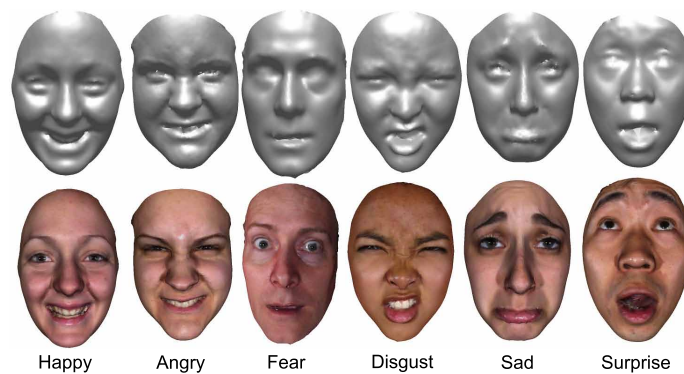| | Happy | Angry | Fear | Disgust | Sad | Surprise |

FIGURE 4.4: Examples of 3D facial expression models (first row 3D shape models, second row 3D textured models) of the BU-3DFE database.

For the goal of performing identity-independent facial expression recognition, the experiments were conducted on the BU-3DFE static database. A dataset captured from 60 subjects were used, half (30) of them were female and the other half (30) were male, corresponding to the high and highest intensity levels 3D expressive models (03-04). These data are assumed to be scaled to the true physical dimensions of the captured human faces. Following a similar setup as in (Gong et al., 2009b), we randomly divided the 60 subjects into two sets, the training set containing 54 subjects (648 samples), and the test set containing 6 subjects (72 samples). To drive the classification experiments, we arbitrarily choose a set of

six reference subjects with its six basic facial expressions. We point out that the selected reference scans do not appear neither in the training nor in the testing set.

II. Discussions of the Results: Using the *Waikato Environment for Knowledge Analysis (Weka)* (Hall et al., 2009), we applied the Multiboost algorithm with three weak classifiers, namely, Linear Discriminant Analysis (LDA), Naive Bayes (NB), and Nearest Neighbor (NN), to the extracted features, and we achieved average recognition rates of 98.81%, 98.76% and 98.07%, respectively. We applied the SVM linear classifier as well, and we achieved an average recognition rate of 97.75%. We summarize the resulting recognition rates in Table 4.1.

TABLE 4.1: Classification results using local shape analysis and several classifiers.

Classifier	Multiboost-LDA	Multiboost-NB	Multiboost-NN	SVM-Linear
Recognition rate	**98.81%**	98.76%	98.07%	97.75%

We note that these rates are obtained by averaging the results of the 10 independent and arbitrarily run experiments (10-fold cross validation) and their respective recognition rate obtained using the Multiboost-LDA classifier. We note that different selections of the reference scans do not affect significantly the recognition results and there is no large variations in recognition rates values. The reported results represent the average over the six runned experiments. The Multiboost-LDA classifier achieves the highest recognition rate and shows a better performance in terms of accuracy than the other classifiers. This is mainly due to the capability of the LDA-based classifier to transform the features into a more discriminative space and, consequently, result in a better linear separation between facial expression classes.

The average confusion matrix relative to the the best performing classification using Multiboost-LDA is given in Table 4.2.

TABLE 4.2: Average confusion matrix given by Multiboost-LDA classifier.

%	AN	DI	FE	HA	SA	SU
AN	**97.92**	1.11	0.14	0.14	0.69	0.0
DI	0.56	**99.16**	0.14	0.0	0.14	0.0
FE	0.14	0.14	**99.72**	0.0	0.0	0.0
HA	0.56	0.14	0.0	**98.60**	0.56	0.14
SA	0.28	0.14	0.0	0.0	**99.30**	0.28
SU	0.14	0.56	0.0	0.0	1.11	**98.19**

In order to better understand and explain the results mentioned above, we apply the Multiboost algorithm on feature vectors built from distances between patches for each class of expression. In this case, we consider these features as weak classifiers. Then, we look at the early iterations of the Multiboost algorithm and the selected patches in each iteration.
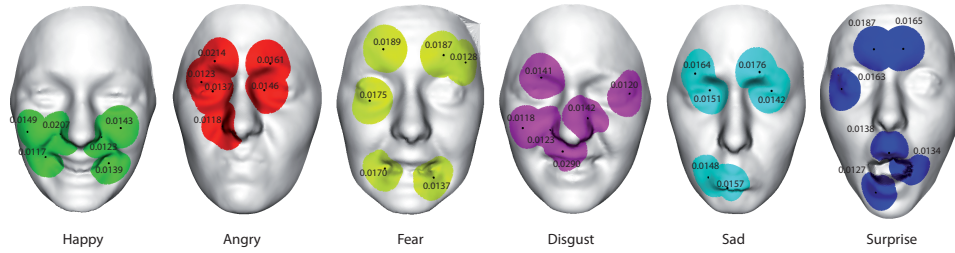
FIGURE 4.5: Selected patches at the early few iterations of Multiboost classifier for the six facial expressions (Angry, Disgust, Fear, Happy, Sadness, Surprise) with their associated weights.

Figure 4.5 illustrates for each class of expression the most relevant patches. Notice that, for example, for the Happy expression the selected patches are localized in the lower part of the face, around the mouth and the chin. As for the Surprise expression, we can see that most relevant patches are localized around the eyebrows and the mouth region. It can be seen that patches selected for each expression lie on facial muscles that contribute to this expression.

III. Comparison with Related Work: In Table 4.3 results of our approach are compared against those reported in (Tang and Huang, 2008b), (Soyel and Demirel, 2007), and (Wang et al., 2006b), on the same experimental setting (54-versus-6-subject partitions) of the BU-3DFE database. The differences between approaches should be noted: Tang et al. (Tang and Huang, 2008b) performed automatic feature selection using normalized Euclidean distances between 83 landmarks, Soyel et al. (Soyel and Demirel, 2007) calculated six distances using a distribution of 11 landmarks, while Wang et al. (Wang et al., 2006b) derived curvature estimation by locally approximating the 3D surface with a smooth polynomial function. In comparison, our approach capture the 3D shape information of local facial patches to derive shape analysis. For assessing how the results of their statistical analysis will generalize to an independent dataset, in (Wang et al., 2006b) a 20-fold cross-validation technique was used, while in (Tang and Huang, 2008b) and (Soyel and Demirel, 2007) the authors used 10-fold cross-validation to validate their approach.

TABLE 4.3: Comparison of this work with respect to previous work (Tang and Huang, 2008b), (Soyel and Demirel, 2007) and (Wang et al., 2006b).

Method	10-fold Average Recognition Rate (%)	20-fold ARR (%)
(Wang et al., 2006b)	—	83.6%
(Tang and Huang, 2008b)	95.1%	—
(Tang and Huang, 2008b)	95.1%	—
(Soyel and Demirel, 2007)	91.3%	—
(Maalej et al., 2011)	**98.81%**	**92.75%**

IV. Impact of Head Pose Variation: In real world situations, frontal view facial scans may not be always available. Thus, non-frontal view facial expression recognition is a challenging issue that needs to be treated. We were interested in evaluating our approach on facial scan under large pose variations. By rotating the 3D shape models in the y-direction, we generate facial scans under six different non-frontal views corresponding to $15°$, $30°$, $45°$, $60°$, $75°$ and $90°$ rotation. We assume that shape information is

unavailable for the occluded facial regions due to the face pose. For each view, we perform facial patches extraction around the visible landmarks in the given scan. In cases where a landmark is occluded, or where the landmark is visible, but the region nearby is partially occluded, we treat it as a missing data problem for all faces sharing this view. In these cases, we are not able to compute the geodesic path between corresponding patches. The corresponding entries in the distance matrix are blank and we fill them using an imputation technique (Batista and Monard, 2003). In our experiments we employed the mean imputation method, which consists of replacing the missing values by the means of values already calculated in frontal-view scenario obtained from the training set. Let $d_{ijk} = d_{\mathcal{S}^{[0,\lambda_0]}}(P_i^k, P_j^k)$ be the geodesic distance between the kth patch belonging to subjects i and j ($i \neq j$). In case of frontal view (fv), the set of instances X_i^{fv} relative to the subject i need to be labeled and is given by:

$$\mathbf{X_i^{fv}} = \begin{pmatrix} d_{i11} & \ldots & d_{i1k} & \ldots & d_{i1N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{ij1} & \ldots & d_{ijk} & \vdots & d_{ijN} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{iJ1} & \ldots & d_{iJk} & \ldots & d_{iJN} \end{pmatrix}$$

where N is the number of attributes. In case of non-frontal view (nfv), if an attribute k is missing, we replace the kth column vector in the distance matrix X_i^{nfv} by the mean of geodesic distances computed in the frontal-view case, with respect to the kth attribute and given by: $m_k^{fv} = \frac{\sum_{j=1}^{J} d_{ijk}}{J}$, where J is the total number of instances.

$$\mathbf{X_i^{nfv}} = \begin{pmatrix} d_{i11} & \ldots & m_k^{fv} & \ldots & d_{i1N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{ij1} & \ldots & m_k^{fv} & \vdots & d_{ijN} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{iJ1} & \ldots & m_k^{fv} & \ldots & d_{iJN} \end{pmatrix}$$

To evaluate the robustness of our approach in a context of non-frontal views, we derive a view-independent facial expression recognition. Error recognition rates are evaluated throughout different testing facial views using the four classifiers trained only on frontal-view facial scans. Figure 4.6 shows the average error rates of the four classification methods. The Multiboost-LDA shows the best performance for facial expression classification on the chosen database. From the figure, it can be observed that the average error rates increase with the rotation angle (values from 0 to 90 degrees of rotation are considered), and the Multiboost-LDA is the best performing methods also in the case of pose variations. As shown in this figure, recognition accuracy remains acceptable, even only 50% of data (half face) are available when we rotate the 3D face by 45 degree in y-direction.

In this section we presented a novel approach for identity-independent facial expression recognition from 3D facial shapes. Our idea was to describe the change in facial expression as a deformation in the vicinity of facial patches in 3D shape scan. An automatic extraction of local curve based patches
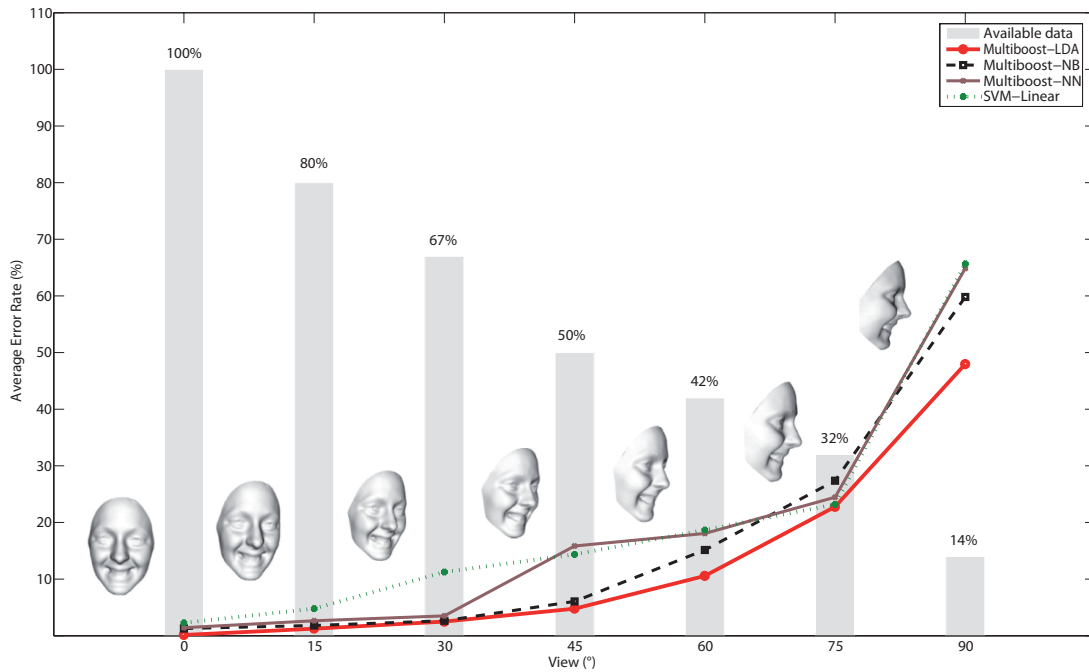
FIGURE 4.6: The average error rates of six expressions with different choice of views corresponding to the best reference and using different classifiers.

within the 3D facial surfaces was proposed. These patches were used as local shape descriptors for facial expression representation. A Riemannian framework was applied to compute the geodesic path between corresponding patches. Qualitative (inter and intra-geodesic paths) and quantitative (geodesic distances) measures of the geodesic path were explored to derive shape analysis. The geodesic distances between patches were labeled with respect to the six prototypical expressions and used as samples to train and test machine learning algorithms. Using Multiboost algorithm for multi-class classification, we achieved a 98.81% average recognition rate for six prototypical facial expressions on the BU-3DFE database. We demonstrated the robustness of the proposed method to pose variations. In fact, the obtained recognition rate remain acceptable (over 93%) even half of the facial scan is missed.

The major limitation of the approach presented in this section is that the 68 landmarks we used to define the facial patches were manually labeled. Several studies were interested detecting and tracking facial feature points, as proposed in (Gupta et al., 2010), (Sun et al., 2010b), for automatic 3D facial expression recognition. We will adopt a different approach (landmark-free) for the analysis of dynamic flows of facial surfaces.

4.3 Expression Classification from 3D *Dynamic* (4D) Faces

In his study, Ekman also showed that facial expressions can be coded through the movement of face points (temporal evolution of the face) as described by a set of **action units** (Ekman and Friesen, 1977).

These results, in turn, inspired many researchers to analyze facial expressions in video data, by tracking facial features and measuring the amount of facial movements in video frames (Zeng et al., 2009). This body of work demonstrates a collective knowledge that facial expressions are highly dynamical processes, and looking at sequences of face instances can help to improve the recognition performance. We further emphasize that, rather than being just a static or dynamic 2D image analysis, it is more natural to analyze expressions as spatio-temporal deformations of 3D faces, caused by the actions of facial muscles. In this approach, the facial expressions can be studied comprehensively by analyzing temporal dynamics of 3D face scans (3D plus time is often regarded as 4D data). From this perspective the relative immunity of 3D scans to lighting conditions and pose variations give support for the use of 3D and 4D data. Motivated by these considerations, there has been a progressive shift from 2D to 3D in performing facial shape analysis for recognition (Chapter 3), and expression recognition as described in Section 4.2. In particular, this latter research subject is gaining momentum thanks to the recent availability of public 3D datasets, like the *Binghamton University* 3D Facial Expression database (BU-3DFE) (Yin et al., 2006a), and the *Bosphorus* 3D Face Database (Savran et al., 2008).

At the same time, advances in 3D imaging technology (Di3D, 2006, 3DMD, 2010) have permitted collections of large datasets that include temporal sequences of 3D scans (i.e., 4D datasets), such as the *Binghamton University* 4D Facial Expression database (BU-4DFE) (Yin et al., 2008), the 4D dataset constructed at *University of Central Lancashire* (Hi4D-ADSIP) (Matuszewski et al., 2011, 2012), and the dynamic 3D FACS dataset (D3DFACS) for facial expression research (Cosker et al., 2011), which also includes fully coded FACS. This trend has been strengthened further by the introduction of inexpensive acquisition devices, such as the consumer 3D cameras like Kinect or Asus (Kinect, 2010, Point grey, 2010, Asus, 2010), that provide fast albeit low-resolution streams of 3D data to a large number of users, thus opening new opportunities and challenges in 3D face recognition and facial expression recognition (Berretti et al., 2012, Li et al., 2013). Motivated by these facts, we focus in this section on the problem of expression recognition from dynamic sequences of 3D facial scans. We propose a new framework for temporal analysis of 3D faces that combines scalar fields modeling of face deformations with effective classifiers. To motivate our solution and to relate it to the state of the art, next we provide an overview of existing methods for 4D facial expression recognition (see also the recent work in (Sandbach et al., 2012b) for a comprehensive survey on this subject), then we give a general overview of our approach.

4.3.1 Related Work

The use of 4D data for face analysis is still at the beginning, with just a few works performing face recognition from sequences of 3D face scans (Li et al., 2013, Benedikt et al., 2008, 2010), and some works focussing on facial expression recognition. In particular, the first approach addressing the problem of facial expression recognition from dynamic sequences of 3D scans was proposed by Sun et al. (Sun et al., 2010a, Sun and Yin, 2008). Their approach basically relies on the use of a generic deformable 3D model

whose changes are tracked both in space and time in order to extract a spatio-temporal description of the face. In the temporal analysis, a vertex flow tracking technique is applied to adapt the 3D deformable model to each frame of a 3D face sequence, and the vertex flow estimation is derived by establishing point-to-point correspondences between 3D meshes based on a conformal mapping approach. Correspondences between vertices across the 3D dynamic facial sequences provide a set of motion trajectories (vertex flow) of 3D face scans. Consequently, the vertex flow can be depicted on the adapted generic model (tracking model) through the estimation of the displacement vector from the tracked points of the current frame to the corresponding points of the first frame (assumed to have a neutral expression). A facial motion vector is then obtained to describe the dynamics of facial expression across a 3D frame sequence. In the spatial analysis, an automatic surface labelling approach is applied on the tracked locations of the depth models in order to classify the 3D primitive surface features into eight basic categories. As a result, each depth scan in the sequence can be represented by a spatio-temporal feature vector that describes both shape and motion information and provides a robust facial surface representation. Once spatio-temporal features are extracted, a two-dimensional Hidden Markov Model (HMM) is used for classification. In particular, a spatial HMM and a temporal HMM were used to model the spatial and temporal relationships between the extracted features. Exhaustive analysis was performed on the BU-4DFE database. The main limit of this solution resides in the use of the 83 manually annotated landmarks of the BU-4DFE that are not released for public use. The approach proposed by (Sandbach et al., 2011) exploits the dynamics of 3D facial movements to analyze expressions. This is obtained by first capturing motion between frames using Free-Form Deformations and extracting motion features using a quad-tree decomposition of several motion fields. GentleBoost classifiers are used in order to simultaneously select the best features to use and perform the training using two classifiers for each expression: one for the onset temporal segment, and the other for the offset segment. Then, HMMs are used for temporal modeling of the full expression sequence, which is represented as the composition of four temporal segments, namely, neutral, onset, apex, offset. These model a sequence with an initial neutral segment followed by the activation of the expression, the maximum intensity of the expression, deactivation of the expression and closing of the sequence again with a neutral expression. Experiments were reported for three prototypical expressions (i.e., happy, angry and surprise) of the BU-4DFE database. An extension of this work has been presented in (Sandbach et al., 2012b), where results on the BU-4DFE database using the six universal facial expressions are reported. In (Le et al., 2011) a level curve based approach is proposed to capture the shape of 3D facial models. The level curves are parameterized using the arclength function. The Chamfer distance is applied to measure the distances between the corresponding normalized segments, partitioned from these level curves of two 3D facial shapes. These features are then used as spatio-temporal features to train HMM, and since the training data were not sufficient for learning HMM, the authors proposed to apply the universal background modeling to overcome the over-fitting problem. Results were reported for the happy, sad and surprise sequences of the BU-4DFE database. Fang et al. (Fang et al., 2011) propose a fully automatic 4D facial expression recognition approach with a particular emphasis on 4D data registration and dense correspondence between 3D meshes along the temporal line. The variant of the Local Binary Patterns (LBP) descriptor proposed in (Zhao and Pietikäinen, 2007), which computes

LBP on three orthogonal planes is used as face descriptor along the sequence. Results are provided on the BU-4DFE database for all expressions and for the subsets of expressions used in (Sandbach et al., 2011) and (Le et al., 2011), showing improved results with respect to competing solutions. In (fan, 2012), the same authors propose a similar methodology for facial expression recognition from dynamic sequences of 3D scans, with an extended analysis and comparison of different 4D registration algorithms, including ICP and more sophisticated mesh matching algorithms, as Spin Images and MeshHOG. However, 12 manually annotated landmarks were used in this study. Recently, Reale et al. (Reale et al., 2013) have proposed a new 4D spatio-temporal feature named *Nebula* for facial expressions and movement analysis from a volume of 3D data. After fitting the volume data to a cubic polynomial, a histogram is built for different facial regions using geometric features, as curvatures and polar angles. They have conducted several recognition experiments on the BU-4DFE database for posing expressions, and on a new database published in (Zhang et al., 2013) for spontaneous expressions. However, manual intervention is used to detect the onset frame and just 15 frames from the onset one are used for classification, and these frames correspond to the most intense expression.

From the discussion above, it becomes clear that solutions specifically tailored for 4D facial expression recognition from dynamic sequences are still preliminary, being semi-automatic, or are capable of discriminating between only a subset of expressions.

4.3.2 Geometric Facial Deformation

The elastic Riemannian metric used in our methodology provides a physical interpretation of measuring deformations between facial curves using a combination of stretching and bending. These elastic deformations are captured by the Dense Scalar Fields used for expression classifications.
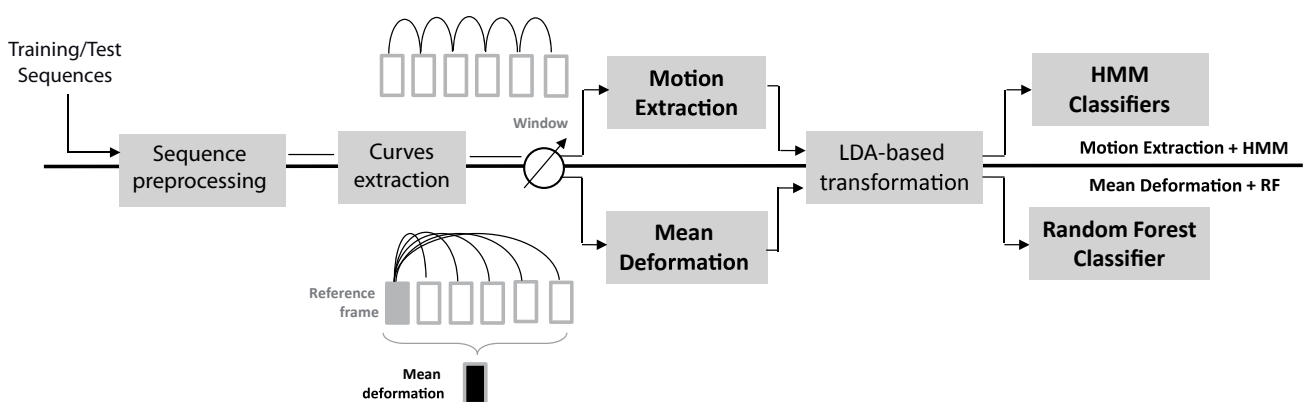

FIGURE 4.7: Overview of the proposed approach. Four main steps are shown: Sequence preprocessing and extraction of the radial curves; Motion extraction and Mean deformation computation; Dimensionality reduction with LDA; HMM- and Random-Forest-based classification. Note that both train and test sequences can go through the upper and lower path in the block-diagram.

Based on these premises, in this work we propose an automatic approach for facial expression recognition that exploits the facial deformations extracted from 3D facial videos. An overview of the proposed approach is given in Figure 4.7. In the preprocessing step, the 3D mesh in each frame is first aligned to the previous one and then cropped. From the obtained subsequence, the 3D deformation is captured based on a Dense Scalar Fields (DSFs) that represents the 3D deformation between two frames. Linear Discriminant Analysis (LDA) is used to transform derived feature space to an optimal compact space to better separate different expressions. Finally, the expression classification is performed in two ways: (1) using the HMM models for temporal evolution; and (2) using mean deformation along a window with Random Forest classifier. Experimental results show that the proposed approaches are capable of improving the state of art performance on the BU-4DFE database. There are three main contributions in this work,

↪ Novel Dense Scalar Fields (DSFs) defined on radial curves of 3D faces using Riemannian analysis in shape spaces of curves. These scalar fields accurately capture deformations occurring between 3D faces represented as collections of radial curves;

↪ A new approach for facial expression recognition from 3D dynamic sequences, that combines the high descriptiveness of DSFs extracted from successive 3D scans of a sequence with the discriminant power of LDA features using HMM and multi-class Random Forest;

↪ An extensive experimental evaluation that compares the proposed solution with the state of the art methods using a common dataset and testing protocols. Results show that our approach outperforms the published state of the art results.

One basic idea to capture facial deformations across 3D video sequences is to track mesh vertices densely along successive 3D frames. Since, as the resolution of the meshes varies across 3D video frames, establishing a dense matching on consecutive frames is necessary. For this purpose, (Sun and Yin, 2008) proposed to adapt a generic model (a tracking model) to each 3D frame using a set of 83 predefined facial landmarks to control the adaptation based on radial basis functions. The main limitation of this approach is that the 83 landmarks are manually annotated in the first frame of each sequence. Moreover, the adaptation decreases the accuracy of expression recognition when emotions are manifested by subtle changes of the face. A second solution is presented by (Sandbach et al., 2012a, 2011), where the authors used an existing non-rigid registration algorithm (FFD) (Rueckert et al., 1999a) based on B-splines interpolation between a lattice of control points. In this case, dense matching is a preprocessing step used to estimate a motion fields between 3D frames t and t-1. The problem of quantifying subtle deformations along the sequence still remains a challenging task, and the results presented in (Sandbach et al., 2011) are limited to just three facial expressions: *happy, angry* and *surprise*. In order to capture and model deformations of the face induced by different facial expressions, we propose to represent the facial surface through a set of parameterized radial curves that originate from the tip of the nose. Approximating the facial surface by an ordered set of radial curves, which locally captures its shape can be seen as a

parameterization of the facial surface. Indeed, similar parameterizations of the face have shown their effectiveness in facial biometrics (Drira et al., 2010a). The mathematical setup for the shape theory offered here comes from Hilbert space analysis (refer to Chapter 2 for DSFs computation).
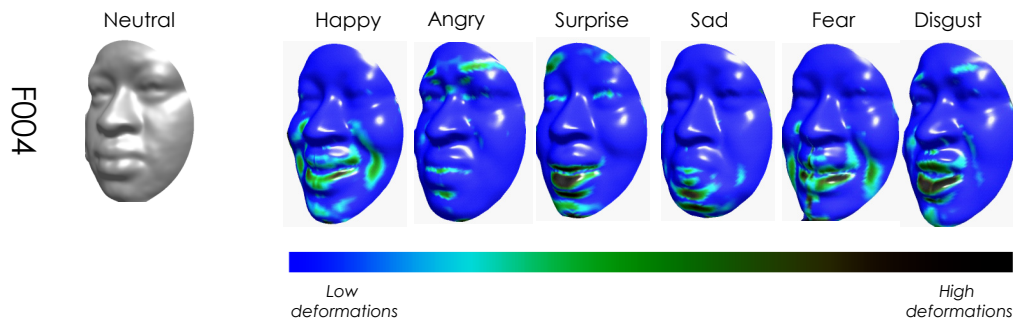

FIGURE 4.8: Optimal deformations (represented by the DFSs) computed between a neural face (on the left) and six apex (maximum expression) frames taken from the BU-4DFE database, each represent one expression from the set (happy, angry,surprise, sad, fear and disgust).

In our experiments, we represent each face with 100 radial curves, and T=50 sampled points on each curve, so that the DSFs between two 3D faces is expressed by a 5000-dimensional feature vector. In Figure 4.8 examples of the deformation fields computed between a neutral face of a given subject and the apex frames of the sequences of the six universal expressions of the same subject are shown. The values of the scalar fields to be applied on the neutral face to convey the six different prototypical expressions are reported using a color scale (blue to black). In particular, colors from green to black represent the highest deformations, whereas the lower values of the dense scalar fields are represented in cyan/blue. As it can be observed, for different expressions, the high deformations are located in different regions of the face. For example, as intuitively expected, the corners of the mouth and the cheeks are mainly deformed for happiness expression, whereas the eyebrows are also strongly deformed for the angry and disgust expressions.

In order to highlight the benefits of the proposed DSFs against other methods for extracting dense deformation features, we selected the Free-Form Deformation approach, which has been originally defined in Rueckert et al. (Rueckert et al., 1999b) for medical images, and later on successfully applied to the problem of 3D dynamic facial expression recognition by (Sandbach et al., 2012a, 2011). In particular, FFD is a method for non-rigid registration based on B-spline interpolation between a lattice of control points. In addition, we also compared our approach with respect to a baseline solution, which uses the point-to-point Euclidean distance between frames of a sequence. Figure 4.9 reports the results for an example case, where a frame of a happy sequence is deformed with respect to the first frame of the sequence. The figure shows quite clearly as the DSFs proposed in this work is capable to capture the face deformations with smooth variations that include, in the example, the mouth, the chin and the cheeks.
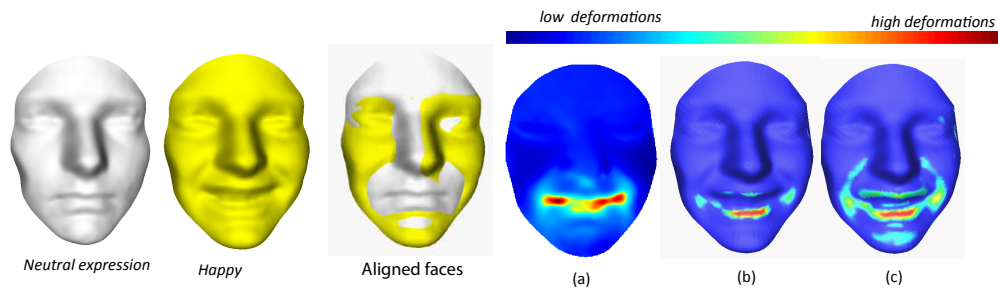
FIGURE 4.9: Capturing the facial deformations between a neutral face and an expressive face trough three methods: (a) the Free Form-based Deformations (FFD); (b) the point-to-point Euclidean distances; and (c) the DSFs deformations.

4.3.3 Expression Recognition using DSFs

Deformations due to facial expressions across 3D video sequences are characterized by subtle variations induced mainly by the motion of facial points. These subtle changes are important to perform effective expression recognition, but they are also difficult to be analyzed due to the face movements. To handle this problem, as described in the previous section, we propose a curve-based parametrization of the face that consists in representing the facial surface by a set of radial curves. According to this representation, the problem of comparing two facial surfaces, a reference facial surface and a target one, is reduced to the computation of the DSFs between them. In order to make possible to enter the expression recognition system at any time and make the recognition process possible from any frame of a given video, we consider subsequences of n frames. Thus, we chose the first n frames as the first subsequence. Then, we chose n-consecutive frames starting from the second frame as the second subsequence. This process is repeated by shifting the starting index of the sequence every one frame till the end of the sequence. In order to classify the resulting subsequences, we propose two different feature extraction and classification framework based on the DSFs:

↬ **Mean Deformation-based features associated to Random Forest classifier**. The first frame of the subsequence is considered as a reference frame and the deformation is calculated from each of the remaining frames to the first one using the DSFs. The average deformation of the n-1 resulting DSFs represents the feature vector in this classification scheme and is presented, after dimensionality reduction, to multi-class Random Forest classifiers;

↬ **3D Motion features combined with HMM classifiers**. The deformation between successive frames in a subsequence are calculated using the DSFs and presented to an HMM classifier preceded by LDA-based dimensionality reduction.

In the following, we shall describe the dynamic features derived from the DSFs.

4.3.3.1 Mean Shape Deformation with the Random Forest Classifier

The idea here is to capture a mean deformation of the face in the sliding window on the 3D expression sequence. In order to get this feature, the first frame of each subsequence is considered as the reference one, and the dense deformation is computed from this frame to each of the remaining frames of the subsequence. Let F_{ref} denote the reference frame of a subsequence and F_i the i-th successive frame in the subsequence; the successive frame, for example, is denoted by F_1. The DSFs is calculated between F_{ref} and F_i, for different values of i ($i = 1, \ldots, n-1$), and the mean deformation is then given by:

$$\overline{DSF} = \frac{1}{n-1} \sum_{i=1}^{n-1} DSF(F_{ref}, F_i). \tag{4.6}$$

Figure 4.10 illustrates one subsequence for each expression with $n = 6$ frames. Each expression is illustrated in two rows: The upper row gives the reference frame of the subsequence and the n-1 successive frames of the subsequences. Below, the corresponding Dense Scalar Fields computed for each frame are shown. The mean deformation fields is reported on the right of each plot and represents the feature vector for each subsequence. The feature vector for this subsequence is built based on the mean deformation of the n-1 calculated deformations. Thus, each subsequence is represented by a feature vector of size equal to the number of points on the face (i.e., the number of points used to sample the radial curves of the face). In order to provide a visual representation of the scalar fields, an automatic labeling scheme is applied: Warm colors (red, yellow) are associated with high $DSF(F_{ref}, F_t)$ values and correspond to facial regions affected by high deformations. Cold colors are associated with regions of the face that remain stable from one frame to another. Thus, this dense deformation fields summarize the temporal changes of the facial surface when a particular facial expression is conveyed.

According to this representation, the deformation of each subsequence is captured by the mean $\overline{DSF}$ defined in Eq. (4.6). The main motivation for using the mean deformation, instead of the maximum deformation for instance, is related to its greater robustness to the noise. In the practice, the mean deformation resulted more resistant to deformations due to, for example, inaccurate nose tip detection or the presence of acquisition noise. In Figure 4.10, for each subsequence, the mean deformation fields illustrate a smoothed pattern better than individual deformation fields in the same subsequence. Since the dimensionality of the feature vector is high, we use LDA-based transformation to map the present feature space to an optimal one that is relatively insensitive to different subjects, while preserving the discriminating expression information. LDA defines the within-class matrix S_w and the between-class matrix S_b. It transforms a n-dimensional feature to an optimized d-dimensional feature, where $d < n$. In our experiments, the discriminating classes are the 6 expressions, thus the reduced dimension d is 5.

For the classification, we used the multi-class Random Forest algorithm. The algorithm was proposed in (Breiman, 2001) and defined as a meta-learner comprised of many individual trees. It was designed to operate quickly over large datasets and more importantly to be diverse by using random samples to
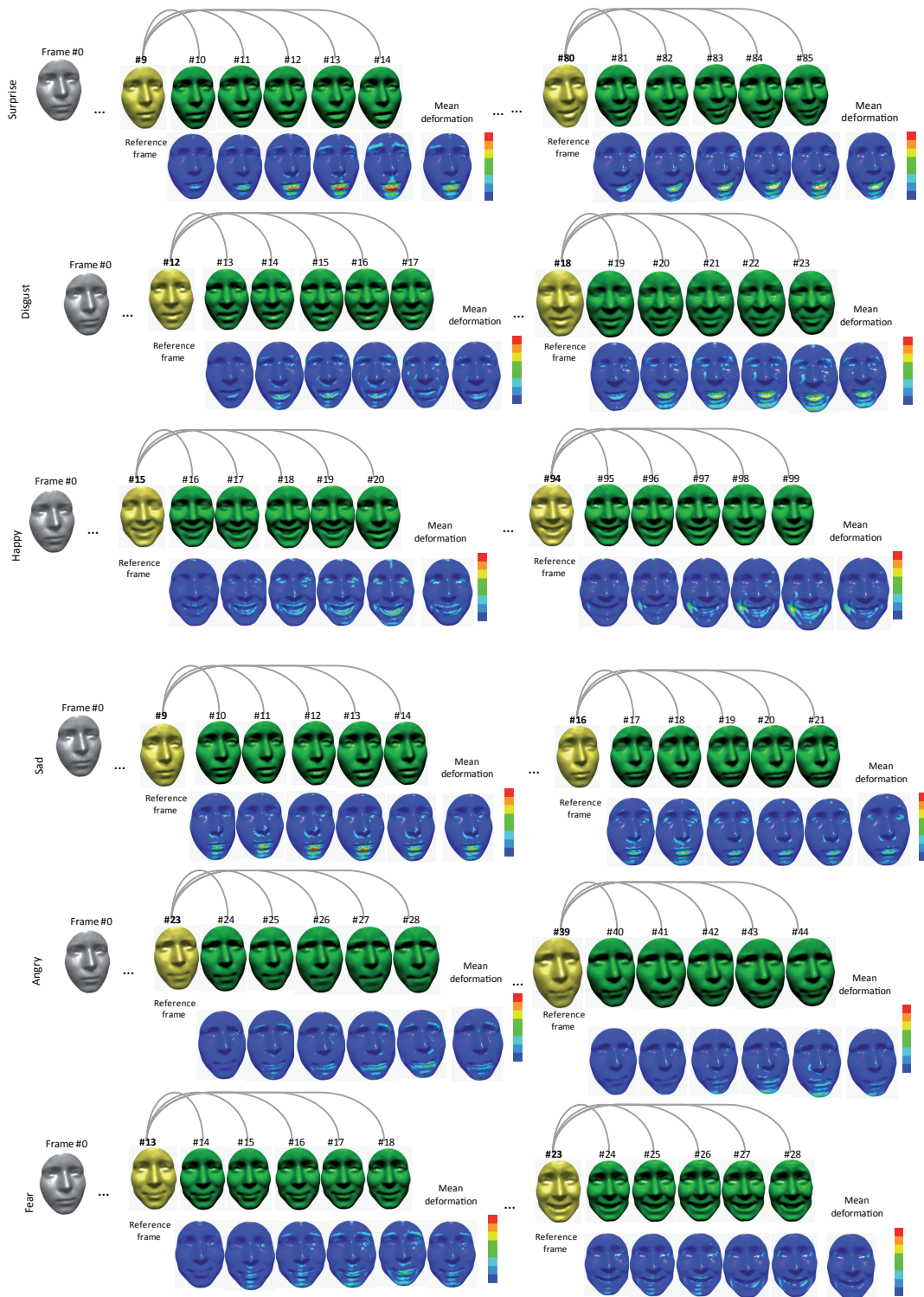
FIGURE 4.10: Computation of dynamic shape deformation on different subsequences taken from the BU-4DFE database. Each expression is illustrated by two rows: the upper one gives the reference frame of the subsequence and the *n*-1 successive frames. The corresponding deformation fields computed for each frame with respect to the reference one are illustrated in the lower row. The mean deformation fields are given on the right of each lower row.

build each tree in the forest. A tree achieves highly non-linear mappings by splitting the original problem into smaller ones, solvable with simple predictors. Each node in the tree consists of a test, whose result directs a data sample towards the left or the right child. During training, the tests are chosen in order to group the training data in clusters where simple models achieve good predictions. Such models are stored at the leaves, computed from the annotated data, which reached each leaf at train time. Once trained, a Random Forest is capable to classify a new expression from an input feature vector by putting it down each of the trees in the forest. Each tree gives a classification decision by voting for that class. Then, the forest chooses the classification having the most votes (over all the trees in the forest).

4.3.3.2 Motion Extraction with the HMM Classifier

The DSFs, can also be applied for expression recognition according to a different classification scheme. The deformations between successive frames in the subsequence are calculated using the DSFs. In particular, the deformation between two successive 3D frames is obtained by computing the pairwise Dense Scalar Fields $DSF(F_{t-1}, F_t)$ of correspondent radial curves. Based on this measure, we are able to quantify the motion of face points along radial curves and thus capture the changes in facial surface geometry.
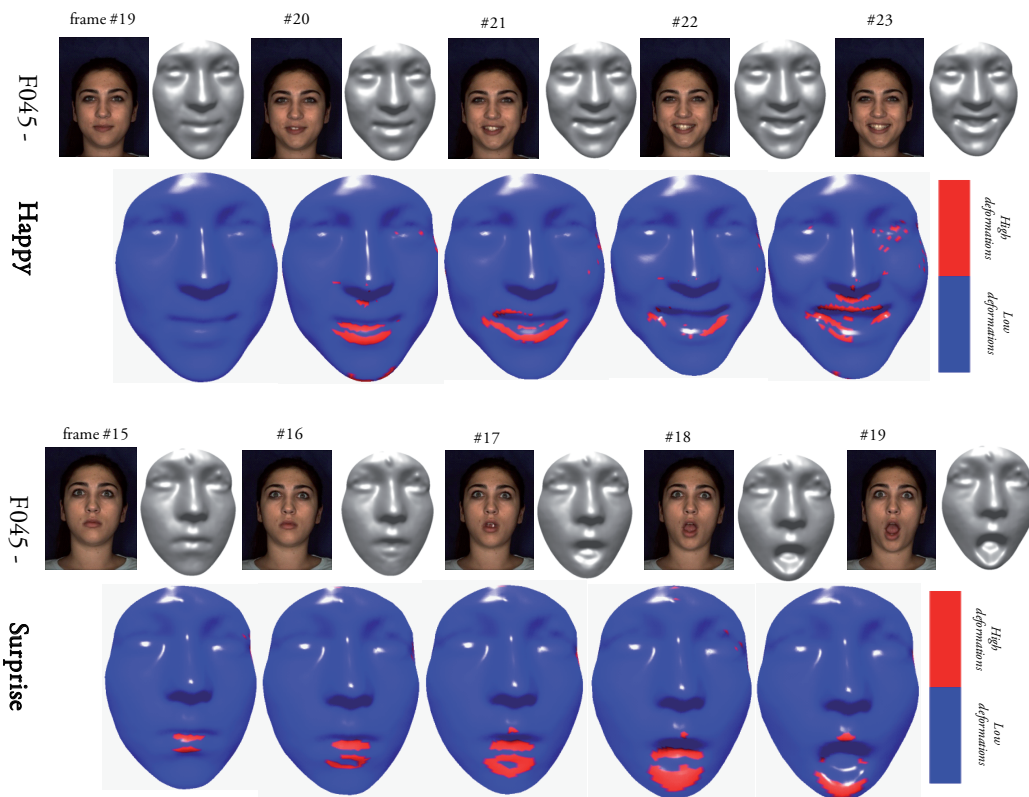

FIGURE 4.11: Examples of DSFs (deformations) between subsequent frames of 3D video sequences: Happy and surprise expressions are shown, respectively, on the left and right.

Figure 4.11 illustrates a direct application of the $DSF(F_{t-1}, F_t)$ and its effectiveness in capturing deformation between one facial surface to another belonging to two consecutive frames in a 3D video sequence. This figure shows two subsequences extracted from videos in the BU-4DFE database (happy and surprise expressions are shown on the left and on the right, respectively). For each sequence, the 2D image and the 3D scans of some frames are shown in the upper row. In the lower row, the deformation scalar fields $DSF(F_{t-1}, F_t)$ computed between consecutive frames (i.e., the current frame and the previous one) in the subsequence is reported. In order to provide a visual representation of the scalar fields, an automatic labeling scheme is applied that includes only two colors: The red color is associated with high $DSF(F_{t-1}, F_t)$ values and corresponds to facial regions affected by high deformations. The blue color is associated with regions that remain more stable from one frame to another. As illustrated in Figure 4.11, for different expressions, different regions are mainly deformed, showing the capability of the deformation fields to capture relevant changes of the face due to the facial expression. In particular, each deformation is expected to identify an expression, for example, as suggested by the intuition, the corners of the mouth and the cheeks are mainly deformed for the happiness expression.

With the proposed approach, the feature extraction process starts by computing for each 3D frame in a given video sequence the Dense Scalar Fields with respect to the previous one. In this way, we obtain as many fields as the number of frames in the sequence (decreased by one), where each field contains as many scalar values as the number of points composing the collection of radial curves representing the facial surface. In practice, the size of $DSF(F_{t-1}, F_t)$ is 1×5000, considering 5000 points on the face, similarly to the feature vector used in the first scheme of classification (mean deformation-based). Since the dimensionality of the resulting feature vector is high, also in this case we use LDA to project the scalar values to a 5-dimensional feature space, which is sensitive to the deformations induced by different expressions. The 5-dimensional *feature vector* extracted for the 3D frame at instant t of a sequence is indicated as f^t in the following. Once extracted, the feature vectors are used to train HMMs and to learn deformations due to expressions along a temporal sequence of frames.

In our case, sequences of 3D frames constitute the temporal dynamics to be classified, and each prototypical expression is modeled by an HMM (a total of 6 HMMs λ^{expr} is required, with $expr \in \{an, di, fe, ha, sa, su\}$). Four states per HMM are used to represent the temporal behavior of each expression. This corresponds to the idea that each sequence starts and ends with a neutral expression (state S_1). The frames that belong to the temporal intervals where the face changes from neutral to expressive and back from expressive to neutral are modeled by the *onset* (S_2) and *offset* (S_4) states, respectively. Finally, the frames corresponding to the highest intensity of the expression are captured by the apex state (S_3). This solution has proved its effectiveness in clustering the expressive states of a sequence also in other works (Sandbach et al., 2012a). Figure 4.12 exemplifies the structure of the HMMs used in our framework.

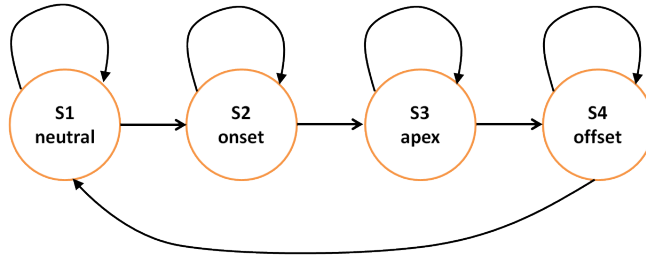The training procedure of each HMM is summarized as follows:

FIGURE 4.12: Structure of the HMMs modeling a 3D facial sequence. The four states model, respectively, the *neutral*, *onset*, *apex* and *offset* frames of the sequence. As shown, from each state it is possible to remain in the state itself or move to the next one (this is known as *Bakis* or left-right HMM).

↬ Feature vectors f^t of the training sequences are first clustered to identify a *codebook* of symbols using the standard LBG algorithm (Linde et al., 1980). This provides the required mapping between multidimensional feature vectors, taking values in a continuous domain, with the alphabet of symbols emitted by the HMM states;

↬ Expression sequences are considered as observation sequences $O = \{O^1, O^2, \ldots, O^T\}$, where each observation O^t at time t is given by the feature vector f^t;

↬ Each HMM λ^{expr} is initialized with random values and the *Baum-Welch* algorithm (Rabiner, 1989) is used to train the model using a set of training sequences. This estimates the model parameters, while maximizing the conditional probability $P(O|\lambda^{expr})$.

Given a 3D sequence to be classified, it is processed, so that each feature vectors f^t corresponds to a *test* observation $O = \{O^1 \equiv f^1, \ldots, O^T \equiv f^T\}$. Then, the test observation O is presented to the six HMMs λ^{expr} that model different expressions, and the *Viterbi* algorithm is used to determine the best *path* $\bar{Q} = \{\bar{q}^1, \ldots, \bar{q}^T\}$, which corresponds to the state sequence giving a maximum of likelihood to the observation sequence O. The likelihood along the best path, $p^{expr}(O, \bar{Q}|\lambda^{expr}) = \bar{p}^{expr}(O|\lambda^{expr})$ is considered as a good approximation of the true likelihood given by the more expensive *forward* procedure (Rabiner, 1989), where all the possible paths are considered instead of the best one. Finally, the sequence is classified as belonging to the class corresponding to the HMM whose log-likelihood along the best path is the greatest one.

4.3.4 Experimental Results

The proposed framework for facial expression recognition from dynamic sequences of 3D face scans has been experimented on the BU-4DFE database. Main characteristics of the database and results are reported in the following sections.

BU-4DFE Database: Description and Preprocessing

To investigate the usability and performance of 3D dynamic facial sequences for facial expression recognition, a dynamic 3D facial expression database has been created at *Binghamton University* (Yin et al., 2008). The Dimensional Imaging's 3D dynamic capturing system (Di3D, 2006), has been used to capture a sequence of stereo images and produce the depth map of the face according to a passive stereo-photogrammetry approach. The range maps are then combined to produce a temporally varying sequence of high-resolution 3D images with an RMS accuracy of 0.2mm. At the same time, 2D texture videos of the dynamic 3D models are also recorded. Each participant (subject) was requested to perform the six prototypical expressions (i.e., *angry*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*) separately. Each expression sequence contains neutral expressions in the beginning and the end, so that each expression was performed gradually from neutral appearance, low intensity, high intensity, and back to low intensity and neutral. Each 3D sequence captures one expression at a rate of 25 frames per second and each 3D sequence lasts approximately 4 seconds with about 35,000 vertices per scan (i.e., 3D *frame*). The database consists of 101 subjects (58 female and 43 male, with an age range of 18-45 years old) including 606 3D model sequences with 6 prototypical expressions and a variety of ethnic/racial ancestries (i.e., 28 Asian, 8 African-American, 3 Hispanic/Latino, and 62 Caucasian). More details on the BU-4DFE can be found in (Yin et al., 2008). An example of a 3D dynamic facial sequence of a subject with "happy" expression is shown in Figure 4.13, where 2D frames (not used in our solution) and 3D frames are reported. From left to right, the frames illustrate the intensity of facial expression passing from *neutral* to *onset*, *offset*, *apex* and *neutral* again.
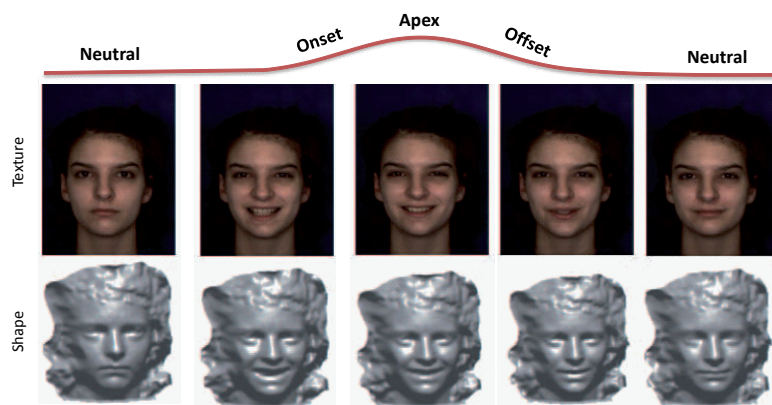

FIGURE 4.13: Examples of 2D and 3D frames extracted from a dynamic 3D video sequence of the BU-4DFE dataset.

It can be observed that the 3D frames present a near-frontal pose with some slight changes occurring mainly in the azimuthal plane. The scans are affected by large outliers, mainly acquired in the hair, neck and shoulders regions (see Figure 4.13). In order to remove these imperfections from each 3D frame the preprocessing pipeline described in Chapter 3 is performed.

Mean deformation-based Expression Classification

Following the experimental protocol proposed in (Sun and Yin, 2008), a large set of subsequences are extracted from the original expression sequences using a sliding window. The subsequences have been defined in (Sun and Yin, 2008) with a length of 6 frames with a sliding step of one frame from one subsequence to the following one. For example, with this approach, a sequence of 100 frames originates a set of $6 \times 95 = 570$ subsequences, each subsequence differing for one frame from the previous one. Each sequence is labelled to be one of the six basic expressions, thus extracted subsequences have the same label. This accounts for the fact that, in general, the subjects can enter the system not necessarily starting with a neutral expression, but with an arbitrary expression. The classification of these short sequences is regarded as an indication of the capability of the expression recognition framework to identify individual expressions. According to this, we first compute for each subsequence the Mean Deformation, which is then presented to multi-class Random Forest, as outlined in Sect. 4.3.3.

Thus, in the following we consider 40 trees and we report detailed results (confusion matrix) with this number of trees in Tab. 4.4. We recall that the rates are obtained by averaging the results of the 10-independent runs (10-fold cross validation). It can be noted that the largest confusions are between the *disgust (Di)* expression and the *angry (An)* and *Fear (Fe)* expressions. Interestingly, these three expressions capture negative emotive states of the subjects, so that similar facial muscles can be activated. The best classified expressions are *happy (Ha)* and *Surprise (Su)* with recognition accuracy of 95.47% and 94.53%, respectively. The standard deviation from the average performance is also reported in the table. The value of this statistical indicator suggests that small variations are observed between different folds.

TABLE 4.4: Confusion matrix for Mean Deformation and Random Forest classifier (for 6-frames window).

%	An	Di	Fe	Ha	Sa	Su
An	**93.11**	2.42	1.71	0.46	1.61	0.66
Di	2.3	**92.46**	2.44	0.92	1.27	0.58
Fe	1.89	1.75	**91.24**	1.5	1.76	1.83
Ha	0.57	0.84	1.71	**95.47**	0.77	0.62
Sa	1.7	1.52	2.01	1.09	**92.46**	1.19
Su	0.71	0.85	1.84	0.72	1.33	**94.53**
Average recognition rate =**93.21±0.81%**						

I. Effect of the Subsequence Size: We have also conducted additional experiments when varying the temporal size of the sliding window used to define the subsequences. In Figure 4.14, we report results for a window size equal to 2, 5 and 6, and using the whole length of the sequence (on average this is about 100 frames). From the figure, it clearly emerges that the recognition rate of the six expressions increases when increasing the temporal length of the window. This reveals the importance of the temporal dynamics and shows that the spatio-temporal analysis outperforms a spatial analysis of the frames. By considering the

whole sequences for the classification, the result reach 100%. In the work, we decided to report detailed results when considering a window length of 6-frames to allow comparisons with previous studies.
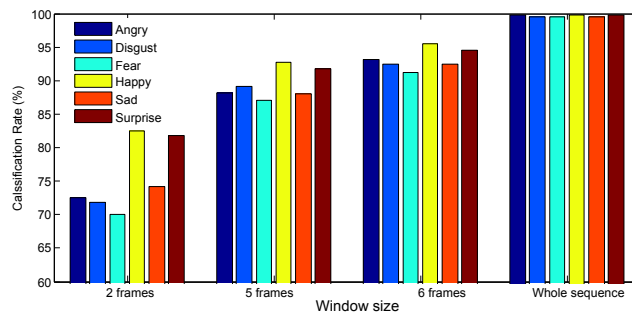

FIGURE 4.14: Effect of the temporal size of the sliding window on the results. The classification rates increase when increasing the length of the temporal window.

II. Effect of the Spatial Resolution of 3D Faces: In the proposed face representation, the DSFs are computed for the points of a set of radial curves originating from the nose tip. Due to this, the density of the scalar fields depend on the number of radial curves and the number of points per curve. So, the resolution used for the number of curves and points per curve can affect the final effectiveness of the representation. To investigated this aspect, we have conducted experiments when varying the spatial resolution of the 3D faces (i.e., the number of radial curves and the number of points per curve). Figure 4.15 expresses quantitatively the relationship between the expression classification accuracy (on the BU-4DFE) and the number of radial curves and the number of points per curve. This can give an indication of the expected decrease in the performance in the case the number of radial curves or points per curve is decreased due to the presence of noise and spikes in the data. From these results, we can also observe that the resolution in terms of number of curves has more importance than the resolution in terms of points per curve.
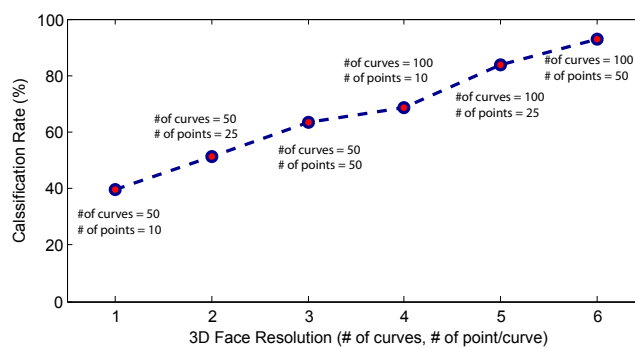

FIGURE 4.15: Effects of varying the 3D face resolution on the classification results.

HMM-based Expression Classification

Following the same setup as in previous section (originally defined in (Sun and Yin, 2008)), for this experiment we trained the HMMs on 6 frames subsequences constructed as discussed above. The 4-state structure of the HMMs resulted adequate to model the subsequences. Also in this experiment, we performed 10-folds cross validation on the overall set of subsequences derived from the 60×6 sequences (31970 in total). The achieved results by classifying individual subsequences of the expression sequences (*frame-by-frame* experiment) are reported in the confusion matrix of Tab. 4.5. Values in the table have been obtained by using features of 6-frames subsequences as input to the 6 HMMs and using the maximum emission probability criterion as decision rule. It is clear that the proposed approach is capable to accurately classify individual frames by analyzing the corresponding subsequence of previous 5 frames. The average recognition rate is equal to 93.83%, slightly higher than the one displayed by Mean Deformation plus Random Forest classification schema (though the standard deviation among different folds shows a greater value in this case). It can also be noted that, compared to the previous classifier, the same tendency of recognition rates is in general achieved. In fact, correct classification of *angry* is high despite the difficulty of this expression analysis. This learning scheme achieved better recognition than the first one for *angry (An)* expression. Actually, whereas the *angry (An)* expression is known for its subtle motions, our classifier achieved 93.95% of correct classification, which demonstrates the ability of the proposed DSFs to capture subtle deformations across the 3D sequences. These similar good achievements are mainly the effect of the proposed deformation scalar fields.

TABLE 4.5: Confusion matrix for Motion extraction and HMM classifiers (for 6-frames window).

%	An	Di	Fe	Ha	Sa	Su
An	**93.95**	1.44	1.79	0.28	2.0	0.54
Di	3.10	**91.54**	3.40	0.54	1.27	0.15
Fe	1.05	1.42	**94.55**	0.69	1.67	0.62
Ha	0.51	0.93	1.65	**94.58**	1.93	0.40
Sa	1.77	0.48	1.99	0.32	**94.84**	0.60
Su	0.57	0.38	3.25	0.38	1.85	**93.57**
*Average recognition rate = **93.83±1.53%***						

4.3.5 Discussion and Comparative Evaluation

To the best of our knowledge, the works reporting results on expression recognition from dynamic sequences of 3D scans are those in (Le et al., 2011, Sun et al., 2010a, Sandbach et al., 2012b, fan, 2012), and recently (Reale et al., 2013). These works have been evaluated on the BU-4DFE dataset, but the testing protocols used in the experiments are sometimes different, so that a direct comparison of the results reported in these papers is not immediate. In the following, we discuss these solutions with respect to our proposal, also evidencing the different settings under which the expression recognition results are obtained.

TABLE 4.6: Comparison of this work to earlier studies. Protocol description: #subjects (S), #expressions (E), Win size (Win). T: temporal only/S-T: spatio-temporal. Accuracy on sliding window/whole sequence (or subsequence).

Authors	Method	Features	Classification	Protocol	T/S-T	RR (%)
(Sun and Yin, 2008)	MU-3D	12 Motion Units	HMM	60 S, 6 E, Win=6	T	70.31, —
(Sun and Yin, 2008)	T-HMM	Tracking model	HMM	60 S, 6 E, Win=6	T	80.04, —
(Sun and Yin, 2008)	P2D-HMM	Curvature + Tracking model	T-HMM + S-HMM	60 S, 6 E, Win=6	S-T	82.19, —
(Sun and Yin, 2008)	R-2DHMM	Curvature + Tracking model	2D-HMM	60 S, 6 E, Win=6	S-T	90.44, —
(Sandbach et al., 2011)	3D Motion-based	FFD + Quad-tree	GentleBoost + HMM	—, 3 E, Win=4	T	73.61, 81.93
(Sandbach et al., 2012b)	3D Motion-based	FFD + Quad-tree	GentleBoost + HMM	—, 6 E, variable Win	T	64.6, —
(Le et al., 2011)	Level curve-based	pair- and segment-wise distances	HMM	60 S, 3 E, —	S-T	—, 92.22
(Fang et al., 2011, fan, 2012)	AFM Fitting	LBP-TOP	SVM-RBF	100 S, 6 E, —	T	—, 74.63
(Fang et al., 2011, fan, 2012)	AFM Fitting	LBP-TOP	SVM-RBF	100 S, 3 E, —	T	—, 96.71
(Reale et al., 2013)	Spatio-temporal volume	"Nebula" Feature	SVM-RBF	100 S, 6 E, Win=15	S-T	—, 76.10
(Ben Amor et al., 2014a)	Geometric Motion Extraction	3D Motion Extraction	LDA-HMM	60 S, 6 E, Win=6	T	93.83, —
(Ben Amor et al., 2014a)	Geometric Mean Deformation	Mean Deformation	LDA-Random Forest	60 S, 6 E, Win=6	T	93.21, —

Table 4.6 summarizes approaches and results reported previously on the BU-4DFE dataset, compared to those obtained in this work. The testing protocols used in the experiments are quite different especially the number of verified expressions, all the six basic expressions in (Fang et al., 2011, Sun et al., 2010a, Sun and Yin, 2008, fan, 2012), and (Reale et al., 2013) whereas (Le et al., 2011, Sandbach et al., 2011) reported primary results on only three expressions. The number of subjects considered is 60, except in (Sandbach et al., 2011) where the number of subjects is not specified. In general, sequences in which the required expressions are acted accurately are selected, whereas in (Fang et al., 2011) and (fan, 2012) 507 sequences out of the 606 total are used for all subjects. In our experiments, we conducted tests by following the same setting proposed by the earliest and more complete evaluation described in (Sun and Yin, 2008). The training and the testing sets were constructed by generating subsequences of 6-frames from all sequences of 60 selected subjects. The process were repeated by shifting the starting index of the sequence every one frame till the end of the sequence. We note that the proposed approaches outperforms state-of-the-art solutions following similar experimental settings. The recognition rates reported in (Sun and Yin, 2008) and (Sun et al., 2010a) based on temporal analysis only was 80.04% and spatio-temporal analysis was 90.44%. In both studies subsequences of constant window width including 6-frames ($win = 6$) is defined for experiments. We emphasize that their approach is not completely automatic requiring 83 manually annotated landmarks on the first frame of the sequence to allow accurate model tracking. The method proposed in (Sandbach et al., 2011) and (Sandbach et al., 2012b) is fully automatic with respect to the processing of facial frames in the temporal sequences, but uses *supervised* learning to annotate individual frames of the sequence in order to train a set of HMMs. Though performed off-line, supervised

learning requires manual annotation and counting on a consistent number of training sequences that can be a time consuming operation. In addition, a drawback of this solution is the computational cost due to Free-Form Deformations based on B-spline interpolation between a lattice of control points for nonrigid registration and motion capturing between frames. Preliminary tests were reported on three expressions: *(An)*, *(Ha)* and *(Su)*. Authors motivated the choice of the happiness and anger expressions with the fact that they are at either ends of the valence expression spectrum, whereas surprise was also chosen as it is at one extreme of the arousal expression spectrum. However, these experiments were carried out on a subset of subjects accurately selected as acting out the required expression. Verification of the classification system was performed using a 10-fold cross-validation testing. On this subset of expressions and subjects, an average expression recognition rate of 81.93% is reported. In (Sandbach et al., 2012b), the same authors have reported 64.6% classification rate when in their evaluation they consider all the six basic expressions. In (Le et al., 2011) a fully automatic method is also proposed, that uses an *unsupervised* learning solution to train a set of HMMs (i.e., annotation of individual frames is not required in this case). Expression recognition is performed on 60 subjects from the BU-4DFE database for the expressions of *happiness*, *sadness* and *surprise*. The recognition accuracy averaged on 10 rounds of 10-fold cross-validation show an overall value of 92.22%, with the highest performance of 95% obtained for the happiness expression. However, the authors reported recognition results on whole facial sequences, but this hinders the possibility of the methods to adhere to a real-time protocol. In fact, reported recognition results depends on the preprocessing of whole sequences unlike our approach and the one described in (Sun and Yin, 2008), which are able to provide recognition results when processing very few 3D frames. In (Fang et al., 2011) and (fan, 2012), results are presented for expression recognition accuracy on 100 subjects picked out from BU-4DFE database. However, 507 sequences are selected manually according to the following criteria: (1) the 4D sequence should start by neutral expression, and (2) sequences containing corrupted meshes are discarded. In addition, to achieve recognition rate of 75.82%, whole sequences should be analyzed. The authors reported highest recognition rates when only (Ha), (An), and (Su) expressions (96.71%) or (Ha), (Sa) and (Su) (95.75%) are considered. The protocol used in (Reale et al., 2013) is quite different from the others. First, the onset frame for each of the six canonical expressions has been marked manually on each sequence of the BU-4DFE database. Then, a fixed size window of 15 frames starting from the onset frame has been extracted from each expression of 100 subjects. So, although sequences from 100 subjects are used by this approach, it also uses a manual intervention to detect the onset frame and just 15 frames from the onset one are used for the classification (and these typically correspond to the most intense expression, including the apex frames).

According to this comparative analysis, the proposed framework compares favorably with state-of-the-art solutions. It consists of two geometric deformation learning schemes with a common feature extraction module (DSFs). This demonstrates the effectiveness of the novel mathematical representation called Dense Scalar Fields (DSFs), under the two designed schemes.

4.4   Conclusions

In this chapter, our interest was to effectively quantify the facial deformations to perform facial expression classification. Two different methodologies have been presented – **(1)** a **landmark-dependent** approach using 3D facial patches extracted around the landmarks and grounding on shape analysis of the patched using iso-level closed curves representation; and **(2)** a **landmark-free** approach based on shape analysis of open radial curves. While in the first approach, the use of static 3D data is explored, the geometric deformations from dynamic data is investigated in the second approach. From a methodological point-of-view, (1) proposed such representation using facial patches to cover important Action Units (AU) on the face, then classify the expression according to similar AU movements. Here the correspondence between facial patches is given explicitly by the correspondence between landmarks. In (2) an holistic representation of the face through multiple radial elastic (open) curves handle with the problem of correspondence and allow dense correspondence and tracking across the 3D video. Then two learning schemes of the geometric deformations serve to perform expression recognition.

Chapter 5

Facial Soft-biometrics Estimation

In this Chapter[1], we explore the contribution of the 3D facial geometry to the task of recognizing facial attributes, such as the gender and the age. There is a growing interest to soft biometrics recognition as means of improving automated face recognition since they hold the promise of significantly reducing recognition errors by filtering the number of entries in a database to be searched. For example, if the user can somehow be identified as a middle-aged Asian female, the search can be restricted only to the subjects with this profile enrolled in the database. There are other uses of automatic soft-biometrics estimation ranging from Human-Machine Interaction (HMI) to customize interfaces or services regarding the user's age or gender. It includes also the commercial use when limiting underage alcohol buyer, for instance, and suggest customized advertising. Most of the previous studies investigated the use of color data to make solft-biometrics estimation, only a few work have explored the role of the geometry (or the shape). In this chapter, we demonstrate the richness and the relevance of use 3D shape data to tackle these problems. Using Morphology-inspired descriptions related to the human differences extracted using our Riemannian framework presented in Chapter 2. To our knowledge, we conducted the first study on age estimation (Xia et al., 2014c) and joint soft-biometrics estimation (Xia et al., 2014d)(**[J3-s]**) from 3D faces. Our methodology to quantify the morphological divergences between age-groups or genders is one of the most original points in this work.

5.1 Introduction

Gender, Ethnicity, and Age are natural recognizable attributes in human faces. In their daily life, human beings are performing their estimation naturally and effectively, from the face. In anthropometry studies[2] (Ziqing et al., 2010), it has also been revealed that significant facial morphology differences

---

[1]The content of this chapter is related to our published conference papers (Xia et al., 2013b) and (Xia et al., 2014c) and submitted journal papers (Xia et al., 2014b)(**[J1-s]**) and (Xia et al., 2014d)(**[J3-s]**).

[2]refers to the measurement of the human individual.

exist in different gender, ethnicity and age groups. In sexual dimorphism studies[3] (Vicki et al., 1993), researchers have found that male faces usually possess more prominent features than female faces. Male's face usually has a more protuberant nose, eyebrows, more prominent chin and jaws. The forehead is more backward sloping, and the distance between top-lip and nose-base is longer. Research presented in (Ziqing et al., 2010) have also demonstrated that females are smaller in all the concerned anthropometric measurements. When studying the ethnicity differences in face (Farkas LG, 2005), researchers found that compared with the North America Whites, the Asian population usually have broader faces and noses, farther apart eyes, and exhibit the greatest difference in the anatomical orbital regions (around eyes and eyebrows). The clinical study reported in (Alphonse et al., 2013) have revealed that the Caucasians has significantly lower fetal frontomaxillary facial angle (FMFA) measurements than Asians. In (Ziqing et al., 2010), 16 anthropometric measurements have been recognized as significantly different between Asian and Caucasian faces. In face aging studies (Rhodes, 2009, Ramanathan et al., 2009), researchers have concluded that, the craniofacial growth is the main change in baby and adolescent faces, which results in the re-sizing and redistribution of facial features. In this period, the bigger the size of the face, the larger the estimated age. When the craniofacial growth stops at about 18 years old, the face contour and texture changes become the dominant changes. In addition, young adults tend to have a more triangle shaped face with a small amount of wrinkles. In contrast, older adults are usually associated with a U-shaped face with significant wrinkles.

Besides the perception of these soft-biometrics[4] in the face, gender, ethnicity and age also interact with each other in characterizing the face shape (Ziqing et al., 2010). For example, according to the anthropometric studies above, the shape of the nose is influenced by all the three attributes. Female faces usually look smoother and younger than male faces, and the Asian faces usually look younger than others (Yukio et al., 2003). In (Vignali et al., 2003), the authors have demonstrated both visually and quantitatively that ethnicity and gender are correlated to some extend in 3D face. In (Gao and Ai, 2009), the authors found that the gender classifier trained on a specific ethnicity could not perform better generalization ability on other ethnicity.

5.2 Related Work

From the analysis above, it is clear that the face demonstrates significant cues to recognize gender, ethnicity and age. They are also correlated with each other. In the literature of facial soft-biometrics recognition, abundant works have been proposed for gender classification, ethnicity classification and age estimation, taken separately. These works are done either with features from the 2D face texture (2D texture-based), or features from the 3D face shape (3D shape-based), or features from both of the texture and the shape of face (texture and shape-based). By definition, the face texture represents the reflection

---

[3] http://www.virtualffs.co.uk/
[4] A.K. Jain defined Soft-biometrics as a set of traits providing information about an individual, though these traits are not able to individually authenticate the subject because they lack in distinctiveness and permanence (Jain et al., 2004).

and absorption effects of external illumination caused by the facial skin, while the 3D face shape define the border which distinguishes the face and the environment.

In 2D-based gender classification, (Ylioinas et al., 2011) combine the Contrast Information (strength of patterns) and the Local Binary Patterns (LBP). (Yang et al., 2011) combine Active Appearance Models (AAM) and LBP with sequence selection algorithms. (Shan, 2012) selects discriminating LBP features with Adaboost. (Kumar et al., 2008) use describable visual attributes in face as features. (Wang et al., 2012) enhance the performance of LBP with one of its variants, named Local Circular Patterns (LCP). (Makinen and Raisamo, 2008) compare different gender classification methods, and reveal that the database, normalization, hair, and experimental settings account more for the results, than the classifiers. (Gao and Ai, 2009), learn ethnicity-specific gender classifiers and achieve higher overall classification results on a collection of 2D images. (Giovanna and Jean-Luc, 2012) perform ethnicity-specific gender classification on the 2D FERET and TRECVID datasets. Unexpectedly, they found that the ethnicity information is not helpful in gender classification. Using the 3D shape, (Liu and Palmer, 2003) extract features from the height and orientation differences on symmetrical facial points. (Vignali et al., 2003) use the 3D coordinates of 436 face landmark points as features. (Han et al., 2009) extract geometric features with the volume and area information of faces. In (Hu et al., 2010), the authors divide each face into four regions, and find that the upper part is the most discriminating for gender. Recently, (Toderici et al., 2010) obtain features with the wavelets and the MDS (Multi Dimensional Scaling). In our previous work conducted in (Ballihi et al., 2012c), we propose to select salient geometrical facial features (radial and iso-level facial curves) using adaptive Boosting. The gender classification, is performed by a Nearest Neighbor classifier to Male and Female templates. (Gilani et al., 2013) automatically detect the biologically significant facial landmarks and calculate the Euclidean and geodesic distances between them as facial features. Combining shape and texture (texture and shape-based), (Lu et al., 2006b) fuse the posterior probabilities generated from the range and intensity images using SVM (Support Vector Machine). (Wu et al., 2007) combine shape and texture implicitly with needle maps recovered from intensity images. (Huynh et al., 2012) fuse the Gradient-LBP from range image and the Uniform LBP features from the gray image.

Compared to gender, ethnicity is less explored in face studies. In 2D texture-based works, the first work was tested on the FERET dataset in (Srinivas et al., 1998), with a hybrid architecture of the Ensemble Radial Basis Function (ERBF) networks and the decision trees. (Srinivas et al., 1998) classify ethnicity with SVM on pixel intensity and Biologically Inspired Model (BIM) features. (Zhiguang and Haizhou, 2007) perform ethnicity classification with LBP features on images of a snapshot dataset and FERET dataset. (Srinivas et al., 1998) classify 2D images into Asian and non-Asian groups with Multi-Scale LDA classifiers. (Giovanna and Jean-Luc, 2012) perform gender-specific ethnicity classification with the 2D FERET and TRECVID datasets. While, again unexpectedly, they found that the gender information is not helpful in ethnicity classification. For the shape-based works, in (Srinivas et al., 1998), the authors perform fuzzy ethnicity recognition of eastern and western groups on 3D FRGC2.0 dataset (Phillips et al., 2005), with the Learned Visual Codebook (LVC) derived from histograms of Gabor features. (Toderici

et al., 2010) extract features using the wavelets and the MDS on the Asian and White subsets of the 3D FRGC2.0 dataset. The only texture and shape-based work is presented in (Lu et al., 2006b), where the authors used the ensemble of 2D and 3D scans from the UND and MSU datasets, considering Asian and non-Asian groups.

In age estimation literature, all existing approaches are 2D-based. As pointed in (Rhodes, 2009, Ramanathan et al., 2009), the mathematical cardioid strain model was first proposed for modeling face aging in young ages. After this, face-appearance-based approaches and intrinsic/extrinsic-factor-aided approaches make the two main branches of research in automatic age estimation. Considering that different people have similar aging patterns in face, researchers have employed the Active Appearance Model (AAM) (Lanitis et al., 2004, 2002, Wu et al., 2012), the manifold embedding parameters (Wu et al., 2012, Guo et al., 2008a,b, Li et al., 2012), and the bio-inspired Features (BIF) (Guodong et al., 2009) for revealing the public-level aging patterns. Beyond this, based on the observation that similar faces tend to age similarly, (Lanitis et al., 2004, 2002) first cluster face images into groups according to their inner similarity, and then perform age estimation within each group. These works confirm that the group-level aging patterns exist and are useful in age estimation. Considering that different people aged differently, (Xin et al., 2007, 2006) proposed the Aging-Pattern-Subspace (AGES) constructed from a temporal sequence of individual images, to study the individual-level aging patterns. Moreover, considering different face components age differently, (Jinli et al., 2010) studies the component-level aging patterns with a hierarchical And-Or Graph, and found that the forehead and eye regions are the most informative. The intrinsic/extrinsic-factor-aided approaches study age estimation together with face appearance and the intrinsic factors (permanent factors like gene, gender, ethnicity, identity, etc.), and the extrinsic factors (temporary factors like lifestyle, health, sociality, expression, pose, illumination, etc.). Thinking that faces age differently in different age range, age-specific approaches are adopted by (Lanitis et al., 2004, Guo et al., 2008a,b), where age estimation is obtained by using a global age estimator first, then adjusting the estimated age by a local estimator that operates within a specific age range. Considering that different gender age differently, (Ramanathan et al., 2009, Guo et al., 2008b, Lakshmiprabha et al., 2011, Kazuya et al., 2010) estimate age on male and female groups separately. (Lanitis et al., 2002) encodes the individual lifestyle information in age estimation and demonstrate its importance in determining the most appropriate aging function of the individual. (Kazuya et al., 2010) gives weights to different lighting conditions for illumination-robust age estimation. (Li et al., 2012) gives consideration of the feature redundancy and uses feature selection to enhance age estimation. All of the previous work considering the intrinsic/extrinsic factors have gained better age estimation performance in comparison of using face appearance only.

From the analysis above, we notice that most of the work in literature deals with gender, ethnicity and age recognition separately. The correlations among these soft-biometrics have attracted little consideration. Only few works perform ethnicity-specific gender classification (Gao and Ai, 2009, Giovanna and Jean-Luc, 2012), gender-specific ethnicity classification (Giovanna and Jean-Luc, 2012), and gender-specific age estimation (Xia et al., 2014c, Ramanathan et al., 2009, Guo et al., 2008b, Lakshmiprabha et al., 2011,

Kazuya et al., 2010). To the best of our knowledge, the age-specific gender or ethnicity recognition, and ethnicity-specific age estimation has never been addressed in the literature. Also, no consensus has been achieved regarding the correlation between gender and ethnicity. The experimental results in (Giovanna and Jean-Luc, 2012) show gender and ethnicity are not helpful in each others' recognition tasks. While, the experimental results of (Gao and Ai, 2009) demonstrate that ethnicity-specific gender classifiers achieve higher overall performances. Since both of these works are 2D texture image based, we think that, in addition to their sensitivity to illumination and pose changes, the face images incorporate distorted shape information. Thus, in this work, we propose to explore the use of the 3D shape of the face to better study these attributes as well as their correlation. In (Yuxiao et al., 2010), the authors have demonstrated that, with the 3D scans, human observers perform better on both gender recognition and ethnicity recognition tasks than with 2D face images. In biological studies (Harold et al., 1995), researchers also found that, when considering gender and ethnicity recognition tasks, the usage of 2D face images is limited to full-face view, while the 3D scans are proved to be adaptable to angled views (non-frontal poses). The proposed study gives a thorough study of the correlations among the three attributes. In particular, we are going to study the following questions :**(1)** Can 3D face shape reveal your gender, ethnicity and age? **(2)** Can the gender, ethnicity and age information be useful in each others recognition task? **(3)** How much they are correlated in the facial shape? and **(4)** Are their correlations useful in real-world like applications?

5.3   Methodology and Contributions

To address the over-mentioned questions, two consecutive steps are designed in our approach: the feature extraction step and the classification/regression step. Within the first step, we first pre-process the 3D scans to limit the facial region, and then extract four types of features from the 3D shape of the face. These four descriptions reflect different perspectives of face perception. Withing the classification/regression step, we feed the features to Random Forest for Gender or Ethnicity classification, and Age estimation. To enhance the performance, we have also introduced two additional steps in our approach, the *Feature Selection* and the *Fusion*. The feature selection method is used for highlighting the salient subset of features which contains the information of gender, ethnicity and age. The Fusion method merges all the information from the four types of features to account for the different perspectives. The main contributions of this work could be summarized as follows:

↪ We propose four different and complementary facial descriptions based on shape analysis of facial radial curves (Chapter 2), with which we demonstrate that the 3D shape of the face can reveal our gender, ethnicity and age. Extensive evaluations on the challenging FRGC2.0 dataset (Phillips et al., 2005) demonstrate the effectiveness of the proposed facial attributes recognition approach and its robustness to facial expressions.

↪ We demonstrate that gender, ethnicity and age are correlated in the 3D face, and the correlations are helpful in each others' estimation. Our conclusion is significantly different with (Giovanna and Jean-Luc, 2012), which claims that gender and ethnicity are not helpful for each other's recognition. We also discover that the correlation between ethnicity and age is the strongest among the correlations.

↪ We demonstrate that the correlation of these soft-biometrics can be useful in real-world applications.

↪ This is the first work in the literature which gives a thorough study of the correlations among Gender, Ethnicity and Age with 3D faces. It is also the first work which studies 3D face based age estimation (Xia et al., 2014a).

In the following, we will describe our Morphology-inspired descriptions extracted based on our Riemannian approach for 3D face analysis.

5.4 Geometrical Features Extraction

In this section, we describe four different morphological descriptions extracted from the 3D shape of the face. These descriptions are densely extracted using shape analysis of 3D radial curves of the face. Based on this representation, earlier studies on 3D face recognition (Drira et al., 2013b) and 4D expression recognition (Ben Amor et al., 2014b) have been successfully conducted. Again, we demonstrate here how this representation is convenient to quantify the morphological divergences of gender, ethnicity and age groups.

5.4.1 Features Computation

In this section, different facial morphology descriptions will be presented. Each of them reflects face morphological properties based on the 3D face geometry. Through the facial representation using multiple curves, one can define an efficient way to deform one shape into another (using geodesics) and effectively quantify their divergence using the Dense Scalar Fields, introduced in Chapter 2. A possible interpretation of the DSFs is the local deformations needed to go from one shape to another through a dense and accurate correspondence between the shapes. We have demonstrated in Chapter 4 a first use of the DSFs to capture geometric deformation across sequences of 3D faces. In this chapter, the use of DSFs is different as it will be used to derive four morphological descriptions from static 3D faces, as described in the following:

↪ **The 3D-Symmetry description (3D-sym.)** shown in Figure 5.2 (a) **3D-sym.** captures densely the deformation between bilateral symmetrical curves (β_α^S and $\beta_{2\pi-\alpha}^S$). This description allows to study the relationship between the bilateral symmetry and the 3D facial attributes, Age, Gender
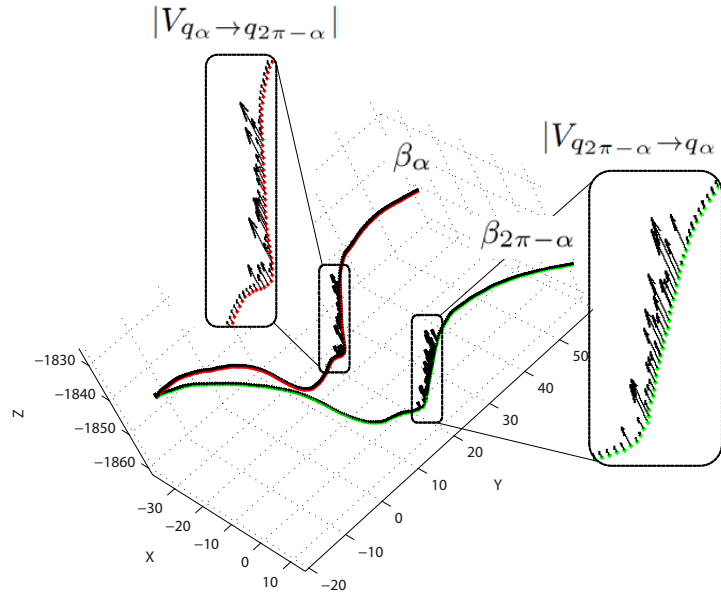
FIGURE 5.1: 3D-Symmetry description from pairwise radial curves. Illustration of the obtained direction needed to deform β_α to fit $\beta_{2\pi-\alpha}$ ($|V_{q_\alpha \rightarrow q_{2\pi-\alpha}}|$) and inversely ($|V_{q_{2\pi-\alpha} \rightarrow q_\alpha}|$). q_α and $q_{2\pi-\alpha}$ denote SRVF of bilateral symmetrical curves β_α and $\beta_{2\pi-\alpha}$, respectively extracted from an arbitrary face.

and Ethnicity. In Figure 5.1, we illustrate the DSFs computation between arbitrary symmetric curves on the face. Its is clear from this figure that although the geodesics lengths connecting q_α and $q_{2\pi-\alpha}$ and inversely, are same, the DSFs vectors $|V_{q_\alpha \rightarrow q_{2\pi-\alpha}}|$ and $|V_{q_{2\pi-\alpha} \rightarrow q_\alpha}|$ are different ($|V_{q_\alpha \rightarrow q_{2\pi-\alpha}}| \neq |V_{q_{2\pi-\alpha} \rightarrow q_\alpha}|$), in general.

↪ **The 3D-Averageness description (3D-avg.)** shown in Figure 5.2 (b) compute DSFs between pairwise curves with the same angle index α, $\beta_\alpha^{\boldsymbol{S}}$ and $\beta_\alpha^{\boldsymbol{T}}$ extracted from the face and a template face, respectively. The template $\boldsymbol{T}$ (presented in 5.2(b)) is defined as the middle point of geodesic path from a representative male face to a representative female face. The 3D-Averageness description is a way to explore the idea that different population groups deviate differently from a given mean facial shape. Faces of different morphology could show different deformations to reach the template face.

↪ **The 3D-Spatial description (3D-spat.)** shown in Figure 5.2 (c) captures the deformation of a curve β_α to the middle-up curve β_0, emanating from the nose tip and goes over the nose and the forehead in $\boldsymbol{S}$. As β_0 is the most rigid curve in the face, the 3D-Spatial description captures the smooth intrinsic deformation from the most rigid part of the face.

↪ **The 3D-Gradient description (3D-grad.)** shown in Figure 5.2 (d) captures the differences between pairwise neighboring curves ($\beta_\alpha^{\boldsymbol{S}}$ and $\beta_{\alpha+\Delta\alpha}^{\boldsymbol{S}}$). It captures the local deformations on the face. This description is useful to detect wrinkles on the face , for example, by performing a kind of "derivation" across the indexed curves.
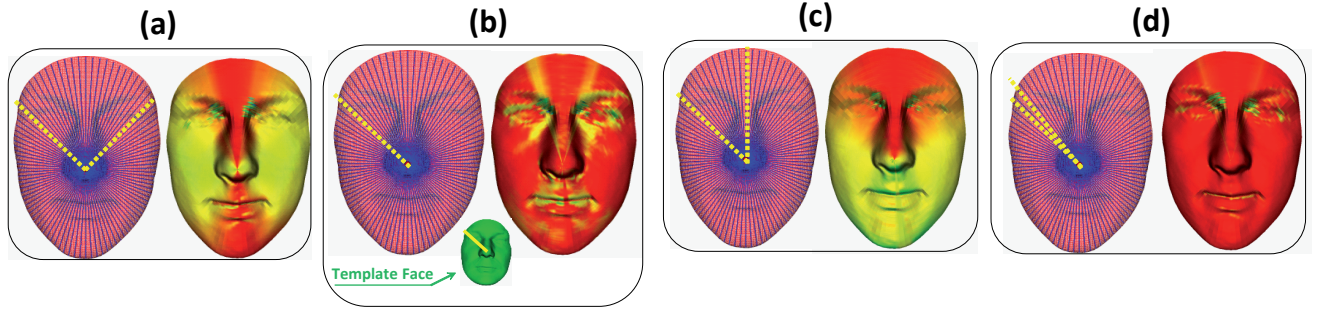
FIGURE 5.2: Illustrations of different features on 3D shape of the face $\boldsymbol{S}$. (a) 3D-Symmetry Description: shape differences from each radial curve $\beta_\alpha^{\boldsymbol{S}}$ to its symmetrical curve $\beta_{2\pi-\alpha}^{\boldsymbol{S}}$; (b) 3D-Averageness Description: shape differences from radial curve $\beta_\alpha^{\boldsymbol{S}}$ in a preprocessed face to radial curve $\beta_\alpha^{\boldsymbol{T}}$ in face template (with same angle index α); (c) 3D-Spatial Description: shape differences from radial curve $\beta_\alpha^{\boldsymbol{S}}$ to the middle-up radial curve $\beta_0^{\boldsymbol{S}}$ in the forehead; (d) 3D-Gradient Description: shape differences from radial curve $\beta_\alpha^{\boldsymbol{S}}$ to its neighbor curve $\beta_{\alpha+\Delta\alpha}^{\boldsymbol{S}}$

In each panel of Figure 5.2, the left part illustrates the extracted radial curves and the curve comparison strategy, the right part shows the corresponding features as color-map on the face. On each point of the face, the warmer the color, the lower the deformation magnitude and the higher the description magnitude. For example, the most symmetrical part of the face is located around the symmetrical plane (red colors), further away the symmetry plane, increasing asymmetry is observable on the face (colder colors). It can be seen, from panel (a) in Figure 5.2 that the local asymmetry emerges around the eyes, mouth, nose, chin, etc. Panel (b) shows clearly that faces shapes differences reside on these features (yellow-green colors). The rest of the facial region shows similarities across the faces (warmer colors). In panel (c), a smooth changing map shows the shape deviation from the upper profile. In panel (d), the gradient highlights local changes (details) around the facial features (yellow-green colors).

5.4.2 Relationship with Facial Morphology

Our goal in this section is to show the relationship between the computed features and the face morphology differences across age, ethnicity and gender groups. First, the **3D-sym.** descriptor densely characterizes the bilateral asymmetry of the face. Also, this description can be employed to distinguish younger and older people, as the latter generally have less symmetric faces. The **3D-sym.** description offers an effective way to study the effect of facial bilateral symmetry on the studied attributes. Second, the morphological differences between shapes can be densely captured using the **3D-avg.** descriptor. This description is very important to distinguish Male and Female faces, as stated in many previous studies (Ziqing et al., 2010, Vicki et al., 1993). In general, Male face has a more protuberant nose, eyebrows, more prominent chin and jaws. The forehead is more backward sloping, and the distance between top-lip and nose-base is longer. It has been shown that differences are observable between Asian and non-Asian faces (Farkas LG, 2005). The Asian population usually has broader faces and noses, farther apart eyes, and exhibit the greatest difference in the anatomical orbital regions (around the eye and eyebrow). It
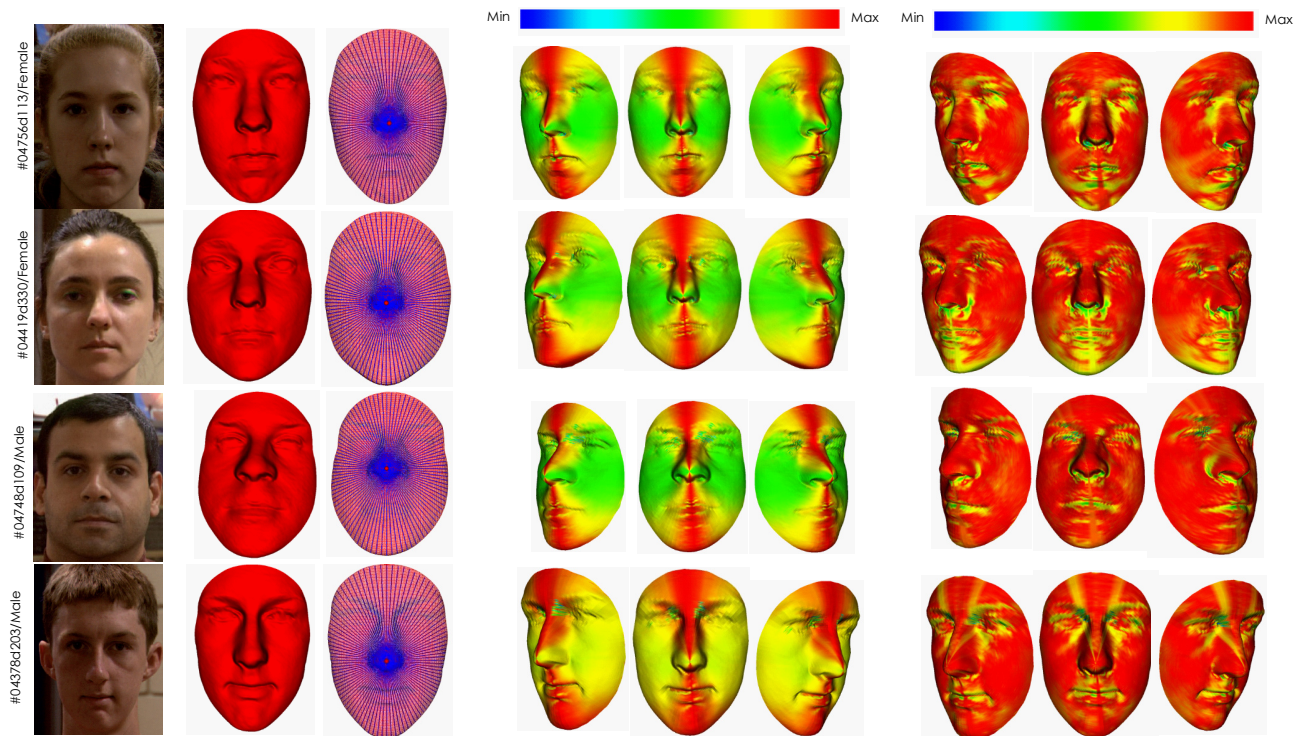
FIGURE 5.3: Examples of bilateral symmetry DSFs and averageness DSFs computed on two female faces (top rows) and two male faces (bottom row). The color, the 3D shape and the representation using radial curves are shown. Notice that in the color bars Max (warmer colors) mean maximum symmetry or averagness. Inversely, the Min in color bars reflect maximum asymmetry or distinctiveness (colder colors).

is also well known that the craniofacial growth is an important change in the baby and adolescent face, which results in re-sizing and redistribution of facial features (Rhodes, 2009, Ramanathan et al., 2009). By measuring the deformation from the face to a defined template face, the **3D-avg.** descriptor is able to capture the morphological differences, in term of features sizes, etc. Third, the **3D-grad.** description is useful to capture the smoothness of the human face. It has been demonstrated, in (Yukio et al., 2003) for example, that female faces usually look smoother and younger than male faces, and the Asian faces usually look younger than non-Asian faces. In addition, the **3D-grad.** can be viewed as a gradient operator over the face and can highlight the wrinkles. Finally, the **3D-spat.** descriptor which measures, over the face, the deviation to the most rigid part (the upper profile) and produce a smooth map of those deviations. It reflects global shape changes on the static face. Shown in Figure 5.3 are four examples of the 3D-Sym. and 3D-Avg. descriptions viewed from three different angles for better visibility. When the first two rows illustrate examples of female faces (taken from FRGC2.0 dataset), in the last two rows are given the descriptions for male faces.

The four descriptions allow to capture densely different morphological perspectives from the facial shapes. Accordingly, we seek to know what clues are relevant to each of the studied facial attribute (gender, ethnicity and age)? In Figure 5.4, we show the magnitude of the correlation between the facial features
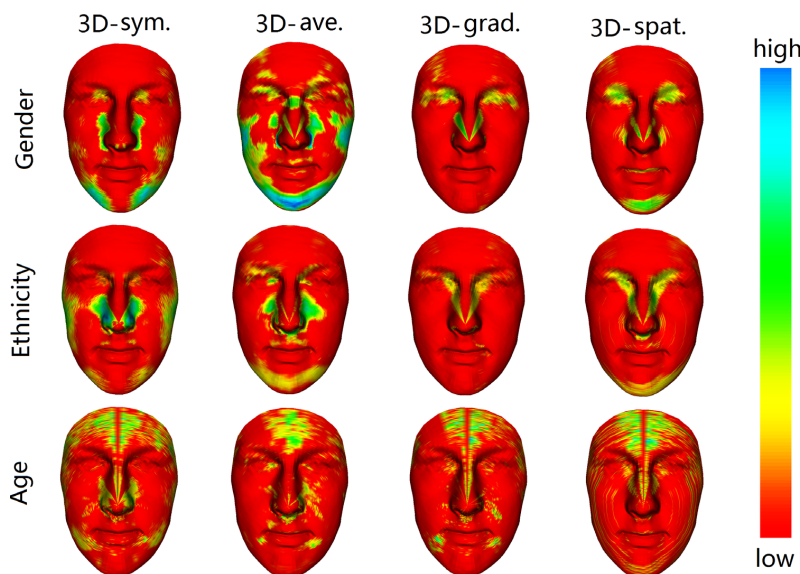
FIGURE 5.4: Correlation between facial features and facial attributes on each point of the face. *3D-sym.* : Symmetry — *3D-ave.* : Averageness — *3D-grad.* : Gradient — *3D-spat.* : Spatial.

and the facial attributes as color-map on each point of the face. The correlations are calculated with the 466 earliest scans of FRGC2.0 dataset. Formally, for all the 466 scans, the ith feature of a face description makes one dimensional vector $D_i = (d_i^1, d_i^2, \ldots, d_i^{466})$, where d_i^m denotes the ith feature of the mth face. Then for a given facial attribute, the labels for 466 scans make a vector $L_a = (L_1, L_2, \ldots, L_{466})$, where L_m denotes the attribute label of the mth face. Then, the correlation between the ith feature of the description and the studied attribute is given by the *Pearson Correlation Coefficient* between D_i and L_a. This coefficient is defined as the covariance of the two variables divided by the product of their standard deviations. Formally, for two variables X and Y, the *Pearson correlation coefficient* $p_{X,Y}$ is defined as $p_{X,Y} = cov(X,Y)/(\sigma_X \sigma_Y)$, where cov denotes the covariance, σ denotes the standard deviation. The absolute value of *Pearson Correlation Coefficient* ranges from 0 to 1. The higher the value, the higher the linear correlation between the two variables. In our case, the higher the correlation between a feature and an attribute, the more informative is this feature for discriminating the concerned attribute. Each column of Figure 5.4 shows the correlation between a face description and each facial attribute, each row shows the correlation between each face description and a facial attribute. The higher the correlation between a facial feature and an attribute, the colder the color on the corresponding facial point.

According to these results, we can confirm three things. First, the green and blue colors in the figure show that the face features are considerably correlated with the facial attributes. With this, we confirm that the 3D face shape is informative for Gender, Ethnicity and Age. Secondly, we confirm that the four descriptions in our approach give different and complementary perspectives for perception of Gender, Ethnicity an Age. For gender, the *3D-sym.* description 'sees' the inner eye corners, the conjecture area of

the nose and the cheeks, and the chin-side regions. The *3D-avg.* description 'looks' at the eyebrows, the eyes, the nose, the lips, and gives big attention to the cheek-sides and the chin. The *3D-grad.* description emphasizes the eyes and the dorsal nasals of the nose, while the *3D-spat.* description also considers the chin and the sides of the nostrils. For Ethnicity, the *3D-sym.* description 'sees' the nose regions and the cheeks. The *3D-ave.* description 'looks' at the conjecture area of the nose and the cheek-sides, and the chin region. The *3D-grad.* description emphasizes the inner eye-corners and the dorsal nasals of the nose, while the *3D-spat.* description also emphasizes the chin and the area around the nostrils. For Age, the *3D-sym.* description 'sees' the whole forhead, the nose, the outer eye corners, and the the regions besides the mouth corners. The *3D-ave.* description mainly 'looks' at the center part of the forehead, the inner eye corners, and the nose surrounding regions. The *3D-grad.* description emphasizes the center part of the forehead, the eye corners, the nose bridge, and the mouth corners, while the *3D-spat.* description emphasizes the center part of the forehead, the eye corners, and the nose bridge. Thus, according to these observations, the four descriptions 'see' different and complimentary clues of Gender, Ethnicity and Age in the face. Thirdly, it confirms the relevance of the facial geometry to the Gender, Ethnicity and Age. For Gender, the eyes, nose, cheek-sides, lips and the chin are particularly informative. In somehow, it matches with the previous findings in sexual dimorphism (Vicki et al., 1993), which claim that males have protuberant nose, eyebrows, chin and jaws than females, and the distance between top-lip and nose-base is longer. For Ethnicity, the eyes, nose, cheek-sides and chin are more informative. This echoes the findings in (Farkas LG, 2005, 17) which stated that the non-Asians have broader faces and noses, farther apart eyes, and lower fetal front-maxillary facial angle (FMFA) measurements than Asians. For both Gender and Ethnicity, the forehead gives little information. While for Age, the forehead, together with the nose, the eye corners and the mouth corners, show the strongest hints. It naturally goes with the knowledge that wrinkles usually are developed on the forehead, around the eyes, the nose and the mouth regions.

5.5   Facial Soft-biometrics Recognition

In this section, we present the remaining steps in our facial attributes recognition method. First, we perform the correlation-based Feature Selection on our descriptions. Then, we feed these selected features to Random Forest for attribute recognition. We note that these two steps are common and useful steps in many recognition tasks and contributions in these fields are out of the scope of this work.

5.5.1   Correlation-based Feature Selection

Feature subset selection is the process of identifying and removing as much irrelevant and redundant information as possible (Hall, 1999). There are mainly two types of feature selection methods, ***the filter*** methods which use heuristics based on general characteristics of the data to evaluate the merit of feature subsets, and ***the wrapper*** methods which use an induction algorithm along with a statistical re-sampling

technique such as cross-validation to estimate the final accuracy of feature subsets (Kohavi, 1995). We choose a filter method for feature selection, named the Correlation-based-Feature-Selection (CFS) (Hall, 1999), because the filters operate independently of learning algorithm and are generally much faster than wrappers. The chosen CFS filter comprises of two parts, a feature correlation measure using the Pearson's correlation coefficient, and a Best-First heuristic search algorithm which moves through the search space by greedy hill-climbing augmented with a backtracking facility. In practice, we perform feature selection for all the three facial attributes, the gender (labeled as male and female), the ethnicity (labeled as Asian and non-Asian) and the age (labeled in two groups, more than 25 and less than 26). After Feature selection, we retain a subset of 10 to 20% of the original features, for each description. Thus, the feature selection procedure significantly reduces the size of the original features.

5.5.2 Random Forest Classification/Regression

Random Forest is an ensemble learning method that grows many decision trees $t \in \{t_1, \ldots, t_T\}$ considering an attribute (Breiman, 2001). To estimate the attribute from a new instance represented as a feature vector, each tree gives a decision result and the forest does the overall estimation. In growing of each tree, two types of randomness are introduced. First, to make the training set, a number of N instances are sampled randomly with replacement from the original data. Then at each node of the tree, a constant number of m $(m<<M)$ variables are randomly selected, and the best split on these m variables is used to split the node. The process goes on until the resulted subsets of the node are totally purified in the label. The performance of the forest depends on the correlation between any two trees, and the strength of each individual tree. The forest error rate increases when the correlation decreases, or the strength increases. Reducing m reduces both the correlation and the strength. Increasing it increases both. Thus, an optimal m is needed for the trade-off between the correlation and the strength. In Random Forest, the optimal value of m is found by using the oob-error rate (out-of-bag-error rate). For making the overall decision, in classification work, the forest predicts the attribute with majority voting. The classification mode of Random Forest is designed for instances with discrete class labels, such as the Gender and Ethnicity labels. While in regression tasks, it takes the average of predictions. The regression mode of Random Forest is designed for instances with continuous class labels, such as the Age labels. Thus, in our work, we use Random Forest in classification mode for Gender and Ethnicity recognition, and in regression mode for Age estimation.

5.6 Experiments

Our experiments are carried out on the Face Recognition Grand Challenge 2.0 dataset (Phillips et al., 2005). The dataset contains 4007 3D near-frontal face scans of 466 participants, where 203 are female and 263 are male, 102 are Asian and 354 are non-Asian. Their age ranges from 18 to 70, with 107 subjects

under 26 years old and 359 subjects above 25. With the FRGC2.0, two types of experiments are carried out:

↪ **Expression-Dependent Experiment** uses the 466 earliest scans from FRGC2.0 for training and testing. The majority of the scans in this subset are neutral. This data subset leads to a possible study of the facial attribute recognition when imposing a neutral expression.

↪ **Expression-Independent Experiment** involves the whole 4007 scans of FRGC2.0 (where more than 40% of the scans are expressive). This makes possible to test the robustness of the approach against facial expression variations.

We use the *Leave-One-Person-Out (LOPO)* cross-validation approach in all our experiments, where each time the scans of one subject are used for testing, and the scans of the rest subjects are used for training. Each subject is tested equally only once and the experiments are conducted in a subject-independent fashion. Thus, there are altogether 466 folds in the cross-validation. There are two major reasons for choosing this experimental setting, the first is its similarity to real-world applications, and the second is it allows training with a maximum number of scans. To establish a comparative study with the previous studies, we will report also the results using person-independent 10-fold cross-validation.

Gender Classification Experiments

Gender classification is to automatically label a query, a 3D face scan in our case, into Male and Female. In the experiments, we first perform LOPO gender classification with Random Forest on the original extracted features. Then, we perform correlation-based feature selection on the original features and carry out the experiments with the selected features. After that, we perform the experiments in consideration of ethnicity and age in the *Ethnicity-specific* and *Age-specific* settings, respectively. For *Ethnicity-specific*, we separate the 466 subjects into Asian group (112 subjects) and non-Asian group (354 subjects) first, and then perform the LOPO experiments on these subsets, separately. For *Age-specific*, we separate the 466 subjects into an older group ($\geq$26 years, 107 subjects) and younger group ($\leq$25 years, 359 subjects) first, and then perform LOPO experiments on the younger and older groups, separately. We have also done the experiments considering both ethnicity and age. In the *Eth.& Age-specific* setting, we perform experiments within the scans of the same ethnicity and age groups, described above.

The Gender classification accuracy are shown as bar-plots in Figure 5.5. The y-axis shows the classification rate following the LOPO settings. The x-axis shows the different experiments. The left panel corresponds to the Expression-Dependent experiments, the right one shows the results when tolerate facial expression variations (Expression-Independent experiments). With the 466 scans in the Expression-Dependent experiments, the original features achieve $> 85\%$ gender classification rate for each description. With the feature selection step, the results are improved by $2 - 5\%$. Now, when considering the ethnicity

(a) Expression-dependant gender classification results

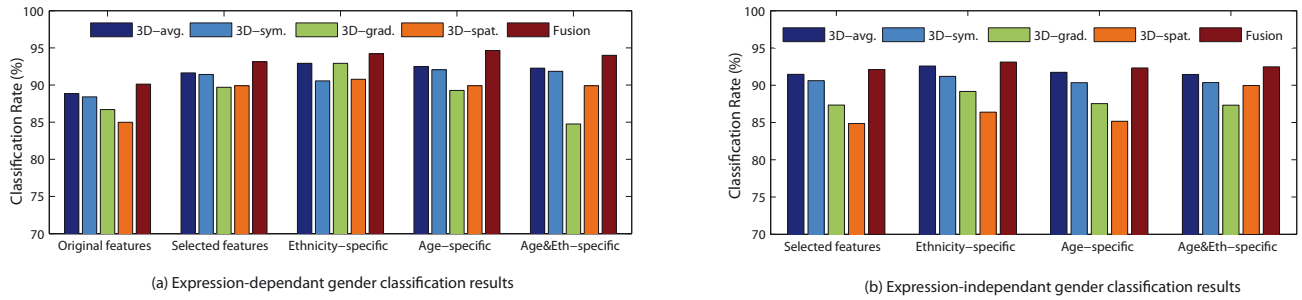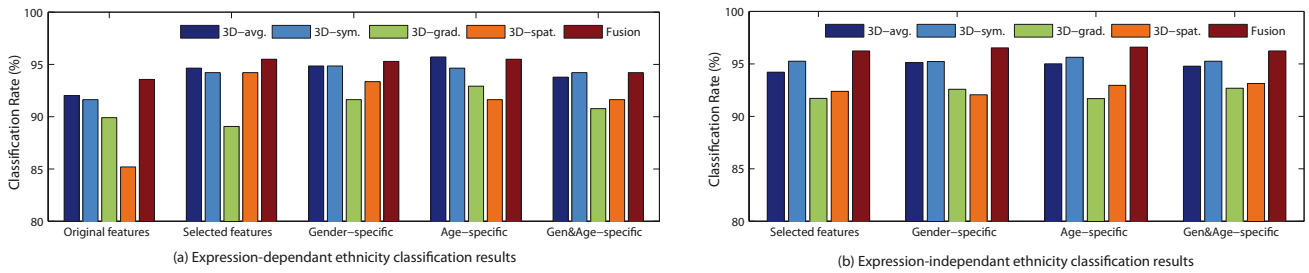(b) Expression-independant gender classification results

FIGURE 5.5: Gender classification results under Expression-dependent and Expression-Independent settings. Features, *3D-avg.*: Averageness — *3D-sym.*: Bilateral Symmetry — *3D-grad.*: Gradient — *3D-spat.*: Spatial — *Fusion*: their fusion by concatenation. Features processing, *Original features*: No feature selection applied — *Selected features*: Correlation-based features selection applied before classification. Settings, *Ethnicity-specific*: Selected features within each ethnicity group — *Age-specific*: Selected features within each age group — *Age&Eth-specific*: Selected features within the same ethnicity and age group.

information in the *Ethnicity-specific* classification setting, the results are improved, in general. These results indicate that **Asian and non-Asian show different gender patterns**. When considering the age information in the *Age-specific* protocol, a more significant improvement is shown. This result shows that **people of different age have different gender related patterns**. In other worlds, human beings' gender patterns correlate to their age. When considering both *Ethnicity-specific* and *Age-specific*, termed *Age&Eth-specific*, the accuracy is generally higher than the Feature Selection and quite comparable to *Ethnicity-specific* and *Age-specific*. With these results, we also find that the fusion of the features always outperforms the descriptions taken individually. The highest gender classification rates, 94.64% and 94.21%, are achieved by the *Fusion* under the *Age-specific* and *Ethnicity-specific* settings, respectively. These findings are furthermore confirmed in the Expression-Independent settings. As shown on the right panel of Figure 5.5, for each description, the gender classification performance is always higher when considering Ethnicity and Age information. The fusion of these features always outperforms each individual description, and achieves an 93.13% accuracy in the *Ethnicity-specific* setting. These results show also that the expression variations affect slightly the performances. Among the four descriptions, generally, the *3D-avg.* descriptor, which captures the shape difference to a given template face, achieves the highest performance. This confirms the studies on sexual dimorphism (Vicki et al., 1993) which claim that Male and Female faces present different morphological features, in their facial shapes. Also, the *3D-sym.* descriptor, gives a confirmation on asymmetry differences related to the gender groups (Liu and Palmer, 2003). In addition to these confirmations, this experiment shows the relationship between the gender patterns to two influencing factors, the age and the ethnicity.

Ethnicity Classification Experiments

Ethnicity classification consists on automatically label a query instance into corresponding ethnicity class (Asian or non-Asian groups in the present study). Similarly to the experiments conducted in section 5.6 for Gender classification, an LOPO experiment is conducted using the original features, then the selected features and a study on *Gender-specific* and *Age-specific* are conducted under the following Expression-Dependent and Expression-Independent settings.

FIGURE 5.6: Ethnicity classification results under Expression-dependent and Expression-Independent settings. Features, *3D-avg.*: Averageness — *3D-sym.*: Bilateral Symmetry — *3D-grad.*: Gradient — *3D-spat.*: Spatial — *Fusion*: their fusion by concatenation. Features processing, *Original features*: No feature selection applied — *Selected features*: Correlation-based features selection applied before classification. Settings, *Gender-specific*: Selected features within each gender group — *Age-specific*: Selected features within each age group — *Age&Gen-specific*: Selected features within the same gender and age group.

The ethnicity classification results are shown in Figure 5.6. As shown in the left panel, under the Expression-Dependent settings, the results from the original features are always higher than 85%. The feature selection process improves the results with $2\% - 7\%$ compared to the previous results. It shows the capability of the feature selection method in ethnicity classification, as outlined in gender classification. Under the *Gender-specific* and *Age-specific* settings, the results are slightly higher than the previous result. This shows that **the Male and the Female have different ethnicity related patterns**. In addition, it indicates that **people of different age have different ethnicity related patterns**. Here, the highest ethnicity classification rates of 95.71% and 95.49% are achieved by the *3D-avg.* description and the fusion, respectively. Again, like in gender classification, the fusion of these features always outperforms the individual description. These results are confirmed in the right panel of Figure 5.6 with a higher accuracy of 96.6%. This demonstrates the robustness of the proposed approach against the facial expressions in ethnicity classification. Roughly speaking, according to the results presented above, Ethnicity (Asian and non-Asian) classification is influenced by gender and age factors. Here also, the 3D-avg. description achieves higher accuracy, compared to the remaining descriptions. It confirms the findings of previous studies (Farkas LG, 2005)(Alphonse et al., 2013) that a significant morphological differences exist between Asian and non-Asian faces. In addition, according to the results in Figure 5.6, the bilateral asymmetry (3D-sym.) can play an important role in Ethnicity classification.

Age Estimation Experiments

Given a query, estimating its Age consists of automatic label with an estimated age or an age group/range. In the following, we use Random Forest in regression mode to estimate the exact age of a query instance. Again, The LOPO experimental protocol is used for training and testing. Similarly to the previous experiments for Gender and Ethnicity classifications, here again, we compare the results achieved with the original features, results from the selected features and results reported under *Gender-specific* and *Ethnicity-specific* settings. In feature selection, we use a different age group partition here than in *Age-specific* experiments. The 466 subjects are divided into two age groups, one $>= 23$ years with 162 subjects, and another group < 22 years with 304 subjects. This partition can better balance the number of scans in the two groups. It also goes with the idea that the craniofacial growth stops at the age of about 20, and faces convey different aging morphology before and after this age. For *Gender-specific* setting, the 466 subjects are separated into Male (263 subjects) and Female groups (203 subjects) first, and then we experiment on each group separately. For *Ethnicity-specific* setting, we separate the 466 scans into Asian group (112 subjects) and Non-Asian group (354 subjects) first, and then do experiments on them separately. The age estimation accuracy is typically measured by the mean absolute error (MAE) and the cumulative score (CS). The MAE is defined as the average of the absolute errors between the estimated age and the ground truth age, while the CS criteria, proposed firstly by (Xin et al., 2007) in age estimation, shows the percentage of cases among the test set where the absolute age estimation error is less than a threshold. In this work, the experimental results are shown as MAEs in Figure 5.7.
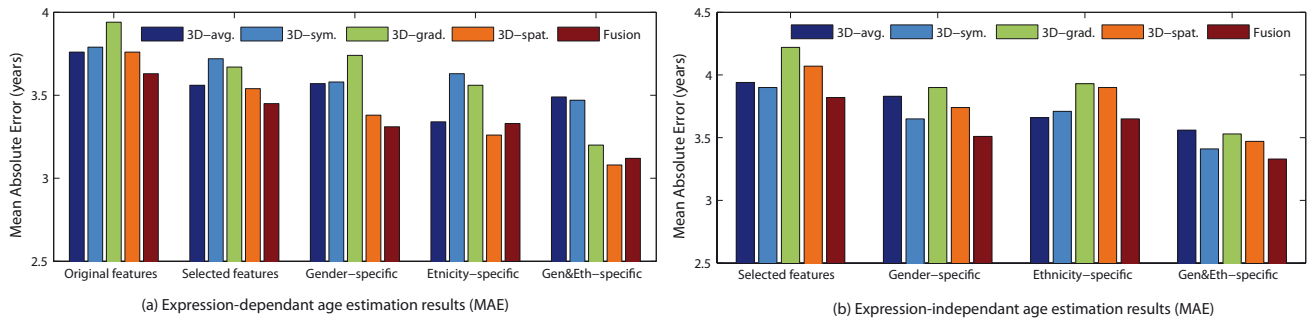

FIGURE 5.7: Age estimation accuracy under Expression-dependent and Expression-independent settings. Features, *3D-avg.*: Averageness — *3D-sym.*: Bilateral Symmetry — *3D-grad.*: Gradient — *3D-spat.*: Spatial — *Fusion*: their fusion by concatenation. Features processing, *Original features*: No feature selection applied — *Selected features*: Correlation-based features selection applied before classification. Settings, *Gender-specific*: Selected features within each gender group — *Ethnicity-specific*: Selected features within each age group — *Eth&Gen-specific*: Selected features within the same gender and ethnicity group.

Under the Expression-Dependent settings, the MEAs for all the descriptions are always < 4 years. After feature selection, the errors decrease which confirms again the usage of feature selection in our approach. Again, the errors decreases when testing under the *Ethnicity-specific* and the *Gender-specific* protocols. It means that **different ethnicity have different aging patterns**, and also, **Male and Female**

people has different aging patterns. The overall effect of gender and ethnicity information in age estimation is evaluated in the *Gen.&Eth.-specific* setting. The MAEs are even lower than the ones under the *Ethnicity-specific* and *Gender-specific* settings. It demonstrates that **the combination of gender and ethnicity information gives the strongest improvement to age estimation performance**. The lowest MAEs are achieved in the *Gen.&Eth.-specific* setting, by **3D-spat.** and **Fusion** with 3.08 and 3.12 years, respectively. Under the Expression-Independent experiments, same observations are shown in the right panel of Fig 5.7, where arbitrary expressions could be conveyed by the participants. We notice, despite the facial expression variations, which affect significantly the facial shape, our algorithm still provides high accuracy. The Expression-Independent experiments have also confirmed that the improvement of performance is the most significant when using the gender and ethnicity information together. Also, the *Fusion* of the descriptions always shows better performance compared to separate descriptions. It yields a 3.33 years MAE under the Gen&Eth-specific setting on the whole FRGC2.0 dataset.

TABLE 5.1: Mean Absolute Errors of the fusion for different age groups under Expression-dependent and -independent settings.

Fusion/Age group	≤ 20	(20,30]	(30,40]	> 40	All
Expression-dependent					
Original features	3.93	2.29	7.03	24.45	3.63
Selected features	3.99	2.11	6.48	24.50	3.45
Gender-specific	3.74	2.07	6.25	22.48	3.31
Ethnicity-specific	3.70	2.05	6.18	23.71	3.33
Gender and Ethnicity-specific	**3.57**	**1.88**	**6.23**	**21.97**	**3.12**
Expression-independent					
Selected features	3.90	2.34	6.69	23.83	3.82
Gender-specific	3.78	2.12	6.48	21.25	3.51
Ethnicity-specific	3.48	2.26	6.14	23.17	3.65
Gender and Ethnicity-specific	**3.44**	**2.04**	**5.98**	**20.26**	**3.33**
# of Subjects	185	246	20	15	466

To explore in details these results, we show in Table 5.1 the age estimation accuracy in each age group. We find that, no matter in Expression-Dependent or Expression-Independent experiments, the MAEs in this age group are always lower when considering Gender or Ethnicity information, than without such consideration (with the selected features). When considering both Gender and Ethnicity in age estimation, the MAEs in each age group almost reach the lowest (marked in bold). Thus, by giving consideration to Gender and Ethnicity, we have successfully enhanced the age estimation performance for all the age groups. In Table 5.1, we also find that our algorithm performs much better in younger age groups, than in older age groups. Considering the number of training subjects as shown in the last row, it is probably due to the fact that for older age groups, very limited scans are available in the training. To the best of our knowledge, this is the first work which addresses the problem of age estimation from

3D facial shapes. It is also the first work studying the correlations between the three attributes and perform joint estimation, based on three-dimensional data.

Comparison to previous work

As shown in Table 5.2, for gender classification, the works closely related to ours (in term of evaluation settings) are presented in (Toderici et al., 2010), (Ballihi et al., 2012c),(Gilani et al., 2013), and (Wang and Kambhamettu, 2013), which are also tested on the FRGC2.0 dataset. Using all the 4007 scans of FRGC2.0, we achieve 93.11% gender classification rate. Our result is 3% lower than (Gilani et al., 2013) (96.12%). Their approach required accurate landmark detection to compute Surface and Euclidean distances between them, used as features. Our result is comparable to the results reported in (Toderici et al., 2010) (93.5%), and (Wang and Kambhamettu, 2013) (93.7%). For ethnicity classification, the nearest works with ours are presented in (Cheng et al., 2009), and in (Toderici et al., 2010). Using all the 4007 scans of FRGC2.0, we achieve 96.45% ethnicity classification rate. This result is much higher than our previous results reported in (Cheng et al., 2009) (82.38%). Compared to the result reported in (Toderici et al., 2010), our ethnicity classification rate is 2.4% lower. However, their result is based on only the 3676 Asian and White scans of FRGC2.0. In contrast, we work with all the 4007 scans of FRGC2.0, which encounters much more complicated ethnicity challenges.

Further Experiments

The previous experimental section demonstrates that gender, ethnicity and age information are correlated, and their relationship are helpful in each others recognition task. Following this, two questions rise up: (1) To how much extend are they correlated? and (2) How to benefit from their correlations in real-world applications where the ground-truth is unavailable of the other attributes? We address these questions in the following two subsections.

How much are Gender, Ethnicity and Age correlated ? Recall that with feature selection, we have obtained the salient subsets of features for Gender, Ethnicity and Age for each of the four descriptions. Thus, we take these subsets as representations of the Gender, Ethnicity and Age information. With this, we explore two ways to quantify their mutual correlation. The first way is to represent a feature subset as a one-dimensional vector with which the optimal class separation is obtained. Then we measure the correlation directly between such vectors. It will provide a single value for each two attributes, which represents their correlation in the ***Decision Level***. To this end, we first apply the Linear Discriminant Analysis (LDA) on each subset. The LDA method is a supervised dimensionality reduction method which projects the data into a subspace where optimal class separation could be obtained. The dimension of the projected subspace equals to the number of classes minus one. In our case, for each of the three attributes, we have always two classes (Male and Female for Gender, Asian and non-Asian for ethnicity, < 23 years and > 22 years for Age). Thus, after LDA projection, for each description, we get one

TABLE 5.2: Comparison with previous studies.

Gender Classification (Male / Female)

Method	Dataset	Auto	Features	Classifiers	Setting	Results	Modality
(Ballihi et al., 2012c)	466 scans of FRGC2.0	Yes	Facial curves	Adaboost	10-fold CV	86.05%	shape
(Toderici et al., 2010)	4007 scans of FRGC2.0	Yes	Wavelets	Polynomial-SVM	10-fold CV	93.5%	shape
(Xia et al., 2013c)	4007 scans of FRGC2.0	No	DSF+LBP features	Random Forest	LOPO	93.27%	shape + texture
(Wang and Kambhamettu, 2013)	192 scans of HFB	No	3D coordinates	RBF-SVM	5-fold CV	94.8%	shape + texture
(Xia et al., 2013b)	4007 scans of FRGC2.0	No	3D coordinates	RBF-SVM	5-fold CV	93.7%	shape + texture
	4007 scans of FRGC2.0	Yes	DSF features	Random Forest	10-fold CV	90.99%	shape
	466 scans of FRGC2.0	Yes	DSF features	Random Forest	10-fold CV	88.12%	shape
(Gilani et al., 2013)	4007 scans of FRGC2.0	Yes	landmark distances	LDA classifier	10-fold CV	97.05%	shape
	466 scans of FRGC2.0	Yes	landmark distances	LDA classifier	10-fold CV	96.12%	shape
(Lu et al., 2006b)	1240 scans in UND and MSU	No	Grid elements	SVM	10-fold CV	91%	shape + texture
(Hu et al., 2010)	729 UND scans + 216 private scans	Yes	Curvature	RBF-SVM	5-fold CV	94.03%	shape
(Wu et al., 2007)	260 needle maps of UND	No	PGA features	Posteriori Probabilities	200 training / 60 testing (6 times)	93.6%	shape + texture
(Vignali et al., 2003)	120 3D scans UND	No	3D coordinates	LDA classifier	LOPO	95%	shape
(Liu and Palmer, 2003)	111 full 3D scans	No	Variance Ratio	Linear classifier	half training / half testing (100 times)	96.22%	shape
(Han et al., 2009)	61 scans in GavabDB	No	Geometry features	RBF-SVM	5-fold CV	82.56%	shape
Our work	466 scans of FRGC2.0	Yes	DSF features	Random Forest	10-fold CV	95.06% ± 0.027	shape
	4007 scans of FRGC2.0	Yes	DSF features	Random Forest	10-fold CV	93.11% ± 0.035	shape

Ethnicity Classification (Asian / non-Asian)

Author	Dataset	Auto	Features	Classifiers	Setting	Results	Modality
(Toderici et al., 2010)	3676 scans of FRGC2.0	Yes	Wavelets	Polynomial-SVM	10-fold CV	99%	shape
(Cheng et al., 2009)	4007 scans of FRGC2.0	No	LVC features	membership probability	(no CV)	82.38%	shape
(Lu et al., 2006b)	1240 scans in UND and MSU	No	Grid elements	SVM	10-fold CV	98%	shape + texture
Our work	466 scans of FRGC2.0	Yes	DSF features	Random Forest	10-fold CV	96.78% ± 0.023	shape
	4007 scans of FRGC2.0	Yes	DSF features	Random Forest	10-fold CV	96.45% ± 0.033	shape

dimension vector for each attribute. In Figure 5.8, we show the distribution of the projected LDA features for each attribute in each description, on the 466 earliest scans of FRGC2.0. Each row of the subfigures shows the distributions of projected LDA features for a facial attribute, and each column of subfigures show the distributions in a facial description. In Figure 5.8, clear separation of different classes is shown for all the three attributes in all the descriptions. It confirms that the feature selection method is able to keep the relevant information. And also, the figure shows the LDA projected features are able to characterizing the three attributes.
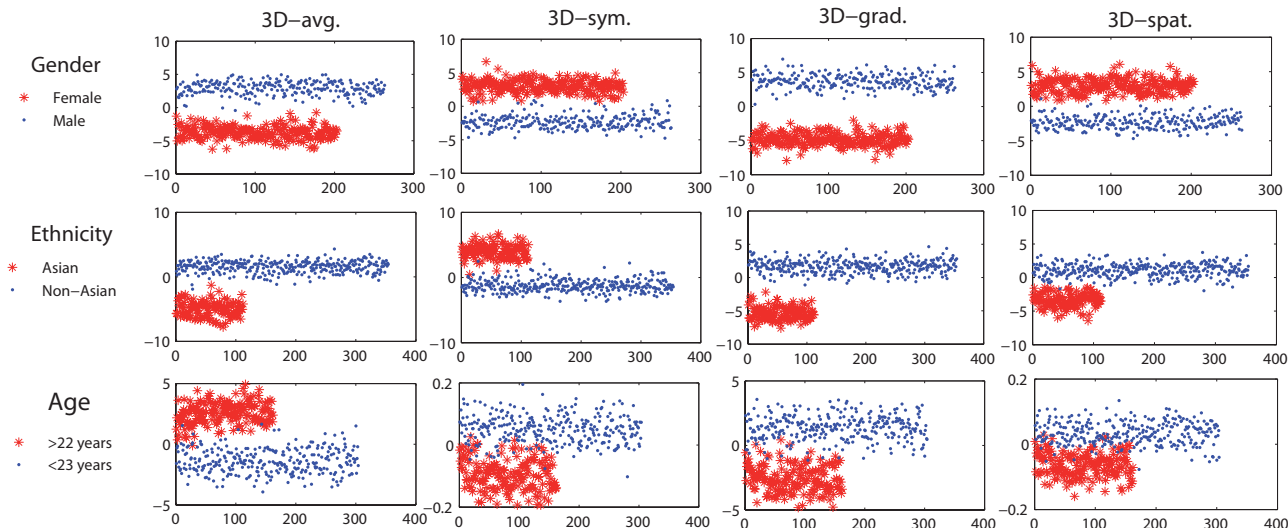
FIGURE 5.8: Distribution of the projected LDA features for gender, ethnicity and age for each face description.

How to benefit from their correlation in real-world applications ?

In realistic applications, we do not have the ground truth of Gender, Ethnicity and Age of the testing instances. While, benefited from the effective recognition performance in the previous section, we can use the recognition results as the gender, ethnicity and age information in the experiments. Thus, instead of using the ground truth, we enroll gender, ethnicity and age information with the predicted gender, ethnicity and age labels given in the previous recognition tasks in the *Feature Selection* setting with the *Fusion* description. The predicted information are termed with * as *Gender**, *Ethnicity** and *Age**. The recognition results are reported in Figure 5.9.

In Figure 5.9 (a) and (b), the gender classification results are presented. In the Expression-dependent experiments, except for the 3D-spat. description, the gender classification results are always higher when considering ethnicity and age information, than without such consideration (in the *Selected features* setting). In the Expression-independent experiments, except for the 3D-grad. description, the gender classification results are always higher when considering ethnicity and age information. With the Fusion description in the *Ethnicity**-specific* setting, we achieve 94.85% Gender classification rate in the

FIGURE 5.9: Experimental results using automatic recognition results of gender, ethnicity and age.

(a) Expression-dependant gender classification results

(b) Expression-independant gender classification results

(c) Expression-dependant ethnicity classification results

(d) Expression-independant ethnicity classification results

(e) Expression-dependant age estimation results (MAE)

(f) Expression-independant age estimation results (MAE)
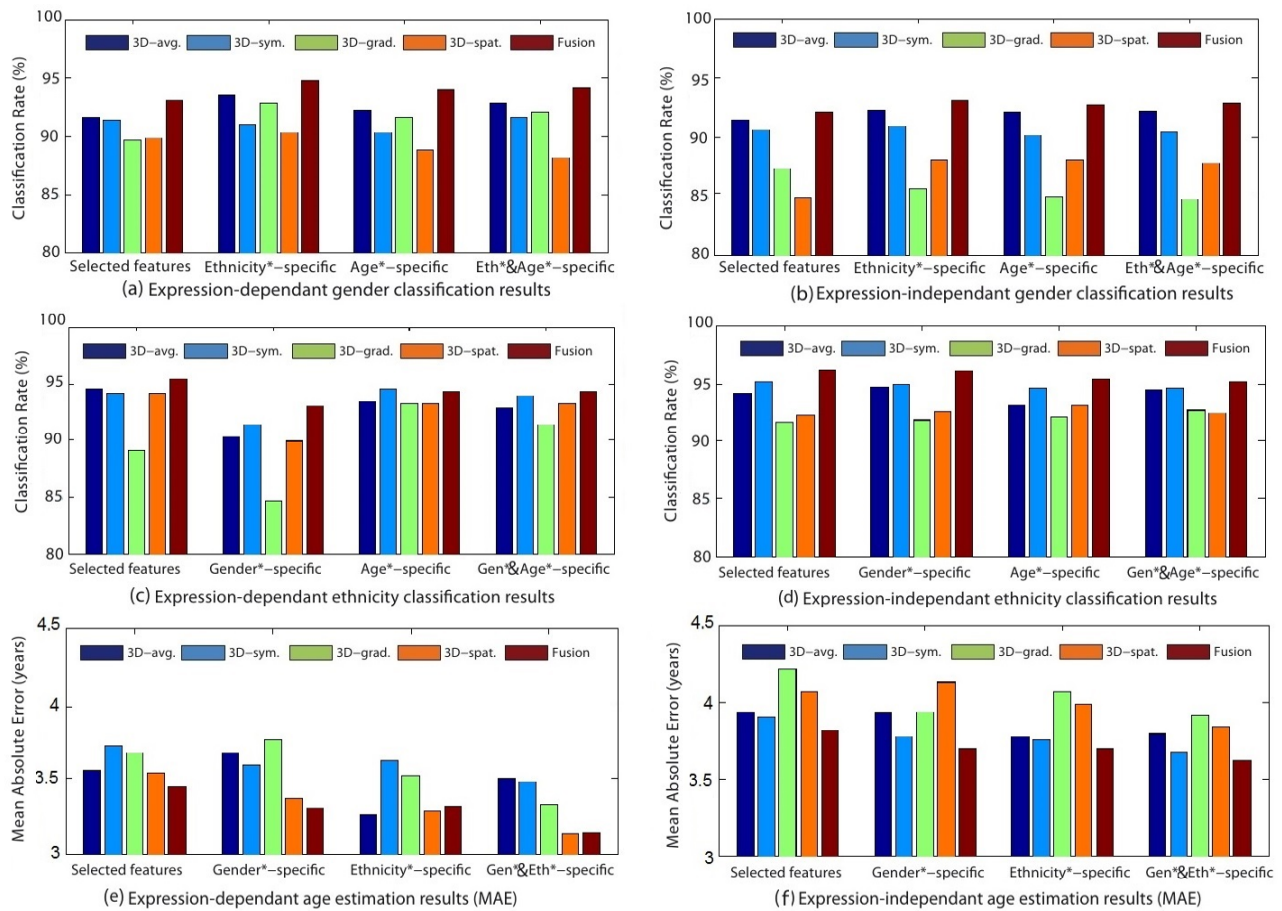
Expression-dependent experiments, and 93.08% Gender classification rate in the Expression-independent experiments. The experimental results of ethnicity classification are presented in Figure 5.9 (c) and (d). In both the Expression-dependent and the Expression-independent experiments, the results considering Gender and Age information are comparable or slightly lower, than without such consideration (in the *Selected features* setting). With the Fusion description, we achieve 94.42% Ethnicity classification rate in the *Age*-specific* and the *Gen*&Age*-specific* settings in the Expression-dependent experiments, and 96.18% Ethnicity classification rate in the *Gen*-specific* settings in the Expression-independent experiments. For Age estimation, the results are shown in Figure 5.9 (e) and (f). Compared to the *Feature Selection* setting, the MAEs are significantly reduced when considering Gender and Ethnicity information in both the Expression-dependent and the the Expression-independent experiments. With the Fusion description in the *Gen*&Eth*-specific* setting, we achieve an MAE of 3.13 years in the Expression-dependent experiments, and 3.62 years MAE in the Expression-independent experiments. In summary of these experiments, it is clear that, for Gender and Age recognition, we have obtained higher performance with the automatically recognized information. For Ethnicity recognition, since the training scans are reduced significantly and the classification rate reaches as high as 94.42% (Expression-dependent) and 96.18% (Expression-independent) when considering gender or age information, we still think that using

the automatically recognized information of gender and age is a good strategy in real world application.

5.7 Conclusions

In this chapter, we have presented morphological characteristics that can be extracted using our shape analysis framework of 3D faces. We have explored their relationship with the facial soft-biometrics (gender, ethnicity and age). Again, we have demonstrated the effectiveness of the proposed computational shape analysis framework and its capability to accurately register facial shapes and quantify their divergence. The study presented here provides (1) the first age estimation approach based on the 3D shape of the face[5]; (2) the first joint estimation of the three attributes using four morphological descriptions and their fusion (Xia et al., 2014d) (**[J3-s]**). We utilize here disjoint ideas from computer vision – Shape Analysis using differential geometry and Machine Learning techniques – and demonstrate the relevance of the proposed approaches in facial attributes estimation.

---

[5]Our paper (Xia et al., 2014a) won the best paper award in the area of IMAGE AND VIDEO UNDERSTANDING at the International Conference on Computer Vision Theory and Applications, 2014.

Chapter 6

Research Project and Perspectives

Throughout this *habilitation*, we have presented our contributions to the field of 3D shape analysis based on Riemannian methods, with applications to classical problems in pattern recognition. The proposed mathematical representations of 3D shapes which rely on multiple curves are elements of a shape space defined as the invariants under the actions of transformation groups (quotient space). Due to the non-linearity of the underlying manifolds, tools from differential geometry are used to provide geometric interpretations such as the notion of geodesic (or its length) and its relevance to find the most efficient way to deform one shape into another (or to define the shortest distance between shapes under rigid and non-rigid transformations). The fundamental ingredient of 3D shape analysis is accurate and dense correspondence between shapes before quantifying their difference. The definition of an elastic model allows to achieve jointly accurate registration of 3D shapes and derive an elastic metric which account for bending and stretching of the shapes. In addition, the notion of shooting vector (its norm or its magnitude) along geodesic paths between shapes, captures effectively the deformations between them, through accurate registration. Statistical computation on manifolds, such as the definition of a sample mean of a set of shapes, a sample covariance and explicit statistical models on the tangent space of a sample mean to model the shape class and the variability within the class are suitable for shapes classification and clustering and data completion. Through the multiple experiments conducted on publicly available benchmarks such as FRGC, BU-3DFE, BU-4DFE, Bosphorus, etc., we have demonstrated the interest of the proposed methodology. We have shown also the interest to combine ideas from differential geometry (extraction of Euclidean representations derived from geometric features) with Machine Learning techniques to design efficient classifiers.

However, there are still several question marks and open research avenues from theoretical and practical perspectives. The move to 4D data analyses is still at the beginning, hence, there is an outstanding need to develop processing tools and computational algorithms to process these data and exploit the temporal dimension. From a theoretical point-of-view, the move to analyze dynamic 3D sequences induces fundamental questions such as (1) Which suitable mathematical representation of shapes and

their dynamics (their animations)? (2) How to exploit the dynamics in the analyses? (3) How to compute statistics on shapes (static representations) and animations (dynamic representations)? (4) How to make the representation insensitive to temporal variations? (5) How to design classification and clustering algorithms based on statistical models of representations and their animations? Responses to these questions should bring solutions to make the analyses of shapes and of their motions insensitive (or robust) to,

↪ **rigid transformations** (translation, scaling and rotation) in order to roughly register 3D frames across and between the 3D sequences. Filter out these changes is required for comparing 3D static frames and to define proper distances between frames and sequences;

↪ **non-rigid transformations** to finely register 3D frames across the sequences. It is important to propose accurate vertex-level tracking approaches to accurately quantify the motion along the sequences and achieve efficient way to deform one frame into another;

↪ **temporal variations** caused mainly by the difference in the execution rate and the starting and ending points of the events (activity, emotion, gesture, etc.). Comparing two sequences translates to measure distances between each frame in a **rate-invariant** fashion. The definition of such metric allows accurately comparing, summarizing, and modeling 3D dynamic sequences;

↪ **missing and noisy data** caused by self-occlusions, pose variations and the 3D (depth) acquisition techniques. Static frame representations should account for this data variability.

It is around these issues that my research project will be organized. Recent advances in 4D facial imaging systems (e.g the Di4D[1] passive stereo photogrammetry technology) allow acquisition of high-resolution (in both spatial and temporal dimensions) facial data. At the same time, the collection of new 4D datasets, such as the *Binghamton University* 4D Facial Expression databases (Yin et al., 2008) (Zhang et al., 2013), the 4D dataset constructed at *University of Central Lancashire* (Hi4D-ADSIP) (Matuszewski et al., 2011, 2012), and the dynamic 3D FACS dataset (D3DFACS) for facial expression research (Cosker et al., 2011), open the doors to study the facial deformations and to develop inference strategies for 3D facial shapes and animations. Several applications would benefit from these tools, for example understanding facial expressions, in particular, spontaneous and subtle expressions (called also micro-expressions). Their study requires high-resolution spatio-temporal acquisitions and accurate algorithms to track the facial movements. These tools would also allow to quantify facial distortions from dynamic data, to improve clinical diagnostics and decisions regarding the pathology (or diseases) affecting the face. Following this line, it is required to propose mathematical representations of static shapes and derive special manifolds for 3D frame (static) analysis. Before specify statistical models for full animations, it is necessary to consider some choices for capturing variability in individual shapes. On the underlying representation spaces, one seek to impose dynamical models that capture the variability in shape evolution. The main

---

[1] www.di4d.com

difficulty here is the **temporal variability of the observations**. In fact, the animation data are typically corrupted by variability in execution rates within a class of animation and, thus, causing major problems in the inference process.

From a practical point-of-view, the 3D meshes used in earlier studies, often come from laser-rangefinder scanners or Stereo Photogrammetry cameras of high-resolution which achieve accurate depth measurements. Their use could be a serious limitation in the new research trends called *face analysis in the wild*, as pointed out in recent studies (Wang et al., 2013)(Simonyan et al., 2013)(Sagonas et al., 2013)(Tai, 2014)(Asthana et al., 2014), and *in spontaneous facial behavior analysis* as pointed out in (Nicolaou et al., 2011)(Bousmalis et al., 2013) and recently in the SFBA[2] and the FFER[3] workshops. With the introduction of a new generation of cost-effective 3D sensors (IR Structured-Light or Time-of-Flight) (Zhang, 2012) capable of dynamic acquisitions, human behavior analysis from dynamic flows under unconstrained conditions is become possible. For instance, the release of Microsoft's Kinect sensors including their Software Development Kit (SDK) have provided a commercially viable approach and hardware platform to capture 3D data in real-time. However, these devices suffer from two major limitations regarding the depth quality that they measure: (1) the low-resolution of depth images due to the large field-of-view of the cameras, and (2) the low-accuracy of depth measurements which results in noisy data. To be able to use these data, one should use appropriate methods and tools to tackle these important issues. Later, we shall introduce our approach to overcome these limitations using subspace methods applied to dynamic flows of 3D data. Our target applications here are face recognition in advertise conditions and analysis of spontaneous emotions from dynamic 3D sequences. Another advantage of using the depth data is its robustness to illumination and pose variations and the ability of the algorithms to track accurately the facial landmarks (eye corners, nose tip, eyebrows, mouth, etc.) as well as the body joints or 3D skeletons (Shotton et al., 2011). The availability of such **sparse shape representations** (registered landmarks) could be explored as a first level of analyses of human dynamics (including the face movements). At a second level, the **dense shape representations** (sequence of surfaces) would benefit of the results of the previous level, in temporal alignment of sequences for example.

Regarding my experience in the field of face analysis using three-dimensional data and the challenges still open (Phillips et al., 2009), it is natural to continue our investigation using dynamic facial data. Modern face recognition approaches target successful person identification in realistic scenarios, where uncooperative subjects are captured under unconstrained imaging conditions. With the introduction of new 3D sensors capable of dynamic 3D acquisitions, the trend of face analysis from video data is now emerging in 2D as well as in 3D (and sometimes in RGB-D, where RGB refers to the color or 2D and D refers to the depth 3D, respectively). Motivated by these considerations, in this work we propose an effective framework to address face recognition from 3D temporal sequences acquired in adverse conditions, which include internal and external occlusions, pose and expression variations, and talking. Due to the novelty of the proposed scenario, a new database has been collected using a single-view

---

[2] http://www.ee.oulu.fi/~gyzhao/ECCV_workshop/
[3] http://www.vap.aau.dk/ffer14/

structured-light dynamic camera, which allows free movements of the acquired subjects (Figure 6.1), in the field-of-view of the camera. The new database includes several 3D video records (panel A. of Figure 6.1) with freely pose variations, expressions (or talking), external occlusions (glasses, hand, scarf), and full 3D models as shown in panel B. of the same Figure. Despite the over-mentioned limitations regarding the noise and low-resolution data, adding the temporal dimension to 3D acquisitions is motivated by the observation that the face is a deformable 3D surface changing over time, so that using the temporal component can be useful to improve the recognition, especially under adverse acquisition conditions.

FIGURE 6.1: The new 4D face dataset recently collected for unconstrained 4D face recognition. A.) illusrtaes 3D sequences with expressions.

At least two alternatives could be explored to overcome these limitations and benefit from the available dynamic data: **(1) A super-resolution approach** which consists on fusing several frames of the 3D video to increase the 3D shape resolution then perform the analysis as proposed recently in (Berretti et al., 2014); or **(2) A subspace-based representation** (Turaga et al., 2011) which permits to smooth the effect of noisy data, at the same time showing robustness with respect to acquisition variations. In our ongoing research, we adopt and elaborate more on the second solution in the context of two applications: 4D face recognition and analysis of spontaneous emotions from depth-consumer cameras,

↪ **Unconstrained 4D Face Recognition** – As stated earlier, a subspace-based modeling approach is investigated, where the spatio-temporal 3D data is mapped on a finite-dimensional manifold. Each 3D video fragment is considered as an element of a Grassmann manifold (the set of real

k-dimensional linear subspaces). This formulation has interesting aspects. First, comparing two subspaces is cheaper than comparing two 3D dynamic fragments; Second, it is more robust to noise and missing data, which are common in realistic scenarios and data from depth sensors. Here, we consider a holistic face descriptor based on shape normals, which do not require any manual/automatic Landmarking step. The facial motion is also modeled and exploited in the recognition process. The 3D temporal sequences are divided into fragments each modeled as a linear subspace in order to embody the shape and the motion of the facial surfaces. In virtue of the Riemannian geometry of the Grassmann manifold, a formulation of the matching between 3D temporal sequences is developed. The underlying representation allows to use differential geometry to develop relevant tools to analyze dynamic 3D data. For example, the use of geodesics to interpolate between subspaces could be an intuitive solution for domain adaptation (Gopalan et al., 2011)(Ni et al., 2013) useful when the test distributions are different from the training distributions. Also, under this representation, it is possible to perform unsupervised clustering over the manifold and compute statistics from populations of dynamic fragments of the same pose, same person, same emotions, same age, or same gender. The shape representation with normals is proposed due to their invariance to location (translation) and resolution (scaling). One important issue to be addressed using this representation is the dense correspondence/tracking of normals across the 3D video, under the variation of expression, pose, missing data, occlusions, etc. Preliminary promising results[4] were obtained using the proposed framework applied on the new 4D face dataset and on BU-4DFE dataset following the experimental settings of (Sun et al., 2010a).

↪ **Analysis of Spontaneous Emotions from Depth-consumer Cameras** – Human emotion understanding (Zeng et al., 2009) from facial dynamics (expressions and head movements) and the human body interactions/reactions (Kleinsmith and Bianchi-Berthouze, 2013) is a challenging problem and presents increasing need in many applications. In this research work, we investigate the use of dynamic depth data acquired using cost-effective sensors for emotion analysis. Most of the existing approaches used the facial data to analyze facial expressions and perform classification in discrete space of six universal facial expressions as defined by (Ekman and Friesen, 1978). The datasets used to develop and test these approaches often consider posed (sometimes exaggerated) expressions, i.e the participants are asked to perform one of the six universal expressions. The use of these approaches in real-world applications is difficult, when the user conveys spontaneous emotions involving in general the upper part of the body including his face and hand gestures. We propose to explore the contribution of the upper-body part and its dynamics to perform emotion classification. Again, we adopt the subspace representations of dynamic map-flows to capture the history of the body motions used for emotion (or a group of emotion) detection. Regarding the test datasets, it is very important that psychologists implement the data acquisition settings and perform sequences segmentation. For these reasons, in our ongoing research, we try to use

---

[4]T. Alashkar, B. Ben Amor, S. Berretti and M. Daoudi, *A Grassmannian Framework for Face Recognition of 3D Dynamic Sequences with Challenging Conditions*, accepted in Sixth Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (NORDIA), 2014.

datasets collected by specialized groups such as the challenging corpus described in (Mahmoud et al., 2011). Recently, we have conducted negative/positive emotion classification on the above-mentioned 3Dcam dataset[5].

Automated analysis of human behavior through visual data has attracted a tremendous interest in the computer vision community. This is due to its huge potential in a wide spectrum of areas, such as human-machine interaction, psychology (Kleinsmith and Bianchi-Berthouze, 2013), surveillance security, health-care and social assistance, and gaming. Beside the verbal communication, the visual data is one of the most important cues in developing systems for understanding human behavior (Mehrabian and Wiener, 1967), ranging from tracking daily activities to classifying emotional states, as well as detecting abnormal and suspicious activities. In this future direction, our goal is to develop computational solutions for human behavior analysis from dynamic depth data. In literature, a vast majority of the features that are extracted to solve this problem often tend to be non-Euclidean manifolds. This means that traditional ideas of comparison and classification need to be generalized to account for the geometry of the manifolds.

↪ **Human Motions as Trajectories on Manifolds** – Earlier studies in such analyses used videos produced by conventional cameras (Aggarwal and Ryoo, 2011)(Abdelkader et al., 2011)(Chaquet et al., 2013)(Ke et al., 2013)(Veeraraghavan et al., 2009). They analyzed videos in order to exploit visual cues related to the human 'form' (or silhouette) in temporal evolution of images and to capture its dynamics (e.g., using optical flows). However, real-life applications of this approach encounter several difficulties, coming from self-occlusions, pose variations, and the complexity of both imaging environments and activities themselves. Additionally, there are two major challenges which also negatively affect video analysis, (i) the temporal variability caused by differences in execution rates of the motions and periods of them, including arbitrary starting/ending observation times, and (ii) the spatial variability due to interpersonal differences when exhibit similar motion. These were difficult issues when dealing with videos from conventional cameras. The depth maps are independent of human appearance (textures) and provide a more complete human 'form' compared to the silhouette information, explored in the past. They can also help in mitigating the issue of pose variability due to availability of 3D registration methods and the existence of pose-independent features. The relevant question in using depth maps is: How to exploit the depth information in an efficient and robust way? Recently, several approaches have been published in action and activity recognition from depth cameras (Li et al., 2010)(Wang et al., 2013)(Ellis et al., 2013)(Oreifej and Liu, 2013)(Vemulapalli et al., 2014)(Devanne et al., 2014)(Slama et al., 2014). In [**J2-s**][6], we study the problem of classifying actions of human subjects using depth movies generated by the Kinect or other depth sensors. Representing human body as dynamical skeletons, we study the evolution of their (skeletons') shapes as trajectories on Kendall's shape manifold (reviewed in Chapter 2).

---

[5] L. Ballihi, A. Lablack, B Ben Amor, M. Bilasco, and M. Daoudi, *Positive/Negative Emotion Detection from RGB-D upper Body Images*, accepted for publication in 1st International Workshop on Face and Facial Expression Recognition from Real World Videos (FFER), 2014.

[6] under major revision with IEEE Trans. on Pattern Analysis and Machine Intelligence.

The action data is typically corrupted by large variability in execution rates within and across subjects and, thus, causing major problems in statistical analyses. To address that issue, we adopt a recently-developed framework of (Su et al., 2013) to this problem domain. Here the execution rate changes correspond to re-parameterizations of trajectories, and one uses a parameterization-invariant metric for aligning, comparing, averaging, and modeling trajectories. The key idea is to use a combination of transported square-root vector fields (T-SRVFs) of trajectories and the standard Euclidean norm. Moreover, we have developed a suite of computational tools for this application domain: smoothing and denoising skeletal trajectories using median filtering, up- and down-sampling actions in time domain, simultaneous temporal-registration of multiple actions, and extracting **invertible Euclidean representations** of actions. Due to invertibility these Euclidean representations allow both discriminative and generative models for statistical analysis. Action recognition from 3D skeletal data was an excellent application to illustrate our methodology, however, in the future we intend to extend our study to other applications. We will emphasize on facial expression classification from automatic detected 2D and 3D landmarks. The Fig. 2.4, in Chapter 2 provides an example of geodesic path between sets of facial landmarks. We will consider trajectories of shapes to study the facial behavior across time and study the application of our mathematical framework.

↪ **Statistical Models for (landmark-based) Shapes and Trajectories** – As stated in the previous item, I shall emphasize in my future work on (3D landmark-based) shapes and their motion. The specification of a shape manifold (Kendall's space) and the corresponding metric enables to compare arbitrary trajectories in terms of their shapes, in a rate-invariant way. Additional shape tools, such as the computation of sample mean and sample covariance statistics, and the transfer of deformations using parallel transports becomes straightforward. However, in classification we often need to define statistical models that are tailored to the geometry of that underlying shape space and to specify some probability models for shapes along trajectories. The **definition of probability models on shape manifolds and trajectory manifolds** will be also explored to perform sequential analysis.

Automated processing and analysis of 3D dynamic data will also open the door to the **large-scale recognition** challenges in which it is important to consider different trade-offs with respect to algorithms efficiency and recognition performance. This topic has received much attention in the computer vision community in the last few years (Taigman et al., 2014), in particular in processing large datasets of images and videos. Hence, there is an outstanding need for developing new computational strategies to adapt the Riemannian methods to fit these requirements. In fact, when the local descriptions provide rich information (high-dimensional features usually represented in Euclidean spaces) and are cheap to compute, the Riemannian methods (involving non-linear spaces) are more costly and provide robustness to several transformations. As presented previously, one possible future direction is to consider two data analysis levels, (1) shape analysis of sparse representations (or landmarks) and (2) shape analysis of dense

representations (or the complete shapes) to tackle the computational cost problem. In human behavior analysis for example, one can exploit the landmark-based representations provided by the depth-consumer cameras such as the skeletal data of the body joints or the 3D Active Shape Model (provided with the Kinect SDK) of the face to derive efficient discriminative and generative models, as a first step. In a second step, the recognition will be restricted to a subset of the large dataset, in which a more complete 3D shape representations are involved.

Applications

In summary, our focus goes to address human (including the face) behavior analysis and recognition in real-world conditions. We explore the use of dynamic depth-flows acquired by cost-effective sensors and high quality 3D sequences of meshes collected using more sophisticated scanners. Despite the temporal dimension, the 3D sequences are independent of human appearance and provide a more complete human 'form' (meshes, depth-maps, landmarks, etc.) compared to the 2D silhouette information, used in the past (Abdelkader et al., 2011). The extracted representations often live on non-Euclidean spaces (Kendall's shape space, shape space of parametrized curves or surfaces, Matrix manifolds, etc.) which required the use of tools from differential geometry to process shapes and their animations. Through the geometric modeling, computing geodesics and interpolate between the shapes is an essential ingredient for comparing, deforming and averaging these elements.

Many applications would be targeted with the developed tools and strategies, in the future. For example, diagnostic, therapy and tracking of diseases like Autism. It is well established nowadays the important role of motion-controlled gaming to improve autistic kid's life by limiting their disables and attract (force) them to communicate with others by sharing space, etc. The above-mentioned ideas could be adapted to track the autism behavior evolution across weeks, months and perhaps years and can provide quantitative statistics, to help doctors and families get clearer picture on the temporal evolution of the children with autism. Other applications exploiting the hand gestures could be also addressed using our recent developments. In fact, new ToF cameras like Softkinetic[7] cameras (or Creative Senz3D[8] cameras) and their middle-ware tend to extract and track accurately skeletal data from hands which makes possible the analyses of dynamic shapes as we have proposed in [**J2-s**]. This open a new avenue to sign language interpretation using the depth data (or the skeletal data) that the cameras provide, for example. We expect that the landmark-based representations automatically estimated and tracked by the depth-cameras will play an important role in shape analysis. In fact, the explicit correspondence between these sparse representations and the dense information (point clouds, meshes, surface patches, etc.) could be exploited to design up-down shape analysis approach linking the shape manifolds of **sparse shape representation** and the **dense shape representations**. Finally, our aim is to see our methodologies implemented for those end-users applications!

---

[7] www.softkinetic.com/
[8] us.creative.com/p/web-cameras/creative-senz3d

Bibliography

Boulbaba Ben Amor, Mohsen Ardabilian, and Liming Chen. New experiments on icp-based 3d face recognition and authentication. In *ICPR (3)*, pages 1195–1199, 2006a.

Boulbaba Ben Amor, Mohsen Ardabilian, and Liming Chen. Enhancing 3D face recognition by mimics segmentation. *Intelligent Systems Design and Applications, International Conference on*, 3:150–155, 2006b. doi: http://doi.ieeecomputersociety.org/10.1109/ISDA.2006.24.

Boulbaba Ben Amor, Mohsen Ardabilian, and Liming Chen. Toward a region-based 3d face recognition approach. In *ICME*, pages 101–104, 2008.

David G. Kendall. Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, March 1984. ISSN 1469-2120. doi: 10.1112/blms/16.2.81. URL http://dx.doi.org/10.1112/blms/16.2.81.

Chafik Samir, Anuj Srivastava, and Mohamed Daoudi. Three-dimensional face recognition using shapes of facial curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1853–1863, November 2006.

Chafik Samir, Anuj Srivastava, Mohamed Daoudi, and Eric Klassen. An intrinsic framework for analysis of facial surfaces. *Int. Journal of Computer Vision*, 82(1):80–95, April 2009.

Boulbaba Ben Amor, Hassen Drira, Lahoucine Ballihi, Anuj Srivastava, and Mohamed Daoudi. An experimental illustration of 3D facial shape analysis under facial expressions. *Annals of telecommunications*, 64(5-6):369–379, 2009.

Hassen Drira, Boulbaba Ben Amor, Mohamed Daoudi, and Anuj Srivastava. Nasal region contribution in 3D face biometrics using shape analysis framework. In *Third International Conference on Biometrics*, pages 357–366, Alghero, Italy, June 2-5 2009a.

Stefano Berretti, Boulbaba Ben Amor, Mohamed Daoudi, and Alberto del Bimbo. 3D facial expression recognition using sift descriptors of automatically detected keypoints. *The Visual Computer*, 27(11): 1021–1036, November 2011.

Shantanu H. Joshi, Eric Klassen, Anuj Srivastava, and Ian Jermyn. A novel representation for riemannian analysis of elastic curves in r^n. In *CVPR*, 2007.

Anuj Srivastava, Eric Klassen, Shantanu H. Joshi, and Ian H. Jermyn. Shape analysis of elastic curves in euclidean spaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(7):1415–1428, 2011.

Hassen Drira, Boulbaba Ben Amor, Mohamed Daoudi, and Anuj Srivastava. Pose and expression-invariant 3D face recognition using elastic radial curves. In *Proc. British Machine Vision Conference*, pages 1–11, Aberystwyth, UK, August 2010a.

Hassen Drira, Boulbaba Ben Amor, Anuj Srivastava, Mohamed Daoudi, and Rim Slama. 3d face recognition under expressions, occlusions, and pose variations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(9):2270–2283, 2013a.

Ahamed Maalej, Boulbaba Ben Amor, Mohamed Daoudi, Anuj Srivastava, and Stefano Berretti. Local 3D shape analysis for facial expression recognition. In *Proc. 20th Int. Conf. on Pattern Recognition*, pages 4129–4132, Istanbul, Turkey, August 2010.

Ahmed Maalej, Boulbaba Ben Amor, Mohamed Daoudi, Anuj Srivastava, and Stefano Berretti. Shape analysis of local facial patches for 3D facial expression recognition. *Pattern Recognition*, 44(8):1581–1589, August 2011.

Hassen Drira, Boulbaba Ben Amor, Mohamed Daoudi, Anuj Srivastava, and Stefano Berretti. 3D dynamic expression recognition based on a novel deformation vector field and random forest. In *Proc. Int. Conf. on Pattern Recognition*, pages 1104–1107, Tsukuba, Japan, November 2012.

Boulbaba Ben Amor, Hassen Drira, Stefano Berretti, Mohamed Daoudi, and Anuj Srivastava. 4D Facial Expression Recognition by Learning Geometric Deformations. *IEEE Transactions on Cybernetics*, page 1, February 2014a. URL http://hal.archives-ouvertes.fr/hal-00949002.

Lahoucine Ballihi, Anuj Srivastava, Boulbaba Ben Amor, Mohamed Daoudi, and Driss Aboutajdine. Which 3d geometric facial features give up your identity? In *ICB*, pages 119–124, 2012a.

Lahoucine Ballihi, Boulbaba Ben Amor, Mohamed Daoudi, Anuj Srivastava, and Driss Aboutajdine. Boosting 3-d-geometric features for efficient face recognition and gender classification. *IEEE Transactions on Information Forensics and Security*, 7(6):1766–1779, 2012b.

Baiqiang Xia, Boulbaba Ben Amor, Hassen Drira, Mohamed Daoudi, and Lahoucine Ballihi. Gender and 3D facial symmetry: What's the relationship? In *FG*, pages 1–6, 2013a.

Baiqiang Xia, Boulbaba Ben Amor, Mohamed Daoudi, and Hassen Drira. Can 3D shape of the face reveal your age? In *VISAPP (2)*, pages 5–13, 2014a.

Ian L. Dryden and K.V. Mardia. *Statistical shape analysis*. Wiley series in probability and statistics: Probability and statistics. J. Wiley, 1998. ISBN 9780471958161.

U. Grenander. *General Pattern Theory*. Oxford University Press, 1993.

E. Klassen, A. Srivastava, W. Mio, and S.H. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(3):372–383, March 2004. ISSN 0162-8828.

K. W. Bowyer, K. I. Chang, and P. J. Flynn. A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, January 2006a.

Baiqiang Xia, Boulbaba Ben Amor, Hassen Drira, Mohamed Daoudi, and Lahoucine Ballihi. Combining Face Averageness and Symmetry for 3D-based Gender Classification. *Minor Revisions in Pattern Recognition*, 2014b.

P. Jonathon Phillips, Patrick J. Flynn, W. Todd Scruggs, Kevin W. Bowyer, Jin Chang, Kevin J. Hoffman, Joe Marques, Jaesik Min, and William J. Worek. Overview of the face recognition grand challenge. In *CVPR (1)*, pages 947–954, 2005.

Y. Wang, G. Pan, Z. Wu, and Y. Wang. Exploring facial expression effects in 3D face recognition using partial ICP. In *Asian Conference on Computer Vision*, pages 581–590, Hyderabad, India, January 2006a.

T. Fang, X. Zhao, S.K. Shah, and I.A. Kakadiaris. 4D facial expression recognition. In *Proc. IEEE Int. Conf. on Computer Vision Workshop*, pages 1594–1601, Barcelona, Spain, November 2011.

B. Gong, Y. Wang, J. Liu, and X. Tang. Automatic facial expression recognition on a single 3D face by exploring shape deformation. In *Proc. ACM Int. Conf. on Multimedia*, pages 569–572, Beijing, China, October 2009a.

Vuong Le, Hao Tang, and Thomas S. Huang. Expression recognition from 3D dynamic faces using robust spatio-temporal shape features. In *IEEE Conference on Automatic Face and Gesture Recognition*, pages 414–421, Santa Barbara, CA, March 2011.

I. Mpiperis, S. Malassiotis, and M. G. Strintzis. Bilinear models for 3-D face and facial expression recognition. *IEEE Transactions on Information Forensics and Security*, 3(3):498–511, September 2008a.

H. Tang and T. S. Huang. 3D facial expression recognition based on automatically selected features. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 1–8, Anchorage, AK, June 2008a.

Yi Sun, Xiaochen Chen, Matthew J. Rosato, and Lijun Yin. Tracking vertex flow and model adaptation for three-dimensional spatiotemporal face analysis. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 40(3):461–474, 2010a.

Georgia Sandbach, Stefanos Zafeiriou, Maja Pantic, and Daniel Rueckert. Recognition of 3D facial expression dynamics. *Image and Vision Computing*, 30(10):762–773, 2012a.

Alexander M. Bronstein, Michael M. Bronstein, and Ron Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, 64(1):5–30, 2005a.

Alexander M. Bronstein, Michael M. Bronstein, and Ron Kimmel. Expression-invariant representations of faces. *IEEE Transactions on Image Processing*, 16:188–197, 2007a.

Georgia Sandbach, Stefanos Zafeiriou, Maja Pantic, and Lijun Yin. Static and dynamic 3D facial expression recognition: A comprehensive survey. *Image and Vision Computing*, 30(10):683–697, 2012b.

A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, 64(1):5–30, 2005b.

S. Berretti, A. Del Bimbo, and P. Pala. 3D face recognition using iso-geodesic stripes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2162–2177, December 2010a.

Ioannis A. Kakadiaris, Georgios Passalis, George Toderici, Mohammed N. Murtuza, Yunliang Lu, Nikolaos Karampatziakis, and Theoharis Theoharis. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):640–649, 2007.

X. Lu, A. K. Jain, and D. Colbry. Matching 2.5d face scans to 3D models. *IEEE Pattern Transactions on Pattern Analysis and Machine Intelligence*, 28(1):31–43, 2006a.

A. S. Mian, M. Bennamoun, and R. Owens. Keypoint detection and local feature matching for textured 3D face recognition. *Int. Journal of Computer Vision*, 79(1):1–12, August 2008.

Di Huang, M. Ardabilian, Yunhong Wang, and Liming Chen. A novel geometric facial representation based on multi-scale extended local binary patterns. In *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 1 –7, march 2011.

Timothy C. Faltemier, Kevin W. Bowyer, and Patrick J. Flynn. A region ensemble for 3-d face recognition. *IEEE Transactions on Information Forensics and Security*, 3(1):62–73, 2008. ISSN 1556-6013.

A. Colombo, C. Cusano, and R. Schettini. Gappy pca classification for occlusion tolerant 3D face detection. *Journal of Mathematical Imaging and Vision*, 35(3):193–207, November 2009.

Laurent Younes. Spaces and manifolds of shapes in computer vision: An overview. *Image and Vision Computing*, 30(6–7):389 – 397, 2012. ISSN 0262-8856. doi: http://dx.doi.org/10.1016/j.imavis.2011.09.009. URL http://www.sciencedirect.com/science/article/pii/S0262885611001028.

D.A.W. Thompson. *On Growth and Form.* University Press, 1917.

Fred L. Bookstein. Size and shape spaces for landmark data in two dimensions. *Statistical Science*, 1(2):181–222, 05 1986. doi: 10.1214/ss/1177013696.

U. Grenander, Y. Chow, and D.M.R. Keenan. *Hands: A Pattern Theoretic Study of Biological Shapes.* Research notes in neural computing. Springer-Verlag, 1991. ISBN 9783540973867. URL http://books.google.com/books?id=6nRtQgAACAAJ.

Laurent Younes. Computable elastic distance between shapes. *SIAM Journal of Applied Mathematics*, 58:565–586, 1998.

Ronen Basri, Luiz Costa, Davi Geiger, and David Jacobs. Determining the similarity of deformable shapes. *Vision Research*, 38(15-16):2365–2385, August 1998. URL http://dx.doi.org/10.1016/s0042-6989(98)00043-1.

Peter W. Michor and David Mumford. Riemannian geometries on spaces of plane curves. *J. Eur. Math. Soc. (JEMS*, pages 1–48, 2003.

Washington Mio, Anuj Srivastava, and Shantanu Joshi. On shape of plane elastic curves. *International Journal of Computer Vision*, 73:307–324, 2007. ISSN 0920-5691. URL http://dx.doi.org/10.1007/s11263-006-9968-0. 10.1007/s11263-006-9968-0.

Anuj Srivastava, Shantanu H. Joshi, Washington Mio, and Xiuwen Liu. Statistical shape analysis: Clustering, learning, and testing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(4):590–602, 2005.

Peter W. Michor and David Mumford. An overview of the riemannian metrics on spaces of curves using the hamiltonian approach. *Applied and Computational Harmonic Analysis*, 23(1):74 – 113, 2007. ISSN 1063-5203. doi: http://dx.doi.org/10.1016/j.acha.2006.07.004. URL http://www.sciencedirect.com/science/article/pii/S1063520307000243. Special Issue on Mathematical Imaging.

Martin Bauer and Peter W. Michor. Sobolev metrics on shape space of surfaces. 3, 2011.

Sebastian Kurtek, Eric Klassen, John C. Gore, Zhaohua Ding, and Anuj Srivastava. Elastic geodesic paths in shape space of parameterized surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(9):1717–1730, 2012. doi: 10.1109/TPAMI.2011.233. URL http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.233.

P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

F. Ghorbel. A complete and stable set of invariant fourier descriptors for a random planar shape. In *Image Processing and its Applications, 1992., International Conference on*, pages 278–281, Apr 1992.

F. Ghorbel, M. Daoudi, A Mokadem, O. Avaro, and H. Sanson. Global planar rigid motion estimation applied to object-oriented coding. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 2, pages 641–645 vol.2, Aug 1996. doi: 10.1109/ICPR.1996.546902.

Mohamed Daoudi, Faouzi Ghorbel, A. Mokadem, Olivier Avaro, and Henri Sanson. Shape distances for contour tracking and motion estimation. *Pattern Recognition*, 32(7):1297–1306, 1999.

Faouzi Ghorbel. Towards a unitary formulation for invariant image description: application to image coding. *Annales des Télécommunications*, 53(5-6):242–260, 1998. doi: 10.1007/BF02997680. URL http://dx.doi.org/10.1007/BF02997680.

Laurent Younes. Optimal matching between shapes via elastic deformations. *Image and Vision Computing*, 17(5–6):381 – 389, 1999. ISSN 0262-8856. doi: http://dx.doi.org/10.1016/S0262-8856(98)00125-5. URL http://www.sciencedirect.com/science/article/pii/S0262885698001255.

Peter W. Michor and David Mumford. Riemannian geometries on spaces of plane curves. *Journal of the European Mathematical Society*, 8:1–48, 2006.

Richard Bellman. *Dynamic Programming.* Princeton University Press, Princeton, NJ, USA, 1 edition, 1957.

Daniel T. Robinson. *Functional Data Analysis and Partial Shape Matching in The Square Root Velocity Framework.* PhD thesis, Florida State University, 2012.

T. Heseltine, N. Pears, and J. Austin. Three-dimensional face recognition: an eigensurface approach. In *International Conference on Image Processing*, volume 2, pages 1421–1424, Singapore, October 2004.

X. Lu, D. C., and A. K. Jain. Three-dimensional model based face recognition. In *ICPR (1)*, pages 362–366, 2004.

A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three dimensional face recognition. *International Journal of Computer Vision*, 64(1):5–30, August 2005c.

S. Berretti, A. Del Bimbo, and P. Pala. Description and retrieval of 3D face models using iso-geodesic stripes. In *ACM International Workshop on Multimedia Information Retrieval*, pages 13–22, Santa Barbara, CA, October 2006.

H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977. doi: 10.1002/cpa.3160300502.

Hassen Drira, Boulbaba Ben Amor, Anuj Srivastava, and Mohamed Daoudi. A Riemannian analysis of 3D nose shapes for partial human biometrics. In *IEEE International Conference on Computer Vision*, pages 2050–2057, 2009b.

Hassen Drira, Boulbaba Ben Amor, Mohamed Daoudi, and Anuj Srivastava. Pose and expression-invariant 3D face recognition using elastic radial curves. In *Proceedings of the British Machine Vision Conference*, pages 1–11. BMVA Press, 2010b. ISBN 1-901725-38-3. doi:10.5244/C.24.90.

Hassen Drira, Boulbaba Ben Amor, Mohamed Daoudi, and Anuj Srivastava. Elastic radial curves to model 3D facial deformations. In *3-D Object Retrieval ACM Multimedia Workshop*, pages 75–80, Florence, Italy, October 25 2010c.

Kevin W. Bowyer, Kyong Chang, and Patrick Flynn. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Comput. Vis. Image Underst.*, 101(1):1–15, 2006b. ISSN 1077-3142. doi: http://dx.doi.org/10.1016/j.cviu.2005.05.005.

Georgios Passalis, Panagiotis Perakis, Theoharis Theoharis, and Ioannis A. Kakadiaris. Using facial symmetry to handle pose variations in real-world 3D face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):1938–1951, 2011.

F. ter Haar and R. C. Velkamp. Expression modeling for expression-invariant face recognition. *Computers and Graphics*, 34(3):231–241, 2010.

Xiaoguang Lu and A.K. Jain. Deformation modeling for robust 3D face matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8):1346 –1357, aug. 2008.

Yonguk Lee, Hwanjong Song, Ukil Yang, Hyungchul Shin, and Kwanghoon Sohn. Local feature based 3D face recognition. In *Proceedings of Audio- and Video-Based Biometric Person Authentication, AVBPA*, pages 909–918, 2005.

Shalini Gupta, J. K. Aggarwal, Mia K. Markey, and Alan C. Bovik. 3D face recognition founded on the structural diversity of human faces. In *Proceeding of Computer Vision and Pattern Recognition, CVPR*, 2007.

Gaile Gordan. Face recognition based on depth and curvature features. In *Proceedings of Conference on Computer Vision and Pattern Recognition, CVPR*, pages 108–110, 1992.

Xiaoxing Li, Tao Jia, and Hao Zhang. Expression-insensitive 3D face recognition using sparse representation. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:2575–2582, 2009. doi: http://doi.ieeecomputersociety.org/10.1109/CVPRW.2009.5206613.

Yueming Wang, Jianzhuang Liu, and Xiaoou Tang. Robust 3D face recognition by local shape difference boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:1858–1870, 2010. ISSN 0162-8828. doi: http://doi.ieeecomputersociety.org/10.1109/TPAMI.2009.200.

R. McKeon and T. Russ. Employing region ensembles in a statistical learning framework for robust 3D facial recognition. In *Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, pages 1–7, 2010.

D. Huang, G. Zhang, M Ardabilian, Wang, and L. Chen. 3D face recognition using distinctiveness enhanced facial representations and local feature hybrid matching. In *Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, pages 1–7, 2010.

A.K. Moorthy, A. Mittal, S. Jahanbin, K. Grauman, and A.C. Bovik. 3D facial similarity: automatic assessment versus perceptual judgments. In *Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, pages 1–7, 2010.

Luuk Spreeuwers. Fast and accurate 3D face recognition using registration to an intrinsic coordinate system and fusion of multiple region classifiers. *International Journal of Computer Vision*, 93(3): 389–414, 2011.

Chaua C. Queirolo, Luciano Silva, Olga R.P. Bellon, and Mauricio Pamplona Segundo. 3D face recognition using simulated annealing and the surface interpenetration measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:206–219, 2010. ISSN 0162-8828. doi: http://doi.ieeecomputersociety.org/10.1109/TPAMI.2009.14.

N. Alyuz, B. Gokberk, and L. Akarun. A 3D face recognition system for expression and occlusion invariance. In *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, 29 2008.

Alexander M. Bronstein, Michael M. Bronstein, and Ron Kimmel. Expression-invariant representations of faces. *IEEE Transactions on Image Processing*, 16(1):188–197, 2007b.

I. Mpiperis, S. Malassiotis, and M.G. Strintzis. Bilinear models for 3-d face and facial expression recognition. *Information Forensics and Security, IEEE Transactions on*, 3(3):498 –511, sept. 2008b. ISSN 1556-6013. doi: 10.1109/TIFS.2008.924598.

Stefano Berretti, Alberto Del Bimbo, and Pietro Pala. 3D face recognition using isogeodesic stripes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2162–2177, 2010b.

A. B. Moreno and A. Sanchez. Gavabdb: A 3D face database. In *Workshop on Biometrics on the Internet*, pages 77–85, 2004.

Ana Belen Moreno, Angel Sanchez, Jose Fco. Velez, and Fco. Javier Díaz. Face recognition using 3D local geometrical features: Pca vs. svm. In *Int. Symp. on Image and Signal Processing and Analysis*, 2005.

Mohammad H. Mahoor and Mohamed Abdel-Mottaleb. Face recognition based on 3D ridge images obtained from range data. *Pattern Recognition*, 42(3):445–451, 2009.

Mir Hashem Mousavi, Karim Faez, and Amin Asghari. Three dimensional face recognition using svm classifier. In *ICIS '08: Proceedings of the Seventh IEEE/ACIS International Conference on Computer and Information Science*, pages 208–213, Washington, DC, USA, 2008. ISBN 978-0-7695-3131-1. doi: http://dx.doi.org/10.1109/ICIS.2008.77.

Alessandro Colombo, Claudio Cusano, and Raimondo Schettini. Three-dimensional occlusion detection and restoration of partially occluded faces. *Journal of Mathematical Imaging and Vision*, 40(1):105– 119, 2011.

Thomas Hofmann and Joachim M. Buhmann. Active data clustering. In *In Advances in Neural Information Processing Systems 10*, pages 528–534, 1998.

Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *EuroCOLT '95: Proceedings of the Second European Conference on Computational Learning Theory*, pages 23–37, London, UK, 1995. Springer-Verlag. ISBN 3-540-59119-2.

Albert Mehrabian and Morton Wiener. Decoding of inconsistent communications. *Journal of Personality and Social Psychology*, 6(1):109–114, May 1967.

P. Ekman. Universals and cultural differences in facial expressions of emotion. In *Proc. Nebraska Symposium on Motivation*, volume 19, pages 207–283, Lincoln, NE, 1972.

P. Ekman and W. V. Friesen. *Manual for the the Facial Action Coding System.* Consulting Psychologist Press, Palo Alto, CA, 1977.

Z. Zeng, M. Pantic, G. Roisman, and T. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31 (1):39–58, January 2009.

L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato. A 3D facial expression database for facial behavior research. In *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 211–216, Southampton, UK, April 2006a.

A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gö, B. Sankur, and L. Akarun. Bosphorus database for 3D face analysis. In *Proc. First COST 2101 Workshop on Biometrics and Identity Management*, May 2008.

Di3D. http://www.di3d.com. 2006.

3DMD. http://www.3dmd.com. 2010.

L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A high-resolution 3D dynamic facial expression database. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FGR08)*, pages 1–6, Amsterdam, The Netherlands, September 2008.

B.J. Matuszewski, W. Quan, and L.-K. Shark. High-resolution comprehensive 3-D dynamic database for facial articulation analysis. In *Proc. IEEE Int. Conf. on Computer Vision Workshops*, pages 2128–2135, Barcelona, Spain, November 2011.

Bogdan J. Matuszewski, Wei Quan, Lik-Kwan Shark, Alison S. McLoughlin, Catherine E. Lightbody, Hedley C.A. Emsley, and Caroline L. Watkins. Hi4D-ADSIP 3-D dynamic facial articulation database. *Image and Vision Computing*, 30(10):713–727, 2012 2012.

D. Cosker, E. Krumhuber, and A. Hilton. A FACS valid 3d dynamic action unit database with applications to 3D dynamic morphable facial modeling. In *Proc. IEEE Int. Conf, on Computer Vision*, pages 2296–2303, Barcelona, Spain, November 2011.

Kinect. http://www.xbox.com. 2010.

Point grey. http://www.ptgrey.com. 2010.

Asus. http://www.asus.com. 2010.

S. Berretti, A. Del Bimbo, and P. Pala. Superfaces: A super-resolution model for 3D faces. In *Proc. of Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, pages 73–82, Florence, Italy, October 2012.

Y. Li, Ajmal Mian, W. Lu, and A. Krishna. Using kinect for face recognition under varying poses, expressions, illumination and disguise. In *Proc. IEEE Workshop on Applications of Computer Vision*, pages 186–192, Tampa, FL, January 2013.

M. Pantic and L. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, December 2000.

A. Samal and Prasana A. Iyengar. Automatic recognition and analysis of human faces and facial expressions: a survey. *Pattern Recognition*, 25(1):65–77, 1992. ISSN 0031-3203.

J. Whitehill and C. W. Omlin. Local versus global segmentation for facial expression recognition. In *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, pages 357–362, 2006.

P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, 1978. Consulting Psychologists Press.

J. Wang, L. Yin, X. Wei, and Y. Sun. 3d facial expression recognition based on primitive surface feature distribution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1399–1406, 2006b.

H. Soyel and H. Demirel. Facial expression recognition using 3d facial feature distances. *International Conference on Image Analysis and Recognition (ICIAR)*, pages 831–838, 2007.

H. Tang and T.S. Huang. 3d facial expression recognition based on automatically selected features. *In First IEEE Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB)*, pages 1–8, 2008b.

E. Klassen and A. Srivastava. Geodesics between 3D closed curves using path-straightening. In *Proceedings of ECCV, Lecture Notes in Computer Science*, pages I: 95–106, 2006.

Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J. Rosato. A 3d facial expression database for facial behavior research. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, FGR '06, pages 211–216, 2006b.

B. Gong, Y. Wang, J. Liu, and X. Tang. Automatic facial expression recognition on a single 3D face by exploring shape deformation. In *Proceedings of the ACM International Conference on Multimedia*, pages 569–572, Beijing, China, Oct 2009b.

M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and Ian H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl*, 11:10–18, 2009.

G. Batista and M. C. Monard. An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence*, 17:519–533, 2003.

S. Gupta, M. K. Markey, and A. C. Bovik. Anthropometric 3d face recognition. *International Journal of Computer Vision*, 2010.

Y. Sun, X. Chen, M. Rosato, and L. Yin. Tracking vertex flow and model adaptation for three-dimensional spatiotemporal face analysis. *IEEE Transactions on Systems, Man, and Cybernetics–Part A*, 40(3): 461–474, 2010b. ISSN 1083-4427.

L. Benedikt, V. Kajic, D. Cosker, P.L. Rosin, and D. Marshall. Facial dynamics in biometric identification. In *Proc. British Machine Vision Conf.*, pages 1–10, Leeds, UK, September 2008.

Lanthao Benedikt, Darren Cosker, Paul L. Rosin, and David Marshall. Assessing the uniqueness and permanence of facial actions for use in biometric applications. *IEEE Transactions on Systems, Man and Cybernetics - Part A*, 40(3):449–460, May 2010.

Y. Sun and L. Yin. Facial expression recognition based on 3D dynamic range model sequences. In *Proc. Eur. Conf. on Computer Vision*, pages 58–71, Marseille, France, October 2008.

Georgia Sandbach, Stefanos Zafeiriou, Maja Pantic, and Daniel Rueckert. A dynamic approach to the recognition of 3D facial expressions and their temporal models. In *Proc. IEEE Conf. on Automatic Face and Gesture Recognition*, pages 406–413, Santa Barbara, CA, March 2011.

G. Zhao and M. Pietikäinen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):915–928, June 2007.

3d/4d facial expression analysis: An advanced annotated face model approach. *Image and Vision Computing*, 30(10):738 – 749, 2012. ISSN 0262-8856.

Michael Reale, Xing Zhang, and Lijun Yin. Nebula feature: A space-time feature for posed and spontaneous 4d facial behavior analysis. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 0:1–8, 2013.

Xing Zhang, Lijun Yin, Jeffrey F. Cohn, Shaun J. Canavan, Michael Reale, Andy Horowitz, and Peng Liu. A high-resolution spontaneous 3d dynamic facial expression database. In *FG*, pages 1–6, 2013.

D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations:application to breast mr images. *IEEE Transactions on medical imaging*, 18(8):712–721, 1999a.

D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999b.

Leo Breiman. Random forests. 45(1):5–32, October 2001.

Y. Linde, A. Buzo, and R.M. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, 28(1):84–94, January 1980.

L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of IEEE*, 77(2):257–286, February 1989.

Baiqiang Xia, Boulbaba Ben Amor, Hassen Drira, Mohamed Daoudi, and Lahoucine Ballihi. Gender and 3D facial symmetry: What's the relationship? In *IEEE Conference on Automatic Face and Gesture Recognition*, 2013b.

Baiqiang Xia, Boulbaba Ben Amor, Mohamed Daoudi, and Hassen Drira. Can 3d shape of the face reveal your age? In *International Conference on Computer Vision Theory and Applications*, 2014c.

Baiqiang Xia, Boulbaba Ben Amor, and Mohamed Daoudi. Joint Gender, Ethnicity and Age Estimation from 3D Faces – An Experimental Illustration of their Mutual Correlation. *Submitted to the International Journal of Computer Vision*, 2014d.

Zhuang Ziqing, Landsittel Douglas, Benson Stacey, Roberge Raymond, and Shaffer Ronald. Facial anthropometric differences among gender, ethnicity, and age groups. 54(4):391–402, 2010.

Bruce Vicki, Burton A. Mike, Hanna Elias, Healey Pat, Mason Oli, and Coombes Anne. Sex discrimination: How do we tell the difference between male and female faces? In *Perception*, volume 22(2), pages 131–152, 1993.

Forrest CR. Farkas LG, Katic MJ. International anthropometric study of facial morphology in various ethnic groups/races. *Journal of Craniofacial Surgery*, 16(4):615–646, 2005.

Jennifer Alphonse, Jennifer Cox, Jill Clarke, Philip Schluter, and Andrew McLennan. The effect of ethnicity on 2d and 3d frontomaxillary facial angle measurement in the first trimester. In *Obstetrics and Gynecology International*, 2013.

Matthew G Rhodes. Age estimation of faces: a review. In *Appl. Cognit. Psychol*, volume 23, pages 1–12, 2009.

Narayanan Ramanathan, Rama Chellappa, and Soma Biswas. Computational methods for modeling facial aging: A survey. In *Journal of Visual Languages & Computing*, volume 20, pages 131–144. Elsevier, 2009.

Anil K. Jain, Sarat C. Dass, Karthik Nandakumar, and Karthik N. Soft biometric traits for personal recognition systems. In *Proceedings of International Conference on Biometric Authentication, Hong Kong*, pages 731–738, 2004.

Shirakabe Yukio, Suzuki Yoshiro, and Lam SamuelM. A new paradigm for the aging asian face. In *Aesthetic Plastic Surgery*, volume 27, pages 397–402, 2003.

Guillaume Vignali, Harold Hill, and Eric Vatikiotis Bateson. Linking the structure and perception of 3d faces: Gender, ethnicity ,and expressive posture. In *International Conference on Audio-Visual Speech Processing (AVSP)*, 2003.

Wei Gao and Haizhou Ai. Face gender classification on consumer images in a multiethnic environment. In *Advances in Biometrics*, pages 169–178, 2009.

J. Ylioinas, A. Hadid, and M. Pietikäinen. Combining contrast information and local binary patterns for gender classification. In *Image Analysis*, volume 6688, pages 676–686, 2011.

W. Yang, C. Chen, K. Ricanek, and C. Sun. Gender classification via global-local features fusion. In *Biometric Recognition*, volume 7098, pages 214–220, 2011.

C. Shan. Learning local binary patterns for gender classification on real-world face images. In *Pattern Recognition Letters*, volume 33, pages 431–437, 2012.

N. Kumar, A. Berg, P.N. Belhumeur, and S. Nayar. Describable visual attributes for face verification and image search. In *Pattern Analysis and Machine Intelligence*, volume 33, pages 1962 –1977, 2008.

C. Wang, D. Huang, Y. Wang, and G. Zhang. Facial image-based gender classification using local circular patterns. In *International Conference on Pattern Recognition*, 11 2012.

E. Makinen and R. Raisamo. An experimental comparison of gender classification methods. In *Pattern Recognition Letters*, volume 29, pages 1544–1556, 2008.

Farinella Giovanna and Dugelay Jean-Luc. Demographic classification: Do gender and ethnicity affect each other? In *International Conference on Informatics Electronics and Vision (ICIEV)*, pages 383–390, 2012.

Y. Liu and J. Palmer. A quantified study of facial asymmetry in 3D faces. In *Analysis and Modeling of Faces and Gestures*, pages 222–229, 2003.

X. Han, H. Ugail, and I. Palmer. Gender classification based on 3D face geometry features using svm. In *CyberWorlds*, pages 114–118, 2009.

Y. Hu, J. Yan, and P. Shi. A fusion-based method for 3D facial gender classification. In *Computer and Automation Engineering (ICCAE)*, volume 5, pages 369–372, 2010.

G. Toderici, S. O'Malley, G. Passalis, T. Theoharis, and I. Kakadiaris. Ethnicity- and gender-based subject retrieval using 3-D face-recognition techniques. In *International Journal of Computer Vision*, volume 89, pages 382–391, 2010.

L. Ballihi, B. Ben Amor, M. Daoudi, A. Srivastava, and D. Aboutajdine. Boosting 3D-geometric features for efficient face recognition and gender classification. In *IEEE Transactions on Information Forensics and Security*, volume 7, pages 1766–1779, 2012c.

Syed Zulqarnain Gilani, Faisal Shafait, and Ajmal S. Mian. Biologically significant facial landmarks: How significant are they for gender classification? In *DICTA*, pages 1–8, 2013.

X. Lu, H. Chen, and A. Jain. Multimodal facial gender and ethnicity identification. In *International conference on Advances in Biometrics*, pages 554–561, 2006b.

J. Wu, W.A.P. Smith, and E.R. Hancock. Gender classification using shape from shading. In *International Conference on Image Analysis and Recognition*, pages 499–508, 2007.

T. Huynh, R. Min, and J-L. Dugelay. An efficient lbp-based descriptor for facial depth images applied to gender recognition using rgb-d face data. In *ACCV 2012, Workshop on Computer Vision with Local Binary Pattern Variants*, 2012.

Gutta Srinivas, Wechsler Harry, and Phillips P Jonathon. Gender and ethnic classification of face images. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 194–199, 1998.

Yang Zhiguang and Ai Haizhou. Demographic classification with local binary patterns. In *Advances in Biometrics*, pages 464–473. 2007.

Andreas Lanitis, Chrisina Draganova, and Chris Christodoulou. Comparing different classifiers for automatic age estimation. In *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,*, volume 34, pages 621–628, 2004.

Andreas Lanitis, Christopher J. Taylor, and Timothy F. Cootes. Toward automatic simulation of aging effects on face images. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 24, pages 442–455, 2002.

Tao Wu, P. Turaga, and R. Chellappa. Age estimation and face verification across aging using landmarks. volume 7, pages 1780–1788, 2012.

Guodong Guo, Yun Fu, Thomas S Huang, and Charles R Dyer. Locally adjusted robust regression for human age estimation. In *IEEE Workshop on Applications of Computer Vision, 2008. WACV 2008*, pages 1–6, 2008a.

Guodong Guo, Yun Fu, Charles R Dyer, and Thomas S Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. 17(7):1178–1188, 2008b.

Changsheng Li, Qingshan Liu, Jing Liu, and Hanqing Lu. Learning ordinal discriminative features for age estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2570–2577, 2012.

Guo Guodong, Mu Guowang, Fu Yun, and Huang Thomas S. Human age estimation using bio-inspired features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 112–119, 2009.

Geng Xin, Zhou Zhi-Hua, and Smith-Miles Kate. Automatic age estimation based on facial aging patterns. 29(12):2234–2240, 2007.

Geng Xin, Zhou Zhi-Hua, Zhang Yu, Li Gang, and Dai Honghua. Learning from facial aging patterns for automatic age estimation. In *ACM international conference on Multimedia*, pages 307–316, 2006.

Suo Jinli, Zhu Song-Chun, Shan Shiguang, and Chen Xilin. A compositional and dynamic model for face aging. volume 32, pages 385–401, 2010.

NS Lakshmiprabha, J Bhattacharya, and S Majumder. Age estimation using gender information. In *Computer Networks and Intelligent Computing*, pages 211–216, 2011.

Ueki Kazuya, Sugiyama Masashi, and Ihara Yasuyuki. Perceived age estimation under lighting condition change by covariate shift adaptation. In *International Conference on Pattern Recognition (ICPR)*, pages 3400–3403, 2010.

Hu Yuxiao, Fu Yun, Tariq Usman, and Huang Thomas S. Subjective experiments on gender and ethnicity recognition from different face representations. In *Advances in Multimedia Modeling*, volume 5916, pages 66–75, 2010.

Hill Harold, Bruce Vicki, and Akamatsu Shigeru. Perceiving the sex and race of faces: The role of shape and colour. In *Proceedings Of The Royal Society Of London Series B Biological Sciences*, volume 261(1362), pages 367–373, 1995.

Hassen Drira, Boulbaba Ben Amor, Anuj Srivastava, Mohamed Daoudi, and Rim Slama. 3d face recognition under expressions, occlusions, and pose variations. 35(9):2270–2283, 2013b.

Boulbaba Ben Amor, Hassen Drira, Stefano Berretti, Mohamed Daoudi, and Anuj Srivastava. 4D Facial Expression Recognition by Learning Geometric Deformations. *IEEE Transactions on Cybernetics, accepted for publication*, February 2014b.

Mark A. Hall. Correlation-based feature subset selection for machine learning. In *PhD thesis, Department of Computer Science, University of Waikato*, 1999.

R. Kohavi. Wrappers for performance enhancement and oblivious decision graphs. In *PhD thesis, Stanford University*, 1995.

Xiaolong Wang and Chandra Kambhamettu. Department of computer and information sciences, university of delaware, newark, usa. In *Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1077–1080. IEEE, 2013.

Zhong Cheng, Sun Zhenan, and Tan Tieniu. Fuzzy 3d face ethnicity categorization. In *Advances in Biometrics*, pages 386–393. 2009.

Baiqiang Xia, Boulbaba Ben Amor, Di Huang, Mohamed Daoudi, Yunhong Wang, and Hassen Drira. Enhancing gender classification by combining 3d and 2d face modalities. In *European Signal Processing Conference (EUSIPCO)*, 2013c.

Hai Wang, Bongnam Kang, and Daijin Kim. Pfw: A face database in the wild for studying face identification and verification in uncontrolled environment. In *Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on*, pages 356–360, Nov 2013.

K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher Vector Faces in the Wild. In *British MACHINE VISION Conference*, 2013.

C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Proceedings of IEEE Int'l Conf. on Computer Vision (ICCV-W 2013), 300 Faces in-the-Wild Challenge (300-W)*, Sydney, Australia, December 2013.

DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. Incremental face alignment in the wild. In *Proceedings of IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR 2014, Accepted)*, 2014.

Mihalis A. Nicolaou, Hatice Gunes, and Maja Pantic. Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *T. Affective Computing*, 2(2):92–105, 2011.

Konstantinos Bousmalis, Marc Mehu, and Maja Pantic. Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools. *Image Vision Comput.*, 31(2):203–221, 2013.

Zhengyou Zhang. Microsoft kinect sensor and its effect. *IEEE Multimedia*, 19(2):4–10, 2012. ISSN 1070-986X.

J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1297–1304, 2011. ISBN 978-1-4577-0394-2.

P.J. Phillips, Scruggs T., A.J. O'Toole, , P.J. Flynn, K.W. Bowyer, and C.T. Schott. Frvt 2006 and ice 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:831 – 846, 2009. ISSN 0162-8828.

S. Berretti, P. Pala, and A del Bimbo. Face recognition by super-resolved 3d models from consumer depth cameras. *Information Forensics and Security, IEEE Transactions on*, 9(9):1436–1449, Sept 2014.

Pavan K. Turaga, Ashok Veeraraghavan, Anuj Srivastava, and Rama Chellappa. Statistical computations on grassmann and stiefel manifolds for image and video-based recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(11):2273–2286, 2011.

Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. Domain adaptation for object recognition: An unsupervised approach. In *IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011*, pages 999–1006, 2011.

Jie Ni, Qiang Qiu, and Rama Chellappa. Subspace interpolation via dictionary learning for unsupervised domain adaptation. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '13, pages 692–699, 2013. ISBN 978-0-7695-4989-7.

Andrea Kleinsmith and Nadia Bianchi-Berthouze. Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing*, 4(1):15–33, 2013. ISSN 1949-3045.

Marwa Mahmoud, Tadas Baltrusaitis, Peter Robinson, and Laurel D. Riek. 3d corpus of spontaneous complex mental states. In *ACII (1)*, pages 205–214, 2011.

J.K. Aggarwal and M.S. Ryoo. Human activity analysis: A review. *ACM Comput. Surv.*, 43(3):16:1–16:43, April 2011. ISSN 0360-0300.

Mohamed F. Abdelkader, Wael Abd-Almageed, Anuj Srivastava, and Rama Chellappa. Silhouette-based gesture and action recognition via modeling trajectories on riemannian shape manifolds. *Comput. Vis. Image Underst.*, 115(3):439–455, March 2011. ISSN 1077-3142.

Jose M. Chaquet, Enrique J. Carmona, and Antonio Fernández-Caballero. A survey of video datasets for human action and activity recognition. *Comput. Vis. Image Underst.*, 117(6):633–659, June 2013. ISSN 1077-3142.

Shian-Ru Ke, Le Uyen Thuc Hoang, Yong-Jin Lee, Jenq-Neng Hwang, Jang-Hee Yoo, and Kyoung-Ho Choi. A review on video-based human activity recognition. *Computers*, 2(2):88–131, 2013.

Ashok Veeraraghavan, Anuj Srivastava, Amit K. Roy Chowdhury, and Rama Chellappa. Rate-invariant recognition of humans and their activities. *IEEE Transactions on Image Processing*, 18(6):1326–1339, 2009.

Wanqing Li, Zhengyou Zhang, and Zicheng Liu. Action recognition based on a bag of 3d points. In *IEEE Inter. Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB)*, page 9–14, 2010.

Chris Ellis, Syed Zain Masood, Marshall F. Tappen, Joseph J. Laviola, Jr., and Rahul Sukthankar. Exploring the trade-off between accuracy and observational latency in action recognition. *Int. J. Comput. Vision*, 101(3):420–436, February 2013. ISSN 0920-5691.

Omar Oreifej and Zicheng Liu. Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 716–723, 2013.

Raviteja Vemulapalli, Felipe Arrate, and Rama Chellappa. Human action recognition by representing 3d skeletons as points in a lie group. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

M. Devanne, H. Wannous, S. Berretti, P. Pala, M. Daoudi, and A Del Bimbo. 3-d human action recognition by shape analysis of motion trajectories on riemannian manifold. *Cybernetics, IEEE Transactions on*, accepted for publication 2014.

R. Slama, H. Wannous, and M. Daoudi. Accurate action recognition using learning on the grassmann manifold. *Pattern Recognition*, accepted for publication 2014.

Jingyong Su, Sebastian Kurtek, Eric Klassen, and Anuj Srivastava. Statistical analysis of trajectories on riemannian manifolds: Bird migration, hurricane tracking, and video surveillance. *Annals of Applied Statistics*, 2013.

Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

Part III

Selected Publications

4-D Facial Expression Recognition by Learning Geometric Deformations

Boulbaba Ben Amor, *Member, IEEE*, Hassen Drira, *Member, IEEE*, Stefano Berretti, *Member, IEEE*, Mohamed Daoudi, *Senior Member, IEEE*, and Anuj Srivastava, *Senior Member, IEEE*

Abstract—In this paper, we present an automatic approach for facial expression recognition from 3-D video sequences. In the proposed solution, the 3-D faces are represented by collections of radial curves and a Riemannian shape analysis is applied to effectively quantify the deformations induced by the facial expressions in a given subsequence of 3-D frames. This is obtained from the dense scalar field, which denotes the shooting directions of the geodesic paths constructed between pairs of corresponding radial curves of two faces. As the resulting dense scalar fields show a high dimensionality, Linear Discriminant Analysis (LDA) transformation is applied to the dense feature space. Two methods are then used for classification: 1) 3-D motion extraction with temporal Hidden Markov model (HMM) and 2) mean deformation capturing with random forest. While a dynamic HMM on the features is trained in the first approach, the second one computes mean deformations under a window and applies multiclass random forest. Both of the proposed classification schemes on the scalar fields showed comparable results and outperformed earlier studies on facial expression recognition from 3-D video sequences.

Index Terms—4-D data, expression, recognition, Hidden Markov model (HMM), random forest, Riemannian geometry, temporal analysis.

I. INTRODUCTION

OVER THE last few years, automatic recognition of facial expressions emerged as a field of active research, with applications in several different areas, such as human–machine interaction, psychology, computer graphics, transport security (by detecting driver fatigue, for example), and such.

The importance of facial expressions was first realized and investigated by psychologists, among others. In a seminal paper by Mehrabian and Wiener [1], the relative importance of verbal and nonverbal messages in communicating feelings and attitude is described. In particular, they have provided evidence that face-to-face communication is governed by the 7%–38%–55% rule that balances the relevance of verbal, vocal, and visual elements, respectively, in communications. Despite this rigid quantification has since been refuted in later papers, it still provides an indication that the words and tone of the voice form only a part of human communication. The nonverbal elements related to the body language (e.g., gestures, postures, facial expressions) also play an important role. Starting from a different point of view, Ekman [2] conducted the first systematic studies on facial expressions in the late 70s. Through his experiments, it is demonstrated that there is an universal set of facial expressions representing anger, disgust, fear, happiness, sadness, and surprise, plus the neutral one, that are universally recognized and remain consistent across different ethnicity and cultures. The presence of these prototypical facial expressions is now widely accepted for scientific analysis. Ekman and Friesen [3] also showed that facial expressions can be coded through the movement of face points as described by a set of action units.

These results, in turn, inspired many researchers to analyze facial expressions in video data, by tracking facial features and measuring the amount of facial movements in video frames [4]. This section of paper demonstrates a collective knowledge that facial expressions are highly dynamical processes, and looking at sequences of face instances can help to improve the recognition performance. We further motivate that, rather than being just a static or dynamic 2-D image analysis, it is more natural to analyze expressions as spatio-temporal deformations of 3-D faces, caused by the actions of facial muscles. In this approach, the facial expressions can be studied comprehensively by analyzing temporal dynamics of 3-D face scans (3-D plus time is often regarded as 4-D data). From this perspective, the relative immunity of 3-D scans to lighting conditions and pose variations give support to the use of 3-D and 4-D data. Motivated by these considerations, there has been a progressive shift from 2-D to 3-D in performing facial shape analysis for recognition [5]–[9], and expression recognition [10], [11]. In particular, this latter research subject is gaining momentum thanks to the recent availability of public 3-D datasets, like the Binghamton University 3-D facial expression database (BU-3-DFE) [12], and the Bosphorus 3-D face database [13]. At the same time, advances in 3-D imaging technology have permitted collections of large datasets that include temporal sequences of 3-D scans (i.e., 4-D datasets), such as the Binghamton University 4-D facial expression database (BU-4DFE) [14], the 4-D dataset constructed at University of Central Lancashire (Hi4D-ADSIP) [15], [16], and the dynamic 3-D FACS dataset (D3DFACS) for facial expression research [17], which also includes fully coded FACS. This trend has been strengthened further by the introduction of inexpensive acquisition devices, such as the consumer 3-D cameras like Kinect or Asus that provide fast albeit low-resolution streams of 3-D data to a large number of users, thus opening new opportunities and challenges in 3-D face recognition and facial expression recognition [18], [19].

Motivated by these facts, we focus in this paper on the problem of expression recognition from dynamic sequences of 3-D facial scans. We propose a new framework for temporal analysis of 3-D faces that combines scalar field modeling of face deformations with effective classifiers. To motivate our solution and to relate it to the state-of-the-art, we provide an overview of existing methods for 4-D facial expression recognition (see the recent paper in [20] for a comprehensive survey on this subject), then we give a general overview of our approach.

A. Related Work

The use of 4-D data for face analysis is still at the beginning, with just a few papers performing face recognition from sequences of 3-D face scans [19], [21], [22], and some papers focussing on facial expression recognition.

In particular, the first approach addressing the problem of facial expression recognition from dynamic sequences of 3-D scans was proposed by Sun *et al.* [23], [24]. Their approach basically relies on the use of a generic deformable 3-D model whose changes are tracked both in space and time in order to extract a spatio-temporal description of the face. In the temporal analysis, a vertex flow tracking technique is applied to adapt the 3-D deformable model to each frame of a 3-D face sequence. Correspondences between vertices across the 3-D dynamic facial sequences provide a set of motion trajectories (vertex flow) of 3-D face scans. As a result, each depth scan in the sequence can be represented by a spatio-temporal feature vector that describes both shape and motion information and provides a robust facial surface representation. Once spatio-temporal features are extracted, a 2-D Hidden Markov Model (HMM) is used for classification. In particular, a spatial HMM and a temporal HMM were used to model the spatial and temporal relationships between the extracted features. Exhaustive analysis was performed on the BU-4DFE database. The main limit of this solution resides in the use of 83 manually annotated landmarks of the BU-4DFE that are not released for public use.

The approach proposed by Sandbach *et al.* [25] exploits the dynamics of 3-D facial movements to analyze expressions. This is obtained by first capturing motion between frames using free-form deformations (FFD) and extracting motion features using a quad-tree decomposition of several motion fields. GentleBoost classifiers are used in order to simultaneously select the best features to use and perform the training using two classifiers for each expression: one

for the onset temporal segment, and the other for the offset segment. Then, HMMs are used for temporal modeling of the full expression sequence that is represented as the composition of four temporal segments, such as neutral, onset, apex, and offset. These model a sequence with an initial neutral segment, followed by the activation of the expression, the maximum intensity of the expression, deactivation of the expression and closing of the sequence again with a neutral expression. Experiments were reported for three prototypical expressions (i.e., happy, angry, and surprise) of the BU-4DFE database. An extension of this paper has been presented in [20], where results on the BU-4DFE database using the six universal facial expressions are reported.

In [26], a level curve based approach is proposed by Le *et al.* to capture the shape of 3-D facial models. The level curves are parameterized using the arclength function. The Chamfer distance is applied to measure the distances between the corresponding normalized segments, partitioned from these level curves of two 3-D facial shapes. These features are then used as spatio-temporal features to train HMM, and since the training data were not sufficient for learning HMM, the authors proposed to apply the universal background modeling to overcome the overfitting problem. Results were reported for the happy, sad, and surprise sequences of the BU-4DFE database.

Fang *et al.* [27] propose a fully automatic 4-D facial expression recognition approach with a particular emphasis on 4-D data registration and dense correspondence between 3-D meshes along the temporal line. The variant of the local binary patterns (LBP) descriptor proposed in [28], which computes LBP on three orthogonal planes is used as face descriptor along the sequence. Results are provided on the BU-4DFE database for all expressions and for the subsets of expressions used in [25] and [26], showing improved results with respect to the competing solutions. Fang *et al.* [29] propose a similar methodology for facial expression recognition from dynamic sequences of 3-D scans, with an extended analysis and comparison of different 4-D registration algorithms, including ICP and more sophisticated mesh matching algorithms, like spin images and MeshHOG. However, 12 manually annotated landmarks were used in this paper.

Recently, Reale *et al.* [30] have proposed a new 4-D spatio-temporal feature named Nebula for facial expressions and movement analysis from a volume of 3-D data. After fitting the volume data to a cubic polynomial, a histogram is built for different facial regions using geometric features, such as curvatures and polar angles. They have conducted several recognition experiments on the BU-4DFE database for posing expressions, and on a new database published in [31] for spontaneous expressions. However, and just 15 frames from the onset one are used for classification, and these frames correspond to the most intense expression.

From the discussion above, it becomes clear that the solutions specifically tailored for 4-D facial expression recognition from dynamic sequences are still preliminary, being semiautomatic, or are capable of discriminating between only a subset of expressions.
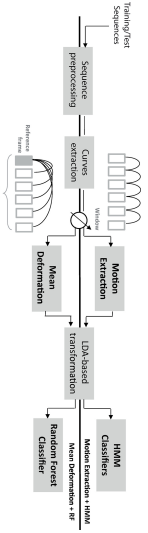
Fig. 1. Overview of the proposed approach. Four main steps are shown: sequence preprocessing and extraction of the radial curves; motion extraction and Mean deformation computation; dimensionality reduction with LDA; HMM and random forest-based classification. Note that both train and test sequences can go through the upper and lower path in the block diagram.

B. Our Method and Contributions

Due to the increasing importance of shape analysis of objects in different applications, including 3-D faces, a variety of mathematical representations and techniques have been suggested, as described above [20], [24], [29]. The difficulty in analyzing shapes of objects comes from the following facts.

1) Shape representations, metrics, and models should be invariant to certain transformations that are termed shape-preserving. For instance, rigid motions and reparameterizations of facial surfaces do not change their shapes, and any shape analysis of faces should be invariant to these transformations.

2) Registration of points across objects is an important ingredient in shape analysis. Specifically, in comparing the shapes of faces, it makes sense that similar biological parts are registered to each other across different faces.

Furthermore, it is important to use techniques that allow a joint registration and comparisons of surfaces in a comprehensive framework, rather than in two separate steps. These two issues, invariance and registration, are naturally handled using Riemannian methods where one can choose metrics that are invariant to certain transformations and form quotient spaces (termed shape spaces) by forming equivalence classes of objects that have the same shape. The elastic Riemannian metric used in this paper provides a nice physical interpretation of measuring deformations between facial curves using a combination of stretching and bending. These elastic deformations are captured by the dense scalar field (DSF) features used in this paper for classifications. In summary, the main motivation of using a Riemannian approach is to perform registration, and obtain deformation fields that are physically interpretable.

Based on these premises, in this paper we extend the ideas presented in [32] to propose an automatic approach for facial expression recognition that exploits the facial deformations extracted from 3-D facial videos. An overview of the proposed approach is given in Fig. 1. In the preprocessing step, the 3-D mesh in each frame is first aligned to the previous one and then cropped. From the obtained subsequence, the 3-D deformation is captured based on a DSF that represents the 3-D deformation between two frames. Linear Discriminant Analysis (LDA) is used to transform derived feature space to an optimal compact space to better separate different expressions. Finally, the expression classification is performed

in the following two ways: (1) Using the HMM models for temporal evolution; and (2) using mean deformation along a window with random forest classifier. Experimental results show that the proposed approaches are capable of improving the state-of-the-art performance on the BU-4DFE database. There are three main contributions in this paper which are as follows.

1) Novel DSFs defined on radial curves of 3-D faces using Riemannian analysis in the shape spaces of curves. These scalar fields accurately capture deformations occurring between 3-D faces represented as collections of radial curves.

2) A new approach for facial expression recognition from 3-D dynamic sequences that combines the high descriptiveness of DSFs extracted from successive 3-D scans of a sequence with the discriminant power of LDA features using HMM and multiclass random forest.

3) An extensive experimental evaluation that compares the proposed solution with the state of the art methods using a common dataset and testing protocols. The results show that our approach outperforms the published state-of-the-art results.

The rest of the paper is organized as follows. In Section II, we present a face representation model that captures facial features relevant to categorizing expression variations in 3-D dynamic sequences. In Section III, the dynamic shape deformation analysis using LDA and classification using HMM and multiclass random forest are addressed. The main characteristics of the BU-4DFE and the preprocessing operations performed on the face scans are described in Section IV. A joint experimental evaluation and the comparative evaluation performed on the BU-4DFE database reported and discussed in the same Section. Finally, conclusions and future research directions are outlined in Section V.

II. GEOMETRIC FACIAL DEFORMATION

One basic idea to capture facial deformations across 3-D video sequences is to track mesh vertices densely along successive 3-D frames. Since, as the resolution of the meshes varies across 3-D video frames, establishing a dense matching on consecutive frames is necessary. For this purpose, Sun et al. [23] proposed to adapt a generic model (a tracking model) to each 3-D frame using a set of 83 predefined facial landmarks to control the adaptation based on

radial basis functions. A second solution is presented by Sandbach et al. [25], [33], where the authors used an existing nonrigid registration algorithm (FFD) [34] based on B-splines interpolation between a lattice of control points. In this case, dense matching is a preprocessing step used to estimate a motion vector field between 3-D frames, t and $t-1$. The problem of quantifying subtle deformations along the sequence still remains a challenging task, and the results presented in [25] are limited to just three facial expressions: happy, angry, and surprise.

In order to capture and model deformations of the face induced by different facial expressions, we propose to represent the facial surface through a set of parameterized radial curves that originate from the tip of the nose. Approximating the facial surface by an ordered set of radial curves that locally captures its shape can be seen as a parameterization of the facial surface. Indeed, similar parameterizations of the face have shown their effectiveness in facial biometrics [35]. The mathematical setup for the shape theory offered here comes from Hilbert space analysis. A facial surface is represented by a collection of radial curves and a Riemannian framework is used to study shapes of these curves. We start by representing facial curves as absolute continuous maps from $\beta : [0, 1] \to \mathbb{R}^3$ and our goal is to analyze shapes represented by these maps. The problem in studying shapes using these maps directly is that they change with reparameterizations of curves. If γ is a reparameterization function (typically a diffeomorphism from [0, 1] to itself), then under the standard L^2 norm, the quantity $\|\beta_1 - \beta_2\| \neq \|\beta_1 \circ \gamma - \beta_2 \circ \gamma\|$, which is problematic. The solution comes from choosing a Riemannian metric under which this inequality becomes equality and the ensuring analysis simplifies. As described in [36], we represent the facial curves using a new function q, called the square-root velocity function (SRVF) [see (1)]. The advantage of using SRVF representation is that under this representation the elastic metric becomes the standard L^2 metric and an identical reparameterization of curves preserves the L^2 norm of between their SRVFs. The mapping from a curve β to q is a bijection (up to a translation), and the space of all SRVFs is the Hilbert space of all square-integrable maps of the type $q : [0, 1] \to \mathbb{R}^3$. This space under the natural L^2 inner product is actually a vector space and geodesics between points in this space are straight lines.

With the proposed representation, a facial surface is approximated by an indexed collection of radial curves β_α, where the index α denotes the angle formed by the curve with respect to a reference radial curve. In particular, the reference radial curve (i.e., the curve with $\alpha = 0$) is chosen as oriented along the vertical axis, while the other radial curves are separated from each other by a fixed angle and are ordered in a clockwise manner. As an example, Fig. 2(a) shows the radial curves extracted for a sample face with happy expression. To extract the radial curves, the nose tip is accurately detected and each face scan is rotated to the upright position so as to establish a direct correspondence between radial curves having the same index in different face scans (the preprocessing steps, including nose tip detection and pose normalization are discussed in more detail in Section IV-A). In Fig. 2(b) and
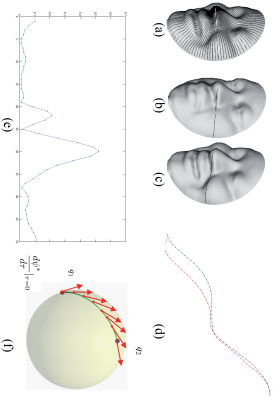

Fig. 2. Figure illustration. (a) Extracted radial curves. (b) and (c) Radial curve on a neutral face and the correspondent curve on the same face with happy expression, respectively. (d) Two radial curves plotted together. (e) Values of the magnitude of $\frac{dq}{d\tau}\big|_{\tau=0}(k)$ computed between the curves in (d) are reported for each point k of the curve. (f) Parallel vector field across the geodesic between q_1 and q_2 in the space of curves C.

(c), two radial curves at $\alpha = 90°$ in the neutral and happy scans of the same subject are shown. As emerged Fig. 2(d), facial expressions can induce consistent variations in the shape of corresponding curves. These variations change in strength from expression to expression and for different parts of the face. In order to capture these variations effectively, a DSF is mathematically represented through the square-root velocity function (SRVF) [36], [37], denoted by $q(t)$, according to

$$q(t) = \frac{\dot\beta(t)}{\sqrt{\|\dot\beta(t)\|}} \quad t \in [0, 1]. \quad (1)$$

This specific representation has the advantage of capturing the shape of the curve and makes the calculus simpler. Let us define the space of the SRVFs as $C = \{q : I \to \mathbb{R}^3, \|q\| = 1\} \subset L^2(I, \mathbb{R}^3)$, with $\|\cdot\|$ indicating the L^2 norm. With the L^2 metric on its tangent space, C becomes a Riemannian manifold. Basically, with this parameterization each radial curve is represented on the manifold C by its SRVF. According to this, given the SRVFs q_1 and q_2 of two radial curves, the shortest path ψ^* on the manifold C between q_1 and q_2 (called geodesic path) is a critical point of the following energy function:

Considering a generic radial curve β of the face, it can be parameterized as $\beta : I \to \mathbb{R}^3$, with $I = [0, 1]$. With the L^2 metric on its tangent space, C becomes a Riemannian manifold.

$$E(\psi) = \frac{1}{2}\int \|\dot\psi(\tau)\|^2 d\tau \quad (2)$$

where ψ denotes a path on the manifold C between q_1 and q_2. τ is the parameter for traveling along the path ψ, $\dot\psi \in T_\psi(C)$ is the tangent vector field on the curve ψ on C, and $\|\cdot\|$ denotes the L^2 norm on the tangent space.

Since the elements of C have a unit L^2 norm, C is a hypersphere in the Hilbert space $L^2(I, \mathbb{R}^3)$. As a consequence, the geodesic path between any two points $q_1, q_2 \in C$ is simply

given by the minor arc of the great circle connecting them on this hypersphere, $\psi^*: [0, 1] \to C$. This is given by

$$\psi^*(\tau) = \frac{1}{\sin(\theta)} (\sin((1-\tau)\theta)q_1 + \sin(\theta\tau)q_2) \quad (3)$$

where $\theta = d_c(q_1, q_2)$. We point out that $\sin(\theta) = 0$ if the distance between the two curves is zero, in other words $q_1 = q_2$. In this case, for each τ, $\psi^*(\tau) = q_1 = q_2$. The tangent vector field on this geodesic is then written as $\frac{d\psi^*}{d\tau}: [0, 1] \to T_\psi(C)$, and is obtained by the following equation:

$$\frac{d\psi^*}{d\tau} = \frac{-\theta}{\sin(\theta)}(\cos((1-\tau)\theta)q_1 - \cos(\theta\tau)q_2). \quad (4)$$

Knowing that on geodesic path, the covariant derivative of its tangent vector field is equal to 0, $\frac{d\psi^*}{d\tau}$ is parallel along the geodesic ψ^* and one can represent it with $\frac{d\psi^*}{d\tau}|_{\tau=0}$ without any loss of information. Accordingly, (4) becomes

$$\frac{d\psi^*}{d\tau}\Big|_{\tau=0} = \frac{\theta}{\sin(\theta)}(q_2 - \cos(\theta)q_1) \ (\theta \neq 0). \quad (5)$$

A graphical interpretation of this mathematical representation is given in Fig. 2. In Fig. 2(a), we show a sample face with happy expression and all the extracted radial curves. In Fig. 2(b) and (c), two corresponding radial curves (i.e., radial curves at the same angle α) on a neutral and a happy face of the same person are highlighted, respectively. These curves are reported together in Fig. 2(d), where the amount of deformation between them can be appreciated, although the two curves lie at the same angle α and belong to the same person. The amount of deformation between the two curves is calculated using (5) and the plot of the magnitude of this vector at each point of the curve is reported in Fig. 2(e) (i.e., 50 points are used to sample each of the two radial curves as reported on the x-axis, while the magnitude of the vector field is reported on the y-axis). Finally, Fig. 2(f) illustrates the idea to map the two radial curves on the hypersphere C in the Hilbert space through their SRVFs q_1 and q_2, and shows the geodesic path connecting these two points on the hypersphere. The tangent vectors of this geodesic path represent a vector field whose covariant derivative is zero. According to this, $\frac{d\psi^*}{d\tau}|_{\tau=0}$ becomes sufficient to represent this vector field with the remaining vectors obtained by parallel transport of $\frac{d\psi^*}{d\tau}|_{\tau=0}$ along the geodesic ψ^*.

Based on the above representation, we define a DSF capable to capture deformations between two corresponding radial curves β_α^1 and β_α^2 of two faces approximated by a collection of radial curves.

Definition 1: DSF: Let $x_\alpha(t) = \|\frac{d\psi^*}{d\tau}|_{\tau=0}(t)\|$ be the values of the magnitude computed for each point t of the curves q_α^1 and q_α^2; let T be the number of sampled points per curve, and $|\Lambda|$ be the number of curves used per face. According to this, we define the function f by

$$f(q_\alpha^1, q_\alpha^2) = (x_\alpha^1, \ldots, x_\alpha^k, \ldots, x_\alpha^T)^\tau,$$

Algorithm 1 Computation of the DSF

Input: Facial surfaces F^1 and F^2; T, number of sample points on a curve; $\Delta\alpha$, angle between successive radial curves; $|\Lambda|$, number of curves per face

Output: $DSF(F^1, F^2)$, the DSF between the two faces

procedure COMPUTEDSF($F^1, F^2, T, \Delta\alpha, |\Lambda|$)
  $n \leftarrow 0$
  while $n < |\Lambda|$ **do**
    $\alpha = n \cdot \Delta\alpha$
    for $i \leftarrow 1, 2$ **do**
      extract the curve β_α^i
      compute the SRVF for β_α^i:
    end for
    compute the distance between q_α^1 and q_α^2:
    $\theta = d_c(q_\alpha^1, q_\alpha^2) = cos^{-1}(\langle q_\alpha^1, q_\alpha^2 \rangle)$
    compute the deformation vector $\frac{d\psi^*}{d\tau}|_{\tau=0}$ using Eq. (5) as:
    $f(q_\alpha^1, q_\alpha^2) = (x_\alpha(1), x_\alpha(2), \ldots, x_\alpha(T)) \in \mathbb{R}_+^T$
    $x_\alpha(t) = \frac{1}{\sin(\theta)}(q_\alpha^2 - \cos(\theta)q_\alpha^1), t = 1, \ldots, T$
  end while
  compute $DSF(F^1, F^2)$ as the magnitude of $\frac{d\psi^*}{d\tau}|_{\tau=0}(k)$:
  $DSF(F^1, F^2) = (f(q_{\alpha 1}^1, q_{\alpha 1}^2), \ldots, f(q_{|\Lambda|}^1, q_{|\Lambda|}^2))$
  return DSF
end procedure

Assuming that $\{\beta_\alpha^1 | \alpha \in \Lambda\}$ and $\{\beta_\alpha^2 | \alpha \in \Lambda\}$ be the collections of radial curves associated with the two faces F^1 and F^2 and let q_α^1 and q_α^2 be their SRVFs, the DSF vector is defined by

$$DSF(F^1, F^2) = (f(q_0, q_0^2), \ldots, f(q_\alpha, q_\alpha^2), \ldots, f(q_{|\Lambda|}, q_{|\Lambda|}^2)).$$

The dimension of the DSF vector is $|\Lambda| \times T$. The steps to compute the proposed DSF are summarized in Algorithm 1.

The first step to capture the deformation between two given 3-D faces F^1 and F^2 is to extract the radial curves originating from the nose tip. Let β_α^1 and β_α^2 denote the radial curves on faces F^1 and F^2, respectively. The initial tangent vector to ψ^*, called also the shooting direction, is computed using (5). Then, we consider the magnitude of this vector at each point t of the curve in order to construct the DSFs of the facial surface. In this way, the DSF quantifies the local deformation between points of radial curves β_α^1 and β_α^2, respectively, of the faces F^1 and F^2. In practice, we represent each face with 100 radial curves, and T=50 sampled points on each curve, so that the DSFs between two 3-D faces is expressed by a 5000-D vector.

In Fig. 3, the examples of the deformation fields computed between a neutral face of a given subject and the apex frames of the sequences of the six prototypical expressions of the same subject are shown. The values of the scalar field to be applied on the neutral face to convey the six different prototypical expressions are reported using a color scale. In particular, colors from green to red represent the highest values of the dense scalar
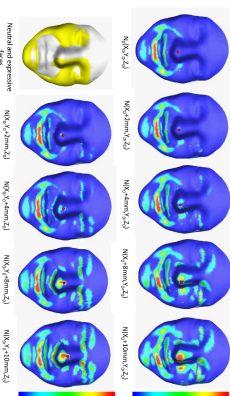
Fig. 4. Effect of the nose tip placement inaccuracy on the shooting directions of the geodesic paths (or the DSFs computation). The first row illustrates DSFs when varying the nose tip position along the X-direction. The second row shows DSFs when the variation is performed along the Y-direction.

field are represented in cyan/blue. As it can be observed, for different expressions, the high deformations are located in different regions of the face. For example, as intuitively expected, the corners of the mouth and the cheeks are mainly deformed for happiness expression, whereas the eyebrows are also strongly deformed for the angry and disgust expressions.

A. Effect of the Nose Tip Localisation Inaccuracy on the DSF Computation

In the following, we present a study on the effects that possible inaccuracies in the detection of the nose tip can have on the computation of the proposed DSF. In particular, we consider the effects on the radial curves originating from the nose tip and the radial curves originating from the nose tip.

We have changed the nose tip coordinates in the X and Y directions and have reported the DSFs computation results (using colormaps on the expressive faces) in Fig. 4. As illustrated in this figure, a large localization error (> 4mm) of the nose tip generates false deformations, which could impact negatively the performance of the approach. In fact, our method is based on learning such geometric deformations to build HMMs or random forest classifiers. However, the left side of the figure illustrates the fact that the DSFs computation tolerates errors quite well, up to 4mm.
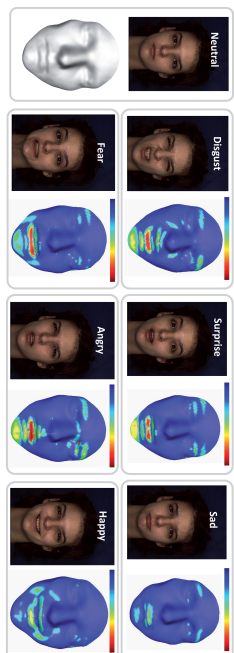

Fig. 3. Deformation Scalar Fields computed between a neutral face of a given subject and the apex frames of the sequences of the six prototypical expressions of the same subject. The neutral scan is shown on the left. Corresponding texture images are also illustrated with each DSFs colormap.

B. DSF Compared to Other Features

In order to compare the proposed DSF feature against other methods for extracting dense deformation features, we selected the FFD approach, which has been originally defined in [38] for medical images, and later on successfully applied to the problem of 3-D dynamic facial expression recognition by Sandbach et al. [25], [33]. In particular, FFD is a method for nonrigid registration based on B-spline interpolation between a lattice of control points. In addition, we also compared our approach with respect to a baseline solution that uses the point-to-point Euclidean distance between frames of a sequence. Fig. 5 reports the results for an example case, where a frame of a happy sequence is deformed with respect to the first frame of the sequence. The figure shows quite clearly as the DSF proposed in this paper is capable to capture the face deformations with smooth variations that include the mouth, the chin and the cheek, in the example. This result is important to discriminate between different expressions whose effects are not limited to the mouth region. Differently, the variations captured by the other two solutions are much more concentrated in the mouth region of the face.

III. EXPRESSION RECOGNITION USING DSFs

Deformations due to facial expressions across 3-D video sequences are characterized by subtle variations induced mainly by the motion of facial points. These subtle changes are important to perform effective expression recognition, but they are also difficult to be analyzed due to the face movements. To handle this problem, as described in Section II, we propose a curve-based parametrization of the face that consists in representing the facial surface by a set of radial curves. According to this representation, the problem of comparing two facial surfaces, a reference facial surface and a target one, is reduced to the computation of the DSF between them.

In order to make possible to enter the expression recognition system at any time and make the recognition process possible from any frame of a given video, we consider subsequences of n frames. Thus, we chose the first n frames as the first subsequence. Then, we chose n-consecutive frames starting from the second frame as the second subsequence. This process is repeated by shifting the starting index of the sequence

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

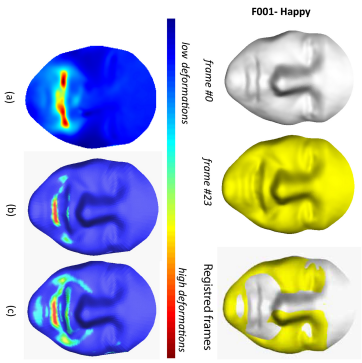BEN AMOR et al.: 4-D FACIAL EXPRESSION RECOGNITION

7

Fig. 5. Comparison of the different features extracted between two frames taken from subject 001 for the happy expression. (a) Free form-based deformations (FFD). (b) Point-to-point Euclidean distances (c) DSFs deformations.

every one frame till the end of the sequence. In order to classify the resulting subsequences, we propose two different feature extraction and classification framework based on the DSF.

1) *Mean Deformation-based Features Associated to Random Forest Classifier:* The first frame of the subsequence is considered as a reference frame and the deformation is calculated from each of the remaining frames to the first one using the DSF. The average deformation of the n-1 resulting DSFs represents the feature vector in this classification scheme and is presented, after dimensionality reduction, to multiclass random forest classifiers.

2) *3-D Motion Features Combined With HMM Classifiers:* The deformation between successive frames in a subsequence are calculated using the DSFs and presented to an HMM classifier preceded by LDA-based dimensionality reduction.

A. Mean Shape Deformation with Random Forest Classifier

The idea here is to capture a mean deformation of the face in the sliding window on the 3-D expression sequence. In order to get this feature, the first frame of each subsequence is considered as the reference one, and the dense deformation is computed from this frame to each of the remaining frames of the subsequence. Let F_{ref} denote the reference frame of a subsequence and F_i the ith successive frame in the subsequence; the successive frame, for example, is denoted by computed from the annotated data, which reached each leaf F_i. The DSF is calculated between F_{ref} and F_i, for different values of i ($i = 1, \ldots, n - 1$), and the mean deformation is then given by

$$DSF = \frac{1}{n-1} \sum_{i=1}^{n-1} DSF(F_{ref}, F_i). \qquad (6)$$

Fig. 6 illustrates one subsequence for each expression with $n = 6$ frames. Each expression is illustrated in two rows: The upper row gives the reference frame of the subsequence and the n-1 successive frames of the subsequences. Below, the corresponding DSFs computed for each frame are shown. The mean deformation field is reported on the right of each plot and represents the feature vector for each subsequence. The feature vector for this subsequence is built based on the mean deformation of the n-1 calculated deformations. Thus, each subsequence is represented by a feature vector of size equal to the number of points on the face (i.e., the number of points used to sample the radial curves of the face). In order to provide a visual representation of the scalar fields, an automatic labeling scheme is applied; warm colors (red, yellow) are associated with high $DSF(F_{ref}, F_i)$ values and correspond to facial regions affected by high deformations, and cold colors are associated with regions of the face that remain stable from one frame to another. Thus, this dense deformation field summarizes the temporal changes of the facial surface when a particular facial expression is conveyed.

According to this representation, the deformation of each subsequence is captured by the mean DSF defined in (6). The main motivation for using the mean deformation, instead of the maximum deformation for instance, is related to its greater robustness to the noise. In the practice, the mean deformation resulted more resistant to deformations due to, for example, inaccurate nose tip detection or the presence of acquisition noise. In Fig. 6, for each subsequence, the mean deformation field illustrates a smoothed pattern better than individual deformation fields in the same subsequence. Since the dimensionality of the feature vector is high, we use LDA-based transformation to map the present feature-space to an optimal one that is relatively insensitive to different subjects, while preserving the discriminating expression information. LDA defines the within-class matrix S_w and the between-class matrix S_b. It transforms a n-D feature to an optimized d-dimensional feature, where $d < n$. In our experiments, the discriminating classes are the six expressions, thus the reduced dimension d is five.

For the classification, we used the multiclass random forest algorithm. The algorithm was proposed by Leo Breiman [39] and defined as a meta-learner comprised of many individual trees. It was designed to operate quickly over large datasets and more importantly to be diverse by using random samples to build each tree in the forest. A tree achieves highly nonlinear mappings by splitting the original problem into smaller ones, solvable with simple predictors. Each node in the tree consists of a test, whose result directs a data sample toward the left or the right child. During training, the tests are chosen in order to group the training data in clusters where simple models achieve good predictions. Such models are stored at the leaves, computed from the annotated data, which reached each leaf at train time. Once trained, a random forest is capable to classify a new expression from an input feature vector by putting it down each of the trees in the forest. Each tree gives a classification decision by voting for that class. Then, the forest chooses the classification having the most votes (over all the trees in the forest).

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8

IEEE TRANSACTIONS ON CYBERNETICS

B. 3-D Motion Extraction with HMM Classifier

The DSF features described in Section II can also be applied for expression recognition according to a different classification scheme. The deformations between successive frames in the subsequence are calculated using the DSF. In particular, the deformation between two successive 3-D frames is obtained by computing the pairwise DSF $DSF(F_{i-1}, F_i)$ of correspondent radial curves. Based on this measure, we are able to quantify the motion of face points along radial curves and thus capture the changes in facial surface geometry.

Fig. 7 illustrates a direct application of the $DSF(F_{i-1}, F_i)$ and its effectiveness in capturing deformation between one
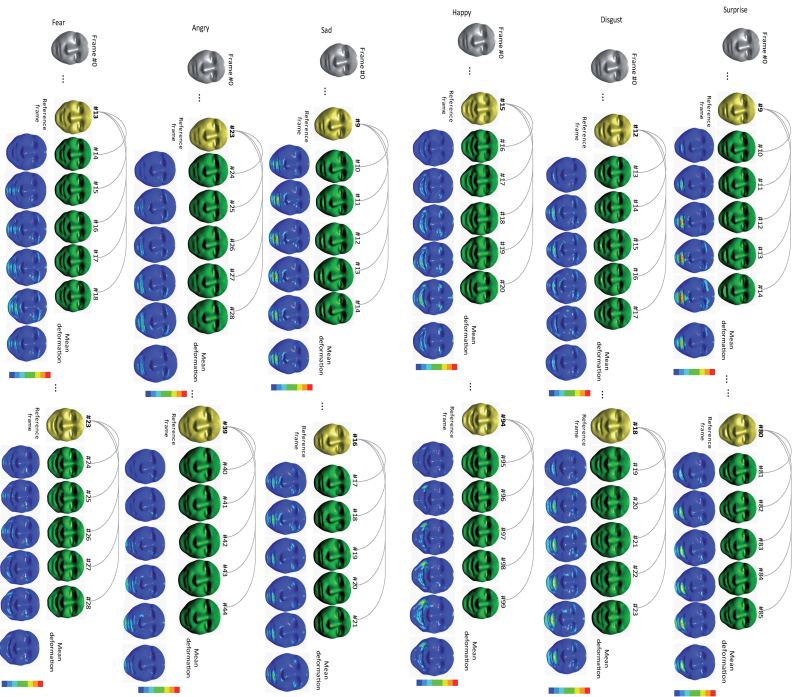

Fig. 6. Computation of dynamic shape deformation on different subsequences taken from the BU-4DFE database. Each expression is illustrated by two rows. The upper one gives the reference frame of the subsequence and the n-1 successive frames. The mean deformation field is given on the right of each lower row.

facial surface to another belonging to two consecutive frames in a 3-D video sequence. This figure shows two subsequences extracted from videos in the BU-4DFE database (happy and surprise expressions are shown on the left and on the right, respectively). For each sequence, the 2-D image and the 3-D scans of some frames are shown in the upper row. In the lower row, the deformation scalar field $DSF(F_{i-1}, F_i)$ computed between consecutive frames (i.e., the current frame and the previous one) in the subsequence is reported. In order to provide a visual representation of the scalar field, an automatic labeling scheme is applied that includes only two colors: The red color is associated with high $DSF(F_{i-1}, F_i)$ values and corresponds to facial regions affected by high deformations,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

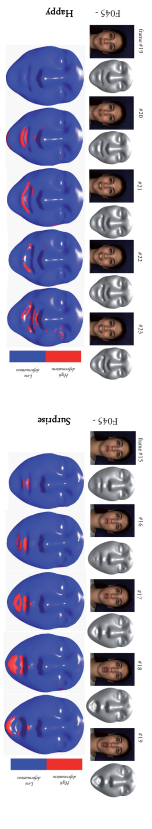BEN AMOR et al.: 4-D FACIAL EXPRESSION RECOGNITION

9

Fig. 7. Examples of DSF deformations between subsequent frames of 3-D video sequences: Happy and surprise expressions are shown, respectively, on the left and right.
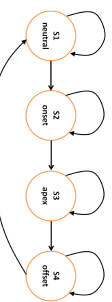
Fig. 8. Structure of the HMMs modeling a 3-D facial sequence. The four states model neutral, onset, apex, and offset frames of the sequence. As shown, from each state it is possible to remain in the state itself or move to the next one (this is known as Bakis or left-right HMM).

The blue color is associated with the regions that remain more stable from one frame to another. As illustrated in Fig. 7, for different expressions, different regions are mainly deformed, showing the capability of the deformation fields to capture relevant changes of the face due to the facial expression. In particular, each deformation is expected to identify an expression, for example, as suggested by the intuition, the corners of the mouth and the cheeks are mainly deformed for the happiness expression.

With the proposed approach, the feature extraction process starts by computing for each 3-D frame in a given video sequence the DSF with respect to the previous one. In this way, we obtain as many fields as the number of frames in the sequence (decreased by one), where each field contains as many scalar values as the number of points composing the collection of radial curves representing the facial surface. In practice, the size of $DSF(F_{r-1}, F_r)$ is 1×5000, considering 5000 points on the face, similarly to the feature vector used in the first scheme of classification (mean deformation-based). Since the dimensionality of the resulting feature vector is high, also in this case we use LDA to project the scalar values to a 5-D feature space, which is sensitive to the deformations induced by different expressions. The 5-D feature vector extracted for the 3-D frame at instant t of a sequence is indicated as f^t in the following. Once extracted, the feature vectors are used to train HMMs and to learn deformations due to expressions along a temporal sequence of frames.

In our case, sequences of 3-D frames constitute the temporal dynamics to be classified, and each prototypical expression is modeled by an HMM (a total of six HMMs χ^{expr} is required, with $expr \in \{an, di, fe, ha, sa, su\}$). Four states per HMM are used to represent the temporal behavior of each expression. This corresponds to the idea that each sequence starts and ends with a neutral expression (state S_1). The frames that belong to the temporal intervals where the face changes from neutral to expressive and back from expressive to neutral are modeled by the onset (S_2) and offset (S_4) states, respectively. Finally, the frames corresponding to the highest intensity of the expression are captured by the apex state (S_3). This solution has proved its effectiveness in clustering the expressive states of a sequence also in other papers [33]. Fig. 8 exemplifies the structure of the HMMs used in our framework.

The training procedure of each HMM is summarized as follows.

1) Feature vectors f^t of the training sequences are first clustered to identify a codebook of symbols using the standard LBG algorithm [40]. This provides the required mapping between multidimensional feature vectors, taking values in a continuous domain, with the alphabet of symbols emitted by the HMM states.

2) Expression sequences are considered as observation sequences $O = \{O^1, O^2, \ldots, O^T\}$, where each observation O^t at time t is given by the feature vector f^t.

3) Each HMM χ^{expr} is initialized with random values and the Baum–Welch algorithm [41] is used to train the model using a set of training sequences. This estimates the model parameters, while maximizing the conditional probability $P(O|\chi^{expr})$.

Given a 3-D sequence to be classified, it is processed as in Section II, so that each feature vector f^t corresponds to a test observation $O^t = \{O^1 = f^1, \ldots, O^T = f^T\}$. Then, the test observation O is presented to the six HMMs χ^{expr} that model different expressions, and the Viterbi algorithm is used to determine the best path $Q = \{\hat{q}^1, \ldots, \hat{q}^T\}$, which corresponds to the observation sequence O. The likelihood along the best path, $p^{expr}(O, \hat{Q}|\chi^{expr}) = \hat{p}^{expr}(O|\chi^{expr})$ is considered as a good approximation of the true likelihood given by the more expensive forward procedure [41], where all the possible paths are considered instead of the best one. Finally, the sequence is classified as belonging to the class corresponding to the HMM whose log-likelihood along the best path is the greatest one.

IV. EXPERIMENTAL RESULTS

The proposed framework for facial expression recognition from dynamic sequences of 3-D face scans has been experimented on the BU-4DFE database. Main characteristics of the database and results are reported in the following sections.
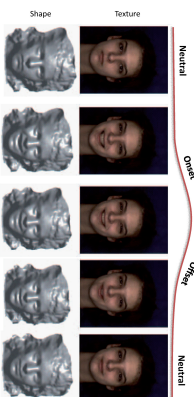
1) Feature vectors f^t of the training sequences are first clustered to identify a codebook of symbols using the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10

IEEE TRANSACTIONS ON CYBERNETICS

Fig. 9. Examples of 2-D and 3-D frames extracted from a dynamic 3-D video sequence of the BU-4DFE dataset.

A. BU-4DFE Database: Description and Preprocessing

To investigate the usability and performance of 3-D dynamic facial sequences for facial expression recognition, a dynamic 3-D facial expression database has been created at Binghamton University [14]. The dimensional imaging's 3-D dynamic capturing system [42], has been used to capture a sequence of stereo images and produce the depth map of the face according to a passive stereo-photogrammetry approach. The range maps are then combined to produce a temporally varying sequence of high-resolution 3-D images with an RMS accuracy of 0.2 mm. At the same time, 2-D texture videos of the dynamic 3-D models are also recorded. Each participant (subject) was requested to perform the six prototypical expressions (i.e., angry, disgust, fear, happiness, sadness, and surprise) separately. Each expression sequence contains neutral expressions in the beginning and in the end, so that each sequence was performed gradually from neutral appearance, low intensity, high intensity, and back to low intensity and neutral. Each 3-D sequence captures one expression at a rate of 25 frames/s and each 3-D sequence lasts approximately four seconds with about 35000 vertices per scan (i.e., 3-D frame). The database consists of 101 subjects (58 female and 43 male, with an age range of 18–45 years old) including 606 3-D model sequences with six prototypical expressions and a variety of ethnic/racial ancestries (i.e., 28 Asian, 8 African-American, 3 Hispanic/Latino, and 62 Caucasian). More details on the BU-4DFE can be found in [14]. An example of a 3-D dynamic facial sequence of a subject with happy expression is shown in Fig. 9, where 2-D frames (not used in our solution) and 3-D frames are reported. From left to right, the frames illustrate the intensity of facial expression passing from neutral to onset, offset, apex, and neutral again.

It can be observed that the 3-D frames present a near-frontal pose with some slight changes occurring mainly in the azimuthal plane. The scans are affected by large outliers, mainly acquired in the hair, neck, and shoulders regions (Fig. 9). In order to remove these imperfections from each 3-D frame, a preprocessing pipeline is performed. The main steps of this pipeline are summarized as follows (Fig. 10).

1) Identify and fill the holes caused, in general, by self-occlusions or open mouth. The holes are identified by locating boundary edges, then triangulating them.
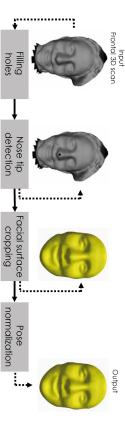
Fig. 10. Outline of the preprocessing steps. A pipeline of four filters is applied to each 3-D sequence: (a) Filling holes, if any. (b) Nose tip detection (for the first frame) and tracking (for remaining frames). (c) Face cropping using a sphere centered on the nose tip and of radius 90 mm. (d) Pose normalization based on the ICP algorithm.
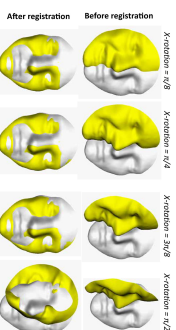
Fig. 11. Registration results using the ICP algorithm when rotating around the x-axis one of the 3-D preprocessed faces. The first row shows the initial rotation applied on the yellow model (before the alignment) and the second row shows the alignment output (after alignment).

2) Detect the nose tip on the face scan in the first frame of the sequence. The nose tip is detected by analyzing the peak point of the face scan in the depth direction. The nose tip is then tracked on all the subsequent frames when the search area is limited to a specific neighborhood around the nose tip detected on the first frame.

3) Crop the facial area using a sphere centered on the detected nose tip with a constant radius set to 90 mm based on some observations.

4) Normalize the pose of a given frame according to its previous frame using the Iterative Closest Point (ICP)-based alignment. We point out that our implementation uses the following parameters to perform the ICP algorithm: a) match the nose tips of the faces first; b) number of vertices considered to find the optimal transformation=50; and c) number of iterations=5. In addition to permit effective alignment, this set of parameters allows also an attractive computational cost.

In a real-world scenario of use, the head can move freely and rotate, whereas in our experiments only near-frontal faces are considered, as the BU-4DFE database does not contain nonfrontal acquisitions. To account for the capability of our approach to cope with non-frontal scans, we report in Fig. 11 some registration results when applying an artificial rotation to one of the 3-D faces to be aligned. It is clear that the registration method is able to handle with moderate pose variations (up to about 30°/45°). Instead, the registration method is not able to register a frontal face with a profile face (right side of the figure).
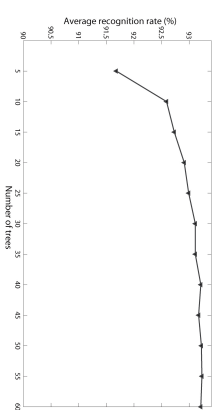
Fig. 12. 4-D expression recognition results using a random forest classifier when varying the number of trees.

TABLE I
CONFUSION MATRIX FOR MEAN DEFORMATION AND RANDOM FOREST CLASSIFIER (FOR 6-FRAMES WINDOW)

%	An	Di	Fe	Ha	Sa	Su
An	**93.11**	2.42	1.71	0.46	1.61	0.66
Di	2.3	**92.46**	2.44	0.92	1.27	0.58
Fe	1.89	1.75	**91.24**	1.5	1.76	1.83
Ha	0.57	0.84	1.71	**95.47**	0.77	0.62
Sa	1.7	1.52	2.01	1.09	**92.46**	1.19
Su	0.71	0.85	1.84	0.72	1.33	**94.53**

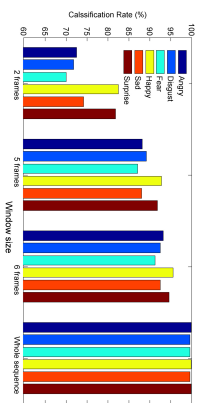Average recognition rate = 93.21±0.81%

Fig. 13. Effect of the temporal size of the sliding window on the results. The classification rates increase with the increase in the length of the temporal window.

In the proposed framework, after preprocessing and DSFs computation across the 3-D sequences, we designed two deformation learning and classification schemes. The first scheme consists on averaging, within a sliding window, the DSF computed on each frame with respect to the first frame of the window. This produces dense deformations across the sliding windows that are learned using a multiclass random forest algorithm (Section III-A). The second scheme consists on a dynamic analysis through the 3-D sequences using conventional temporal HMMs-modeling. Here, the 3-D motion (deformation) is extracted and then learned for each class of expression, as described in Section III-B. In both the cases, a common experimental set up has been used. In particular, the data of 60 subjects have been selected to perform recognition experiments according to the evaluation protocol followed in other papers [23], [26], [27]. The 60 subjects have been randomly partitioned into ten sets, each containing six subjects, and 10-fold cross validation has been used for training/test, where at each round nine of the ten folds (54 subjects) is used for the training, while the remaining fold (six subjects) is used for the test. In the following, we report experimental evaluation and comparative analysis with respect to the previous papers.

B. Mean Deformation-Based Expression Classification

Following the experimental protocol proposed in [23], a large set of subsequences are extracted from the original expression sequences using a sliding window. The subsequences have been defined in [23] with a length of six frames with a sliding step of one frame from one subsequence to the following one. For example, with this approach, a sequence of 100 frames originates a set of 6×95 = 570 subsequences, each subsequence differing for one frame from the previous one. Each sequence is labelled to be one of the six basic expressions, thus extracted subsequences have the same label. This accounts for the fact that, in general, the subjects can enter the system not necessarily starting with a neutral expression, but with an arbitrary expression. The classification of these short sequences is regarded as an indication of the capability of the expression recognition framework to identify individual expressions. According to this, we first compute the Mean Deformation for each subsequence, which is then presented to multiclass random forest, as outlined in Section III.

The performance of random forest classifier varies with the number of trees. Thus, we perform the experiments with different numbers of trees and the results of this experimentation is shown in Fig. 12. As illustrated in this figure, the average recognition rate raises with the increasing number of trees until 40, when the recognition rate reaches 93.21%, and then becomes quite stable. Thus, in the following we consider 40 trees and report detailed results (confusion matrix) with this number of trees in Table I. We recall that the rates are obtained by averaging the results of the ten independent runs (10-fold cross validation). It can be noted that the largest confusions are between the disgust (Di) expression and the angry (An) and Fear (Fe) expressions. Interestingly, these three expressions capture negative emotive states of the subjects, so that similar facial muscles can be activated. The best classified expressions are happy (Ha) and surprise (Su) with recognition accuracy of 95.47% and 94.53%, respectively. The standard deviation from the average performance is also reported in the table. The value of this statistical indicator suggests that small variations are observed between different folds.

Effect of the Subsequence Size: We have also conducted additional experiments when varying the temporal size of the sliding window used to define the subsequences. In Fig. 13, we report results for a window size equal to 2, 5, and 6, and using the whole length of the sequence (on average this is about 100 frames). From the figure, it clearly emerges that the recognition rate of the six expressions increases with the increase in the temporal length of the window. This reveals the importance of the temporal dynamics and shows that the spatio-temporal analysis outperforms a spatial analysis of the frames. By considering the whole sequences for the classification, the result reach 100%. In the paper, we decided to report detailed results when considering a window length of 6-frames to allow comparisons with previous papers.

Effect of the Spatial Resolution of 3-D Faces: In the proposed face representation, the DSF is computed for the
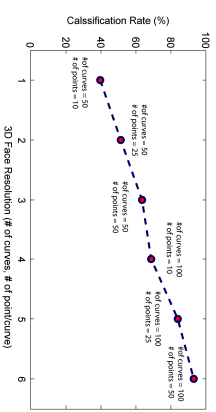
Fig. 14. Effects of varying the 3-D face resolution on the classification results.

points of a set of radial curves originating from the nose tip. Due to this, the density of the scalar field depends on the number of radial curves and the number of points per curve. So, the resolution used for the number of curves and points per curve can affect the final effectiveness of the representation. To investigate this aspect, we have conducted experiments when varying the spatial resolution of the 3-D faces (i.e., the number of radial curves and the number of points per curve), Fig. 14 expresses quantitatively the relationship between the expression classification accuracy (on the BU-4DFE) and the number of radial curves and the number of points per curve. This can give an indication of the expected decrease in the performance in the case the number of radial curves or points per curve is decreased due to the presence of noise and spikes in the data. From these results, we can also observe that the resolution in terms of number of curves has more importance than the resolution in terms of points per curve.

C. HMM-Based Expression Classification

Following the same setup as in Section IV-B (originally defined in [23], for this experiment we trained the HMMs on six frames subsequences constructed as discussed above. The 4-state structure of the HMMs resulted adequate to model the subsequences. Also in this experiment, we performed 10-fold cross validation on the overall set of subsequences derived from the 60 × 6 sequences (31970 in total). The achieved results by classifying individual subsequences of the expression sequences (frame-by-frame experiment) are reported in the confusion matrix of Table II. The values in the table have been obtained by using features of six frames subsequences as input to the 6 HMMs and using the maximum emission probability criterion as decision rule. It is clear that the proposed approach is capable to classify individual frames accurately by analyzing the corresponding subsequence of previous five frames. The average recognition rate is equal to 93.83%, slightly higher than the one displayed by mean deformation plus random forest classification schema (though the standard deviation among different folds shows a greater value in this case). It can also be noted that, compared to the previous classifier, the same tendency of recognition rates is achieved generally. In fact, correct classification of angry is high despite the difficulty of this expression analysis. This learning scheme achieved better recognition than the first one for angry (An) expression. Actually, whereas the angry

TABLE II
CONFUSION MATRIX FOR MOTION EXTRACTION AND HMM CLASSIFIERS (FOR 6-FRAMES WINDOW)

%	An	Di	Fe	Ha	Sa	Su
An	**93.95**	1.44	1.79	0.28	2.0	0.54
Di	8.63	**76.1**	6.65	2.82	4.18	1.62
Fe	3.1	5.5	**80.23**	1.99	6.31	2.87
Ha	1.43	2.02	3.77	**86.12**	5.31	1.35
Sa	5.79	1.4	4.99	0.86	**85.83**	1.13
Su	1.73	2.04	6.1	1.55	3.9	**84.65**

Average recognition rate = 81.9±3.26%

TABLE III
CONFUSION MATRIX FOR FREE-FORM DEFORMATION (FFD) AND HMM CLASSIFIERS (FOR 6-FRAMES WINDOW)

%	An	Di	Fe	Ha	Sa	Su
An	**93.95**	1.44	1.79	0.28	2.0	0.54
Di	3.10	**91.54**	3.40	0.54	1.27	0.15
Fe	1.05	1.42	**94.55**	0.69	1.67	0.62
Ha	1.43	2.02	3.77	**86.12**	5.31	1.35
Sa	5.79	1.4	4.99	0.86	**85.83**	1.13
Su	0.57	0.38	3.25	0.32	1.85	**93.57**

Average recognition rate = 93.83±1.53%

(An) expression is known for its subtle motions, our classifier achieved 93.95% of correct classification, which demonstrates the ability of the proposed DSF to capture subtle deformations across the 3-D sequences. These similar good achievements are mainly the effect of the proposed deformation scalar field.

Comparison With the FFD Feature: The proposed framework can also fit with different deformation fields than the proposed DSF. So, considering alternative features to densely capture the deformation fields on the lattice of points of the radial curves of the face can permit a direct comparison of our DSF feature with different ones. In particular, we considered the FFD [38] feature, which is a standard method for nonrigid registration and has been successfully proved in the context of expression recognition [25] (see also Section II-B). Table III reports the confusion matrix obtained by posing FFD in our classification framework, using the same setting as in the experiments above (i.e., 100 radial curves, with 50 sampled points each, and LDA reduction of the deformation field from a 5000-D vector to a 5-D subspace). The overall result is that the proposed DSF feature provides a finer discriminative capability with respect to the FFD, thus resulting in a better classification accuracy. This can be motivated by the nice invariant properties of the proposed Riemannian framework (as discussed in Section II).

D. Discussion and Comparative Evaluation

To the best of our knowledge, the papers reporting results on expression recognition from dynamic sequences of 3-D scans are those in [20], [24], [26], [29], and recently [30]. These papers have been evaluated on the BU-4DFE dataset, but the testing protocols used in the experiments are sometimes different, so that a direct comparison of the results reported in these papers is not immediate. In the following, we discuss the different settings under which the expression recognition results are obtained.

TABLE IV
COMPARISON OF THIS PAPER TO EARLIER PAPERS. PROTOCOL DESCRIPTION: #SUBJECTS (S), #EXPRESSIONS (E), WIN SIZE (WIN).
T: TEMPORAL ONLY/S-T: SPATIO-TEMPORAL. ACCURACY ON SLIDING WINDOW/WHOLE SEQUENCE (OR SUBSEQUENCE)

Authors	Method	Features	Classification	Protocol	T/S-T	RR (%)
Sun et al. [23]	MU-3D	12 Motion Units	HMM	60 S, 6 E, Win=6	T	70.31, —
Sun et al. [23]	T-HMM	12 Motion Units	HMM	60 S, 6 E, Win=6	T	80.04, —
Sun et al. [23]	P2D-HMM	Curvature+Tracking model	T-HMM+S-HMM	60 S, 6 E, Win=6	S-T	82.19, —
Sun et al. [23]	R-2DHMM	Curvature+Tracking model	2D-HMM	60 S, 6 E, Win=6	S-T	90.44, —
Sandbach et al. [25]	3D Motion-based	FFD+Quad-tree	GentleBoost+HMM	—, 3 E, Win=4	T	73.61, 81.93
Sandbach et al. [20]	3D Motion-based	FFD+Quad-tree	GentleBoost+HMM	—, 6 E, variable Win	T	64.6, —
Le et al. [26]	Level curve-based	pair- and segment-wise distances	HMM	60 S, 3 E, —	S-T	92.22, —
Fang et al. [27], [29]	AFM Fitting	LBP-TOP	SVM-RBF	100 S, 6 E, —	T	74.63, —
Fang et al. [27], [29]	AFM Fitting	LBP-TOP	SVM-RBF	100 S, 3 E, —	T	96.71, —
Reale et al. [30]	"Nebula" Feature	LBP-TOP	SVM-RBF	100 S, 6 E, Win=15	T	76.10, —
This work	Spatio-temporal volume	3D Motion Extraction	LDA-HMM	60 S, 6 E, Win=6	S-T	93.83, —
This work	Geometric Mean Deformation	Mean Deformation	LDA-Random Forest	60 S, 6 E, Win=6	T	93.21, —

Table IV summarizes approaches and results reported previously on the BU-4DFE dataset, compared to those obtained in this paper. The testing protocols used in the experiments are quite different especially by the number of verified expressions, all the six basic expressions in [23], [24], [27], [29], and [30], whereas [25] and [26] reported primary results on only three expressions. The number of subjects considered is 60, except in [25] where the number of subjects is not specified. In general, sequences in which the required expressions are acted accurately are selected, whereas in [27] and [29] 507 sequences out of the 606 total are used for all subjects. On our experiments, we conducted tests by following the same setting proposed by the earliest and more complete evaluation described in [23]. The training and the testing sets were constructed by generating subsequences of 6-frames from all sequences of 60 selected subjects. The process were repeated by shifting the starting index of the sequence every one frame till the end of the sequence.

We note that the proposed approaches outperforms state-of-the-art solutions following similar experimental settings. The recognition rates reported in [23] and [24] based on temporal analysis only was 80.04% and spatio-temporal analysis was 90.44%. In both papers, subsequences of constant window width including 6-frames ($win = 6$) is defined for experiments. We emphasize that their approach is not completely automatic requiring 83 manually annotated landmarks on the first frame of the sequence to allow accurate model tracking.

The method proposed in [25] and [20] is fully automatic with respect to processing of facial frames in the temporal sequences, but uses supervised learning to annotate individual frames of the sequence in order to train a set of HMMs. Though performed offline, supervised learning requires manual annotation and counting on a consistent number of training sequences that can be a time consuming operation. In addition, a drawback of this solution is the computational cost due to free-form deformations based on B-spline interpolation between a lattice of control points for nonrigid registration and motion capturing between frames. Preliminary tests were reported on three expressions: (An), (Ha), and (Su). The Authors motivated the choice of the happiness and anger expressions with the fact that they are at either ends of the valence expression spectrum, whereas surprise was also chosen as it is at one extreme of the arousal expression spectrum. However, these experiments were carried out on

a subset of subjects accurately selected as acting out the required expression. Verification of the classification system was performed using a 10-fold cross-validation testing. On this subset of expressions and subjects, an average expression recognition rate of 81.93% is reported. Sandbach et al. [20] have reported 64.6% classification rate when in their evaluation by considering all the six basic expressions.

In [26], a fully automatic method is also proposed that uses an unsupervised learning solution to train a set of HMMs (i.e., annotation of individual frames is not required in this case). Expression recognition is performed on 60 subjects from the BU-4DFE database for the expressions of happiness, sadness, and surprise. The recognition accuracy averaged on ten rounds of 10-fold cross-validation show an overall value of 92.22% with the highest performance of 95% obtained for the happiness expression. However, the authors reported recognition results on whole facial sequences, but this hinders the possibility of the methods to adhere to a real-time protocol. In fact, the reported recognition results depends on the preprocessing of whole sequences unlike our approach and the one described in [23], which are able to provide recognition results when processing very few 3-D frames.

In [27] and [29], results are presented for expression recognition accuracy on 100 subjects picked out from BU-4DFE database. However, 507 sequences are selected manually according to the following criteria: 1) the 4-D sequence should start by neutral expression and 2) sequences containing corrupted meshes are discarded. In addition, to achieve recognition rate of 75.82%, whole sequences should be analyzed. The authors reported highest recognition rates when only (Ha), (An), and (Su) expressions (96.71%) or (Ha), (Su) and (Su) are considered.

The protocol used in [30] is quite different from the others. First, the onset frame for each of the six canonical expressions has been marked manually on each sequence of the BU-4DFE database. Then, a fixed size window of 15 frames starting from the onset frame has been extracted from each expression of 100 subjects. So, although sequences from 100 subjects are used by this approach, it also uses a manual intervention to detect the onset frame and just 15 frames from the onset one are used for the classification (and these typically correspond to the most intense expression, including the apex frames).

According to this comparative analysis, the proposed framework compares favorably with the state-of-the-art solutions. It

consists of two geometric deformation learning schemes with a common feature extraction module (DSF). This demonstrates the effectiveness of the novel mathematical representation called DSF, under the two designed schemes.

V. CONCLUSION

In this paper, we presented an automatic approach for identity-independent facial expression recognition from 3-D video sequences. Through a facial shapes representation by collections of radial curves, a Riemannian shape analysis was applied to quantify dense deformations and extract motion from successive 3-D frames. Two different classification schema were performed, a HMM-based classifier and a mean deformation-based classifier. An LDA-based transformation was applied to decrease the dimensionality of the resulting feature vectors. The proposed approach outperforms previous ones; it is capable to accurately classify short sequences containing very different 3-D faces with an average accuracy of 93.83% using HMM-based classifier and 93.21% using mean deformation-based classifier, following the state-of-the-art settings on the BU-4DFE dataset. This emphasizes the capability of the proposed geometric shape deformation analysis to capture subtle deformations in 4-D videos.

One limitation of the approach is the nose tip detection in case of nonfrontal views and/or the presence of occlusions (by glasses, hand, hair, and so on). The BU-4DFE database contains frontal 3-D faces without occlusion, however, in realistic scenarios, more elaborated techniques should be applied to detect the nose tip. As future perspectives of the presented paper are, first, its extension to spatio-temporal analysis by introducing a spatial (intraframe) analysis beside the temporal analysis (interframe). Second, its adaptation to low resolution 3-D videos, outputs of the depth-consumer cameras like the Kinect, is a distant goal mainly due to the presence of large noise and the low resolution of the acquired scans.

REFERENCES

[1] A. Mehrabian and M. Wiener, "Decoding of inconsistent communications," J. Personality Social Psychol., vol. 6, no. 1, pp. 109–114, 1967.
[2] P. Ekman, "Universals and cultural differences in facial expressions of emotion," in Proc. NE Symp. Motivation, 1972, pp. 207–283.
[3] P. Ekman and W. V. Friesen, Manual for the the Facial Action Coding System. Palo Alto, CA, USA: Consulting Psychologist Press, 1977.
[4] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, pp. 39–58, Jan. 2009.
[5] I. A. Kakadiaris et al., "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable approach," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 4, pp. 640–649, Apr. 2007.
[6] A. S. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3-D face recognition," Int. J. Comput. Vision, vol. 79, no. 1, pp. 1–12, 2008.
[7] C. Samir, A. Srivastava, M. Daoudi, and E. Klassen, "An intrinsic framework for analysis of facial surfaces," Int. J. Comput. Vision, vol. 82, no. 1, pp. 80–95, 2009.
[8] S. Berretti, A. Del Bimbo, and P. Pala, "3-D face recognition using iso-geodesic stripes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 12, pp. 2162–2177, Dec. 2010.
[9] Y. Wang, J. Liu, and X. Tang, "Robust 3-D face recognition by local shape difference boosting," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 10, pp. 1858–1870, Oct. 2010.
[10] A. Maalej, B. Ben Amor, M. Daoudi, A. Srivastava, and S. Berretti, "Shape analysis of local facial patches for 3-D facial expression recognition," Pattern Recognit., vol. 44, no. 8, pp. 1581–1589, 2011.

[11] S. Berretti, B. Ben Amor, M. Daoudi, and A. Del Bimbo, "3-D facial expression recognition using sift descriptors of automatically detected keypoints," The Visual Comput., vol. 27, no. 11, pp. 1021–1036, 2011.
[12] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato, "A 3-D facial expression database for facial behavior research," in Proc. IEEE Int. Conf. Automat. FGR, 2006, pp. 211–216.
[13] A. Savran et al., "Bosphorus database for 3-D face analysis," in Proc. BIOID, 2008, pp. 47–56.
[14] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3-D dynamic facial expression database," in Proc. Int. Conf. Automat. FGR, Amsterdam, 2008, pp. 1–6.
[15] B. Matuszewski, W. Quan, and L.-K. Shark, "High-resolution comprehensive 3-D dynamic database for facial articulation analysis," in Proc. IEEE ICCV, 2011, pp. 2128–2135.
[16] B. J. Matuszewski et al., "Hi4D-ADSIP 3-D dynamic facial articulation database," Image Vision Comput., vol. 30, no. 10, pp. 713–727, 2012.
[17] D. Cosker, E. Krumhuber, and A. Hilton, "A FACS valid 3-D dynamic action unit database with applications to 3-D dynamic morphable facial modeling," in Proc. IEEE ICCV, 2011, pp. 2296–2303.
[18] S. Berretti, A. Del Bimbo, and P. Pala, "Superfaces: A super-resolution method for 3-D faces," in Proc. 3-D Face, NORDIA, 2012, pp. 73–82.
[19] Y. Li, A. Mian, W. Lu, and A. Krishna, "Using Kinect for face recognition under varying poses, expressions, illumination and disguise," in Proc. WACV, 2013, pp. 186–192.
[20] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3-D facial expression recognition: A comprehensive survey," Image Vision Comput., vol. 30, no. 10, pp. 683–697, 2012.
[21] L. Benedikt, V. Kajic, D. Cosker, P. Rosin, and D. Marshall, "Facial dynamics in biometric identification," in Proc. British Mach. Vision Conf., 2008, pp. 1–10.
[22] L. Benedikt, D. Cosker, P. L. Rosin, and D. Marshall, "Assessing the uniqueness and permanence of facial actions for use in biometric applications," IEEE Trans. Syst., Man, Cybern., A, Syst., Humans, vol. 40, no. 3, pp. 449–460, May 2010.
[23] Y. Sun and L. Yin, "Facial expression recognition based on 3-D dynamic range model sequences," in Proc. ECCV, 2008, pp. 58–71.
[24] Y. Sun, X. Chen, M. J. Rosato, and L. Yin, "Tracking vertex flow and model adaptation for three-dimensional spatiotemporal face analysis," IEEE Trans. Syst., Man, Cybern., A, Syst., Humans, vol. 40, no. 3, pp. 461–474, May 2010.
[25] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, "A dynamic approach to the recognition of 3-D facial expressions and their temporal models," in Proc. IEEE Conf. Automat. FGR, 2011, pp. 406–413.
[26] V. Le, H. Tang, and T. S. Huang, "Expression recognition from 3-D dynamic faces using robust spatio-temporal shape features," in Proc. IEEE Conf. Automat. FGR, 2011, pp. 414–421.
[27] T. Fang, X. Zhao, S. Shah, and I. Kakadiaris, "4-D facial expression recognition," in Proc. IEEE ICCV, 2011, pp. 1594–1601.
[28] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 6, pp. 915–928, Jun. 2007.
[29] T. Fang, X. Zhao, O. Ocegueda, S. Shah, and I. Kakadiaris, "3-D facial expression recognition: A comprehensive annotated face model approach," Image Vision Comput., vol. 30, no. 10, pp. 738–749, 2012.
[30] M. Reale, X. Zhang, and L. Yin, "Nebula feature: A space-time feature for posed and spontaneous 4d facial behavior analysis," in Proc. IEEE Int. Conf. Automat. FGR, 2013, pp. 1–8.
[31] X. Zhang et al., "A high-resolution spontaneous 3-D dynamic facial expression database," in Proc. IEEE FG, 2013, pp. 1–6.
[32] H. Drira, B. Ben Amor, M. Daoudi, A. Srivastava, and S. Berretti, "3-D dynamic expression recognition based on a novel deformation vector field and random forest," in Proc. Int. Conf. Pattern Recognit., 2012, pp. 1104–1107.
[33] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, "Recognition of 3-D facial expression dynamics," Image Vision Comput., vol. 30, no. 10, pp. 762–773, 2012.
[34] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," IEEE Trans. Med. Imaging, vol. 18, no. 8, pp. 712–721, Aug. 1999.
[35] H. Drira, B. Ben Amor, M. Daoudi, and A. Srivastava, "Pose and expression-invariant 3-D face recognition using elastic radial curves," in Proc. British Mach. Vision Conf., 2010, pp. 1–11.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

BEN AMOR et al.: 4-D FACIAL EXPRESSION RECOGNITION

15

[36] A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn, "Shape analysis of elastic curves in Euclidean spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1415–1428, Jul. 2011.

[37] S. Joshi, E. Klassen, A. Srivastava, and I. Jermyn, "A novel representation for Riemannian analysis of elastic curves in ℝ^n," in *Proc. CVPR*, 2007, pp. 1065–6919.

[38] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imaging*, vol. 18, no. 8, pp. 712–721, Aug. 1999.

[39] L. Breiman, "Random forests," *Mach. Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[40] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–94, Jan. 1980.

[41] L. Rabiner, "A tutorial on Hidden Markov models and selected applications in speech recognition," in *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[42] DIMENSIONAL IMAGING. (2006). *Di3-D* [Online]. Available: http://www.di3d.com

Boulbaba Ben Amor (M'10) received the engineering degree in computer science from ENIS, Sfax, Tunisia, in 2002, and the M.S. and Ph.D. degrees in Computer Science from Ecole Centrale de Lyon, Écully, France, in 2003 and 2006, respectively.

He is with the Institut Mines-Télécom Lille/Télécom Lille, Lille, France, as an Associate Professor, since 2007. He is a member of the Computer Science Laboratory of University Lille 1, Lille, France. He has co-authored several papers in refereed journals and proceedings of international conferences and has been involved in French and international projects. He has served as program committee member and reviewer for international journals and conferences. He is currently a Visiting Researcher at the Florida State University, Florida, USA. His current research interest includes statistical 3-D face analysis.

Hassen Drira received the engineering and M.Sc. degrees in computer science from the National School of Computer Science (ENSI), Manouba, Tunisia, in 2006 and 2007, respectively and the Ph.D degree in computer science from University of Lille 1, Lille, France, in 2011.

From 2011 to 2012, he was with the MIIRE Research Group within the Fundamental Computer Science Laboratory of Lille as a Post-Doctoral Fellow. Since 2012, he is an Assistant Professor of Computer Science at Institut Mines-Télécom Lille/Télécom Lille, Lille, France. His current research interests include pattern recognition, statistical analysis, 3-D face recognition, biometrics, and 3-D facial expression recognition. He has published several refereed journal and conference articles in these areas.

Stefano Berretti (M'07) received the Ph.D. degree in information and telecommunications engineering from the University of Florence, Florence, Italy, in 2001.

He was a Visiting Researcher with the Indian Institute of Technology (IIT), in Mumbai, India, in 2000, and a Visiting Professor with Institute TELECOM, TELECOM Lille 1, Lille, France, in 2009 and with Khalifa University, Sharjah, UAE, in 2013. He is currently an Associate Professor at the Department of Information Engineering and at the Media Integration and Communication Center of University of Florence. He has authored more than 100 papers in conference proceedings and international journals in the area of pattern recognition, computer vision, and multimedia. His current research interests include content modeling, retrieval, and indexing of image and 3-D object databases. His recent research has addressed 3-D object retrieval and partitioning, 3-D face recognition, 3-D and 4-D facial expression recognition.

Dr. Berretti is in the program committee of several international conferences and serves as a frequent reviewer of many international journals. He was a Co-Chair of the Fifth Workshop on Nonrigid Shape Analysis and Deformable Image Alignment, held in conjunction with ECCV 2012.

Mohamed Daoudi (SM'96) received the Ph.D. degree in computer engineering from the University of Lille 1, Lille, France, in 1993 and in Habilitation from the University of Littoral, France, in 2000.

He is a Professor of Computer Science at Télécom Lille and LIFL (UMR CNRS 8022). He is the Head of the Computer Science Department at Télécom Lille, France. He was the Founder and the Scientific Leader of MIIRE research group (http://www-rech.telecom-lille.fr/miire/). He has authored several books, such as 3-D processing: Compression, Indexing and Watermarking (Wiley, 2008) and 3-D Face Modeling, Analysis and Recognition (Wiley, 2013). He has published over 140 papers in some of the most distinguished scientific journals and international conferences. His current research interests include pattern recognition, image processing, 3-D analysis and retrieval, and 3-D face analysis and recognition.

Dr. Daoudi was in the program committee of several international conferences. He has Co-Chaired several workshops, including 3-D Objects Retrieval ACM Workshop 2010, 3-D Object Retrieval Eurographics Workshop 2010, and 3-D Face Biometrics 2013 (IEEE FG 2013).

Anuj Srivastava (SM'85) received the Ph.D. degree in electrical engineering from Washington University, St. Louis, MO, USA, in 1996.

From 1996 to 1997, he was a Visiting Research Associate with the Division of Applied Mathematics at Brown University. In 1997, he joined the Department of Statistics at the Florida State University as an Assistant Professor. He was promoted to the Associate Professor position in 2003 and to the full Professor position in 2007. He has been a Visiting Professor to INRIA, France and the University of Lille, Lille, France. He has published more than 175 papers in refereed journals and proceedings of refereed international conferences. His current research interests include statistics on nonlinear manifolds, statistical image understanding, functional analysis, and statistical shape theory.

Dr. Srivastava has been the Associate Editor for the *Journal of Statistical Planning and Inference*, THE IEEE TRANSACTIONS ON SIGNAL PROCESSING, AND THE IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE.

3D Face Recognition under Expressions, Occlusions, and Pose Variations

Hassen Drira, Boulbaba Ben Amor, *Member, IEEE,* Anuj Srivastava, *Senior Member, IEEE,* Mohamed Daoudi, *Senior Member, IEEE,* and Rim Slama

Abstract—We propose a novel geometric framework for analyzing 3D faces, with the specific goals of comparing, matching, and averaging their shapes. Here we represent facial surfaces by radial curves emanating from the nose tips and use elastic shape analysis of these curves to develop a Riemannian framework for analyzing shapes of full facial surfaces. This representation, along with the elastic Riemannian metric, seems natural for measuring facial deformations and is robust to challenges such as large facial expressions (especially those with open mouths), large pose variations, missing parts, and partial occlusions due to glasses, hair, and so on. This framework is shown to be promising from both—empirical and theoretical—perspectives. In terms of the empirical evaluation, our results match or improve upon the state-of-the-art methods on three prominent databases: FRGCv2, GavabDB, and Bosphorus, each posing a different type of challenge. From a theoretical perspective, this framework allows for formal statistical inferences, such as the estimation of missing facial parts using PCA on tangent spaces and computing average shapes.

Index Terms—3D face recognition, shape analysis, biometrics, quality control, data restoration

---

1   INTRODUCTION

DUE to the natural, nonintrusive, and high throughput nature of face data acquisition, automatic face recognition has many benefits when compared to other biometrics. Accordingly, automated face recognition has received growing attention within the computer vision community over the past three decades. Among different modalities available for face imaging, 3D scanning has a major advantage over 2D color imaging in that nuisance variables, such as illumination and small pose changes, have a relatively smaller influence on the observations. However, 3D scans often suffer from the problem of missing parts due to self-occlusions or external occlusions, or some imperfections in the scanning technology. Additionally, variations in face scans due to changes in facial expressions can also degrade face recognition performance. To be useful in real-world applications, these 3D face recognition approach should be able to handle these challenges, i.e., it should recognize people despite large facial expressions, occlusions, and large pose variations. Some examples of face scans highlighting these issues are illustrated in Fig. 1.

1.1   Previous Work

The task of recognizing 3D face scans has been approached in many ways, leading to varying levels of successes. We refer the reader to one of many extensive surveys on the topic, for example, see Bowyer et al. [3]. Below we summarize a smaller subset that is more relevant to our paper.

1.   *Deformable template-based approaches.* There have been several approaches in recent years that rely on deforming facial surfaces into one another, under some chosen criteria, and use quantifications of these deformations as metrics for face recognition. Among these, the ones using nonlinear deformations facilitate local stretching, compression, and bending of surfaces to match each other and are referred to as *elastic* methods. For instance, Kakadiaris et al. [13] utilize an *annotated face model* to study geometrical variability across faces. The annotated face model is deformed elastically to fit each face, thus matching different anatomical areas such as the nose, eyes,

We note that the most recent research on 3D face analysis has been directed toward tackling changes in facial expressions while only a relatively modest effort has been spent on handling occlusions and missing parts. Although a few approaches and corresponding results dealing with missing parts have been presented, none, to our knowledge, has been applied systematically to a full real database containing scans with missing parts. In this paper, we present a comprehensive Riemannian framework for analyzing facial shapes, in the process dealing with large expressions, occlusions, and missing parts. Additionally, we provide some basic tools for statistical shape analysis of facial surfaces. These tools help us to compute a typical or average shape and measure the intraclass variability of shapes, and will even lead to face atlases in the future.

2.   *Local regions/features approaches.* Another common framework especially for handling expression variability is based on matching only parts or regions rather than matching full faces. Lee et al. [15] use ratios of distances and angles between eight fiducial points, followed by an SVM classifier. Similarly, Gupta et al. [11] use euclidean/geodesic distances between anthropometric fiducial points in conjunction with linear classifiers. As stated earlier, the problem of automated detection of fiducial points is nontrivial and hinders automation of these methods. Gordon [10] argues that curvature descriptors have the potential for higher accuracy in describing surface features and are better suited to describing the properties of faces in areas such as the cheeks, forehead, and chin. These descriptors are also invariant to viewing angles. Li et al. [16] design a feature pooling and ranking scheme to collect various types of low-level geometric features, such as curvatures, and rank them according to their sensitivity to facial expressions. Along similar lines, Wang et al. [32] use a signed shape-difference map between two aligned 3D faces as an intermediate representation for shape comparison. McKeon and Russ [19] use a region ensemble approach that is based on Fisherfaces, i.e., face representations are learned using Fisher's discriminant analysis.

In [12], Huang et al. use a multiscale local binary pattern for a 3D face jointly with shape index. Similarly, Moorthy et al. [20] use Gabor features around automatically detected fiducial points. To avoid passing over deformable parts of faces encompassing discriminative information, Faltemier et al. [9] use 38 face regions that densely cover the face classes. They match and fuse scores and decisions after performing ICP on each region. A similar idea is proposed in [29] that uses PCA-LDA for feature extraction, treating the likelihood ratio as a matching score and using the majority voting for face identification.

and mouth. In [25], Passalis et al. use automatic landmarking to estimate the pose and to detect occluded areas. The facial symmetry is used to overcome the challenges of missing data here. Similar approaches, but using manually annotated models, are presented in [31] and [17]. For example, Lu and Jain [17] use manual landmarks to develop a thin-plate-spline-based matching of facial surfaces. A strong limitation of these approaches is that the extraction of fiducial landmarks needed during learning is either manual or semi-automated, except in [13], where it is fully automated.

3.   *Surface distance-based approaches.* There are several papers that utilize surface distances between points in facial surfaces to define features that are eventually used in recognition. (Some papers call it *geodesic distance* but, to distinguish it from our later use of *geodesics* on shape spaces of curves and surfaces, we shall call it *surface distance*.) These papers assume that surface distances are relatively invariant to small changes in facial expressions and therefore help generate features that are robust to facial expressions. Bronstein et al. [4] provide a limited experimental illustration of this invariance by comparing changes in surface distances with the euclidean distances between corresponding points on a canonical face surface. To handle the open mouth problem, they first detect and remove the lip region, and then compute the surface distance in the presence of a hole corresponding to the removed part [5]. The assumption of preservation of surface distances under facial expressions motivates several authors to define distance-based features for facial recognition. Samir et al. [28] use the level curves of the surface distance function (from the tip of the nose) as features for face recognition. Since an open mouth affects the shape of some level curves, this method is not able to handle the problem of missing data due to occlusion or pose variations. A similar polar parameterization of the facial surface is proposed in [24], where the authors study local geometric attributes under this parameterization. To deal with the open mouth problem, they modify the parameterization by disconnecting the top and bottom lips. The main limitation of this approach is the need for detecting the lips, as proposed in [5]. Berretti et al. [2] use surface distances to define facial stripes which, in turn, are used as nodes in a graph-based recognition algorithm.

The main limitation of these approaches, apart from the issues resulting from open mouths, is that they assume that surface distances between facial points are preserved within face classes. This is not valid in the case of large expressions. Actually, face expressions result from stretching or the shrinking of underlying muscles and, consequently, the facial skin is deformed in a nonisometric manner. In other words, facial surfaces are also stretched or compressed locally, beyond a simple bending of parts.

Querolo et al. [26] use surface interpenetration measure as a similarity measure to match two face images. The authentication score is obtained by combining the SIM values corresponding to the matching of four different face regions: circular and elliptical areas around the nose, forehead, and the entire face region. In [1], Alyuz et al. use average region models (ARMs) locally to handle the challenges of missing data and expression-related deformations. They manually divide the facial area into several meaningful components and the registration of faces is carried out by separate dense alignments to the corresponding ARMs. A strong limitation of this approach is the need for manual segmentation of a face into parts that can then be analyzed separately.

- H. Drira, B. Ben Amor, and M. Daoudi are with Laboratoire d'Informatique Fondamentale de Lille (LIFL), (UMR CNRS 8022), Institut Mines-Télécom/ TELECOM Lille 1, Cité Scientifique, Rue G. Marconi, BP 20145, Villeneuve d'Asq 59653 France. E-mail: hassen.drira@telecom-lille1.eu.
- A. Srivastava is with the Department of Statistics, Florida State University, 106D OSB, 600 W College Ave, Tallahassee, FL 32306.
- R. Slama is with Laboratoire d'Informatique Fondamentale de Lille (LIFL), (UMR CNRS 8022), University of Lille 1, Télécom Lille 1, Cité Scientifique, Rue G. Marconi, BP 20145, Villeneuve d'Asq 59653, France.

Fig. 1. Different challenges of 3D face recognition: expressions, missing data, and occlusions.

Fig. 2. Significant changes in both euclidean and surface distances under large facial expressions.

Fig. 3. Overview of the proposed method.

Common stages   **Specified stages**   Approach Components

To demonstrate this assertion, we placed four markers on a face and tracked the changes in the surface and euclidean (straight line) distances between the markers under large expressions. Fig. 2 shows some facial expressions leading to a significant shrinking or stretching of the skin surface and thus causing both euclidean and surface distances between these points to change. In one case, these distances decrease (from 113 to 103 mm for the euclidean distance and from 115 to 106 mm for the surface distance), while in the other two cases they increase. This clearly shows that large expressions can cause stretching and shrinking of facial surfaces, i.e., the facial deformation is elastic in nature. Hence, the assumption of an isometric deformation of the shape of the face is not strictly valid, especially for large expressions. This also motivates the use of elastic shape analysis in 3D face recognition.

1.2 Overview of Our Approach

This paper presents a Riemannian framework for 3D facial shape analysis. This framework is based on elastically matching and comparing radial curves emanating from the tip of the nose and it handles several of the problems described above. The main contributions of this paper are as follows:

- It extracts, analyzes, and compares the shapes of radial curves of facial surfaces.
- It develops an elastic shape analysis of 3D faces by extending the elastic shape analysis of curves [30] to 3D facial surfaces.
- To handle occlusions, it introduces an occlusion detection and removal step that is based on recursive-ICP.
- To handle the missing data, it introduces a restoration step that uses statistical estimation on shape manifolds of curves. Specifically, it uses PCA on tangent spaces of the shape manifold to model the normal curves and uses that model to complete the partially observed curves.

The different stages and components of our method are laid out in Fig. 3. While some basic steps are common to all application scenarios, there are also some specialized tools suitable only for specific situations. The basic steps that are common to all situations include 3D scan preprocessing (nose tip localization, filling holes, smoothing, face cropping), coarse and fine alignment, radial curve extraction, quality filtering, and elastic shape analysis of curves (Component III and quality module in Component II). This basic setup is evaluated on the $FRGCv2$ dataset following the standard protocol (see Section 4.2). It is also tested on the $GAVAB$ dataset where, for each subject, four probe images out of nine have large pose variations (see Section 4.3). Some steps are only useful when one anticipates some data occlusion and missing data. These steps include occlusion detection and missing data
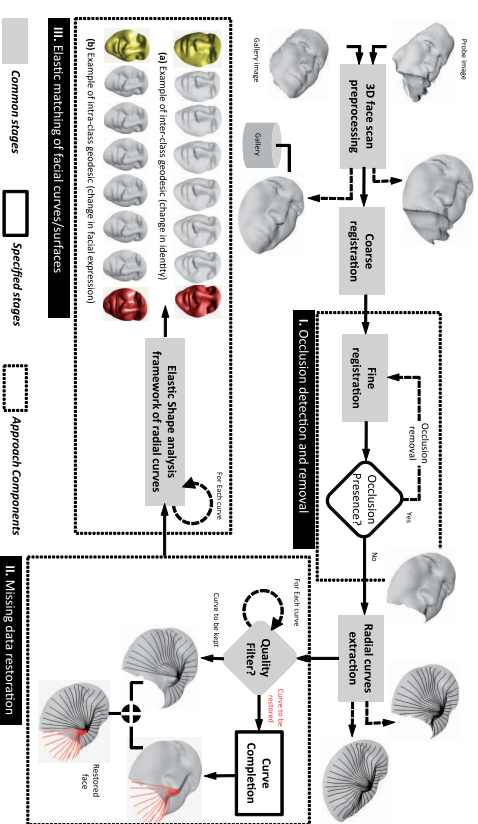
Fig. 4. A smile (see middle) changes the shapes of the curves in the lower part of a the face while the act of surprise changes shapes of curves in the upper part of the face (see right).

missing data restoration (Component II). In these situations, the full processing includes Components I+II+III to process the given probes. This approach has been evaluated on a subset of the $Bosphorus$ dataset that involves occlusions (see Section 4.4). In the last two experiments, except for the manual detection of nose coordinates, the remaining processing is automatic.

2 RADIAL, ELASTIC CURVES: MOTIVATION

An important contribution of this paper is its novel use of radial facial curves studied using elastic shape analysis.

2.1 Motivation for Radial Curves

Why should one use the radial curves emanating from the tip of the nose for representing facial shapes? First, why curves and not other kinds of facial features? Recently, there has been significant progress in the analysis of curves shapes and the resulting algorithms are very sophisticated and efficient [30], [33]. The changes in facial expressions affect different regions of a facial surface differently. For example, during a smile, the top half of the face is relatively unchanged while the lip area changes a lot, and when a person is surprised the effect is often the opposite. If chosen appropriately, curves have the potential to capture regional shapes and that is why their role becomes important. The locality of shapes represented by facial curves is an important reason for their selection. The next question is: Which facial curves are suitable for recognizing people? Curves on a surface can, in general, be defined either as the level curves of a function or as the streamlines of a gradient field. Ideally, one would like curves that maximally separate interclass variability from the intraclass variability (typically due to expression changes). The past usage of the level curves (of the surface distance function) has the limitation that each curve goes through different facial regions and that makes it difficult to isolate local variability. Actually, the previous work on shape analysis of facial curves for 3D face recognition was mostly based on level curves [27], [28].

In contrast, the radial curves with the nose tip as origin have tremendous potential. This is because:

1. The nose is in many ways the focal point of a face. It is relatively easy and efficient to detect the nose tip (compared to other facial parts) and to extract radial curves, with the nose tip as the center, in a completely automated fashion. It is much more difficult to automatically extract other types of curves, for example, those used by sketch artists (cheek contours, forehead profiles, eye boundaries, etc.).
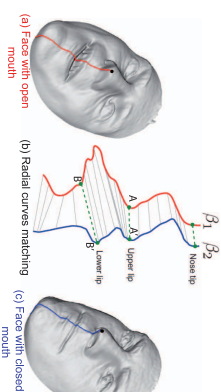
Fig. 5. An example of matching radial curves extracted from two faces belonging to the same person: a curve with an open mouth (on the left) and a curve with a closed mouth (on the right). One needs a combination of stretching and shrinking to match similar points (upper lips, lower lips, etc.)

2. Different radial curves pass through different regions and, hence, can be associated with different facial expressions. For instance, differences in the shapes of radial curves in the upper half of the face can be loosely attributed to the interclass variability, while those for curves passing through the lips and cheeks can largely be due to changes in expressions. This is illustrated in Fig. 4, which shows a neutral face (left), a smiling face (middle), and a surprised face (right). The main difference in the middle face, relative to the left face, lies in the lower part of the face, while for the right face the main differences lie in the top half.

3. Radial curves have a more universal applicability. The curves used in the past have worked well for some specific tasks, for example, lip contours in detecting certain expressions, but they have not been as efficient for some other tasks, such as face recognition. In contrast, radial curves capture the full geometry and are applicable to a variety of applications, including facial expression recognition.

4. In the case of the missing parts and partial occlusion, at least some part of every radial curve is usually available. It is rare to miss a full radial curve. In contrast, it is more common to miss an eye due to occlusion by glasses, the forehead due to hair, or parts of cheeks due to a bad angle for laser reflection. This issue is important in handling the missing data via reconstruction, as will be described later in this paper.

5. Natural face deformations are largely (although not exactly) symmetric and, to a limited extent, are radial around the nose. Based on these arguments, we choose a novel geometrical representation of facial surfaces using radial curves that start from the nose tip.

2.2 Motivation for Elasticity

Consider the two parameterized curves shown in Fig. 5; call them β_1 and β_2. Our task is to automatically match points on these radial curves associated with two different facial expressions. The expression on the left has the mouth open, whereas the expression on the right has the mouth closed. To compare their shapes, we need to register points across those curves. One would like the correspondence to
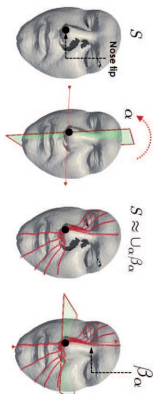
Fig. 6. Extraction of radial curves: Images in the middle illustrate the intersection between the face surface and planes to form two radial curves. The collection of radial curves is illustrated in the rightmost image.

be such that geometric features match across the curves as well as possible. In other words, the lips should match the lips and the chin should match the chin. Clearly, if we force an arc-length parameterization and match points that are at the same distance from the starting point, then the resulting matching will not be optimal. The points A and B on β_1 will not match the points A' and B' on β_2 as they are not placed at the same distances along the curves. For curves, the problem of optimal registration is actually the same as that of optimal reparameterization. This means that we need to find a reparameterization function $\gamma(t)$ such that the point $\beta_1(t)$ is registered with the point $\beta_2(\gamma(t))$ for all t. The question is how to find an optimal γ for an arbitrary β_1 and β_2? Keep in mind that the space of all γs is infinite dimensional because it is a space of functions.

As described in [30], this registration is accomplished by solving an optimizing problem using the dynamic programming algorithm, but with an objective function that is developed from a Riemannian metric. The chosen metric, termed an *elastic metric*, has a special property that the same reparameterization of two curves does not change the distance between them. This in turn enables us to fix the parameterization of one curve arbitrarily and to optimize over the parameterization of the other. This optimization leads to a proper distance (geodesic distance) and an optimal deformation (geodesic) between the shapes of curves. In other words, it results in their elastic comparisons. Please refer to [30] for details.

2.3 Automated Extraction of Radial Curves

Each facial surface is represented by an indexed collection of radial curves that are defined and extracted as follows: Let S be a facial surface obtained as an output of the preprocessing step. The reference curve on S is chosen to be the vertical curve after the face has been rotated to the upright position. Then, a radial curve β_α is obtained by slicing the facial surface by a plane P_α that has the nose tip as its origin and makes an angle α with the plane containing the reference curve. That is, the intersection of P_α with S gives the radial curve β_α. We repeat this step to extract radial curves from S at equally separated angles, resulting in a set of curves that are indexed by the angle α. Fig. 6 shows an example of this process.

If needed, we can approximately reconstruct S from these radial curves according to $S \approx \cup_\alpha \beta_\alpha = \cup_\alpha(S \cap P_\alpha)$. In the later experiments, we have used 40 curves to represent a surface. Using these curves, we will demonstrate that the elastic framework is well suited to modeling

of deformations associated with changes in facial expressions and for handling missing data.

In our experiments, the probe face is first rigidly aligned to the gallery face using the ICP algorithm. In this step, it is useful but not critical to accurately find the nose tip on the probe face. As long as there is a sufficient number of distinct regions available on the face, this alignment can be performed. Next, after the alignment, the radial curves on the probe model are extracted using the plane P_α passing through the nose tip of the *gallery* model at an angle α with the vertical. This is an important point in that only the nose tip of the gallery and a good alignment between the gallery probe is needed to extract good-quality curves. Even if some parts of the probe face are missing, including its nose region, this process can still be performed. To demonstrate this point, we take session #04042222, from the FRGCv2 dataset, in which some parts of the nose are missing and are filled using a linear interpolation filter (top row of Fig. 7). The leftmost panel shows the hole-filled probe face, the next panel shows the gallery face, the third panel shows its registration with the gallery face, the extracted curves on the probe face. The last panel shows that during alignment of the gallery face with the probe face is good despite a linear interpolation of the missing points. Then, we use the gallery nose coordinates to extract radial curves on the probe surface. The gallery face in this example belongs to the same person under the same expression. In the second row, we show an example where the two faces belong to the same person but represent different expressions/poses. Finally, in the last row we show a case where the probe and the gallery faces belong to different people. Since the curve extraction on the probe face is based on the gallery nose coordinates, which belong to another person, the curves may be shifted in this nose region. However, this small inaccuracy in curve extraction is actually helpful since it increases the interclass distances and improves the biometric performance.
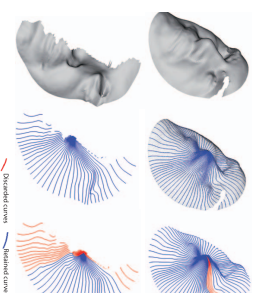

Fig. 7. Curves extraction on a probe face after its rigid alignment with a gallery face. In (a), the nose region of the probe is missing and filled using linear interpolation. The probe and gallery faces are from the same class for (a) and (b), while they are from different classes for (c).

2.4 Curve Quality Filter

In situations involving nonfrontal 3D scans, some curves may be partially hidden due to self-occlusion. The use of these curves in face recognition can severely degrade the recognition performance and therefore they should be identified and discarded. We introduce a quality filter that uses the continuity and the length of a curve to detect such curves. To pass the quality filter, a curve should be one continuous piece and have a certain minimum length, say of 70 mm. The discontinuity or the shortness of a curve results either from missing data or large noise.

We show two examples of this idea in Fig. 8, where we display the original scans, the extracted curves, and then the action of the quality filter on these curves. Once the quality filter is applied and the high-quality curves retained, we can perform face recognition procedure using only the remaining curves. That is, the comparison is based only on curves that have passed the quality filter. Let β denotes a facial curve; we define the Boolean function $quality$: $(quality(\beta) = 1)$ if β passes the quality filter and $(quality(\beta) = 0)$ otherwise. Recall that during the preprocessing step, there is a provision for filling holes. Sometimes the missing parts are too large to be faithfully filled using linear interpolation. For this reason, we need the quality filter that will isolate and remove curves associated with those parts.

— Discarded curves
— Retained curves

Fig. 8. Curve quality filter: examples of detection of broken and short curves (in red) and good curves (in blue).

3 SHAPE ANALYSIS OF FACIAL SURFACES

In this section, we will start by summarizing a recent work in elastic shape analysis of curves and extend it to shape analysis of facial surfaces.

3.1 Background on the Shapes of Curves

Let $\beta : I \to \mathbb{R}^3$ represent a parameterized curve on the face, where $I = [0, 1]$. To analyze the shape of β, we shall represent it mathematically using the *square-root velocity function* (SRVF) [30], denoted by

$$q(t) = \frac{\dot{\beta}(t)}{\sqrt{|\dot{\beta}(t)|}};$$

$q(t)$ is a special function of β that simplifies computations under the elastic metric. More precisely, as shown in [30], an elastic metric for comparing shapes of curves becomes the simple $\mathbb{L}^2$-metric under the SRVF representation. (A

similar metric and representation for curves was also developed by Younes et al. [33], but it only applies to planar curves and not to facial curves.) This point is very important as it simplifies the analysis of curves under the elastic metric to the standard functional analysis. Furthermore, under the $\mathbb{L}^2$-metric, the reparameterization group acts by isometries on the manifold of q functions, which is not the case for the original curve β. To elaborate on the last point, let q be the SRVF of a curve β. Then, the SRVF of a reparameterized curve $\beta \circ \gamma$ is given by $\sqrt{\dot{\gamma}}(q \circ \gamma)$. Here $\gamma : I \to I$ is a reparameterization function and let Γ be the set of all such functions. Now, if q_1 and q_2 are SRVFs of two curves β_1 and β_2, respectively, then it is easy to show that under the $\mathbb{L}^2$ norm, $\|q_1 - q_2\| = \|\sqrt{\dot{\gamma}}(q_1 \circ \gamma) - \sqrt{\dot{\gamma}}(q_2 \circ \gamma)\|$ for all $\gamma \in \Gamma$, while $\|\beta_1 - \beta_2\| \neq \|(\beta_1 \circ \gamma) - (\beta_2 \circ \gamma)\|$ in general. This is one more reason why SRVF is a better representation of curves than β for shape analysis.

Define the preshape space of such curves: $C = \{q : I \to \mathbb{R}^3 | \|q\| = 1\} \subset \mathbb{L}^2(I, \mathbb{R}^3)$, where $\|\cdot\|$ implies the $\mathbb{L}^2$ norm. With the $\mathbb{L}^2$ metric on its tangent spaces, C becomes a Riemannian manifold. Also, since the elements of C have a unit $\mathbb{L}^2$ norm, C is a hypersphere in the Hilbert space $\mathbb{L}^2(I, \mathbb{R}^3)$. Furthermore, the geodesic path between any two points $q_1, q_2 \in C$ is given by the great circle,

$$\psi(\tau) = \frac{1}{\sin(\theta)}(\sin((1-\tau)\theta)q_1 + \sin(\theta\tau)q_2),$$ (1)

and the geodesic length is $\theta = d_c(q_1, q_2) = cos^{-1}(\langle q_1, q_2 \rangle)$.

To study *shapes* of curves, one should identify all rotations and reparameterizations of a curve as an equivalence class. Define the equivalent class of q as

$$[q] = closure\{\sqrt{\dot{\gamma}(t)}Oq(\gamma(t))|O \in SO(3), \ \gamma \in \Gamma\}.$$

The set of such equivalence classes, denoted by $S = \{[q]|q \in C\}$, is called the *shape space* of open curves in $\mathbb{R}^3$. As described in [30], S is a metric space with the metric inherited from the larger space C. To obtain geodesics and geodesic distances between elements of S, one needs to solve the optimization problem:

$$(O^*, \gamma^*)^* = \underset{(O, \gamma) \in SO(3) \times \Gamma}{\text{argmin}} \ d_c(q_1, \sqrt{\dot{\gamma}}Oq_2 \circ \gamma).$$ (2)

For a fixed O in $SO(3)$, the optimization over Γ is done using the dynamic programming algorithm while, for a fixed $\gamma \in \Gamma$, the optimization over $SO(3)$ is performed using SVD. By iterating between these two, we can reach a solution for the joint optimization problem. Let

$$q_2^*(t) = \sqrt{\dot{\gamma}^*(t)}O^*q_2(\gamma^*(t))$$

be the optimal element of $[q_2]$ associated with the optimal rotation O^* and reparameterization γ^* of the second curve; then the geodesic distance between $[q_1]$ and $[q_2]$ in S is $d_s([q_1], [q_2]) = d_c(q_1, q_2^*)$ and the geodesic is given by (1), with q_2 replaced by q_2^*.

3.2 Shape Metric for Facial Surfaces

Now, we extend the framework from radial curves to full facial surfaces. A facial surface S is represented by an
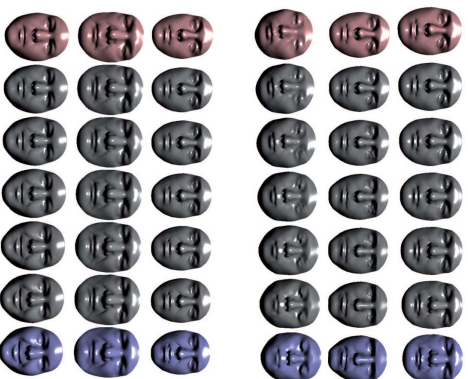
indexed collection of radial curves, indexed by the n uniform angles $A = \{0, \frac{2\pi}{n}, \frac{4\pi}{n}, \ldots, 2\pi\frac{(n-1)}{n}\}$. Thus, the shape of a facial surface can been represented as an element of the set S^n. The indexing provides a correspondence between curves across faces. For example, the curve at an angle α on a probe face is compared with the curve at the same angle α on a gallery face. Thus, the distance between two facial surfaces is $d_S : S^n \times S^n \to \mathbb{R}_{\geq 0}$ given by $d_S(S^1, S^2) = \frac{1}{n}\sum_{\alpha \in A} d_s([q_\alpha^1], [q_\alpha^2])$. Here, q_α denotes the SRVF of the radial curve β_α^i on the ith facial surface. The distance d_S is computed by the following algorithm.

Since we have deformations (geodesic paths) between corresponding curves, we can combine these deformations to obtain deformations between full facial surfaces. In fact, these full deformations can be shown to be formal geodesic paths between faces when represented as elements of S^n. Shown in Fig. 9 are examples of some geodesic paths between source and target faces. The three top rows illustrate paths between faces of different subjects and are termed *interclass geodesics*, whereas the remaining rows illustrate paths between faces of the same person conveying different expressions and are termed *intraclass geodesics*.

These geodesics provide a tangible benefit beyond the current algorithms that provide some kind of a similarity score for analyzing faces. In addition to their interpretation as optimal deformations under the chosen metric, the geodesics can also be used for computing the mean shape and measuring the shape covariance of a set of faces, as illustrated later. To demonstrate the quality of this deformation, we compare it qualitatively for faces with the deformation obtained using a linear interpolation between registered points under an ICP registration of points in Fig. 10. The three rows show, respectively, a

3.3 Computation of the Mean Shape

As mentioned above, an important advantage of our Riemannian approach over many past papers is its ability to compute summary statistics of a set of faces. For example, one can use the notion of Karcher mean [14] to define an average face that can serve as a representative face of a group of faces. To calculate a Karcher mean of facial surfaces $\{S^1, \ldots, S^k\}$ in S^n, we define an objective function: $\mathcal{V} : S^n \to \mathbb{R}$, $\mathcal{V}(S) = \sum_{i=1}^k d_S(S^i, S)^2$. The Karcher mean is then defined by $\bar{S} = \arg\min_{S \in S^n} \mathcal{V}(S)$. The algorithm for computing the Karcher mean is a standard one,

Fig. 9. Examples of geodesics in the shape space. The top three rows illustrate examples of interclass geodesics and the bottom three rows illustrate intraclass geodesics.

Fig. 10. Examples of geodesics in shape space (top row), preshape space (middle row), and a linearly interpolated path after ICP alignment (bottom row).

Algorithm 1. Elastic distance computation.

Input: Facial surfaces S_1 and S_2.
Output: The distance d_S.
for $i \leftarrow 1$ to 2 **do**
  Extract the curve β_α^i;
for $\alpha \leftarrow 0$ to 2Π **do**
  if $quality(\beta_\alpha^1) = 1$ and $quality(\beta_\alpha^2) = 1$ **then**
    Compute the optimal rotation and
    re-parameterization alignment O_α^* and
    γ_α^* using Eqn. 2.
    set $q_\alpha^{2*}(t) = \sqrt{\dot{\gamma}_\alpha^*(t)} O_\alpha^* q_\alpha^2(\gamma_\alpha^*(t))$.
    compute
    $d_s([q_\alpha^1], [q_\alpha^2]) = \cos^{-1}(\langle q_\alpha^1, q_\alpha^{2*} \rangle)$.
  end
end
Compute $d_S = \frac{1}{n}\sum_{\alpha \in A} d_s([q_\alpha^1], [q_\alpha^2])$, where n is the number of valid pairs of curves.
end

geodesic path in the shape space, the corresponding path in the preshape space, and a path using ICP. Algorithm 1 is used to calculate the geodesic path in the shape space. In other words, the optimal matching (reparameterization) between curves is established and thus anatomical points are well matched across the two surfaces. The upper lips match the upper lips, for instance, and this helps produce a natural opening of the mouth, as illustrated in the top row in Fig. 10. However, the optimal matching is not established yet when the geodesic is calculated in the preshape space. This results in an unnatural deformation along the geodesic in the mouth area.

see, for example, [8], and is not repeated here to save space. This minimizer may not be unique and, in practice, one can pick any one of those solutions as the mean face. This mean has a nice geometrical interpretation: $\bar{S}$ is an element of S^n that has the smallest total (squared) deformation from all given facial surfaces $\{S^1, \ldots, S^k\}$. An example of a Karcher mean face for eight faces belonging to different people is shown in Fig. 11.

3.4 Completion of Partially Obscured Curves

Earlier we introduced a filtering step that finds and removes curves with missing parts. Although this step is effective in handling some missing parts, it may not be sufficient when parts of a face are missing due to external occlusions, such as glasses and hair. In the case of external occlusions, the majority of radial curves could have hidden parts that should be predicted before using these curves. This problem is more challenging than self-occlusion because, in addition to detecting the occluding object(s) in parts of the occluding object(s) in the scan. In a non-cooperative situation, where the acquisition is unconstrained, there is a high probability for this kind of occlusion to occur. Once we detect points that belong to the face and points that belong to the occluding object, we first remove the occluding object and use a statistical model in the shape space of radial curves to complete the broken curves. This replaces the parts of face that have been occluded using information from the visible part and the training data.

The core of this problem, in our representation of facial surfaces by curves, is to take a partial facial curve and predict its completion. The sources of information available for this prediction are as follows: 1) the current (partially observed) curve and 2) several (complete) training curves at the same angle that are extracted from full faces. The basic idea is to develop a sparse model for the curve from the training curves and use that to complete the observed curve. To keep the model simple, we use the PCA of the training data in an appropriate vector space to form an orthogonal basis representing training shapes. Then, this basis is used to estimate the coefficients of the observed curve and the coefficients help reconstruct the full curve. Since the shape space of curve S is a nonlinear space, we use the tangent space $T_\mu(S)$, where μ is the mean of the training shapes, to perform PCA. Let α denote the angular index of the occluded curve, and let $q_o, q_o^1, \ldots, q_o^k$ be the SRVFs of the curves taken from the training faces at that angle. As described earlier, we can compute the sample Karcher mean of their shapes $\{[q_o^i] \in S\}$, denoted by μ_o. Then, using the

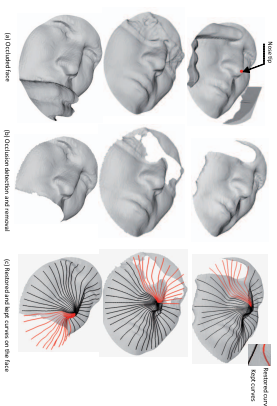Fig. 11. The Karcher mean of eight faces (left) is shown on the right.

Fig. 12. (a) Faces with external occlusion, (b) faces after the detection and removal of occluding parts, and (c) the estimation of the occluded parts using a statistical model on the shape spaces of curves.

geometry of S we can map these training shapes in the tangent space using the inverse exponential map. We obtain $v_{i,\alpha} = \exp^{-1}_{\mu_\alpha}([q_\alpha^i])$, where

$$\exp^{-1}_{q_1}([q_2]) = \frac{\theta}{\sin(\theta)}\left(q_2^* - \cos(\theta)q_1\right), \quad \theta = \cos^{-1}(\langle q_1, q_2^* \rangle),$$

and where q_2^* is the optimal rotation and reparameterization of q_2 to be aligned with q_1, as discussed earlier. A PCA of the tangent vectors $\{v_i\}$ leads to the principal basis vectors $u_{1,\alpha}, u_{2,\alpha}, \ldots, u_{J,\alpha}$, where J represents the number of significant basis elements.

Now, returning to the problem of completing a partially occluded curve, let us assume that this curve is observed for parameter value t in $[0, \tau]$ in $[0, 1]$. In other words, the SRVF of this curve $q(t)$ is known for $t \in [0, \tau]$ and unknown for $t > \tau$. Then, we can estimate the coefficients of q under the chosen basis according to $c_{j,\alpha} = \langle q_1, u_{j,\alpha} \rangle \approx \int_0^\tau q(t), u_{j,\alpha}(t))dt$, and estimate the SRVF of the full curve according to

$$\tilde{q}_\alpha(t) = \sum_{j=1}^J c_{j,\alpha} u_{j,\alpha}(t), \quad t \in [0, 1].$$

We present three examples of this procedure in Fig. 12, with each face corrupted by an external occlusion as shown in column (a). The detection and removal of occluded parts is performed as described in the previous section, and the result of that step is shown in column (b). Finally, the curves passing through the missing parts are restored and shown in column (c).

To evaluate this reconstruction step, we have compared the restored surface (shown in the top row of Fig. 12) with the complete neutral face of that class, as shown in Fig. 13. The small values of both absolute deviation and signed deviation between the restored face and the corresponding face in the gallery demonstrate the success of the restoration process.

In the remainder of this paper, we will apply this comprehensive framework for 3D face recognition using a variety of well-known and challenging datasets. These databases have different characteristics and challenges, and together they facilitate an exhaustive evaluation of a 3D face recognition method.
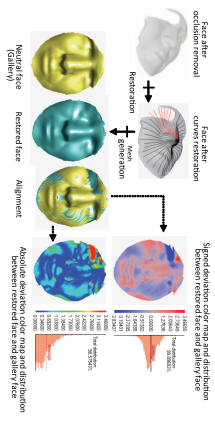
Fig. 13. Illustration of a face with missing data (after occlusion removal) and its restoration. The deviation between the restored face and the corresponding neutral face is also illustrated.
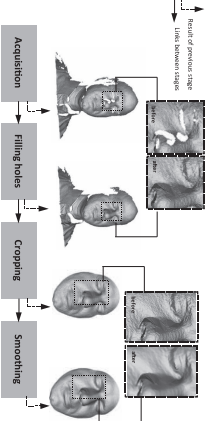
Fig. 14. The different steps of preprocessing: acquisition, filling holes, cropping, and smoothing.

4 EXPERIMENTAL RESULTS

In the following, we provide a comparative performance analysis of our method with other state-of-the-art solutions, using three datasets: the FRGC v2.0 dataset, the GavabDB, and the Bosphorus dataset.

4.1 Data Preprocessing

Since the raw data contains a number of imperfections, such as holes, spikes, and includes some undesired parts, such as clothes, neck, ears, and hair, the data preprocessing step is very important and nontrivial. As illustrated in Fig. 14, this step includes the following items:

- *The hole-filling filter* identifies and fills holes in input meshes. The holes are created either because of the absorption of the laser in dark areas, such as eyebrows and mustaches, or self-occlusion or open mouths. They are identified in the input mesh by locating boundary edges, linking them together into loops, and then triangulating the resulting loops.

- *A cropping filter* cuts and returns parts of the mesh inside a euclidean sphere of radius 75 mm centered at the nose tip to discard as much hair as possible. The nose tip is automatically detected for frontal scans and manually annotated for scans with occlusions and large pose variation.
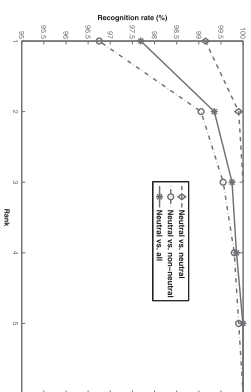
- *A smoothing filter* reduces high-frequency components (spikes) in the mesh, improves the shapes of cells, and evenly distributes the vertices on a facial mesh.

We have used functions provided in the VTK (www.vtk.org) library to develop these filters.

4.2 Comparative Evaluation on the FRGCv2 Dataset

For the first evaluation, we use the FRGCv2 dataset in which the scans have been manually clustered into three categories: neutral expression, small expression, and large expression. The gallery consists of the first scans for each subject in the database, and the remaining scans make up the probe faces. This dataset was automatically preprocessed as described in the Section 4.1. Fig. 15 shows cumulative matching curves (CMCs) of our method under this protocol for the three cases: neutral versus neutral, neutral versus nonneutral, and neutral versus all. Note that this method results in 97.7 percent rank-1 recognition rate in the case of neutral versus all. In the difficult scenario of neutral versus expressions, the rank-1 recognition rate is 96.8 percent, which represents a high performance, while in the simpler case of neutral versus neutral the rate is 99.2 percent.

A comparison of recognition performance of our method with several state-of-the-art results is presented in Table 1. This time, to keep the comparisons fair, we kept all 466 scans in the gallery. Notice that our method achieved a 97 percent rank-1 recognition, which is close to the highest published results on this dataset [29], [26], [9]. Since the scans in FRGCv2 are all frontal, the ability of region-based algorithms such as [9], [26] to deal with the missing data is not tested in this dataset. For that end, one would need a systematic evaluation on a dataset with the missing data issues, for example, the GavabDB. The best recognition score on FRGCv2 is reported by Spreeuwers [29], who use an intrinsic coordi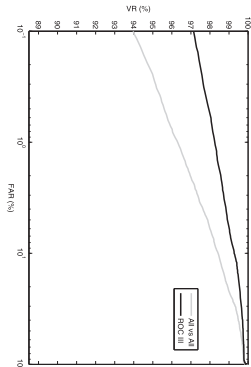nate system based on the vertical symmetry plane through the nose. The missing data due to pose variation and occlusion challenges will be a challenge there as well.

To evaluate the performance of the proposed approach in the verification scenario, the receiver operating characteristic (ROC) curves for the ROC III mask of FRGCv2 and "all-versus-all" are plotted in Fig. 16. For comparison, Table 2 shows the verification results at false acceptance rate (FAR) of 0.1 percent for several methods. For the standard protocol testings, the ROC III mask of FRGC v2, we obtain the verification rates of around 97 percent, which is comparable to the best published results. In the all-versus-all experiment, our method provides 93.96 percent VR at 0.1 percent FAR, which is among the best rates in the table [26], [29], [32]. Note that these approaches are applied to FRGCv2 only. Since scans in FRGCv2 are mostly frontal and have high quality, many methods are able to provide good performance. It is thus important to evaluate a method in other situations where the data quality is not as good. In the next two sections, we will consider those situations with the GavabDB involving the pose variation and the Bosphorus dataset involving the occlusion challenge.

4.3 Evaluation on the GavabDB Dataset

Since GavabDB [21] has many noisy 3D face scans under large facial expressions, we will use that database to help evaluate our framework. This database consists of the Minolta Vi-700 laser range scans from 61 subjects—45 males and 16 females—all of them Caucasian. Each subject was scanned nine times from different angles and under different facial expressions (six with a neutral expression and three with nonneutral expressions). The neutral scans include several frontal scans—one scan while looking up (+35 degrees), one scan from the right side (+90 degrees), and one from the left side (−0 degrees). The nonneutral scans include cases of a smile, a laugh, and an arbitrary expression chosen freely by the subject. We point out that in these experiments the nose tips in profile faces have been annotated manually.

One of the two frontal scans with the neutral expression, for each person is taken as a gallery model, and the remainder are used as probes. Table 3 compares the results of our method with the previously published results following the same protocol. As noted, our approach provides the highest recognition rate for faces with non-neutral expressions (94.54 percent). This robustness comes from the use of radial, elastic curves since: 1) Each curve represents a feature that characterizes local geometry and 2) the elastic matching is able to establish a correspondence with the correct alignment of anatomical facial features across curves.

Fig. 17 illustrates examples of correct and incorrect matches for some probe faces. In each case we show a pair of faces with the probe shown on the right. These pictures also exhibit examples of the variability in facial expressions of the scans

Fig. 15. The CMC curves of our approach for the following scenario: neutral versus neutral, neutral versus expressions, and neutral versus all.

Fig. 16. The ROC curves of our approach for the following scenario: All versus All and the ROC III mask.

Fig. 17. Examples of correct (top row) and incorrect matches (bottom row). For each pair, the probe (on the left) and the ranked-first face from the gallery (on the right) are reported.

TABLE 1
Comparison of Rank-1 Scores on the FRGCv2 Dataset with State-of-the-Art Results

Spreeuwers [29]	Wang et al. [32]	Haar et al. [31]	Berretti et al. [2]	Queirolo et al. [26]	Faltemier et al. [9]	Kakadiaris et al. [13]	Our approach
99%	96.3%	97%	94.1%	98.4%	97.2%	97%	97%

TABLE 2
Comparison of Verification Rates at FAR = 0.1 Percent on the FRGCv2 Dataset with State-of-the-Art Results
(the ROC III Mask and the All versus All Scenario)

Approaches	Kakadiaris et al. [13]	Faltemier et al. [9]	Berretti et al. [2]	Queirolo et al. [26]	Spreeuwers [29]	Wang et al. [32]	Our approach
ROC III	97%	94.8%	81.2%	96.6%	94.6%	98.4%	97.14%
All vs. All	-	93.2%	85.3%	96.5%	94.6%	98.13%	93.96%

TABLE 3
Recognition Results Comparison of the Different Methods on the GavabDB

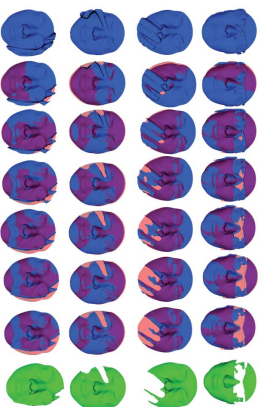|	Lee et al. [16]	Moreno et al. [22]	Mahoor et al. [18]	Berretti et al. [2]	Mousavi et al. [23]	Our method
Neutral	96.67%	90.16%	78%	-	-	100%
Expressive	93.33%	77.9%	72%	81.2%	91%	94.54%
Neutral+expressive	94.68%	-	78%	85.3%	-	95.9%
Rotated looking down	-	-	-	-	-	100%
Rotated looking up	-	-	-	-	-	91%
Overall	-	-	85.3%	88.6%	-	98.36%
Scans from right side	-	-	-	98%	81.67%	96.99%
Scans from left side	-	-	-	-	-	70.49%
	-	-	-	-	-	86.89%

Fig. 18. Gradual removal of occluding parts in a face scan using recursive-ICP.

included in the probe dataset. As far as faces with the neutral expressions are concerned, the recognition accuracy naturally depends on their pose. The performance decreases for scans from the left or right sides because more parts are occluded in those scans. However, for pose variations up to 35 degrees, the performance is still high (100 percent for looking up and 98.36 percent for looking down). Fig. 17 (top row) shows examples of successful matches for up and down looking faces and unsuccessful matches for sideways scans.

Table 3 provides an exhaustive summary of results obtained using GavabDB; our method outperforms the majority of other approaches in terms of the recognition rate. Note that there is no prior result in the literature on 3D face recognition using sideway-scans from this database. Although our method works well on common faces with a range of pose variations within 35 degrees, it can potentially fail when a large part of the nose is missing as it can cause an incorrect alignment between the probe and the gallery. This situation occurs if the face is partially occluded by external objects such as glasses, hair, and so on. To solve this problem, we first restore the data missing due to occlusion.

4.4 3D Face Recognition on the Bosphorus Dataset: Recognition Under External Occlusion

In this section, we will use components I (occlusion detection and removal) and II (missing data restoration) in the algorithm. The first problem we encounter in externally occluded faces is the detection of the external object parts. We accomplish this by comparing the given scan with a template scan, where a template scan is developed using an average of training scans that are complete, frontal, and have neutral expressions. The basic matching procedure between a template and a given scan is recursive ICP, which is implemented as follows: In each iteration, we match the current face scan with the template using ICP and remove those points on the scan that are more than a certain threshold away from the corresponding points on the template. This threshold has been determined using experimentation and is fixed for all faces. In each iteration, additional points that are considered extraneous are incrementally removed and the alignment (with the template) based on the remaining points is further refined.

Fig. 18 shows an example of this implementation. From left to right, each face shows an increasing alignment of the test face with the template, with the aligned parts shown in magenta, and also an increasing set of points labeled as extraneous, drawn in pink. The final result, the original scan minus the extraneous parts, is shown in green at the end.

In the case of faces with external occlusion, we first restore them and then apply the recognition procedure. That is, we detect and remove the occluded part and recover the missing part, resulting in a full face that can be compared with a gallery face using the metric d_S. The recovery is performed using the tangent PCA analysis and Gaussian models, as described in Section 3.4. To evaluate our approach, we perform this automatic procedure on the Bosphorus database [1]. We point out that for this dataset the nose tip coordinates are already provided. The Bosphorus database is suitable for this evaluation as it contains scans of 60 men and 45 women, 105 subjects in total, in various poses, expressions, and in the presence of external occlusions (eyeglasses, hand, hair). The majority of the subjects are aged between 25 and 35. The number of total face scans is 4,652; at least 54 scans each are available for most of the subjects, while there are only 31 scans each for 34 of them. The interesting part is that for each subject there are four scans with occluded parts. These occlusions refer to

1. mouth occlusion by hand,
2. eyeglasses,
3. occlusion of the face with hair, and
4. occlusion of the left eye and forehead regions by hands.

Fig. 19 shows sample images from the Bosphorus 3D database illustrating a full scan on the left and the remaining scans with typical occlusions.

We pursued the same evaluation protocol used in previously published papers: A neutral scan for each person is taken to form a gallery dataset of size 105 and the probe set contains 381 scans that have occlusions. The training is performed using other sessions so that the training and test data are disjoint. The rank-1 recognition rate is reported in Fig. 20 for different approaches depending upon the type of occlusion. As these results show, the process of restoring occluded parts significantly increases the accuracy of recognition. The rank-1 recognition rate is 78.63 percent when we remove the occluded parts and apply the recognition algorithm using the remaining parts, as described in Section 2.4. However, if we perform restoration, the recognition rate is improved to 87.06 percent. Clearly, this improvement in performance is due to the estimation of missing parts on curves. These
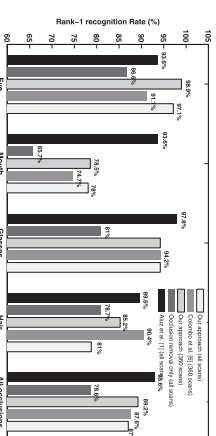

Fig. 19. Examples of faces from the Bosphorus database. The unoccluded face on the left and the different types of occlusions are illustrated.

parts, which include important shape data, were not considered by the algorithm described earlier. Even if the part added with restoration introduces some error, it still allows us to use the shapes of the partially observed curves. Furthermore, during restoration, the shape of the partially observed curve is conserved as much as possible.

Examples of 3D faces recognized by our approach are shown in Fig. 12, along with different steps of the algorithm. The faces in the two bottom rows are examples of incorrectly recognized faces by our algorithm without restoration (as described earlier), but after the restoration step, they are correctly recognized. Aluz et al. [1] reported a 93.69 percent rank-1 recognition rate overall for this database using the same protocol that we have described above. While this reported performance is very good, their processing has some manual components. Actually, the authors partition the face manually and fuse the scores for matching different parts of the face together. To compare with Colombo et al. [6], we reduce the probe dataset to 360 by discarding bad quality scans, as Colombo et al. [6] did. Our method outperforms their approach with an overall performance of 89.25 percent, although individually our performance is worse in the case of occlusion by hair. It is difficult in this case to completely overcome face occlusion. Therefore, during the restoration step, our algorithm tries to keep the majority of parts. This leads to a deformation in the shape of curves and hence affects the recognition accuracy. We present some examples of unrecognized faces in the case of occlusion by hair in Fig. 21. In this instance, the removal of curves passing through occlusion is better than restoring them, as illustrated in Fig. 20.

Fig. 20. Recognition results on the Bosphorus database and comparison with state-of-the-art approaches.

5 DISCUSSION

To study the performance of the proposed approach in the presence of different challenges, we have presented experimental results using three well-known 3D face databases. We have obtained competitive results relative to the state of the art for 3D face recognition in the presence of large expressions, nonfrontal views, and occlusions. As listed in Table 4, our fully automatic results obtained on the FRGCv2 are near the top. Table 4 also reports the computational time of our approach and some state-of-the-art methods on the FRGCv2 dataset. For each approach, we report the time needed for preprocessing and/or feature extraction in the first column. In the second column, we report the time needed to compare two faces. The third column is the sum of the previous computation times for each approach. In the last column, we report the accuracy (recognition rate on FRGCv2) of different approaches. Regarding computational efficiency, parallel techniques can also be exploited to improve performance of our approach since the computation of curve distances, preprocessing, and so on, are independent tasks.

In the case of GavabDB and Bosphorus, the nose tip was manually annotated for nonfrontal and occluded faces. In the future, we hope to develop automatic nose tip detection methods for nonfrontal views and for faces that have undergone occlusion.

6 CONCLUSION

In this paper, we have presented a framework for statistical shape analysis of facial surfaces. We have also presented results on 3D face recognition designed to handle variations of facial expression, pose variations, and occlusions between gallery and probe scans. This method has several properties that make it appropriate for 3D face recognition in noncooperative scenarios. First, to handle pose variation and missing data, we have proposed a local representation by using a curve representation of a 3D face and a quality filter for selecting curves. Second, to handle variations in facial expressions, we have proposed an elastic shape analysis of 3D faces. Finally, in the presence of occlusion, we have proposed to remove the occluded parts, then to recover only the missing data on the 3D scan using statistical shape models. That is, we have constructed a low-dimensional shape subspace for each element of the indexed collection of curves, and then represent a curve (with missing data) as a linear combination of its basis elements.
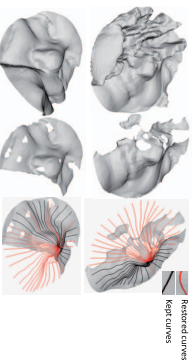

Fig. 21. Examples of nonrecognized faces. Each row illustrates, from left to right, the occluded face, the result of occlusion removal, and the result of restoration.

TABLE 4
Comparative Study of Time Implementations and Recognition Accuracy on FRGCv2 of the Proposed Approach with State of the Art

Approach	preprocessing (s)	Face matching (s)	Comparison time (s)	Accuracy (%)
Wang et al. [32]	1.48	0.65	2.2	98.3%
Spreeuwers [29]	2.5	1/11	2.5	99%
This work	6.18	1.27	7.45	97%
Faltemier et al. [9]	7.52	2.4	9.92	97.2%
Kakadiaris et al. [13]	15	1/1000	15	97%
Hair et al. [31]	3	-	18	97%
Berretti et al. [2]	-	15	-	94.1%
Queirolo et al. [26]	-	4	-	98.4%

REFERENCES

[1] N. Alyuz, B. Gokberk, and L. Akarun, "A 3D Face Recognition System for Expression and Occlusion Invariance," *Proc. Second IEEE Int'l Conf. Biometrics: Theory, Applications and Systems,* 2008.

[2] S. Berretti, A. Del Bimbo, and P. Pala, "3D Face Recognition Using Isogeodesic Stripes," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 32, no. 12, pp. 2162-2177, Dec. 2010.

[3] K.W.K. Bowyer, K. Chang, and P. Flynn, "A Survey of Approaches and Challenges in 3D and Multi-Modal 3D + 2D Face Recognition," *Computer Vision and Image Understanding,* vol. 101, no. 1, pp. 1-15, 2006.

[4] A.M.A. Bronstein, M.M.M. Bronstein, and R. Kimmel, "Three-Dimensional Face Recognition," *Int'l J. Computer Vision,* vol. 64, no. 1, pp. 5-30, 2005.

[5] A.M.A. Bronstein, M.M.M. Bronstein, and R. Kimmel, "Expression-Invariant Representations of Faces," *IEEE Trans. Image Processing,* vol. 16, no. 1, pp. 188-197, Jan. 2007.

[6] A. Colombo, C. Cusano, and R. Schettini, "Three-Dimensional Occlusion Detection and Restoration of Partially Occluded Faces," *J. Math. Imaging and Vision,* vol. 40, no. 1, pp. 105-119, 2011.

[7] H. Drira, B. Ben Amor, M. Daoudi, and A. Srivastava, "Pose and Expression-Invariant 3D Face Recognition Using Elastic Radial Curves," *Proc. British Machine Vision Conf.,* pp. 1-11, 2010.

[8] H. Drira, B. Ben Amor, A. Srivastava, and M. Daoudi, "A Riemannian Analysis of 3D Nose Shapes for Partial Human Biometrics," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 2050-2057, 2009.

[9] T.C.T. Faltemier, K.W.K. Bowyer, and P.J.P. Flynn, "A Region Ensemble for 3D Face Recognition," *IEEE Trans. Information Forensics and Security,* vol. 3, no. 1, pp. 62-73, Mar. 2008.

[10] G. Gordan, "Face Recognition Based on Depth and Curvature Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 108-110, 1992.

[11] S. Gupta, J.K. Aggarwal, M.K. Markey, and A.C. Bovik, "3D Face Recognition Founded on the Structural Diversity of Human Faces," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2007.

[12] D. Huang, G. Zhang, M. Ardabilian, Wang, and L. Chen, "3D Face Recognition Using Distinctiveness Enhanced Facial Representations and Local Feature Hybrid Matching," *Proc. Fourth IEEE Int'l Conf. Biometrics: Theory, Applications and Systems,* pp. 1-7, 2010.

[13] I.A.I. Kakadiaris, G. Passalis, G. Toderici, M.N.M. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-Dimensional Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 4, pp. 640-649, Apr. 2007.

[14] H. Karcher, "Riemannian Center of Mass and Mollifier Smoothing," *Comm. Pure and Applied Math.,* vol. 30, pp. 509-541, 1977.

[15] Y. Lee, H. Song, U. Yang, H. Shin, and K. Sohn, "Local Feature Based 3D Face Recognition," *Proc. Audio- and Video-Based Biometric Person Authentication,* pp. 909-918, 2005.

[16] X. Li, T. Jia, and H. Zhang, "Expression-Insensitive 3D Face Recognition Using Sparse Representation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 2575-2582, 2009.

[17] X. Lu and A. Jain, "Deformation Modeling for Robust 3D Face Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 30, no. 8, pp. 1346-1357, Aug. 2008.

[18] M.H.M. Mahoor and M. Abdel-Mottaleb, "Face Recognition Based on 3D Ridge Images Obtained from Range Data," *Pattern Recognition,* vol. 42, no. 3, pp. 445-451, 2009.

[19] R. McKeon and T. Russ, "Employing Region Ensembles in a Statistical Learning Framework for Robust 3D Facial Recognition," *Proc. Fourth IEEE Int'l Conf. Biometrics: Theory Applications and Systems,* pp. 1-7, 2010.

[20] A. Moorthy, A. Mittal, S. Jahanbin, K. Grauman, and A. Bovik, "3D Facial Similarity: Automatic Assessment versus Perceptual Judgments," *Proc. Fourth IEEE Int'l Conf. Biometrics: Theory Applications and Systems,* pp. 1-7, 2010.

[21] A.B.A. Moreno and A.A. Sanchez, "GavabDB: A 3D Face Database," *Proc. Workshop Biometrics on the Internet,* pp. 77-85, 2004.

[22] A.B. Moreno, A. Sanchez, J.F. Velez, and F.J. Diaz, "Face Recognition Using 3D Local Geometrical Features: PCA versus SVM," *Proc. Int'l Symp. Image and Signal Processing and Analysis,* 2005.

[23] M.H.M. Mousavi, K. Faez, and A. Asghari, "Three Dimensional Face Recognition Using SVM Classifier," *Proc. Seventh IEEE/ACIS Int'l Conf. Computer and Information Science,* pp. 208-213, 2008.

[24] I. Mpiperis, S. Malassiotis, and M.G.M. Strintzis, "3D Face Recognition with the Geodesic Polar Representation," *IEEE Trans. Information Forensics and Security,* vol. 2, no. 3, pp. 537-547, Sept. 2007.

[25] G. Passalis, P. Perakis, T. Theoharis, and I.A.I. Kakadiaris, "Using Facial Symmetry to Handle Pose Variations in Real-World 3D Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 33, no. 10, pp. 1938-1951, Oct. 2011.

[26] C.C.C. Querelo, L. Silva, O.R.O. Bellon, and M.P.M. Segundo, "3D Face Recognition Using Simulated Annealing and the Surface Interpenetration Measure," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 32, no. 2, pp. 206-219, Feb. 2010.

[27] C. Samir, A. Srivastava, and M. Daoudi, "Three-Dimensional Face Recognition Using Shapes of Facial Curves," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 11, pp. 1858-1863, Nov. 2006.

[28] C. Samir, A. Srivastava, M. Daoudi, and E. Klassen, "An Intrinsic Framework for Analysis of Facial Surfaces," *Int'l J. Computer Vision,* vol. 82, no. 1, pp. 80-95, 2009.

[29] L. Spreeuwers, "Fast and Accurate 3D Face Recognition Using Registration to an Intrinsic Coordinate System and Fusion of Multiple Region Classifiers," *Int'l J. Computer Vision,* vol. 93, no. 3, pp. 389-414, 2011.

[30] A. Srivastava, E. Klassen, S.H.S. Joshi, and I.H.I. Jermyn, "Shape Analysis of Elastic Curves in Euclidean Spaces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 33, no. 7, pp. 1415-1428, July 2011.

[31] F. ter Haar and R.C.R. Veltkamp, "Expression Modeling for Expression-Invariant Face Recognition," *Computers and Graphics,* vol. 34, no. 3, pp. 231-241, 2010.

[32] Y. Wang, J. Liu, and X. Tang, "Robust 3D Face Recognition by Local Shape Difference Boosting," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 32, no. 10, pp. 1858-1870, Oct. 2010.

[33] L. Younes, P.W.P. Michor, J. Shah, and D. Mumford, "A Metric on Shape Space with Explicit Geodesics," *Rendiconti Lincei-Matematica e Applicazioni,* vol. 9, pp. 25-57, 2008.

Hassen Drira received the engineering and MSc degrees in computer science from the National School of Computer Science, Manouba, Tunisia, in 2006 and 2007, respectively, and the PhD degree in computer science from the University of Lille 1, France, in 2011. He has been an assistant professor of computer science at the Institut Mines-Télécom/Télécom Lille1, LIFL UMR (CNRS 8022) since September 2012. He spent the year 2011-2012 in the MIIRE research group within the Fundamental Computer Science Laboratory of Lille as a postdoctoral researcher. His research interests include pattern recognition, statistical analysis, 3D face recognition, biometrics, and, more recently, 3D facial expression recognition. He has published several refereed journal and conference articles in these areas.

Boulbaba Ben Amor received the engineer degree in computer science from ENIS, Tunisia, in 2002, and the MS and PhD degrees in computer science both from the Ecole Centrale de Lyon, France, in 2003 and 2006, respectively. He joined the Mines-Télécom/Télécom Lille1 Institute as an associate professor in 2007. Since then, he has also been a member of the Computer Science Laboratory at the University Lille 1 (LIFL UMR CNRS 8022). His research interests include statistical 3D face analysis and recognition and facial expression recognition using 3D. He is a coauthor of several papers in refereed journals and proceedings of international conferences. He has been involved in French and international projects and has served as a program committee member and reviewer for international journals and conferences. He is a member of the IEEE.

Anuj Srivastava received the MS and PhD degrees in electrical engineering from Washington University, St. Louis, Missouri, in 1993 and 1996, respectively. He is a professor of statistics at Florida State University (FSU) in Tallahassee. After spending the year 1996-1997 at Brown University as a visiting researcher, he joined FSU as an assistant professor in 1997. His research interests include pattern theoretic approaches to problems in image analysis, computer vision, and signal processing. Specifically, he has developed computational tools for performing statistical inferences on certain nonlinear manifolds and has published more than 200 refereed journal and conference articles in these areas. He is a senior member of the IEEE.

Mohamed Daoudi received the PhD degree in computer engineering from the University of Lille 1, France, in 1993 and the Habilitation à Diriger des Recherches from the University of Littoral, France, in 2000. He is a professor of computer science at TELECOM Lille 1 and LIFL (UMR CNRS 8022). He is the head of the Computer Science Department at Telecom Lille1. He was the founder and the scientific leader of the MIIRE research group http://www-rech.telecom-lille1.eu/miire/. His research interests include pattern analysis and recognition, and 3D face analysis and recognition. He has published more than 100 papers in some of the most distinguished scientific journals and international conferences. He is the coauthor of the book *3D Processing: Compression, Indexing and Watermarking* (Wiley, 2008). He is a senior member of the IEEE.

Rim Slama received the engineering and MSc degrees in computer science from the National School of Computer Science, Manouba, Tunisia, in 2010 and 2011, respectively. She is currently working toward the PhD degree and is a member of the MIIRE research group within the Fundamental Computer Science Laboratory of Lille, France. Her current research interests include human motion analysis, computer vision, pattern recognition, 3D video sequences people, dynamic 3D human body, shape matching, and their applications in computer vision.

▷ **For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.**

Boosting 3-D-Geometric Features for Efficient Face Recognition and Gender Classification

Lahoucine Ballihi, Boulbaba Ben Amor, Mohamed Daoudi, *Senior Member, IEEE*, Anuj Srivastava, *Senior Member, IEEE*, and Driss Aboutajdine, *Senior Member, IEEE*

Abstract—We utilize ideas from two growing but disparate ideas in computer vision—shape analysis using tools from differential geometry and feature selection using machine learning—to select and highlight salient geometrical facial features that contribute most in 3-D face recognition and gender classification. First, a large set of geometries curve features are extracted using level sets (circular curves) and streamlines (radial curves) of the Euclidean distance functions of the facial surface; together they approximate facial surfaces with arbitrarily high accuracy. Then, we use the well-known Adaboost algorithm for feature selection to derive a composite classifier that achieves high performance with a minimal set of features. This greatly reduced set, consisting of some level curves on the nose and some radial curves in the forehead and cheeks regions, provides a very compact signature of a 3-D face and a fast classification algorithm for face recognition and gender selection. It is also efficient in terms of data storage and transmission costs. Experimental results, carried out using the FRGCv2 dataset, yield a rank-1 face recognition rate of 98% and a gender classification rate of 86% rate.

Index Terms—Face recognition, gender classification, geodesic path, facial curves, machine learning, feature selection.

I. INTRODUCTION

SINCE facial biometrics is natural, contact free, nonintrusive, and psychologically supported, it has emerged as a popular modality in the biometrics community. Unfortunately, the technology for 2-D image-based face recognition still faces difficult challenges, such as pose variations, changes in lighting conditions, occlusions, and facial expressions. Due to the robustness of 3-D observations to lighting conditions and pose variations, face recognition using shapes of facial surfaces has become a major research area in the last few years. Many of the

state-of-the-art methods have focused on the variability caused by facial deformations, e.g., those due to face expressions, and have proposed methods that are robust to such shape variations. At the same time, gender classification is emerging as an interesting problem that can be a useful preprocessing step for face recognition. Gender is similar to other soft biometric traits, such as skin color, age, eyes colors, and so on, used by humans to distinguish their peers. Most existing work on gender classification uses 2-D-images to extract distinctive facial features like hair density and inner morphology of the face, but 3-D shape has not yet been extensively used for gender classification. Several works in psychology have shown that gender has close relationships both with 2-D information and 3-D shape [1], [2], and it motivates the use of 3-D shapes for gender classification.

The development of a practical, high-performance system for automatic face recognition and gender classification is an important issue in intelligent systems. In this work, we focus on a feature selection technique from machine-learning that is fully automatic and versatile enough for different applications like face recognition and gender classification. The features comes from different types of facial curves extracted from facial surfaces in an intrinsic fashion, and comparisons of these curve features is based on latest advances in shape analysis of parameterized curves using tools from differential geometry. In the process we also develop an effective approach for tackling facial expressions variation, an important focus of the face recognition grand challenge. Our approach offers the advantage of classifying either facial identity and/or gender, both independent of the ethnicity. Specifically, the main contributions of this paper include:

- A new geometric feature-selection approach for efficient 3-D face recognition that seeks most relevant characteristics for recognition while handling the challenge of facial expressions. In particular, we are interested in finding those facial curves that are most suitable for 3-D face recognition.

- A new gender classification approach using the 3-D face shape represented by collections of curves. In particular, we are interested in finding those facial curves that are most suitable for gender discrimination.

The rest of the paper is organized as follows. Section II summarizes existing approaches on 3-D face recognition with an emphasis on *facial curve-based* and *facial feature-based* methods. It also presents some progress in 3-D imaging-based gender classification. Section III overviews the proposed approach for both the target applications. In Section IV, we present procedures for extracting facial curves. Section V

recalls the main ideas of the Riemannian geometric shape analysis framework to compare and match facial curves. In Section VI, we give formulations to the classification problem and describes the use of the boosting procedure to achieve the feature selection step, for each of the two applications. Experimental evaluations and comparative studies to previous approaches are given in Section VII. We conclude in the paper with a discussion and summary in Section VIII.[1]

II. RELATED WORK

As the proposed approach combines curve-based face comparison with feature selection techniques, we mainly focus on previous methods that primarily use *local* facial feature-selection and *holistic* facial curves.

A. Feature Selection-Based 3-D Face Recognition

Several methods have been proposed to analyze the discriminative power of different facial regions or features for face recognition. Daniyal *et al.* [4] proposed an algorithm in which a face is represented as a vector of distances between pairs of facial landmarks. They selected the landmarks by exhaustive search over possible combinations of used/unused landmarks, comparing the recognition rates, and concluded that the best selection corresponded to the landmarks located around the eyes and the nose. In the 3-D face recognition approach used by Faltemier *et al.* [5], the nose tip and 28 small regions were selected automatically for improving recognition. More recently, Wang *et al.* [6] computed a signed shape difference map (SSDM) between two aligned 3-D faces as a intermediate representation for the shape comparison. Based on the SSDMs, Haar-like, Gabor, and Local Binary Pattern (LBP) were used to encode both the local similarity and the change characteristics between facial shapes. The most discriminative local features were selected optimally by boosting. Using similar features, Li *et al.* [7] proposed to design a feature pooling and ranking scheme to collect various types of low-level geometric features, such as curvatures, and ranked them according to their sensitivities to facial expressions. They applied sparse representations to the collected low-level features and achieved good results on the GAVAB database. In [8] Ocegueda *et al.* proposed a Markov Random Field model for the analysis of latices (e.g., image or 3-D meshes) in terms of the discriminative information of their vertices. They observed that the nose and the eyes are consistently marked as discriminative regions of the face in a face recognition system. Li *et al.* [9] proposed an expression-robust 3-D face recognition approach by learning weighted sparse representation of encoded normal information, which they called multiscale local normal patterns (MS-LNPs). They utilized the learned average quantitative weights related to different facial physical components to enhancing the robustness of their system to expression variations.

B. Curve-Based Face Representation

The basic idea of these approaches is to represent a surface using an indexed family of curves which provide an

approximate representation of the surface. Samir *et al.* [10], for instance, used the level curves of the height function to define facial curves. Since these curves are planar, they used shape analysis of planar curves, taken from [11], to compare and deform faces; nonlinear matching problem was not studied here (that is, the mapping was fixed to be linear). The authors proposed to compare facial surface by using two metrics: Euclidean mean and geometric mean. However, there were no discussion on how to obtain optimal curves. Later in [12], the same authors used the level curves of the geodesic distance function that resulted in 3-D curves. They used a nonelastic metric and a path-straightening method to compute geodesics between these curves. Here also, the matching was not studied and the correspondence of curves and points across faces was simply linear. In [13], Mpiperis *et al.* proposed a geodesic polar parametrization of the facial surface. With this parametrization, the intrinsics attributes do not change under isometric deformation when the mouth is closed. Otherwise, it violates the isometry assumption and thus they adapt their geodesic polar parametrization by disconnecting the lips. Through this representation, they proposed an elastically deformable model algorithm that establishes correspondence among a set of faces. Then, they construct bilinear models that decouple the identity and facial expression factors. The invariance to facial expressions is obtained by fitting these models to unknown faces. The main limitation of this approach is the need for a large set which should also be annotated with respect to facial expressions. In [14], Drira *et al.* explored the use of shapes of noses for performing partial human biometrics. More recently, in [15], the same authors proposed similar shape analysis approach this time using radial curves. They model elastic deformations of facial surfaces (including opening the mouth) as an optimal reparametrization (or matching) problem that they solve using the dynamic programming algorithm. This approach provided promising results on GAVAB database even where the probe pose is nonfrontal. In [16], Berretti *et al.* used the geodesic distance on the face to extract iso-geodesic facial stripes. Equal width iso-geodesic facial stripes were used as nodes of the graph and edges between nodes were labeled with descriptors, referred to as 3-D Weighted Walkthroughs (3DWWs), that captured mutual relative spatial displacement between all the pairs of points of the corresponding stripes. Face partitioning into iso-geodesic stripes and 3DWWs together provided an approximate representation of local morphology of faces that exhibits smooth variations for changes induced by facial expressions. More recently Ballihi *et al.* [3] propose a new curve selection approach for efficient 3-D face recognition.

C. Gender Classification

The human face presents a clear sexual dimorphism that makes face gender classification an extremely efficient and fast cognitive process [17]. Although a significant progress has been made, the task of automated, robust face gender classification is still a distant goal. 2-D image-based methods are inherently limited by variability in imaging factors such as illumination and pose. An emerging solution is to use laser scanners for capturing three-dimensional (3-D) observations of

human faces, and use this data in performing face gender classification. Bruce et al. [1] performed an interesting experiment in which they tested the human visual system that is accurate at deciding whether faces are male or female, even when cues from hairstyle, makeup, and facial hair are minimized. The authors found that subjects were considerably less accurate when asked to judge the sex of 3-D representations of faces obtained by laser-scanning, compared photographs that were taken with hair concealed and eyes closed. They proved that by having a more protuberant nose/brow and more prominent chin/jaw. The effects of manipulating the shapes of the noses and chins of the laser-scanned heads were assessed and significant effects of such manipulations on the apparent masculinity or femininity of the heads were revealed. In O'Toole et al. [2], the authors assumed that the sex of a face is perhaps its most salient feature. They applied principal components analysis (PCA) separately to the three-dimensional structure and gray level image data from laser-scanned human heads. The results showed that the three-dimensional head data supported more accurate sex classification than the gray level image data, across a range of PCA-compressed (dimensionality-reduced) representations of the heads. Jing et al. [18] investigated gender classification based on 2.5D facial surface normals (facial needle-maps) which can be recovered from 2-D intensity images using a weighted principal geodesic analysis (WPGA) method. They described a weighted principal geodesic analysis (WPGA) method to extract features from facial surface normals. They described a non-Lambertian Shape-from-shading (SFS) method to increase the gender discriminating power in the leading eigenvectors. They adopted a Bayesian method for gender classification. Xiaoguang et al. [19] exploited the range information of human faces for ethnicity identification using a Support Vector Machine (SVM). An integration scheme is also proposed for ethnicity and gender identifications by combining the registered range and intensity images. Yuan et al. [20] proposed a fusion-based gender classification method, based on SVM, for 3-D frontal neutral expression facial. A method for fusion of information from four regions (upper region of the face, the lower region of the face, the nose and the left eye) was proposed.

From the above discussion it is clear that majority of current methods on curve-based 3-D face recognition used a holistic representation/parameterization of facial surfaces. In this paper, we consider curves as geometric features that capture local facial shape and we propose to learn the most relevant curves using adaptive boosting. Thus, we propose to represent a facial surface by two types of facial curves, radials and levels, for 3-D face recognition and gender classification. This strategy raises a few issues : (i) How to define curves on facial surfaces?, (ii) How to compare shapes of facial curves?, and (iii) How to select the most relevant curves for 3-D face recognition and gender classification? To address these issues, our strategy includes the following steps:

1) A facial surface representation by collection of curves of level sets (circular curves) and streamlines (radial curves) of a distance function;
2) A geometric shape analysis framework based on Riemannian geometry to compare pairwise facial curves;

3) A boosting method to highlight geometric features according to the target application;
4) A through experimental evaluation that compares the proposed solution with latest methods on a common data set and common experimental settings.

As demonstrated later, the proposed approach achieves highest performance for the face recognition task, with the additional computational advantage of using a compact signature. Furthermore, it is one of the first approaches to address the gender classification problem using 3-D face images. To the best of our knowledge no previous work has proposed a unique framework for 3-D face recognition and gender classification.

III. OVERVIEW OF THE PROPOSED APPROACH

In this work, we combine ideas from shape analysis using tools from differential geometry and feature selection derived from machine learning to select and highlight salient 3-D geometrical facial features. After preprocessing of 3-D scans, we represent resulting facial surfaces by a finite indexed collections of circular and radial curves. The comparison of pairwise curves, extracted from faces, is based on shape analysis of parameterized curves using differential geometry tools. According to the target application, the extracted features are trained as weak classifiers and the most discriminative features are selected optimally by adaptive boosting. For the case of gender recognition, the classification is formulated as a binary problem (Male/Female classes) and we propose to use the inter- and intrapersonal comparisons formulation to achieve feature selection for face identification, which is basically a Multiclass classification problem. Fig. 1 overviews the proposed approach with the target applications, face recognition and gender classification.

- The **Offline training step**, learns the most salient circular and radial curves from the sets of extracted ones, according to each application in a supervised fashion. In face recognition, for instance, construct feature vectors by comparing pairwise curves extracted from facial surfaces. Next, feed these examples, together with labels indicating if they are interclass or not. Thus, the adaptive boosting selects and learns iteratively the weak classifiers and adding them to a final strong classifier, with suitable weights. As a result of this step, we keep the T-earliest selected features for the testing step.

- The **Online test step**, performs classification of a given test face. In the identity recognition problem, a probe face is compared to the gallery faces using only individual scores computed based on selected features which are fused using the arithmetic mean. In the gender recognition problem, a test face is compared to computed templates of Male and Female classes using curves selected for that purpose. The templates are computed, once for all, within the training stage.

IV. 3-D FACIAL CURVES EXTRACTION

Let S be a facial surface denoting the output of a preprocessing step that crops the mesh, fills holes, removes noise, and prepares the mesh for curve extraction. We extract radial curves emanating from a reference anchor point (the tip of the nose)
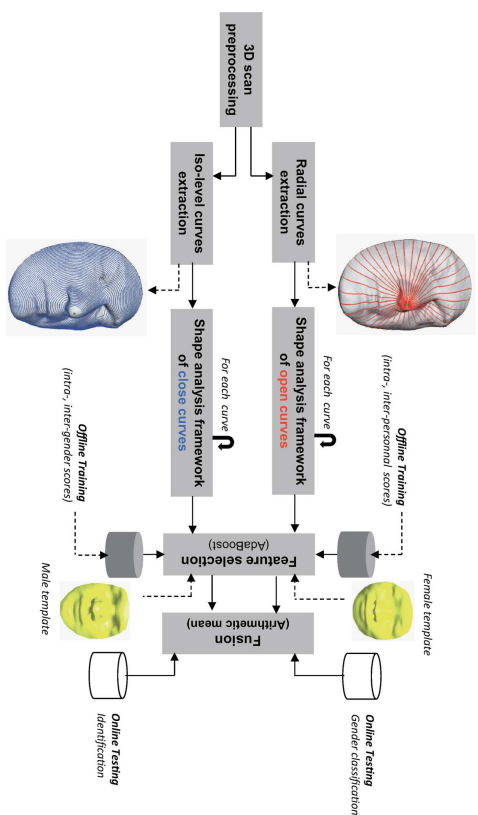
Fig. 1. Overview of the proposed approach, including both stages of training and testing and both target applications: identity recognition and gender classification.
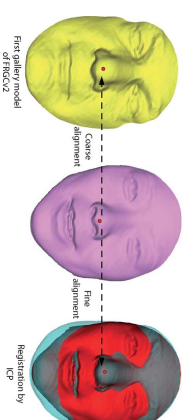
Fig. 2. Probe model pose normalization by registration with the first gallery face of $FRGC$ v2, a coarse alignment is performed by translating the probe face according to the translation vector formed by the tips of the noses. A fine registration is then achieved by the ICP algorithm.

and circular curves having with the same point as the pole, using simple procedures detailed in the following paragraphs.

A. Radial Curves

Let β^r_α denote the radial curve on S which makes an angle α with a reference radial curve; the superscript r denotes that it is a radial curve. The reference curve is chosen to be the vertical curve once the face has been rotated to the upright position. In practice, each radial curve β^r_α is obtained by slicing the facial surface by a plane P_α that has the nose tip as its origin and makes an angle α with the plane containing the reference curve, as shown in Fig. 3. That is, a curve β^r_α is obtained by slicing the facial surface by P_α defined by the angle α with the vertical plane and having as origin the nose tip. We repeat this step to extract radial curves from the facial surface at equal angular separation. Each curve is indexed by the angle α.

To avoid pose variations problem, all probe faces are aligned with the first face model (generic model) of FRGCv2 database. This step is achieved by performing a coarse alignment by translating the probe face to a reference face, using their noses tips. This coarse alignment step is followed by a fine alignment using the ICP algorithm, as illustrated in Fig. 2. The process of curve extraction then follows.

If needed, we can approximately reconstruct S from these radial curves according to $S \approx \cup_\alpha \beta^r_\alpha = \cup_\alpha \{S \cap P_\alpha\}$ as illustrated in Fig. 3. This indexed collection of radial curves captures the shape of a facial surface and forms the first mathematical representation of that surface.

B. Circular Curves

Let β^c_λ denote the circular curve on S which makes a distance λ from the reference point (nose tip). A similar procedure

is employed to extract these curves. The only difference is the slicing function which is now a sphere M_λ having the reference point as center and variable radius λ. The intersection of a given sphere and the facial surface defines equidistant points from the reference point, in the surface. Fig. 4 illustrates results of such extraction procedure. We note that any points ordering is needed for both kind of curves since the slicing procedure kept edges between points. However, a curve subsampling procedure is introduced to achieve the same number of points for all curves (100 points per curve here).

Similarly to radial curves, we can also approximately reconstruct S from these circular curves according to $S \approx \cup_\lambda \beta^c_\lambda = \cup_\lambda \{S \cap M_\lambda\}$ as illustrated in Fig. 4, we describe the geometric framework which allow matching and comparison of curves. Fig. 5 gives some results of facial curves extraction on several 3-D faces.
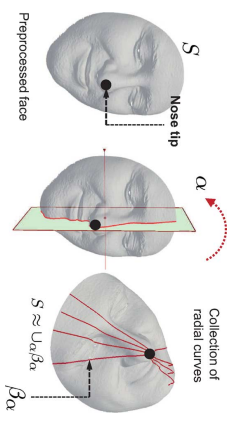
Fig. 3. Procedure for extraction of radial curves, a curve β_α is obtained by slicing the facial surface by P_α, defined by the angle α with the vertical plane and having as origin the nose tip.
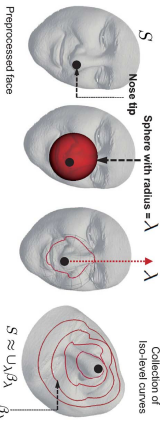
Fig. 4. Procedure for extraction of circular curves, a curve β_λ is obtained by slicing the facial surface by S, defined by the radius λ and having as center the nose tip.
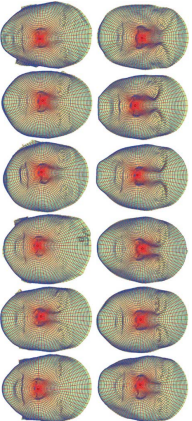
Fig. 5. Examples of facial representation by circular and radial curves. The first row illustrates preprocessed faces of male subjects, the second row gives preprocessed faces of females.
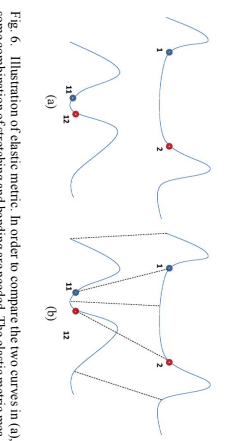
Fig. 6. Illustration of elastic metric. In order to compare the two curves in (a), some combination of stretching and bending are needed. The elastic metric measures the amounts of these deformations. The optimal matching between the two curves is illustrated in (b).

V. GEOMETRIC SHAPE ANALYSIS OF FACIAL CURVES

In the last few years, many approaches have been developed to analyze shapes of 2-D curves. We can cite approaches based on Fourier descriptors, moments or the median axis. More recent works in this area consider a formal definition of shape spaces as a Riemannian manifold of infinite dimension on which they can use the classic tools for statistical analysis. The recent results of Michor and Mumford [21], Klassen et al. [11], and Yezzi and Mennucci [22] show the efficiency of this approach for 2-D curves. Joshi et al. [23], [24] have recently proposed a generalization of this work to the case of curves defined in $\mathbb{R}^n$. We will adopt this work to our problem since our 3-D curves are defined in $\mathbb{R}^3$.

We start by considering a curve β in $\mathbb{R}^3$. While there are several ways to analyze shapes of curves, an elastic analysis of the parameterized curves is particularly appropriate in our application—face analysis under facial expression variations. This is because (1) such analysis uses the square-root velocity function representation which allows us to compare local facial shapes in presence of elastic deformations, (2) this method uses a square-root representation under which the elastic metric reduces to the standard L^2 metric and thus simplifies the analysis, (3) under this metric the Riemannian distance between curves is invariant to the reparametrization. To analyze the shape of β, we shall represent it mathematically using a square-root representation of β as follows; for an interval $I = [0, 1]$, let $\beta : I \longrightarrow \mathbb{R}^3$ be a curve and define $q : I \longrightarrow \mathbb{R}^3$ to be its square-root velocity function (SRVF), given by:

$$q(t) \doteq \frac{\dot{\beta}(t)}{\sqrt{|\dot{\beta}(t)|}}$$ (1)

Here t is a parameter $\in I$ and $|\cdot|$ is the Euclidean norm in $\mathbb{R}^3$. We note that $q(t)$ is a special function that captures the shape of β and is particularly convenient for shape analysis, as we describe next. The classical elastic metric for comparing shapes of curves becomes the L^2-metric under the SRVF representation [24]. This point is very important as it simplifies the calculus of elastic metric to the well-known calculus of functional analysis under the L^2-metric. Also, the squared L^2-norm of q, given by: $\|q\|^2 = \int_I \langle q(t), q(t) \rangle dt = \int_I \|\dot{\beta}(t)\| dt$, is the length of β. If we set $\|q\| = \|\|q\|\| = 1$, implying all curves are rescaled to unit length, then translation and scaling variability have been removed by this mathematical representation of curves.

Consider the two curves in Fig. 6(a), let us fix the parameterization of the top curve to be arc-length, i.e., traverse that curve with a constant speed equal to one. In order to better match that curve with the bottom one, one should know at what rate we are going to move along the bottom curve so that points reached at the same time on two curves are as close as possible under some geometric criterion. In other words, peaks and valleys should be reached at the same time. Fig. 6(b) illustrates the matching where point 1 on the top curve matches to point 11 on the bottom curve. The part between the point 1 and 2 on the top curve shrinks on the second curve. Therefore, the point 2 matches the point 22 on the second curve. An elastic metric is the measure of that shrinking.
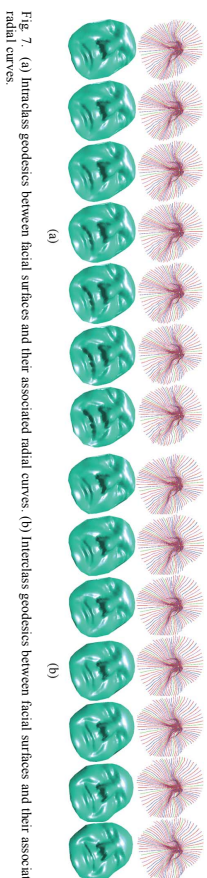
Fig. 7. (a) Intraclass geodesics between facial surfaces and their associated radial curves. (b) Interclass geodesics between facial surfaces and their associated radial curves.

A. Radial Open Curves

The set of all unit-length curves in $\mathbb{R}^3$ is given by $C = \{q : I \longrightarrow \mathbb{R}^3 | \|q\| = 1\} \subset L^2(I, \mathbb{R}^3)$. With the L^2-metric on its tangent spaces, C becomes a Riemannian manifold. Since the elements of C have a unit L^2 norm, C is a hypersphere in the Hilbert space $L^2(I, \mathbb{R}^3)$. In order to compare the shapes of two radial curves, we can compute the distance between them in C under the chosen metric. This distance is found to be the length of the minor arc connecting the two elements in C. Since C is a hypersphere, the formulas for the geodesic and the geodesic length are already well known. The geodesic length between any two points $q_1, q_2 \in C$ is given by:

$$d_c(q_1, q_2) = cos^{-1}(\langle q_1, q_2 \rangle),$$ (2)

and the geodesic path $\alpha : [0, 1] \longrightarrow C$, is given by:

$$\alpha(\tau) = \frac{1}{\sin(\theta)} (\sin((1 - \tau)\theta)q_1 + \sin(\theta\tau)q_2),$$ (3)

where $\theta = d_c(q_1, q_2)$.

It is easy to see that several elements of C can represent curves with the same shape. For example, if we rotate a face in $\mathbb{R}^3$, and thus its facial curves, we get different SRVFs for the curves but their shapes remain unchanged. Another similar situation arises when a curve is reparametrized: a reparameterization changes the SRVF of curve but not its shape. In order to handle this variability, we define orbits of the rotation group $SO(3)$ and the reparameterization group Γ as equivalence classes in C. Here, Γ is the set of all orientation-preserving diffeomorphisms of I (to itself) and the elements of Γ are viewed as reparameterization functions. For example, for a curve $\beta_\alpha : I \longrightarrow \mathbb{R}^3$ and a function $\gamma \in \Gamma$, the curve $\beta_\alpha \circ \gamma$ is a reparameterization of β_α. The corresponding SRVF changes according to $q(t) \longmapsto \sqrt{\dot{\gamma}(t)}q(\gamma(t))$. We define the equivalent class containing q as:

$$[q] = \{\sqrt{\dot{\gamma}(t)}Oq(\gamma(t)) | O \in SO(3), \gamma \in \Gamma\},$$

The set of such equivalence class is called the shape space S of elastic curves [23]. To obtain geodesics and geodesic distances between elements of S, one needs to solve the optimization problem. The resulting shape space is the set of such equivalence classes:

$$S \doteq C/(SO(3) \times \Gamma)$$ (4)

We denote by $d_s(\beta_{\alpha_1}, \beta_{\alpha_2})$ the geodesic distance between the corresponding equivalence classes $[q_1]$ and $[q_2]$ in shape space S. In Fig. 7 we show geodesic paths between radial curves and the facial surfaces obtained by Delaunay triangulation of the set of points of radial curves. In 7(a) we show a geodesic path between two facial surfaces of the same person, while in 7(b) we show the same for faces belonging to different persons.

B. Circular Closed Curves

We will use $\mathcal{B}_\lambda^c$ to denote the circular closed curves. Using SRVF representation as earlier, we can define the set of closed curves in $\mathbb{R}^3$ by $C = \{q : I \longrightarrow \mathbb{R}^3 | \int_I q(t)|q(t)|dt = 0, \|q\| = 1\} \subset L^2(I, \mathbb{R}^3)$. The quantity $\int_I q(t)|q(t)|dt$ is the total displacement in $\mathbb{R}^3$ while moving from the origin of the curve until the end. If it is zero, the corresponding curve is closed. Thus, the set $\tilde{C}$ represents the set of all closed curves in $\mathbb{R}^3$. It is called a preshape space since curves with same shapes but different orientations and reparameterizations can be represented by different elements of C. To define a shape, its representation should be independent of its rotation and reparameterization. This is obtained mathematically by a removing the rotation group $SO(3)$ and the reparameterization group Γ from $\tilde{C}$. As described in [23], [24], we define the orbits of the rotation group $SO(3)$ and the reparameterization group Γ as equivalence classes in $\tilde{C}$. The resulting shape space is:

$$\tilde{S} \doteq \tilde{C}/(SO(3) \times \Gamma)$$ (5)

To define geodesics on preshape and shape spaces we need a Riemannian metric. For this purpose we inherit the standard L^2-metric of the large space $L^2(I, \mathbb{R}^3)$. For any $u, v \in L^2(I, \mathbb{R}^3)$, the standard L^2 inner-product is given by:

$$\langle\langle u, v \rangle\rangle = \int_I \langle u(t), v(t) \rangle dt.$$ (6)

The computation of geodesics and geodesic distances utilize the intrinsic geometries of these spaces. While the detailed description of the geometries of C and S are given in [23], [24], we briefly mention the tangent and normal spaces of $\tilde{C}$. It can be shown that the set of all functions normal to C at a point q are given by:

$$N_q(\tilde{C}) = span \left\{ q(t), \frac{q'(t)}{|q(t)|}q(t) + |q(t)|e^i, i = 1, 2, 3 \right\}$$ (7)

where $\{e^1, e^2, e^3\}$ form an orthonormal basis of $\mathbb{R}^3$. Thus, the tangent space at any point $q \in \tilde{C}$ is given by:

$$T_q(\tilde{C}) = \{v : I \longrightarrow \mathbb{R}^3 | v \perp N_q(\tilde{C})\}$$ (8)
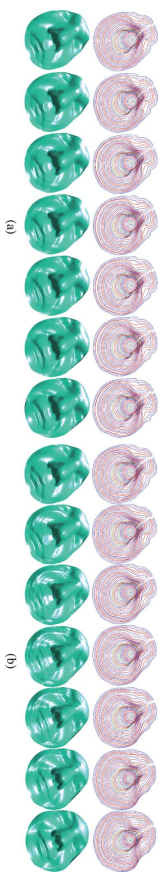
Fig. 8. (a) Intraclass geodesics between facial surfaces and their associated circular curves. (b) Interclass geodesics between facial surfaces and their associated circular curves.

Now, an important tool in our framework is the construction of a geodesic path between two elements of S, under the Riemannian metric given by (6). Given two curves q_1 and q_2, we need to find a geodesic path between the orbits $[q_1]$ and $[q_2]$ in the space S. We use in this context, a numerical method called the *path-straightening* method [25] which connects the two points $[q_1]$ and $[q_2]$; an arbitrary path α and then updates this path repeatedly in the negative direction of the gradient of energy given by:

$$E[\alpha] = \frac{1}{2} \int_s \langle \dot{\alpha}(s), \dot{\alpha}(s) \rangle ds \quad (9)$$

It has been proven in [25] that the critical points of E defined by (9) are geodesic paths in S. We denote by $d_s(\mathcal{B}_{\lambda_1}, \mathcal{B}_{\lambda_2})$ the geodesic distance between the corresponding equivalence classes $[q_1]$ and $[q_2]$ in S. In Fig. 8 we show geodesic paths between circular curves and the facial surfaces obtained by Delaunay triangulation of the set of points of circular curves. In 8(a) we show a geodesic path between two facial surfaces of the same person, while in 8(b) we show the same for faces belonging to different persons.

C. Extension to Facial Surfaces Shape Analysis

In this section we extend our study from shapes of curves to shapes of facial surfaces. We represent the surface of the facial surface S by a collection of 3-D circular and radial curves,

$$S \approx \left\{ \bigcup_{\lambda=1}^{N_\lambda} \mathcal{B}_\lambda^c \right\} \cup \left\{ \bigcup_{\alpha=1}^{N_\alpha} \mathcal{B}_\alpha^r \right\}. \quad (10)$$

where $\mathcal{B}_\lambda^c$ represents the circular curves, $\mathcal{B}_\alpha^r$ represents the radial curves. N_λ is the cardinality of the set of circular curves, N_α is the cardinality of the set of radial curves. Two shapes of facial surfaces are compared by comparing their corresponding facial curves. The distance between two facial surfaces S^1 and S^2 could be defined by:

$$d(S^1, S^2) = \frac{1}{N_\alpha} \sum_{\alpha=1}^{N_\alpha} d_s(\mathcal{B}_\alpha^{r,1}, \mathcal{B}_\alpha^{r,2}) + \frac{1}{N_\lambda} \sum_{\lambda=1}^{N_\lambda} d_s(\mathcal{B}_\lambda^{c,1}, \mathcal{B}_\lambda^{c,2}) \quad (11)$$

VI. Boosting for Geometric Feature Selection

Radial and circular curves capture locally the shape of the faces. However, their comparison under different expressions runs 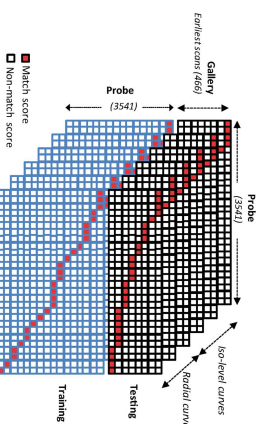into trouble. In fact, their shapes are affected by changes in facial expressions. For that purposes, we introduce a feature selection step to identify (or localize) the most stable and most discriminative curves. We propose to use the well known machine learning algorithm AdaBoost introduced by Freund and Schapire in [26]. Recall that, boosting is based on iterative selection of weak classifiers by using a distribution of training samples. At each iteration, the best weak classifier is provided and weighted by the quality of its classification. In practice, the individual circular curves and radial curves are used as weak classifiers. After M iterations, the most relevant $T(T < M)$ facial curves are returned by the algorithm.

A. Face Recognition

To train and test Adaboost classifier for this application, we use the 3-D face models of FRGCv2 dataset. For each radial and circular curve, we compute the *All versus All* (4007 × 4007) similarity matrix. We then split the matrices as we keep the *Gallery versus Probe* Probe of size 466 × 3541 for testing and uses the remaining *Probe versus Probe* submatrices, of the size of 3541 × 3541, for the training. Thus, we separate the training and the testing samples (the set of individual distances related to facial curves) and these disjoint sets serve as inputs to Adaboost, as illustrated in Fig. 9.

From these areas of the matrices, we extract two kinds of scores (i) the *match scores* (intrapersonal comparisons) and (ii) the *nonmatch scores* (interpersonal comparison). Together these scores form an input to the boosting algorithm. More formally, we consider a set of pairs $(x_n^k, y_n)_{1 \leq n \leq N}$ corresponding to similarity scores between radial or circular curves at the same level, with $k = r$ or $k = c$, y_n can take two values: 0 in the case of *non-match score* and 1 in the case of *match score*. For each circular or radial curve, the weak classifier determines the optimal threshold classification function such that the minimum number of samples are misclassified. Each weak classifier $h_j(x_n^k)$ will take a value of distance computed based on a radial or circular curve, threshold it, and classify the comparison as positive (intraclass) or negative (interclass), depending on the distance value being under or over the threshold θ:

$$h_j(x_n^k) = \begin{cases} 1 & \text{if } x_n^k < \theta \ (\text{intra-personal}) \\ 0 & \text{otherwise.} \ (\text{inter-personal}) \end{cases} \quad (12)$$

where, h_t denotes for the weak hypothesis given by $h_t : \mathcal{X} \rightarrow \{0, 1\}$. The final strong classifier is defined by a set of weak classifiers weighted by a set of weights $\mathcal{W} = \{w_{t,n}\}_{1 \leq n \leq N}$. The pseudocode of AdaBoost algorithm is summarized in algorithm 1.

Algorithm 1 AdaBoost Algorithm

- Input: A set of samples $(x_1^k, y_1), \ldots, (x_N^k, y_N)$ where x_i^k is the score of similarity of the circular or radial curves k ($1 \leq k \leq \lambda_0$) and $y_n = \{0, 1\}$.
- Let m be the number of nonmatch score and l be the number of match score.
  Initialization of weights $w_{1,n} = (1)/(2m), (1)/(2l)$ it depends on the value of y_n where $n \in 1 \ldots N$.
  $\lambda_0 = N_\lambda + N_\alpha$
  For $t = 1, \ldots, M$:
  1- Normalize the weights $w_{t,n}$ such that $\sum_{n=1}^{N} w_{t,n} = 1$.
  2- For each curve β_j (feature), train a weak classifier h_j that uses a single curve. The error ϵ_j of classifier h_j is determined with the corresponding weight $w_{t,1}, \ldots, w_{t,N}$:

$$\epsilon_j = \sum_{n}^{N} w_{t,n} |h_j(x_n^k) - y_n|$$

  3- Choose the classifier h_j with the lowest error.
  4- Update the weights $w_{t+1,n} = w_{t,n} \gamma_t^{1-e_n}$, where $\gamma_t = (\epsilon_t)/(1 - \epsilon_t)$ and $e_n = 0$, if the sample x_n is correctly classified by h_t and 1 else.
- The final hypothesis is a weighted linear combinations of the T hypothesis where the weights are inversely proportional to the training errors. The strong classifier is given by:

$$H(x) = \begin{cases} 1 & \text{if } \sum_{t=1}^{T} \log \frac{1}{\gamma_t} h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \log \left(\frac{1}{\gamma_t} \right) \\ 0 & \text{else.} \end{cases}$$

The set of selected curves returned by Adaboost is shown in Fig. 10. The first row shows the locations of the selected curves on different sessions of the same person with different expressions, whereas, the second row gives curves location on different subjects. We note that the boosting algorithm selects

Fig. 9. *All versus Probe* distance matrices splitting; the Gallery versus Probe distance matrix is kept for testing, as suggested in the FRGC evaluation protocol and the remaining submatrices *Probe versus Probe* are used for training.

iso-curves located on the nasal region, which is stable under expressions and radial curves avoiding two parts. The first one is the lower part of the face since its shape is affected by expressions, particularly when the mouth is open. The second area corresponds to the eye/eyebrow regions. Shapes of radial curves passing throw these regions change when conveying expressions. In contrast, the most stable area cover the nasal/forehead regions.

To demonstrate the usefulness of the curves selection step, different graphs in Fig. 11 plot the rate of False Rejection for different configurations. These curves are produced from the *Probe versus Probe* matrices (i.e., using the training set). As shown in Fig. 11(b), minimum errors are given by fusing scores of selected radial and selected circular curves. We note also that the selection performed on radial curves only or circular curves only minimizes the errors compared to the use of all radial curves or circular curves, respectively.

The online testing step consists on comparing faces S^1 and S^2 by the fusion of scores related to selected curves as following:

$$d(S^1, S^2) = \frac{1}{N_{\alpha_s}} \sum_{\alpha=1}^{N_{\alpha_s}} d_s(\mathcal{B}_\alpha^{r,1}, \mathcal{B}_\alpha^{r,2}) + \frac{1}{N_{\lambda_s}} \sum_{\lambda=1}^{N_{\lambda_s}} d_s(\mathcal{B}_\lambda^{c,1}, \mathcal{B}_\lambda^{c,2}) \quad (13)$$

where N_{α_s} is the cardinality of the set of selected circular curves and N_{α_s} the cardinality of the set of selected radial curves.

B. Gender Classification

For 3-D face-based gender classification task, we first compute Male and Female representative templates using the geometric shape analysis framework for open and closed curves. In fact, this framework allows us to compute intrinsic means (Karcher mean) of curves that we extend to facial surfaces. Then, within the training step, we compute intraclass (same gender) and interclass (different gender) pairwise distances (for each curve index) between sample faces and the templates (for each gender and interclass different sample faces are selected optimally by boosting as done in face recognition application. Finally, the most discriminative geometric features are selected optimally by boosting as done in face recognition application. For the testing step, distances to male and female templates are computed (based only on selected features), and the nearest neighbor algorithm denotes the class (Male/Female) membership. Different steps are detailed in the following:

Fig. 10. The most discriminative radial and circular curves selected by Boosting for face recognition, given on different faces.
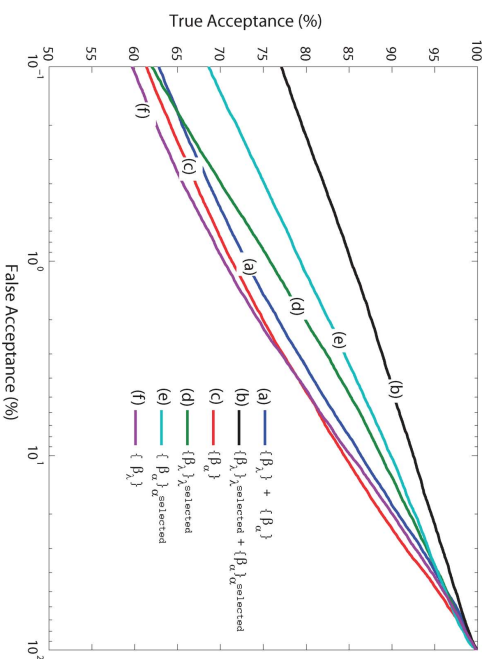
Fig. 11. ROC curves produced from the training set (a) all radial and circular curves, (b) selected radial and selected level curves, (c) all radial curves, (d) selected circular curves, (e) selected radial curves, and (f) all circular curves.

Male/Female Geometric Templates Computation: One advantage of the proposed geometrical framework for shape analysis of curves is to calculate some statistics as the "mean" of facial curves and to extend it to facial surfaces, called *Karcher mean* [24]. The Riemannian structure defined on a Riemannian manifold enables us to perform such statistical analysis for computing faces mean and variance. There are at least two ways of defining a mean value for a random variable that takes values on a nonlinear manifold. The first definition, called *extrinsic mean*, involves embedding the manifold in a larger vector space, computing the Euclidean mean in that space, and then projecting it down to the manifold. The other definition, called the *intrinsic mean* or the *Karcher mean* utilizes the intrinsic geometry of the manifold to define and compute a mean on that manifold. It is defined as follows: Let $d_c(\beta_1^a, \beta_2^a)$ denotes the length of the geodesic path from curves in C. To calculate the Karcher mean of facial curves $\{\beta_1^a, \ldots, \beta_n^a\}$ in C, we define the variance function:

$$\mathcal{V} : C \rightarrow \mathbb{R}, \mathcal{V}(\mu) = \sum_{i=1}^{n} d_c \left(\beta_i^a, \beta^a \right)^2 \qquad (14)$$

The Karcher mean is then defined by:

$$\overline{\beta^a} = \arg \min_{\mu \in C} \mathcal{V}(\mu) \qquad (15)$$

The intrinsic mean may not be unique, i.e., there may be a set of points in C for which the minimizer of $\mathcal{V}$ is obtained. $\overline{\beta^a}$ is an element of C that can be seen as the smallest geodesic path length from all given facial surfaces. We present a commonly

used algorithm for finding Karcher mean for a given set of facial surfaces (by using their curves). This approach, presented in algorithm 2 uses the gradient of $\mathcal{V}$, in the space $T_{\mu}(C)$, to iteratively update the current mean μ. The same pseudo-algorithm will be obtained for radial curves defined in the shape space C.

Algorithm 2 Karcher Mean Algorithm

Set $k = 0$. Choose some time increment $\epsilon \leq (1)/(n)$. Choose a point $\mu_0 \in C$ as an initial guess of the mean. (For example, one could just take $\mu_0 = S_1$.)

1- For each $i = 1, \ldots, n$ choose the tangent vector $f_i \in T_{\mu_k}(C)$ which is tangent to the geodesic from μ_k to S^a. The vector $g = \sum_{i=1}^{i=n} f_i$ is proportional to the gradient at μ_k of the function $\mathcal{V}$.

2- Flow for time ϵ along the geodesic which starts at μ_k and has velocity vector g. Call the point where you end up μ_{k+1}.

3- Set $k = k + 1$ and go to step 1.

the templates for both male and female classes denoted respectively T^{male} and T^{female}. Then, we computed pairwise distances (based on curves) between test images and both of templates. Thus, we obtained a matrix containing feature vectors (distance based on curves) which will be used to train Adaboost algorithm, as illustrated in Table I. From this matrix, we extract two kinds of scores (i) the *match scores* (intragender comparisons) and (ii) the *nonmatch scores* (intergender comparison).

Both score lists represent the input of the boosting algorithm. More formally, we consider a set of pairs $(x_n^{o,\lambda}, y_n)_{1 \leq n \leq N}$ where $x_n^{o,\lambda}$ is a similarity score between two curves at the same level α, λ and y_n can take two values: 0 in the case of *non-match score* and 1 in the case of *match score*. For each curve β_j, the weak learner determines the optimal threshold classification function, such that the minimum number of samples are misclassified. A weak classifier $h_j(x_n^k)$ thus consists of a geometric feature β_j and a threshold θ, such that:

$$h_j(x_n^k) = \begin{cases} 1 & \text{if } x_n^k < \theta \text{ (intra} - \text{gender)} \\ 0 & \text{otherwise. (inter} - \text{gender)} \end{cases} \qquad (16)$$

Male template facial surface is computed by averaging ten males facial surfaces of different person as shown in Fig. 12. Female template facial surface is computed by averaging ten females facial surfaces of different person as shown in Fig. 13.

Geometric Feature Selection: To train and test the boosting algorithm for this application, we use 20 previous 3-D faces from the FRGCv1 dataset for training and 466 subjects of FRGCv2 for testing. Firstly, we selected a subset of faces of men and women (ten from each class) from FRGCv1, to calculate

Fig. 12. Different facial surfaces of different male persons taken from FRGCv1 and their Karcher mean. (a) Sample faces (from FRGCv1) used to compute the template face. (b) Female template face (Karcher mean).
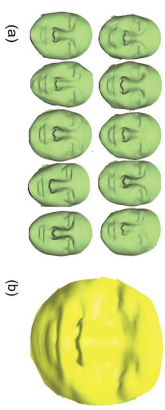

Fig. 13. Different facial surfaces of different female persons taken from FRGCv1 and their Karcher mean. (a) Sample faces (from FRGCv1) used to compute the template face. (b) Female template face (Karcher mean).

TABLE I
INPUT FEATURE VECTORS OF BOOSTING ALGORITHM FOR GENDER CLASSIFICATION

For each curve	Face male1	Face female1	...
T^{male}	$\{x_n^{o,\lambda}, 1\}$	$\{x_n^{o,\lambda}, 0\}$	...
T^{female}	$\{x_n^{o,\lambda}, 0\}$	$\{x_n^{o,\lambda}, 1\}$	...

We note that the boosting algorithm selects circular curves on the cheeks region, which is discriminative shape of 3-D face for gender classification and radial curves between template faces and the 466 test faces. The most stable areas for gender classification cover the cheeks/sellion regions.

Classification: As described in Table I, this time round, we calculated different matrices of distances of all selected circular and radial curves between template faces and the 466 test faces of FRGCv2. The pseudocode of the proposed gender classification algorithm is given in algorithm 3.

Algorithm 3 Gender Classification Algorithm

• Input: A set of curves c_i where c_i is either circular curves or radial curves. N^s is the total number of selected circular and Radial curves.

• For each face request S_k, $k = 1, \ldots, 466$:
  1- Calculate the geodesic distances:

$$\sum_{i=0}^{N^s} d_k (T^{male}, F_k) \text{ and } \sum_{i=0}^{N^s} d_k (T^{female}, F_k)$$

• The final decision $D(F_k)$ for classification is given by:

$$D(F_k)$$
$$= \begin{cases} 1 & \sum_{i=0}^{N^s} d_k (T^{male}, F_k) < \sum_{i=0}^{N^s} d_k (T^{female}, F_k) \\ 0 & \text{else.} \end{cases}$$

Fig. 14. The most discriminative radial and circular curves selected by Boosting for gender classification, given on different male faces.

Fig. 15. The most discriminative radial and circular curves selected by Boosting for gender classification, given on different female faces.

Fig. 15 shows the location of selected curves on different sessions of some male faces whereas, Fig. 14 shows the location of selected curves on different sessions of some female faces.
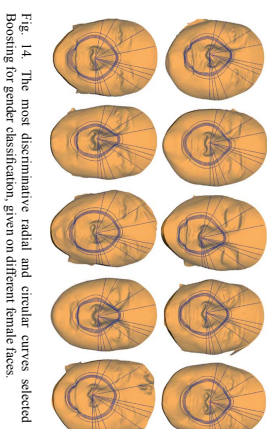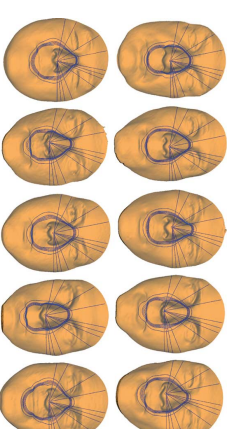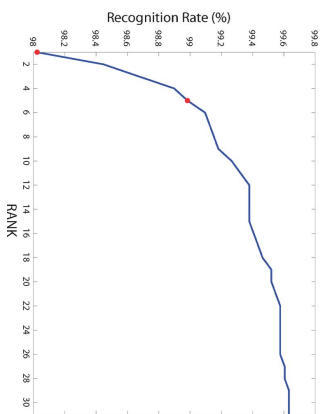
Fig. 16. Cumulative Match Characteristic curve for selected radial and level curves.

VII. EXPERIMENTAL RESULTS

In the following, we present conducted experiments and the obtained results with the proposed methods. In particular, we report 3-D face recognition performances on FRGCv2 [27] and provide a comparative study with state-of-the-art. FRGCv2 dataset contains 4 007 3-D scans of 466 people, in which more than 40% of the models are nonneutral. A standard evaluation protocol for identification and verification biometric scenarios supports this data set. Furthermore, we give gender classification performances achieved by our approach, using the same dataset. We note that the subjects in FRGCv2 dataset are 57% male and 43% female.

A. 3-D Face Recognition Results

1) Identification: For testing on *FRGCv2* dataset, only the identification evaluation was carried out. In fact, as mentioned in Section VI since our approach requires a training stage, it was tested on a subset of this dataset following the *FRGC* evaluation protocol for the identification scenario, as following. We kept, for the test, the *Gallery versus Probe* (of size 466 × 354) similarity matrices. The remaining submatrices (i.e., *Probe versus Probe* similarity scores) were used to train the feature selection step by boosting algorithm. This means that disjointly similarity vectors are used for the training and the test. Following these settings, our approach achieved 98.02% as rank-1 recognition rate and reached 99% in rank-5 as illustrated in the CMC plot (Cumulative Match Characteristic) given in Fig. 16. Following these experiments were computed to build the male and female templates (of FRGCv2) based on their circular and radial curves were computed for all the 10 subsets used exactly once.

1) Classification Results: We conducted experiments by first computing Male (T^{male}) and Female (T^{female}) templates using sample 3-D scans taken from FRGCv1. Then, comparisons between those templates {T^{male}, T^{female}} and the 466 test images (of FRGCv2) based on their circular and radial curves were computed to build the feature vectors. Finally, two experiments, detailed below, were carried out:

- **Experiment 1. *Curve selection-based*:** To select the most relevant combination of curves using Adaboost algorithm, we first use 10 male 3-D face models and 10 female models from FRGCv1 to compute the male and female templates. Then, we compute pairwise curve distances between the

In addition to performance improvement, the curve selection results on a more compact biometric signature which reduce the time-processing of one-to-one face matching.

2) Comparative Study With State-of-the-Art: Following the FRGC standard protocol for the identification scenario, the Table III shows identification results of previous approaches (curve-based, feature selection-based, and others) by keeping the earliest scan of the 466 subjects in the gallery and the remaining for testing. We note that experiments reported in [28] and [16] follow a modified protocol by keeping the earliest neutral scan in the gallery and the remaining as test images. It is clear that the proposed method outperforms the majority of state-of-the-art methods. Only the approach proposed recently by *Wang et al.* [6], based on boosting of descriptors (Haar-like, Gabor, and Local Binary Pattern (LBP)) computed on the *Shape Difference Map* between faces, achieved a better result 98.3% which means that ten more faces are recognized by this approach.

As shown in Table III, the proposed approach outperforms the state-of-the-art except the work of [6].

B. Gender Classification

The proposed gender classification of 3-D face scans has been experimented using the FRGCv2 database. This was motivated by the fact that this dataset contains the largest number of subjects compared to existing 3-D face datasets as Bosphorus, BU-3DFE, etc. To evaluate the proposed approach, we have considered 466 3-D images related to the 466 subjects of FRGCv2 data set. Thus, if several sessions exist in the dataset, we select the earliest (neutral or nonneutral) one for our experiment. We use also few 3-D images taken from FRGCv1 to compute male and female templates, as described in Section VI. The difficulty encountered to compare our approach to related work, is there is no standard protocol to compare gender classification results, unlike FRGC standard protocol for 3-D face recognition. Most of previous approaches [34], [18], [20] reported classification results on a subset taken from FRGCv1 dataset.

1) Classification Results: We conducted experiments by first computing Male (T^{male}) and Female (T^{female}) templates using sample 3-D scans taken from FRGCv1. Then, comparisons between those templates {T^{male}, T^{female}} and the 466 test images (of FRGCv2) based on their circular and radial curves were computed to build the feature vectors. Finally, two experiments, detailed below, were carried out:

- **Experiment 2. *Classifier decision-based*:** In order to evaluate the boosting classifier results, we have conducted, following the same setting, experiments using two machine learning methods, SVM and Neural Network. Instead of feature selection step, we consider the final classifier decision. For example, we consider the binary decision of Adaboost. Table V summarizes the performance using SVM and Neural Network. Training and testing steps are carried out using a 10-fold cross-validation experiment. According to this, the 466 subjects are split into disjoint subsets for training and test. Using 10-fold cross validation, training is repeated 10 times, with each of the 10 subsets used exactly once as the test data. Finally, the results from the ten steps are averaged to produce a single estimation of the performance of the classifier for the experiment. In this way, all observations are used for both training and test, and each observation is used for test exactly once.

Then, we compute pairwise curve distances between the

TABLE II
RANK-1 (COMPUTATION COST (IN SECONDS) FOR DIFFERENT CONFIGURATIONS

Performance	All		Selected	
	Rank-1	Time(s)	Rank-1	Time(s)
Radial	88.65%	1.6	89.04%	0.48
circular	66.51%	1.04	85.65%	0.20
Fusion based on Eq. 13	91.81%	2.64	**98.02%**	**0.68**

TABLE III
COMPARISON WITH STATE-OF-THE-ART APPROACHES ON FRGCv2 (RANK-1 RECOGNITION RATE)

	A. Curve-based representation				B. Feature selection-based			Others	
Methods	ter Haar [29]	Berretti(*) [15]	Mpiperis [12]	Drira [30]	Faltemier [4]	Kakadiaris [31]	**Wang** [5]	Huang [32]	Mian (*) [28]	Cook [33]
Rank-1	97%	94.1%	-	97.2%	97%	**98.3%**	97.2%	91.9%	92.9%
Our					98.02%				

(*) "E-N-S" means the earliest neutral scan in the gallery and the remaining as probes.

TABLE IV
EXPERIMENTAL COMPARISON OF GENDER CLASSIFICATION METHODS USING DIFFERENT TYPES OF SELECTED CURVES

Methods	Curve selection-based 3		
	circular curves	Radial curves	circular and radial curves
Gender Classification Rate	79.40%	80.69%	84.12%

(*) Results for 10-fold cross-validation.

TABLE V
EXPERIMENTAL COMPARISON OF GENDER CLASSIFICATIONS METHODS USING DIFFERENT MACHINE LEARNING TECHNIQUES

Methods	Classifier decision-based		
	AdaBoost(*)	SVM(*)	Neural Network
Gender Classification Rate	86.05%	83.91%	83.05%

2) Comparative Study With State-of-the-Art: Table VI shows gender classification results compared to previous approaches tested on different subsets. In [35] the authors used a subset of FRGCv1, six female subjects and four male subjects, while in [19] a subset of FRGCv1 is used with only 28 female subjects and 80 male subjects. However they used more than one session of each subject. Note that this ad-hoc division does not guarantee that all subjects will have a neutral expression, some FRGCv2 subjects are scanned with arbitrary facial expression.

The analysis of some misclassified examples given by Figs. 17 and 18 shows that there are two major reasons of this misclassification. The first one is the bad quality of 3-D scans, such as some occlusions in relevant regions which affected the shape of curves. The second reason lies in the fact that only based on the shape information, there is some confusion (even for a person), to recognize correctly the gender of the person. One solution of this problem and for improving the proposed approach is to introduce the texture information, which contains complementary (as hair density, etc.) in order to consolidate the classifier decision.

Fig. 19 shows some faces with selected iso-level and radial curves for both applications face recognition and gender classification. The blue curves are selected for 3-D face gender classification, the red curves are selected for 3-D face recognition while the black curves are the common selected curves for both classifications.
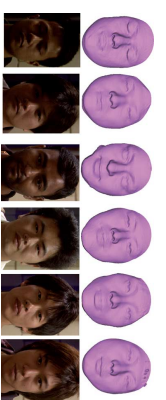

Fig. 17. Different female 3-D faces misclassified by our approach. The first row shows 3-D data; the second row shows the corresponding texture of 3-D data.

Fig. 18. Different female 3-D faces misclassified by our approach. The first row gives 3-D data, the second gives 3-D data; the second row shows the corresponding texture of 3-D data.
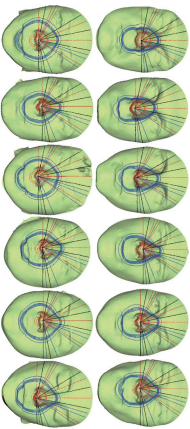
Fig. 19. Examples of selected circular and radial curves for different applications. The first row shows different sessions for different male persons; the second shows different sessions for different female persons.

— Selected Curves for 3D face Gender Classification
— Selected Curves for 3D face Recognition
— Selected Curves for both applications

TABLE VI
GENDER CLASSIFICATION RESULTS, COMPARISON TO STATE-OF-THE-ART RESULTS

Approach	Dataset	Number of Subjects			Classification rate
		All Subjects	Female Subjects	Male Subjects	
[35]	FRGCv1	10	6	4	69.7%
[19]	FRGCv1	376	139	237	85.4%
[20]	FRGCv1	108	28	80	94.3%
Our	FRGCv2	466	202	264	86.05%/84.12%

VIII. CONCLUSION

In this paper, by combining tools from Riemannian geometry and the well-known Adaboost algorithm, we have proposed to select the most discriminative curves for facial recognition and gender classification. The experiments, carried out on FRGCv2 including neutral and nonneutral images, demonstrate the effectiveness of the proposed approach. Based on 17 curves, including 12 radial curves and 5 circular curves, our fully automatic approach achieved rank-1 recognition rate of 98.02%. For gender classification, 19 curves with 12 radial and 7 circular, were selected and achieved a classification rate of 86.05%. The algorithm computation time was on the order of 0.68 second (for recognition rate) and 0.76 second (for gender classification) to compare two faces with selected curves instead of 2.64 second with all curves. The boosting selects those curves passing through stable regions on the face for different applications. This approach is efficient in terms of computation time, data storage and transmission costs.

The proposed approach can be extended in different directions in order to improve the performances and to address other important classification tasks. For example, the texture information could be associated to shape information to consolidate the classifier decision. In fact, texture channel provides additional informations which could be complementary to the shape. Additionally, more facial attributes recognition, such as ethnicity and age estimation could be addressed using the same framework.

REFERENCES

[1] V. Bruce, A. M Burton, E. Hanna, P. Healey, O. Mason, A. Coombes, R. Fright, and A. Linney, "Sex discrimination: How do we tell the difference between male and female faces?," *Perception*, vol. 22, no. 2, pp. 131–152, 1993.

[2] A. J. O'toole, T. Vetter, N. F. Troje, and H. H. Bülthoff, "Sex classification is better with three-dimensional head structure than with image intensity information," *Perception*, vol. 26, no. 1, pp. 75–84, 1997.

[3] L. Ballihi, B. B. Amor, M. Daoudi, A. Srivastava, and D. Aboutajdine, "Selecting 3D curves on the nasal surface using adaboost for person authentication," in *Proc. 3DOR*, 2011, pp. 101–104.

[4] F. Danyal, P. Nair, and A. Cavallaro, "Compact signatures for 3D face recognition under varying expressions," in *Proc. 6th IEEE Int. Conf. Advanced Video and Signal Based Surveillance, 2009 (AVSS'09)*, Sep. 2009, pp. 302–307.

[5] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "A region ensemble for 3-d face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 62–73, Jan. 2008.

[6] Y. Wang, J. Liu, and X. Tang, "Robust 3D face recognition by local shape difference boosting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1858–1870, Oct. 2010.

[7] X. Li, T. Jia, and H. Z. 0002, "Expression-insensitive 3D face recognition using sparse representation," in *Proc. CVPR*, 2009, pp. 2575–2582.

[8] O. Ocegueda, S. K. Shah, and I. A. Kakadiaris, "Which parts of the face give out your identity?," in *Proc. CVPR*, 2011, pp. 641–648.

[9] H. Li, D. Huang, J.-M. Morvan, and L. Chen, "Learning weighted sparse representation of encoded facial normal information for expression-robust 3D face recognition," in *Proc. Int. Joint Conf. Biometrics*, 2011, vol. 0, pp. 1–7.

[10] C. Samir, A. Srivastava, and M. Daoudi, "Three-dimensional face recognition using shapes of facial curves," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1847–1857, Nov. 2006.

[11] E. Klassen, A. Srivastava, W. Mio, and S. Joshi, "Analysis of planar shapes using geodesic paths on shape spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 3, pp. 372–383, Mar. 2004.

[12] C. Samir, A. Srivastava, M. Daoudi, and E. Klassen, "An intrinsic framework for analysis of facial surfaces," *Int. J. Comput. Vis.*, vol. 82, no. 1, pp. 80–95, 2009.

[13] I. Mpiperis, S. Malassiotis, and M. G. Strintzis, "3-D face recognition with the geodesic polar representation," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pt. 2, pp. 537–547, Sep. 2007.

[14] H. Drira, B. B. Amor, A. Srivastava, and M. Daoudi, "A Riemannian analysis of 3D nose shapes for partial human biometrics," in *Proc. IEEE Int. Conf. Computer Vision*, 2009, pp. 2050–2057.

[15] H. Drira, B. B. Amor, M. Daoudi, and A. Srivastava, "Pose and expression-invariant 3D face recognition using elastic radial curves," in *Proc. BMVC*, 2010, pp. 1–11.

[16] S. Berretti, A. D. Bimbo, and P. Pala, "3D face recognition using isogeodesic stripes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2162–2177, Dec. 2010.

[17] A. M. Burton, V. Bruce, and N. Dench, *What is the Difference Between Men and Women? Evidence From Facial Measurement.* London: Perception, 1993, vol. 22, pp. 153–176, (Print), no. 2.

[18] J. Wu, W. A. P. Smith, and E. R. Hancock, "Facial gender classification using shape from shading and weighted principal geodesic analysis," in *Proc. ICIAR*, 2008, pp. 925–934.

[19] X. Lu, H. Chen, and A. K. Jain, "Multimodal facial gender and ethnicity identification," in *Proc. ICB*, 2006, pp. 554–561.

[20] J. Hu, J. Yan, and P. Shi, "A fusion-based method for 3D facial gender classification," in *Proc. 2010 The 2nd Int. Conf. Computer and Automation Engineering (ICCAE)*, Feb 2010, vol. 5, pp. 369–372.

[21] P. W. Michor and D. Mumford, "Riemannian geometries on spaces of plane curves," *J. Eur. Math. Soc.*, vol. 8, pp. 1–48, 2006.

[22] A.-J. Yezzi and A. Mennucci, "Conformal metrics and true" gradient flows" for curves," in *Proc. ICCV*, 2005, vol. 1, pp. 913–919.

[23] S. H. Joshi, E. Klassen, A. Srivastava, and I. Jermyn, "A novel representation for Riemannian analysis of elastic curves in r^n," in *Proc. CVPR*, 2007, pp. 1–7.

[24] A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn, "Shape analysis of elastic curves in euclidean spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1415–1428, Jul. 2011.

[25] E. Klassen and A. Srivastava, "Geodesics between 3D closed curves using path-straightening," in *Proc. ECCV, Lecture Notes in Computer Science*, 2006, pp. I: 95–1: 106.

[26] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Proc. 2nd Eur. Conf. Computational Learning Theory (EuroCOLT'95)*, 1995, pp. 23–37, London U.K.: Springer-Verlag.

[27] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Computer Vision and Pattern Recognition*, 2005, pp. 947–954.

[28] A. S. Mian, M. Bennamoun, and R. A. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1927–1943, Nov. 2007.

[29] F. B. ter Haar and R. C. Veltkamp, "A 3D face matching framework for facial curves," *Graph. Models*, vol. 71, no. 2, pp. 77–91, 2009.

[30] H. drira, "Statistical Computing on Manifolds for 3D Face Analysis and Recognition" Ph.D. dissertation, University Lille 1, Lille, France, 2011 [Online]. Available: http://www.telecomlille1.eu/people/drira/dissertation.pdf

[31] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 640–649, Apr. 2007.

[32] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "A novel geometric facial representation based on multi-scale extended local binary patterns," in *Proc. 2011 IEEE Int. Conf. Automatic Face Gesture Recognition and Workshops (FG 2011)*, Mar. 2011, pp. 1–7.

[33] J. A. Cook, V. Chandran, and C. B. Fookes, "3D face recognition using log-gabor templates," in *Proc. British Machine Vision Conf.*, Edinborough, Scotland, 2006.

[34] L. Lu and P. Shi, "A novel fusion-based method for expression-invariant gender classification," in *Proc. ICASSP*, 2009, pp. 1065–1068.

[35] J. Wu, W. A. P. Smith, and E. R. Hancock, "Gender classification using shape from shading," in *Proc. BMVC*, U.K., 2007, pp. 499–508. A. Cook, V. Chandran, and C. B. Fookes, "3D face recognition using Log-Gabor templates," in *Proc. BMVC*, Edinburgh, Scotland, 2006, pp. 769–778.

Lahoucine Ballihi received the Master of Advanced Studies (M.A.S.) degree in computer science and telecommunications from Mohammed V-Agdal University, Faculty of Sciences-Rabat (FSR, Morocco, in 2007 and the PhD degree in computer science from the University of Lille 1 (USTL), France and the Mohammed V-Agdal University, Rabat, Morocco, in 2012.

He is also a member of the Computer Science Laboratory in University Lille 1 (LIFL UMR CNRS 8022) and the Computer Science and Telecommunications Research Laboratory (LRIT, unit associated to CNRST (URAC 29)). His research interests are mainly focused on statistical three-dimensional face analysis and recognition and gender classification using 3-D images. He is also a teaching assistant in (Telecom Lille1), School of Engineering, Currently, he is a Postdoc within the Computer Science Laboratory of the University Lille 1 (LIFL).

Boulbaba Ben Amor received the M.Sc. degree in 2003 and the Ph.D. degree in computer science in 2006, both from Ecole Centrale de Lyon, France.

Currently, he is an associate-professor in Institut Mines-Telecom/Telecom Lille1. He is also a member of the Computer Science Laboratory in University Lille 1 (LIFL UMR CNRS 8022). His research interests are mainly focused on statistical three-dimensional face analysis and recognition and facial expression recognition in 2006 and facial expression recognition in refereed journals and proceedings several papers in refereed journals and proceedings. He has been involved in French and international projects and has served as program committee member and reviewer for international journals and conferences.

Mohamed Daoudi (M'96–SM'08) received the these de doctorat degree in computer engineering from the University of Lille 1 (USTL), France, in 1993 and Habilitation à Diriger des Recherches from the University of Littoral, France, in 2000.

He is Full Professor of Computer Science at TELECOM Lille 1 and LIFL (UMR CNRS 8022). He is the head of the Computer Science Department, Telecom Lille1. He was the founder and the leader of MIIRE Research Group. His research interests include pattern recognition, image processing, three-dimensional analysis and retrieval, and 3-D face analysis and recognition. He has published over 100 papers in some of the most distinguished scientific journals and international conferences. He is the coauthor of the book *3D Processing: Compression, Indexing and Watermarking* (Wiley, 2008). He was a co-organizer and cochair of Eurographics 3-D Object Retrieval (2009) and ACM Workshops on 3-D Object Retrieval (2010) and Multimedia access to 3-D Human Objects MA3HO (2011).

Anuj Srivastava (S'87–M'96–SM'09) received the M.S. and Ph.D. degrees in electrical engineering from Washington University, St. Louis, MO, in 1993 and 1996, respectively.

After spending the year 1996–1997 at Brown University as a visiting researcher, he joined FSU as an assistant professor in 1997. He is currently a professor in Department of Statistics at Florida State University (FSU), Tallahassee. His research is focused on pattern-theoretic approaches to problems in image analysis, computer vision, and signal processing. In particular, he has developed computational tools for performing statistical inferences on certain nonlinear manifolds, including shape manifolds of curves and surfaces. He has published more than 150 peer-reviewed journal and conference articles in these areas.

Driss Aboutajdine (M'89–SM'99) received the Doctorat de 3ème Cycle and the Doctorat d'Etat-es-Sciences degrees in signal processing from the Mohammed V-Agdal University, Rabat, Morocco, in 1980 and 1985, respectively.

He joined Mohammed V-Agdal University, Rabat, Morocco, in 1978, first as an assistant professor, in 1985, as an associate professor in 1990, and full Professor since 1990, where he is teaching Signal/Image Processing and Communications. For over 35 years, he developed teaching and research activities covering various topics of signal and image processing, wireless communication and pattern recognition which allow him to advise more than 50 Ph.D. theses and publish over 200 journal papers and conference communications. He succeeded to found and manage since 1993 and 2001, respectively, the LRIT Laboratory and the Centre of Excellence of Information and Communication Technology (pole de competences STIC) which gathers more than 30 research laboratories from all Moroccan universities and including industrial partners as well. He is organized and chaired several international conferences and workshops. He was elected member of the Moroccan Hassan II Academy of Science and Technology in 2006 and fellow of the TWAS Academy of Sciences in 2007. He cofounded the IEEE Morocco Section in 2005 and he is chairing the Signal Processing chapter he founded in December 2010.

Shape analysis of local facial patches for 3D facial expression recognition

Ahmed Maalej [a,b], Boulbaba Ben Amor [a,b,*], Mohamed Daoudi [a,b], Anuj Srivastava [c], Stefano Berretti [d]

[a] LIFL (UMR CNRS 8022), University of Lille 1, France
[b] Institut TELECOM; TELECOM Lille 1, France
[c] Department of Statistics, Florida State University, USA
[d] Dipartimento di Sistemi e Informatica, University of Firenze, Italy

ABSTRACT

In this paper we address the problem of 3D facial expression recognition. We propose a local geometric shape analysis of facial surfaces coupled with machine learning techniques for expression classification. A computation of the length of the geodesic path between corresponding patches, using a Riemannian framework, in a shape space provides a quantitative information about their similarities. These measures are then used as inputs to several classification methods. The experimental results demonstrate the effectiveness of the proposed approach. Using multiboosting and support vector machines (SVM) classifiers, we achieved 98.81% and 97.75% recognition average rates, respectively, for recognition of the six prototypical facial expressions on BU-3DFE database. A comparative study using the same experimental setting shows that the suggested approach outperforms previous work.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

In recent years, 3D facial expression recognition has received growing attention. It has become an active research topic in computer vision and pattern recognition community, impacting important applications in fields related to human–machine interaction (e.g., interactive computer games) and psychological research. Increasing attention has been given to 3D acquisition systems due to the natural fascination induced by 3D objects visualization and rendering. In addition 3D data have advantages over the 2D data, in that 3D facial data have high resolution and convey valuable information that overcomes the problem of pose/lighting variations and the detail concealment of low resolution acquisition.

In this paper, we present a novel approach for 3D identity-independent facial expression recognition based on a local shape analysis. Unlike the identity recognition task that has been the subject of many papers, only few works have addressed 3D facial expression recognition. This could be explained through the challenge imposed by the demanding security and surveillance requirements. Besides, there has long been a shortage of publicly available 3D facial expression databases that serve the researchers exploring 3D information to understand human behaviors and emotions. The main task is to classify the facial expression of

a given 3D model, into one of the six prototypical expressions, namely Happiness, Anger, Fear, Disgust, Sadness and Surprise. It is stated that these expressions are universal among human ethnicity as described in [1,2].

The remainder of this paper is organized as follows. First, a brief overview of related work is presented in Section 2. In Section 3 we describe the BU-3DFE database designed to explore 3D information and improve facial expression recognition. In Section 4, we summarize the shape analysis framework applied earlier for 3D curves matching by Joshi et al. [3], and discuss its use to perform 3D patches analysis. This framework is further expounded in Section 5, so as to define methods for shapes analysis and matching. In Section 6 a description of the feature vector and used classifiers is given. In Section 7, experiments and results of our approach are reported, and the average recognition rate over 97% is achieved using machine-learning algorithms and SVM. Finally, discussion and conclusion are given in Section 8.

2. Related work

Facial expression recognition has been extensively studied over the past decades especially in 2D domain (e.g., images and videos) resulting in a valuable enhancement. Existing approaches that address facial expression recognition can be divided into three categories: (1) static versus dynamic; (2) global versus local; (3) 2D versus 3D. Most of the approaches are based on feature

extraction/detection as a mean to represent and understand facial expressions. Pantic and Rothkrantz [4] and Samal and Iyengar [5] presented a survey where they explored and compared different approaches that were proposed, since the mid 1970s, for facial expression analysis from either static facial images or image sequences. Whitehill and Omlin [6] investigated on the Local versus Global segmentation for facial expression recognition. In particular, their study is based on the classification of action units (FACS) manual by Ekman and Friesen [7], and designating the elementary muscle movements involved in the bio-mechanical of facial expressions. They reported, in their study on face images, that the local expression analysis showed no consistent improvement in recognition accuracy compared to the global analysis. As for 3D facial expression recognition, the first work related to this issue was presented by Wang et al. [8]. They proposed a novel geometric feature based facial expression descriptor, derived from an estimation of primitive surface feature distribution.

A labeling scheme was associated with their extracted features, and they constructed samples that have been used to train and test several classifiers. They reported that the highest average recognition rate they obtained was 83%. They evaluated their approach not only on frontal view facial expressions of the BU-3DFE database, but they also tested its robustness to non-frontal views. A second work was reported by Soyel and Demirel [9] on the same database. They extracted six characteristic distances between 11 facial landmarks, using Neural Network architecture that analysis the calculated distances, they classified the BU-3DFE facial scans into seven facial expressions including neutral expression. The average recognition rate they achieved was 91.3%. Mpiperis et al. [10] proposed a joint 3D face and facial expression recognition using bilinear model. They fitted both formulations, using symmetric and asymmetric bilinear models to encode both identity and expression. They reported an average recognition rate of 90.5%. They also reported that the facial expressions of disgust and surprise were well identified with an accuracy of 100%. Tang and Huang [11] proposed an automatic feature selection computed from the normalized Euclidean distances between two picked landmarks from 83 possible ones. Using regularized multi-class Adaboost classification algorithm, they reported an average recognition rate of 95.1%, and they mentioned that the surprise expression was recognized with an accuracy of 99.2%.

3. Database description

BU-3DFE is one of the very few publicly available databases of annotated 3D facial expressions, collected by Yin et al. [12] at Binghamton University. It was designed for research on 3D human face and facial expression and to develop a general understanding of the human behavior. Thus the BU-3DFE database is beneficial for several fields and applications dealing with human computer interaction, security, communication, psychology, etc. There are a total of 100 subjects in the database, 56 females and 44 males. A neutral scan was captured for each subject, then they were asked to perform six expressions namely: Happiness (HA), Anger (AN), Fear (FE), Disgust (DI), Sad (SA) and Surprise (SU). The expressions vary according to four levels of intensity (low, middle, high and highest or 01–04). Thus, there are 25 3D facial expression models per subject in the database. A set of 83 manually annotated facial landmarks is associated to each model. These landmarks are used to define the regions of the face that undergo specific deformations due to single muscles movements when conveying facial expression [7]. In Fig. 1, we illustrate examples of the six universal facial expressions 3D models including the highest intensity level.

In this paper, we further investigate the problem of 3D identity-independent facial expression recognition. The main contributions of our approach are the following: (1) We propose a new process for representing and extracting patches on the facial surface scan that cover multiple regions of the face and (2) we apply a framework to derive 3D shape analysis to quantify similarity measure between corresponding patches on different 3D facial scans. Thus, we combine a local geometric-based shape analysis approach of 3D faces and several machine-learning techniques to perform such classification.

4. 3D facial patches-based representation

Most of the earlier work in 3D shape analysis use shape descriptors such as curvature, crest lines, shape index (e.g., ridge, saddle, rut, dome, etc.). These descriptors are defined based on the geometric and topological properties of the 3D object, and are used as features to simplify the representation and thus the
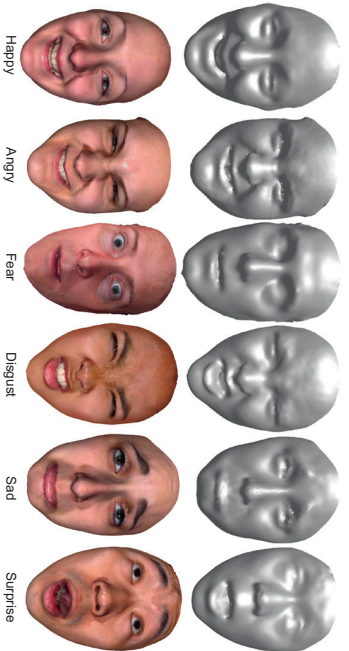

Fig. 1. Examples of 3D facial expression models (first row 3D shape models, second row 3D textured models) of the BU-3DFE database.

Happy   Angry   Fear   Disgust   Sad   Surprise

comparison for 3D shape matching and recognition tasks. Despite their rigorous definition, such features are computed based on numerical approximation that involves second derivatives and can be sensitive to noisy data. In case of 3D range models, the facial surface labeling is a critical step to describe the facial behavior or expression, and a robust facial surface representation is needed. In Samir et al. [13] the authors proposed to represent facial surfaces by an indexed collections of 3D closed curves on faces. These curves are level curves of a surface distance function (i.e. geodesic distance) defined to be the length of the shortest path between a fixed reference point (taken to be the nose tip) and a point of the extracted curve along the facial surface. This being motivated by the robustness of the geodesic distance to facial expressions and rigid motions. Using this approach they were able to compare 3D shapes by comparing facial curves rather than comparing corresponding shape descriptors.

In our work we intend to further investigate on local shapes of the facial surface. We are especially interested in capturing deformations of local facial regions caused by facial expressions. Using a different solution, we compute curves using the Euclidean distance which is sensitive to deformations and thus can better capture differences related to variant expressions. To this end, we choose to consider N reference points (landmarks) $\{r_i\}_{1 \leq i \leq N}$ (Fig. 2(a)) and associated sets of level curves $\{c_{\lambda}^i\}_{1 \leq \lambda \leq \lambda_0}$ (Fig. 2(b)). Here r_i stands for the value of the distance function between the reference point r_i and the point belonging to the curve c_{λ}^i, and λ_0 stands for the maximum value taken by λ. Accompanying each facial model there are 83 manually picked landmarks, these landmarks are practically similar to the MPEG-4 feature points and are selected based on the facial anatomy structure. Given these points the feature region on the face can be easily determined and extracted. We were interested in a subset of 68 landmarks laying within the face area, discarding those marked on the face border. Contrary to the MPEG-4 feature points specification that annotates the cheeks center and bone, in BU-3DFE there were no landmarks associated with the cheek regions. Thus, we add two extra landmarks at both cheeks, obtained by extracting the middle point along the geodesic path between the mouth corner and the outside eye corner.

We propose to represent each facial scan by a number of patches centered on the considered points. Let r_i be the reference point and P_i a given patch centered on this point and localized on the facial surface denoted by S. Each patch will be represented by an indexed collection of level curves. To extract these curves, we

Fig. 2. (a) 3D annotated facial shape model (70 landmarks); (b) 3D closed curves extracted around the landmarks; (c) 3D curve-based patches composed of 20 level curves with a size fixed by a radius $\lambda=20$ mm; (d) extracted patches on the face.

● Discarded landmarks provided by BU-3DFE
● Retained landmarks provided by BU-3DFE
● Automatically added landmarks

use the Euclidean distance function $\|r_i - p\|$ to characterize the length between r_i and any point p on S. Using this function we defined the curves as level sets of

$$\|r_i - p\| : c_{\lambda}^i = \{p \in S | \|r_i - p\| = \lambda\} \subset S, \quad \lambda \in (0, \lambda_0]. \quad (1)$$

Here c_{λ}^i is a closed curve, consisting of a collection of points situated at an equal distance λ from r_i. Fig. 2 resumes the scheme of patches extraction.

5. Framework for 3D shape analysis

Once the patches are extracted, we aim at studying their shape and design a similarity measure between corresponding ones on different scans under different expressions. This is motivated by the common belief that people smile, or convey any other expression, the same way, or more appropriately certain regions taking part in a specific expression undergo practically the same dynamical deformation process. We expect that certain corresponding patches associated with the same given expression will be deformed in a similar way, while those associated with two different expressions will deform differently. The following sections describe the shape analysis of closed curves in $\mathbb{R}^3$, initially introduced by Joshi et al. [3], and its extension to analyze shape of local patches on facial surfaces.

5.1. 3D curve shape analysis

We start by considering a closed curve β in $\mathbb{R}^3$. While there are several ways to analyze shapes of closed curves, an elastic analysis of the parametrized curves is particularly appropriate in 3D curves analysis. This is because (1) such analysis uses a square-root velocity function representation which allows us to compare local facial shapes in presence of elastic deformations, (2) this method uses a square-root representation under which the elastic metric reduces to the standard $\mathbb{L}^2$ metric and thus simplifies the analysis, (3) under this metric the Riemannian distance between curves is invariant to the re-parameterization. To analyze the shape of β, we shall represent it mathematically using a square-root representation of β as follows: for an interval $I=[0,1]$, let $\beta : I \rightarrow \mathbb{R}^3$ be a curve and define $q : I \rightarrow \mathbb{R}^3$ to be its

square-root velocity function (SRVF), given by:

$$q(t) = \frac{\dot{\beta}(t)}{\sqrt{\|\dot{\beta}(t)\|}}. \quad (2)$$

Here t is a parameter $\in I$ and $\|.\|$ is the Euclidean norm in $\mathbb{R}^3$. We note that $q(t)$ is a special function that captures the shape of β and is particularly convenient for shape analysis, as we describe next. The classical elastic metric for comparing shapes of curves becomes the $\mathbb{L}^2$-metric under the SRVF representation [14]. This point is very important as it simplifies the calculus of elastic metric to the well-known calculus of functional analysis under the $\mathbb{L}^2$-metric. Also, the squared $\mathbb{L}^2$-norm of q, given by: $\|q\|^2 = \int_{S^1} \langle q(t), q(t) \rangle > dt = \int_{S^1} \|\dot{\beta}(t)\| dt$, which is the length of the curve β. In order to restrict our shape analysis to closed curves, we define the set: $C = \{q : S^1 \rightarrow \mathbb{R}^3 | \int_{S^1} q(t)|q(t)| dt = 0\} \subset \mathbb{L}^2(S^1, \mathbb{R}^3)$. Notice that the elements of C are allowed to have different lengths. Due to a non-linear (closure) constraint on its elements, C is a non-linear manifold. We can make it a Riemannian manifold by using the metric: for any $u, v \in T_q(C)$, we define:

$$\langle u, v \rangle = \int_{S^1} \langle u(t), v(t) \rangle \, dt. \quad (3)$$

So far we have described a set of closed curves and have endowed it with a Riemannian structure. Next we consider the issue of representing the shapes of these curves. It is easy to see that several elements of C can represent curves with the same shape. For example, if we rotate a curve in $\mathbb{R}^3$, we get a different SRVF but its shape remains unchanged. Another similar situation arises when a curve is re-parameterized: a re-parameterization changes the SRVF of curve but not its shape. In order to handle this variability, we define orbits of the rotation group $SO(3)$ and the re-parameterization group Γ as the equivalence classes in C. Here, Γ is the set of all orientation-preserving diffeomorphisms of S^1 (to itself) and the elements of Γ are viewed as re-parameterization functions. For example, for a curve $\beta : S^1 \rightarrow \mathbb{R}^3$ and a function $\gamma : S^1 \rightarrow S^1$, $\gamma \in \Gamma$, the curve $\beta \circ \gamma$ is a re-parameterization of β. The corresponding SRVF changes according to $q(t) \rightarrow \sqrt{\dot{\gamma}(t)}$ $q(\gamma(t))$. We set the elements of the orbit:

$$[q] = \left\{ \sqrt{\dot{\gamma}(t)} O q(\gamma(t)) | O \in SO(3), \gamma \in \Gamma \right\}, $$ (4)

to be equivalent from the perspective of shape analysis. The set of such equivalence classes, denoted by $S = C/(SO(3) \times \Gamma)$ is called the shape space of closed curves in $\mathbb{R}^3$. S inherits a Riemannian metric from the larger space C due to the quotient structure.

The main ingredient in comparing and analysing shapes of curves is the construction of a geodesic between any two elements of S, under the Riemannian metric given in Eq. (3). Given any two curves β_1 and β_2, represented by their SRVFs q_1 and q_2, we want to compute a geodesic path between the orbits $[q_1]$ and $[q_2]$ in the shape space S. This task is accomplished using a path-straightening approach which was introduced in [15]. The basic idea here is to connect the two points $[q_1]$ and $[q_2]$ by an arbitrary initial path α and to iteratively update this path using the negative gradient of an energy function $E[\alpha] = \frac{1}{2} \int_s \langle \dot{\alpha}(s), \dot{\alpha}(s) \rangle \, ds$. The critical points of E are actually geodesic paths in S. Thus, this gradient-based update leads to a critical point of E shown in [15], the critical points of E are actually geodesic paths derived analytically and can be used directly for updating α. As $\alpha(s)$ denote the geodesic path between the given points. In the remainder of the paper, we will use the notation $d_s(\beta_1, \beta_2)$ to denote the length of the geodesic in the shape space S between the orbits q_1 and q_2, to reduce the notation.

5.2. 3D patches shape analysis

Now, we extend ideas developed in the previous section from analyzing shapes of curves to the shapes of patches. As mentioned earlier, we are going to represent a number of patches of a facial surface S with an indexed collection of the level curves of the $\|r_i - p\|$ function (Euclidean distance from the reference point r_i). That is, $P_i \leftrightarrow \{c_{\lambda}^i, \lambda \in (0, \lambda_0]\}$, where c_{λ}^i is the level set associated with $\|r_i - p\| = \lambda$. Through this relation, each patch has been represented as an element of the set $S^{M_{\lambda 0}}$. In our framework, the shapes of any two patches are compared by comparing their corresponding level curves. Given any two patches P_1 and P_2, and their level curves $\{c_{\lambda}^1, \lambda \in (0, \lambda_0]\}$ and $\{c_{\lambda}^2, \lambda \in (0, \lambda_0]\}$, respectively, our idea is to compare the patches curves c_{λ}^1 and c_{λ}^2, and to accumulate these differences over all λ. More formally, we define a distance $d_{SP_{\lambda 0}}$ given by:

$$d_{SP_{\lambda 0}}(P_1, P_2) = \int_0^{\lambda_0} d_s(c_{\lambda}^1, c_{\lambda}^2) \, d\lambda. \quad (5)$$

In addition to the distance $d_{SP_{\lambda 0}}(P_1, P_2)$, which is useful in biometry and other classification experiments, we also have a geodesic path in $S^{M_{\lambda 0}}$ between the two points represented by P_1 and P_2. This geodesic corresponds to the optimal elastic deformations of these curves and, thus, facial surfaces from one to another. Fig. 3 shows some examples of geodesic paths that are computed between corresponding patches associated with shape models sharing the same expression, and termed intra-class geodesics. In the first column we illustrate the source, which represents scan models of the same subject, but under different expressions. The third column represents the targets as scan models of different subjects. As for the middle column, it shows the geodesic path from the source to the target. The mean curvature mapping representations of the patches along the geodesic paths. In each row we have both the shape and the curvature representation is added to identify concave/convex areas on the source and target patches and equally spaced steps of geodesics. This figure shows that certain patches, belonging to the same class of expression, are deformed in a similar way. In contrast, Fig. 4 shows geodesic paths between patches of different facial expressions. These geodesics are termed inter-class geodesics. Unlike the intra-class geodesics shown in Fig. 3, these patches deform in a different way.

6. Feature vector generation for classification

In order to classify expressions, we build a feature vector for each facial scan. Given a candidate facial scan of a person j, facial patches are extracted around facial landmarks. For a facial patch P_j^i, a set of level curves $\{c_{\lambda}^i\}_i$ are extracted centered on the ith landmark. Similarly, a patch P_{ref}^i is extracted in correspondence to landmarks of a reference scan ref. The length of the geodesic path between each level curve and its corresponding curve on the reference scan is computed using a Riemannian framework for shape analysis of 3D curves (see Sections 5.1 and 5.2). The length of the geodesic path between two patches P_j^i and P_{ref}^i is then formed by the distances computed on all the patches and its dimension is equal to the number of used landmarks $N=70$ (i.e. 68 landmarks provided by BU-3DFE and the two additional cheek points). The ith element of this vector represents the length of the geodesic path that separates each patch to the corresponding one on the reference face scan. All feature vectors computed on the overall dataset will be labeled and used as input data to
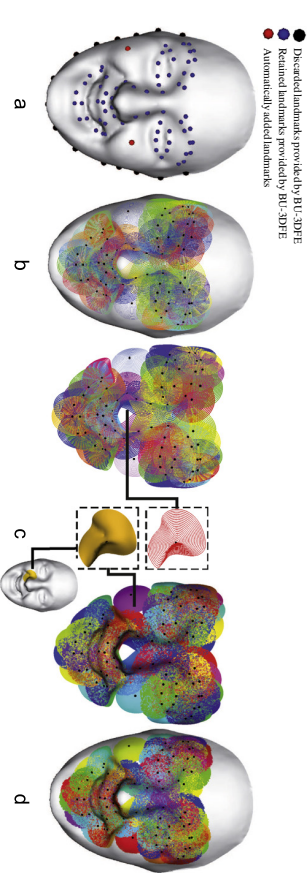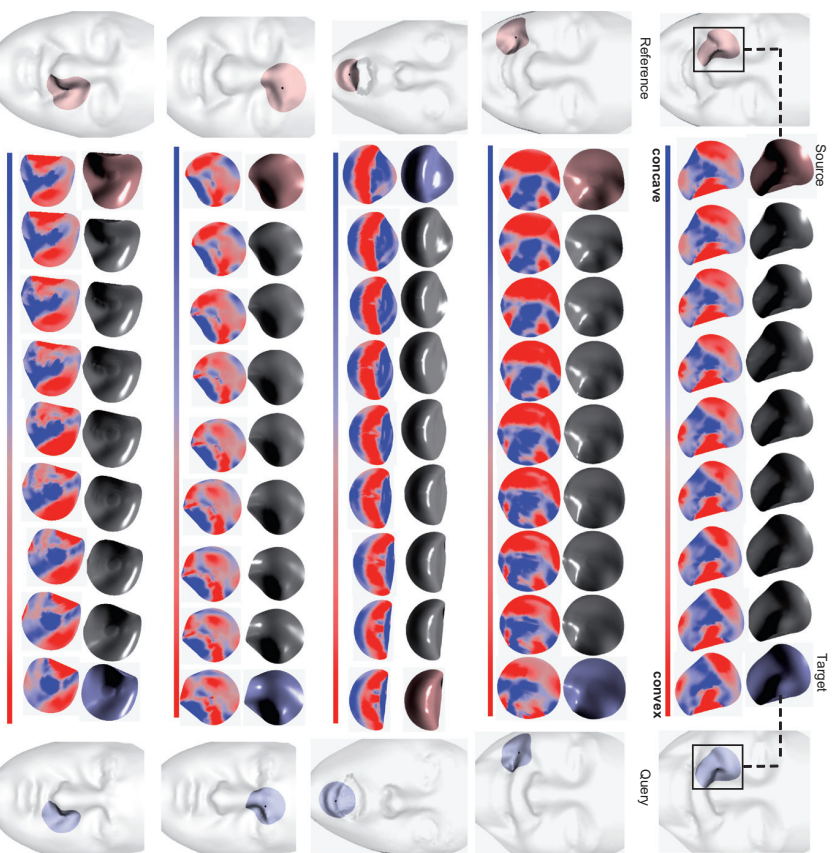
Fig. 3. Examples of intra-class (same expression) geodesic paths with shape and mean curvature mapping between corresponding patches.

machine-learning algorithms such as multiboosting and SVM, where multiboosting is an extension of the successful Adaboost technique for forming decision committees.

7. Recognition experiments

To investigate facial expression recognition, we have applied our proposed approach on a dataset that is appropriate for this task. In this section, we describe the experiments, obtained results and comparisons with related work.

7.1. Experimental setting

For the goal of performing identity-independent facial expression recognition, the experiments were conducted on the BU-3DFE static

database. A dataset captured from 60 subjects were used, half (30) of them were female and the other half (30) were male, corresponding to the high and highest intensity levels 3D expressive models (03–04). These data are assumed to be scaled to the true physical dimensions of the captured human faces. Following a similar setup as in [16], we randomly divided the 60 subjects into two sets, the training set containing 54 subjects (648 samples), and the test set containing six subjects (72 samples).

To drive the classification experiments, we arbitrarily choose a set of six reference subjects with its six basic facial expressions. We point out that the selected reference scans do not appear neither in the training nor in the testing set. These references, shown in Fig. 5, with their relative expressive scans corresponding to the highest intensity level, are taken to play the role of representative models for each of the six classes of expressions. For each reference subject, we derive a facial expression recognition experience.
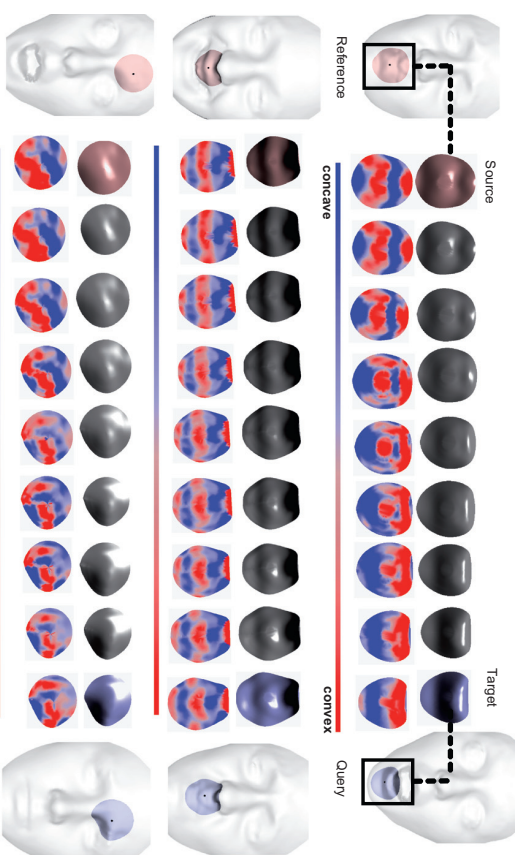

7.2. Discussion of the results

Several facial expression recognition experiments were conducted with changing at each time the reference Fig. 5 illustrates the selected references (neutral scan). Using the *Waikato Environment for Knowledge Analysis (Weka)* [17], we applied the multiboost algorithm with three weak classifiers, namely, Linear Discriminant Analysis (LDA), Naive Bayes (NB), and Nearest Neighbor (NN), to the extracted features, and we achieved average recognition rates of 98.81%, 98.76% and 98.07%, respectively. We applied the SVM linear classifier as well, and we achieved an average recognition rate of 97.75%. We summarize the resulting recognition rates in Table 1.

We note that these rates are obtained by averaging the results of the 10 independent and arbitrarily run experiments (10-fold

cross-validation) and their respective recognition rate obtained using the multiboost-LDA classifier. We note that different selections of the reference scans do not affect significantly the recognition results and there is no large variations in recognition rates values. The reported results represent the average over the six runned experiments. The multiboost-LDA classifier achieves the highest recognition rate and shows a better performance in

Fig. 4. Examples of inter-class (different expressions) geodesic paths between source and target patches.

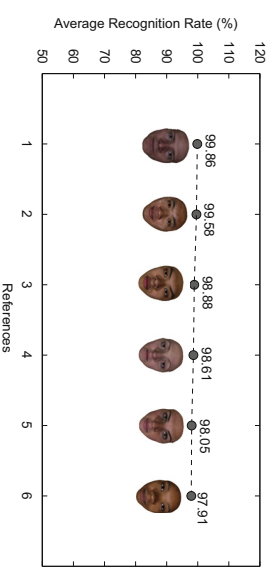**Fig. 5.** Different facial expression average recognition rates obtained using different reference subjects (using multiboost-LDA).

Table 1
Classification results using different classifiers.

Classifier	Multiboost-LDA	Multiboost-NB	Multiboost-NN	SVM-Linear
Recognition rate	**98.81%**	98.76%	98.07%	97.75%

terms of accuracy than the other classifiers. This is mainly due to the capability of the LDA-based classifier to transform the features into a more discriminative space and, consequently, result in a better linear separation between facial expression classes.

The average confusion matrix relative to the best performing classification using multiboost-LDA is given in Table 2.

In order to better understand and explain the results mentioned above, we apply the multiboost algorithm on feature vectors built from distances between patches for each class of expression. In this case, we consider these features as weak classifiers. Then, we look at the early iterations of the multiboost algorithm and the selected patches in each iteration.

Fig. 6 illustrates for each class of expression the most relevant patches. Notice that, for example, for the Happy expression the selected patches are localized in the lower part of the face, around the mouth and the chin. As for the Surprise expression, we perform facial patches extraction around the visible landmarks in the given scan. In cases where a landmark is occluded, or where the landmark is visible, but the region nearby is partially occluded, we treat it as a missing data problem for all faces sharing this view. In these cases, we are not able to compute the geodesic path between corresponding patches. The corresponding entries in the distance matrix are blank and we fill them using an imputation technique [18]. In our experiments we employed the mean imputation method, which consists of replacing the missing values by the means of values already calculated in frontal view scenario by the means of the training set. Let $d_{jk} = g_{geo}(P_j^*, P_k^*)$ be the geodesic distance between the kth patch belonging to subjects i and j ($i \neq j$). In case of frontal view X_i^{fv} relative to the subject i need to be labeled and is given by:

$$X_i^{fv} = \begin{pmatrix} d_{i11} & \cdots & d_{i1k} & \cdots & d_{i1N} \\ \vdots & & \vdots & & \vdots \\ d_{ij1} & \cdots & d_{ijk} & \cdots & d_{ijN} \\ \vdots & & \vdots & & \vdots \\ d_{iJ1} & \cdots & d_{iJk} & \cdots & d_{iJN} \end{pmatrix}$$

where N is the number of attributes. In case of non-frontal view (X_i^{fv}), if an attribute k is missing, we replace the kth column vector in the distance matrix X_i^{nfv} by the mean of geodesic distances computed in the frontal view case, with respect to the kth attribute and given by: $m_k^{fv} = \sum_{j=1}^{J} d_{jk}/J$, where J is the total number of instances.

7.3. Comparison with related work

Table 3 results of our approach are compared against those reported in [11,9,8], on the same experimental setting (54 versus 6 subject partitions) of the BU-3DFE database. The differences between approaches should be noted: Tang et al. [11] performed automatic feature selection using normalized Euclidean distances between 83 landmarks, Soyel et al. [9] calculated six distances using a distribution of 11 landmarks, while Wang et al. [8] derived curvature estimation by locally approximating the 3D surface with a smooth polynomial function. In comparison, our approach capture the 3D shape information of local facial patches to derive shape analysis. For assessing how the results of their statistical analysis will generalize to an independent dataset, in [8] a 20-fold cross-validation technique was used, while

in [11,9] the authors used 10-fold cross-validation to validate their approach.

7.4. Non-frontal view facial expression recognition

In real world situations, frontal view facial scans may not be always available. Thus, non-frontal view facial expression recognition is a challenging issue that needs to be treated. We were interested in evaluating our approach on facial scan under large pose variations. By rotating the 3D shape models in the y-direction, we generate facial scans under six different non-frontal views corresponding to 15°, 30°, 45°, 60°, 75°, and 90° rotation. In cases where a landmark is occluded, or where the landmark is visible, but the region nearby is partially occluded, we treat it as a missing data problem for all faces sharing this view. In these cases, we are not able to compute the geodesic path between corresponding patches. The corresponding entries in the distance matrix are blank and we fill them using an imputation technique [18].

number of instances.

$$X_i^{nfv} = \begin{pmatrix} d_{i11} & \cdots & m_k^{fv} & \cdots & d_{i1N} \\ \vdots & & \vdots & & \vdots \\ d_{ij1} & \cdots & m_k^{fv} & \cdots & d_{ijN} \\ \vdots & & \vdots & & \vdots \\ d_{iJ1} & \cdots & m_k^{fv} & \cdots & d_{iJN} \end{pmatrix}$$

7.5. Sensitivity to landmarks mis-localization

It is known that the automatic 3D facial feature points detection is a challenging problem. The most difficult task remains the localization of points around the eyebrow regions, which appear to play an important role in the expression of emotions. The effect of the mis-localization of the landmarks has been addressed in a specific experiment. We considered the eyebrow regions in that specific experiment. We considered the eyebrow regions in that the points in these regions are expected to be the most difficult to detect automatically. In these regions, we added noise to the landmarks provided with the BU-3DFED. In particular, we added noise to the position of the landmarks by moving them randomly in a region with a radius of 10 mm, as illustrated Fig. 8 by blue circles. Then we performed expression recognition experiments with such noisy landmarks. The results are reported in Fig. 8. It can be noted that with the multiboost-LDA algorithm the lower decrease in the recognition rate is observed, and even with a recognition rate equal to 85.64% the result still outperforms the one reported in Wang et al [8].

To evaluate the robustness of our approach in a context of non-frontal views, we derive a view-independent facial expression recognition. Error recognition rates are evaluated throughout different testing facial views using the four classifiers trained only on frontal view facial scans. Fig. 7 shows the average error rates of the four classification methods. The multiboost-LDA shows the best performance for facial expression classification on the average error rates increase with the rotation angle (values from 0° to 90° of rotation are considered), and the multiboost-LDA recognition rate is acceptable (over 93%) even half of the facial scan is missed.

8. Conclusions

In this paper we presented a novel approach for identity-independent facial expression recognition from 3D facial shapes. Our idea was to describe the change in facial expression as a deformation in the vicinity of facial patches in 3D shape scan. An automatic extraction of local curve-based patches within the 3D facial surfaces was proposed. These patches were used as local shape descriptors for facial expression representation. A Riemannian framework was applied to compute the geodesic path between corresponding patches. Qualitative (inter- and intrageodesic paths) and quantitative (geodesic distances) measures of geodesic distances between patches were labeled with respect to the six prototypical expressions and used as samples to train and test machine-learning algorithms. Using multiboost algorithm for multi-class classification, we achieved a 98.81% average recognition rate for six prototypical facial expressions on the BU-3DFE database. We demonstrated the robustness of the proposed method to pose variations. In fact, the obtained recognition rate remains acceptable (over 93%) even half of the facial scan is missed.

Table 2
Average confusion matrix given by multiboost-LDA classifier.

%	AN	DI	FE	HA	SA	SU
AN	**97.92**	1.11	0.14	0.14	0.69	0.0
DI	0.56	**99.16**	0.14	0.0	0.14	0.0
FE	0.14	0.14	**99.72**	0.0	0.0	0.0
HA	0.56	0.14	0.0	**98.60**	0.56	0.14
SA	0.28	0.14	0.0	0.0	**99.30**	0.28
SU	0.14	0.56	0.0	0.0	1.11	**98.19**

Table 3
Comparison of this work with respect to previous work [11,9,8].

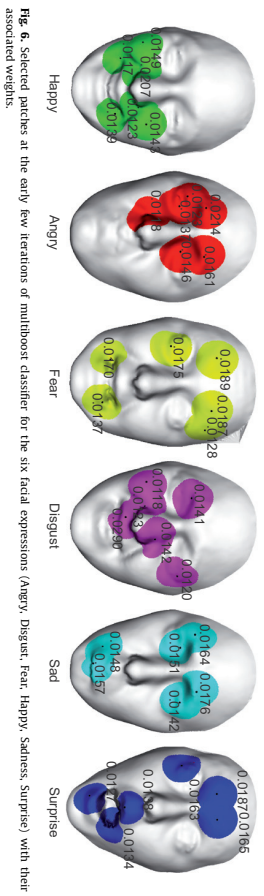| Cross-validation	This work	Tang et al. [11]	Soyel et al. [9]	Wang et al. [8]
10-fold	98.81%	95.1%	91.3%	–
20-fold	92.75%	–	–	83.6%

Fig. 6. Selected patches at the early few iterations of multiboost classifier for the six facial expressions (Angry, Disgust, Fear, Happy, Sadness, Surprise) with their associated weights.
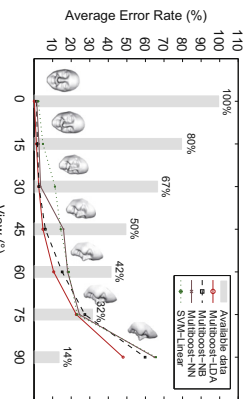

Fig. 7. The average error rates of six expressions with different choices of views corresponding to the best reference and using different classifiers.

Fig. 8. Recognition experiment performed adding noise to the eyebrow landmarks (random displacement). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
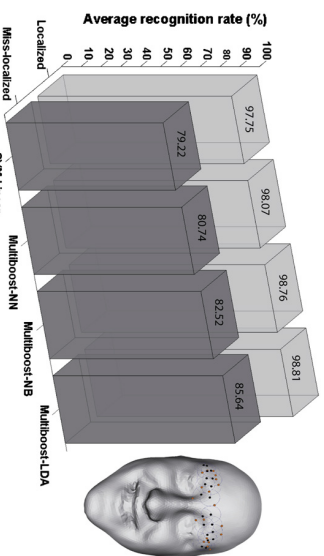
The major limitation of our approach is that the 68 landmarks we used to define the facial patches were manually labeled. For our future work we are interested in detecting and tracking facial feature points, as proposed in [19,20], for automatic 3D facial expression recognition.

References

[1] P. Ekman, T.S. Huang, T.J. Sejnowski, J.C. Hager, Final report to nsf of the planning workshop on facial expression understanding, Technical Report, 1992.

[2] P. Ekman, W.V. Friesen, Constants Across Cultures in the Face and Emotion, 1971.

[3] S. Joshi, E. Klassen, A. Srivastava, I.H. Jermyn, A novel representation for Riemannian analysis of elastic curves in $\mathbb{R}^n$, in: Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR), 2007.

[4] M. Pantic, L. Rothkrantz, Automatic analysis of facial expressions: the state of the art, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1424–1445.

[5] A. Samal, P.A. Iyengar, Automatic recognition and analysis of human faces and facial expressions: a survey, Pattern Recognition 25 (1) (1992) 65–77.

[6] J. Whitehill, C.W. Omlin, Local versus global segmentation for facial expression recognition, in: FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, 2006, pp. 357–362.

[7] P. Ekman, W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, 1978.

[8] J. Wang, L. Yin, X. Wei, Y. Sun, 3D facial expression recognition based on primitive surface feature distribution, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006, pp. 1399–1406.

[9] H. Soyel, H. Demirel, Facial expression recognition using 3d facial feature distances, in: International Conference on Image Analysis and Recognition (ICIAR), 2007, pp. 831–838.

[10] I. Mpiperis, S. Malassiotis, M.G. Strintzis, Bilinear models for 3d face and facial expression recognition, IEEE Transactions on Information Forensics and Security 3 (3) (2008) 498–511.

[11] H. Tang, T. Huang, 3D facial expression recognition based on automatically selected features, in: First IEEE Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB), 2008, pp. 1–8.

[12] L. Yin, X. Wei, Y. Sun, J. Wang, M.J. Rosato, A 3d facial expression database for facial behavior research, in: FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, 2006, pp. 211–216.

[13] C. Samir, A. Srivastava, M. Daoudi, E. Klassen, An intrinsic framework for analysis of facial surfaces, International Journal of Computer Vision 82 (1) (2009) 80–95.

[14] A. Srivastava, E. Klassen, S.H. Joshi, I.H. Jermyn, Shape analysis of elastic curves in Euclidean spaces, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 99, IEEE Computer Society, Los Alamitos, CA, USA, 2010. 〈http://doi.ieeecomputersociety.org/10.1109/TPAMI.2010.184〉.

[15] E. Klassen, A. Srivastava, Geodesics between 3d closed curves using path-straightening, in: ECCV, vol. 1, 2006, pp. 95–106.

[16] B. Gong, Y. Wang, J. Liu, X. Tang, Automatic facial expression recognition on a single 3d face by exploring shape deformation, in: Proceedings of the ACM International Conference on Multimedia, Beijing, China, 2009, pp. 569–572.

[17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The weka data mining software: an update, SIGKDD Explorations Newsletters 11 (2009) 10–18.

[18] G. Batista, M.C. Monard, An analysis of four missing data treatment methods for supervised learning, Applied Artificial Intelligence 17 (2003) 519–533.

[19] S. Gupta, M.K. Markey, A.C. Bovik, Anthropometric 3D face recognition, International Journal of Computer Vision 90 (3) (2010) 331–349.

[20] Y. Sun, X. Chen, M. Rosato, L. Yin, Tracking vertex flow and model adaptation for three-dimensional spatiotemporal face analysis, IEEE Transactions on Systems, Man, and Cybernetics – Part A 40 (3) (2010) 461–474.

Ahmed Maalej is currently a Ph.D. candidate within the Fundamental Computer Science laboratory of Lille (LIFL UMR 8022), France. He obtained the M.S. degree in Telecommunications from the Higher School of Communications of Tunis (SUPCOM), Tunisia, in 2008, and the electrical engineering degree from the National Engineering School of Monastir (ENIM), Tunisia, in 2005. His main research interests focus on 3D facial expression recognition.

Boulbaba Ben Amor received the M.Sc. degree in 2003 and the Ph.D. degree in Computer Science in 2006, both from Ecole Centrale de Lyon, France. Currently, he is an associate-professor in Institut Telecom/Telecom Lille 1. He is also a member of the Computer Science Laboratory in University of Lille 1 (LIFL UMR CNRS 8022). His research interests are mainly focused on statistical three-dimensional face analysis and recognition and facial expression recognition using 3D. He is co-author of several papers in refereed journals and proceedings of international conferences. He has been involved in French and international projects and has served as program committee member and reviewer for international journals and conferences.

Mohamed Daoudi is a Full Professor of Computer Science in the Institut TELECOM ; TELECOM Lille 1. He received the Ph.D. degree in Computer Engineering from the University Lille 1, France in 1993 and Habilitation à Diriger des Recherches (HDR) from the University of Littoral, France, in 2000. He is the founder and the head of the MIIRE research group of LIFL (UMR CNRS 8022). His research interests include pattern recognition, image processing, invariant representation of images and shapes, three-dimensional analysis and retrieval and more recently 3D face recognition. He has published more than 80 papers in refereed journals and proceedings of international conferences. He is the author the book 3D Processing: Compression, Indexing and Watermarking (Wiley, 2008). He has served as a Program Committee member for the International Conference on Pattern Recognition (ICPR) in 2004 and the International Conference on Multimedia and Expo (ICME) in 2004 and 2005. He is a frequent reviewer for IEEE Transactions on Pattern Analysis and Machine Intelligence and for Pattern Recognition Letters. His research has been funded by ANR, RNRT and European Commission grants. He is Senior Member of IEEE.

Anuj Srivastava is a Professor of Statistics at Florida State University in Tallahassee, FL. He obtained his MS and Ph.D. degrees in Electrical Engineering from Washington University in St. Louis in 1993 and 1996, respectively. After spending the year 1996–1997 at Brown University as a visiting researcher, he joined FSU as an Assistant Professor in 1997. He has received the Developing Scholar and the Graduate Faculty Mentor Awards at FSU. His research is focused on pattern theoretic approaches to problems in image analysis, computer vision, and signal processing. He has developed computational tools for performing statistical inferences on certain non-linear manifolds, in particular the shape spaces of curves and surfaces. He has published over 120 journal and conference articles in these areas.

Stefano Berretti received the Laurea Degree in Electronics Engineering, the Laurea Degree in Multimedia Content Design, and the Ph.D. degree in Information and Telecommunications engineering in 1997, 2000, and 2001, respectively, from the University of Florence, Italy, where, since 2002, he has been an assistant professor and teaches "Operating Systems" and "Fundamentals of Computer Programming" in the School of Computer Engineering, and since 2001, he has also been teaching "Database Systems" in the postdoctoral school in "Multimedia Content Design". His scientific interests are pattern recognition, content-based image retrieval, 3D object partitioning and retrieval, and 3D face recognition.

A Riemannian Analysis of 3D Nose Shapes For Partial Human Biometrics

Hassen drira
LIFL (UMR USTL/CNRS 8022)
Université de Lille 1, France.
hassen.drira@lifl.fr

Boulbaba Ben Amor
Institut TELECOM ; TELECOM Lille 1, France.
LIFL (UMR USTL/CNRS 8022)
http://www.telecom-lille1.eu/people/benamor/

Anuj Srivastava
Departement of Statistics
FSU, Tallahassee, FL 32306, USA.
http://stat.fsu.edu/~anuj/

Mohamed Daoudi
Institut TELECOM ; TELECOM Lille 1, France.
LIFL (UMR USTL/CNRS 8022)
http://www.telecom-lille1.eu/people/daoudi/

Abstract

In this paper we explore the use of shapes of noses for performing partial human biometrics. The basic idea is to represent partial human surfaces using indexed collections of iso-curves, and to analyze shapes of noses by comparing their corresponding curves. We extend past work in Riemannian analysis of shapes of closed curves in $\mathbb{R}^3$ to obtain a similar Riemannian analysis for nasal surfaces. In particular, we obtain algorithms for computing geodesics, computing statistical means, and stochastic clustering. We demonstrate these ideas in two application contexts : authentication and identification. We evaluate performances on a large database involving 2000 scans from FRGC v2 database, and present a hierarchical organization of nose databases to allow for efficient searches.

1. Introduction

Human biometrics has become an area of tremendous importance and potential. Although its growth in recent years have been motivated by security applications, one can safely expect an exponential growth in a general use of biometrics in our increasingly digital society. By human biometrics we mean the use of physiological characteristics, of human body parts and their appearances, to identify individual human beings in the course of their daily activities. The appearances of body parts, especially in imaged data, have a large variability and are influenced by their shapes, colors, illumination environment, presence of other parts, and so on. Therefore, the biometrics researchers have focused on body parts and images that try to minimize this variability within class (subjects) and maximize it across classes. Although several modalities, such as fingerprints,

face images, iris, and gait, have been tried, none of them by itself may lead to a generic solution for a large population. Instead, an emerging consensus to develop a suite of different features (characteristics) and modalities, where each feature provides certain a partial information, such that their joint analysis may succeed for large populations. This has led to the notion of **partial biometrics** – the contribution of an individual feature/modality in identifying human beings.

The use of shapes of facial parts is an important example of this idea. Since 2D (visible light) images of faces are greatly susceptible to variations in the imaging environments (camera pose, illumination patterns, etc.), the researchers have argued for the need to use 3D face data, typically collected by laser scanners, for studying shapes of peoples' faces and using this data for biometrics. The output from laser scanners are minimally dependent on external environmental factors and provide faithful measurements of shapes of facial parts. Perhaps the only remaining variability that is manifested within the same class, i.e. within the same person, is the one introduced by the changes in facial expressions. Facial expressions, such as smile, fear, and anger, etc, are prime indicators of the emotional state of a person. While they are important in estimating the mood of a person, for example in developing intelligent ambient systems, they have a lesser role in biometric applications. In fact, changes in facial expressions change the shapes of facial parts to some extent. We argue that this variability has become one of the most important issue in 3D face recognition as described in [2] and [9]. The other important challenge relates to data collection and imperfections introduced in that process. It is difficult to obtain pristine, continuous facial surfaces, or meshes representing such surfaces, with the current laser technology. One typically gets holes in the scanned data in locations of eyes, lips, and outside regions. These factors lead to a decrease

in biometric performance. To handle these issues – shape variability due to facial expressions and presence of holes (missing data), we advocate the use of **facial parts** individually, especially for partial biometrics. The use of facial parts for biometrics is not new. For example, [16] and [3] have studied several different regions on a facial shape does not change with facial expressions. Similarly, Faltemier et al. [5] have studied several different regions on a facial surface and have compared their contributions. We are not aware of any past work in an elastic (Riemannian) analysis of facial parts.

At the outset, the shape of the nose seems like a bad choice of feature for biometrics. The shapes of noses seem very similar to a human observer but we will support this choice using real data and automated techniques for shape analysis. We reiterate that this framework may not be sufficient for identifying human subjects across a vast population, but we argue for its role in shortlisting possible hypotheses so that a reduced hypothesis set can be evaluated using a more elaborate, multi-modal biometric system. The stability of nose data collection, the efficiency of nasal shape analysis, and the invariance of nasal shape to changes in facial expressions make it an important partial biometric.

In recent years, there have been several papers for studying shapes of continuous curves, the earlier papers were mainly concerned with curves in $\mathbb{R}^2$ [11, 12]. In this paper, we will follow the theory laid out by Joshi et al [7, 8] for *elastic shape analysis* of continuous, closed curves in $\mathbb{R}^n$.

2. Riemannian Framework

Our approach is to analyze shapes of nasal surfaces using shapes of iso-curves. In other words, we divide each surface into an indexed collection of simple, closed curves in $\mathbb{R}^3$ and the geometry of a surface is then studied using the geometries of the associated curves. Since these curves, called *nasal curves*, have been defined as level curves of an intrinsic distance function on the surface, their geometries in turn are invariant to the rigid transformation (rotation and translation) of the original surface. At least theoretically, these curves jointly contain all the information about the surface and one can go back-and-forth between the surface and the curves without any ambiguity. In practice, however, some information is lost when one works with a finite subset of these curves rather than the full continuum.

2.1. Nasal Curves

Let N be a nasal surface, output of our preprocessing step (removing spikes, filling holes, and cropping nasal region) illustrated in Figure 1. Although, in practice, N is a triangulated mesh, we start the discussion by assuming that it is a continuous surface. Let $D : N \to \mathbb{R}^+$ be a continuous geodesic map on N. Let c_λ denote the level set of D, also called a **nasal curve**, for the value $\lambda \in D(N)$, i.e. $c_\lambda = \{p \in N | D(p) = GD(r, p) = \lambda\} \subset N$ where r denotes the reference point (in our case the tip of the nose) and $GD(r, p)$ is the length of the shortest path from r to p on the mesh. We can reconstruct N from these level curves according to $N = \cup_\lambda c_\lambda$ (see figure 1).
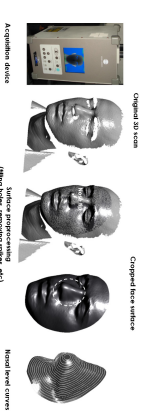

Figure 1. Automatic data preprocessing and nasal curves extraction

We start by considering a closed curve β in $\mathbb{R}^3$. Since it is a closed curve, it is natural to parameterize it using $\beta : \mathbb{S}^1 \to \mathbb{R}^3$. Note that the parameterization is not assumed to be arc-length; we allow a larger class of parametrizations for improved analysis. To analyze the shape of β, we shall represent it mathematically using a *square-root*

The rest of this paper is organized as follows. In Section 2 we summarize some past work on comparing shapes of closed curves in $\mathbb{R}^3$ using an elastic metric, and apply this idea to analyze shapes of nasal surfaces. Also, we apply this classical Karcher mean computation to compute average of nasal surfaces. In Section 3, we study the use of geodesic distances in two biometric scenarios – identification and authentification. These experiments are based on comparing each query shape to each gallery shape. In the Section 4, we seek to make the database search more efficient, $O(\log(n))$ rather than n, by deriving a hierarchical database of gallery shapes.

Our approach for analyzing shapes of nasal surfaces is Riemannian. That is, we define a differentiable manifold, with a suitable Riemannian metric, whose elements represent individual noses. Our choice of representation is based on similar ideas followed previously for full faces – use a collection of iso-curves to represent a surface and to compare surfaces by comparing their corresponding curves [1, 14]. Our goals here are several : (i) Firstly, we want to evaluate the choice of (shape of) nose as a partial metric on a reasonably large database (involving 2000 scans), (ii) Secondly, we want to use Riemannian geometry to organize the database containing noses of a large population into a hierarchical (tree) so that efficient searches can be performed. Towards this end, we will explore the use of geodesic distances and Karcher means for clustering and averaging nasal surfaces.

velocity function (SRVF), denoted by $q(t)$, according to :

$$q(t) \doteq \frac{\dot{\beta}(t)}{\sqrt{\|\dot{\beta}(t)\|}}.$$

The classical elastic metric for comparing shapes of curves is the $\mathbb{L}^2$-metric under this representation [7]. This point is very important and simplifies calculus of elastic metric to the well-known oneunder the $\mathbb{L}^2$-metric. In order to restrict our shape analysis to closed curves, we define the set :

$$C = \{q : S^1 \to \mathbb{R}^3, \int_{S^1} q(t)\|q(t)\|dt = 0\} \subset \mathbb{L}^2(S^1, \mathbb{R}^3). \quad (1)$$

Here $\mathbb{L}^2(S^1, \mathbb{R}^3)$ denotes the set of all functions from S^1 to $\mathbb{R}^3$ that are square integrable. The quantity $\int_{S^1} q(t)\|q(t)\|dt$ denotes the total displacement in $\mathbb{R}^3$ as one traverses along the curve from start to end. Setting it equal to zero is equivalent to having a closed curve. Therefore, C is the set of all closed curves in $\mathbb{R}^3$ each represented by its SRVF. Notice that the elements of C are allowed to have different lengths. Due to a non-linear (closure) constraint on its elements, C is a non-linear manifold. We can make it a Riemannian manifold by using the metric : for any $u, v \in T_q(C)$, we define :

$$\langle u, v \rangle = \int_{S^1} \langle u(t), v(t) \rangle dt. \quad (2)$$

We have used the same notation for the Riemannian metric on C and the Euclidean metric in $\mathbb{R}^3$ hoping that the difference is made clear by the context.

So far we have described a set of closed curves and have endowed it with a Riemannian structure. Next we consider the issue of representing the *shapes* of these curves. It is easy to see that several elements of C can represent curves with the same shape. For example, if we rotate a curve in $\mathbb{R}^3$, we get a different SRVF but its shape remains unchanged. Another similar situation arises when a curve is re-parameterized ; a re-parameterization changes the SRVF of a curve but not its shape. In order to handle this variability, we define orbits of the rotation group $SO(3)$ and the re-parameterization group Γ as the equivalence classes in C. Here, Γ is the set of all orientation-preserving diffeomorphisms of S^1 (to itself) and the elements of Γ are viewed as re-parameterization functions. For example, for a curve $\beta : S^1 \to \mathbb{R}^3$ and a function $\gamma : S^1 \to S^1, \gamma \in \Gamma$, the curve $\beta(\gamma)$ is are-parameterization of β. The corresponding SRVF changes according to $q(t) \mapsto \sqrt{\dot{\gamma}(t)}q(\gamma(t))$. We set the elements of the set :

$$[q] = \{\sqrt{\dot{\gamma}(t)}Oq(\gamma(t))|O \in SO(3), \gamma \in \Gamma\},$$

to be equivalent from the perspective of shape analysis. The set of such equivalence classes, denoted by $S \doteq C/(SO(3) \times \Gamma)$ is called the *shape space* of closed curves in $\mathbb{R}^3$. S inherits a Riemannian metric from the larger space C due to the quotient structure.
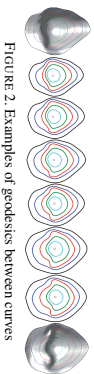
FIGURE 2. Examples of geodesics between curves

The main ingredient in comparing and analysing shapes of curves is the construction of a geodesic between any two elements of S, under the Riemannian metric given in Eqn. 2. Given any two curves β_1 and β_2, represented by their SRVFs q_1 and q_2, we want to compute a geodesic path between the orbits $[q_1]$ and $[q_2]$ in the shape space S. This task is accomplished using a *path straightening approach* which was introduced in [10]. The basic idea here is to connect the two points $[q_1]$ and $[q_2]$ by an arbitrary initial path α and to iteratively update this path using the negative gradient of an energy function $E[\alpha] = \frac{1}{2}\int_s \langle \dot{\alpha}(s), \dot{\alpha}(s)\rangle ds$. The interesting part is that the gradient of E has been derived analytically and can be used directly for updating α. As shown in [10], the critical points of E are actually geodesic paths in S. Thus, this gradient-based update leads to a critical point of E which, in turn, is a geodesic path between given points. Figure 2 shows two different surfaces of two different subjects and some examples of geodesic paths between different curves. The first and the last curves are the ones extracted from the two surfaces, and the intermediate curves denote equally-spaced points on the corresponding geodesic α. These curves have been scaled to the same length to improve display of geodesics. We will use the notation $d([q_1], [q_2])$ to denote the geodesic distance, or the length of the geodesic in S, between representations of the two curves β_1 and β_2.

2.2. Nasal Surfaces

Now we extend the framework from curves to surfaces. As mentioned earlier, we are going to represent a nose region surface S with an indexed collection of the level curves of the D function. That is, $N \mapsto \{c_\lambda, \lambda \in [0, \lambda_0]\}$, where c_λ is the level set associated with the distance D equal to λ. Through this relation, each nasal surface has been represented as an element of the set $C^{[0,\lambda_0]}$. In our framework, the shapes of any two noses are compared by comparing their corresponding nasal curves. Given any two nasal surfaces N^1 and N^2, and their associated curves $\{c_\lambda^1, \lambda \in [0, \lambda_0]\}$ and $\{c_\lambda^2, \lambda \in [0, \lambda_0]\}$, respectively, our idea is to compare the curves c_λ^1 and c_λ^2, and to accumulate these distances over all λ. For that, we define two possible metrics :

- **Arithmetic mean** : d_a ; $d_a : C^{[0,\lambda_0]} \times C^{[0,\lambda_0]} \to \mathbb{R}_{\geq 0}$, given by $d_a(N^1, N^2) = \frac{1}{\lambda_0}\sum_{\lambda=1}^{\lambda_0} d(c_\lambda^1, c_\lambda^2)$.

- **Geometric mean** : d_g ; $d_g : C^{[0,\lambda_0]} \times C^{[0,\lambda_0]} \to \mathbb{R}_{\geq 0}$, given by $d_g(N^1, N^2) = (\prod_{\lambda=1}^{\lambda_0} d(c_\lambda^1, c_\lambda^2))^{(1/\lambda_0)}$.

One advantage of a deformation-based comparisons of shapes of surfaces is that one has the actual deformations in addition to distances. (Also, the distances have important physical interpretations associated with them.) In particular, we have a geodesic path in $C^{[0,\lambda_0]}$ between the two surfaces N^1 and N^2. This geodesic corresponds to the optimal elastic deformations of nasal curves and, thus, nasal surfaces from one to other. Shown in Figure 3 are examples of such geodesic paths – The first row involves nose regions of same subject but different scans, while the other rows show geodesics between nose regions that belong to different people.
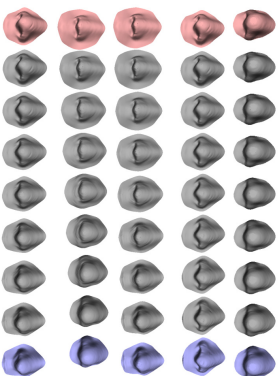
FIGURE 3. Geodesic paths between source and target noses. First two rows : intra-class paths, source and target with different expressions. Last three rows : inter-class paths.

2.3. Mean Nasal Shape

We can now to use our framework to calculate some statistics as the "mean" of surfaces. The Riemannian structure defined on the manifold of nose surfaces $C^{[0,\lambda_0]}$ enables us to perform such statistical analysis for computing noses mean and variance. The Karcher mean utilizes the intrinsic geometry of the manifold to define and compute a mean on that manifold. It is defined as follows : Let $d_s(N^i, N^j)$ denote the length of the geodesic from N^i to N^j in $C^{[0,\lambda_0]}$. To calculate the Karcher mean of nose surfaces $\{N^1, ..., N^n\}$ in $C^{[0,\lambda_0]}$, we define the variance function :

$$\mathcal{V} : C^{[0,\lambda_0]} \to \mathbb{R}, \mathcal{V}(N) = \sum_{i=1}^{n} d_s(N, N^i)^2 \quad (3)$$

The Karcher mean is then defined by :

$$\bar{N} = \arg\min_{N \in C^{[0,\lambda_0]}} \mathcal{V}(N) \quad (4)$$

The intrinsic mean may not be unique, i.e. there may be a set of points in $C^{[0,\lambda_0]}$ for which the minimizer of $\mathcal{V}$ is obtained. To interpret geometrically, $\bar{N}$ is an element of $C^{[0,\lambda_0]}$, that has the smallest total deformation from all given nose surfaces.

We present a commonly used algorithm for finding Karcher mean for a given set of nose surfaces. This approach, presented in Algorithm 1, uses the gradient of $\mathcal{V}$ to iteratively update the current mean μ.

Algorithm 1 Karcher Mean Algorithm

Set $k = 0$. Choose some time increment $\epsilon \leq \frac{1}{n}$. Choose a point $\bar{N}_0 \in C^{[0,\lambda_0]}$ as an initial guess of the mean. (For example, one could just take $\bar{N}_0 = S1$.)

1- For each $i = 1, ..., n$ choose the tangent vector $f_i \in T_{\bar{N}_k}(C^{[0,\lambda_0]})$ which is tangent to the geodesic from $\bar{N}_k$ to N^i. The vector $g = \sum_{i=1}^{n} f_i$ is proportional to the gradient at $\bar{N}_k$ of the function $\mathcal{V}$.

2- Flow for time ϵ along the geodesic which starts at $\bar{N}_k$ and has velocity vector g. Call the point where you end up $\bar{N}_{k+1}$.

3- Set $k = k + 1$ and go to step 1.

Since this is a gradient approach, it only ensures a local minimizer of the variance function $\mathcal{V}$. Several examples of using the Karcher mean to compute average faces are shown in Figures 4.
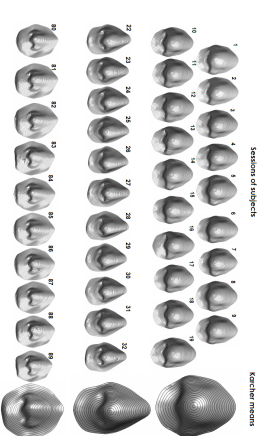
FIGURE 4. Examples of shapes and their karcher means.

What is the motivation for using statistical means of shapes in a biometric application ? There are many possibilities. For example, one can develop a hierarchical organization of a population, where people (or observations) are first clustered into small groups, and these clusters are replaced by their representatives, in this case Karcher means. Then, at the next higher level, one can cluster these means again and find their representative, and so on.

3. Application to Human Biometrics

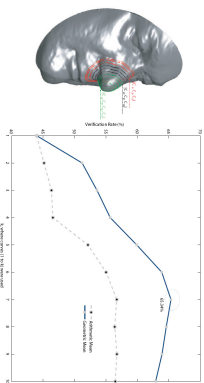In this section we present some experimental results in two different biometric scenario. These experiments use a

large subset, the first 2000 scans, of FRGC v2 dataset [13]. In order to produce results for both identification and authentication scenarios and to explore effect of the presence of facial expressions on performance, a distance matrix between a gallery and a query datasets is computed. The gallery contains 2000 sessions for 209 different subjects. The query set is identical to the gallery. The effectiveness of d_a and d_g in biometric applications generally increases with the number of curves used. But the question is how to choose the curves which can give best results ? Indeed, in the nasal region, there are some areas like nasal cavities which exhibit undesired variability. Therefore, curves passing through these cavities could be eliminated from the experiment. This hypothesis will be tested and discussed in the following experiments. The results of identification and authentication scenarios will be presented separately.

Authentication Scenario : The first experiment involves authentication where the performance is measured using the verification rate (VR) for a given false acceptance rate (FAR) and with a receiver operating characteristic (ROC) curve that shows the trade-off between the verification and the false accept rates.

Figure 5 (right) shows a plot of VR at FAR = 0.1%, computed using both the geometric and arithmetic means. The location of these curves on a nose are shown in the left. The best verification rate is obtained using the first seven curves and the geometric mean metric ; it is approximately 65.34%. As shown in the Figure 5, the verification rate increases with the number of used curves until $k = 7$ curves are used and then it decreases. The curves which decrease the recognition rates pass through the nasal cavities. Moreover, we notice that the verification rates using the geometric mean d_g are much better than those using the arithmetic mean d_a. So we choose this metric for further experiments. Figure 6 shows the ROC curves for experiments involving different number k of curves. For $k < 7$, the ROC curves using $1 - k$ curves are higher for larger k. For $k > 7$, this trend is reversed. So, using a larger number of curves improves the authentication
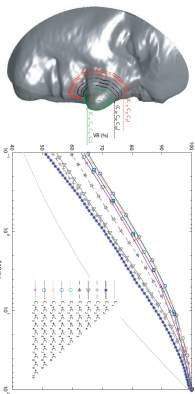
FIGURE 5. Verification Rate at FAR=0.1% using (1 to k) curves

FIGURE 6. ROC curves using 1 to k curves

FIGURE 7. Rank one recognition rates using 1 to k curves

Identification Scenario : The second type of experiment is for identification for which the performance is quoted as a rank-one recognition rate. In this experiment the gallery set consists of n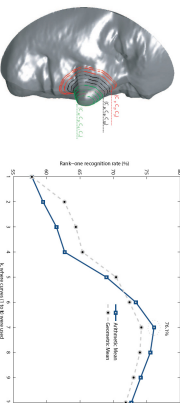eutral faces of 209 subjects. Figure 7 shows the rank-one recognition rates obtained using both geometric and arithmetic mean in the right and the locations of these curves on a nose in the left. We make the following remarks. First, we notice that the best recognition rate is obtained when using the arithmetic mean metric d_a. Actually, we obtain 76.1% using this metric using only seven curves. Second, we see that the recognition rate initially increases with the number of curves, reaches a peak and then starts decreasing, for both the metrics. The later curves which decrease the recognition rates are curves 8,9 and 10 and, as we see in the Figure 7 (left), these curves pass through the nasal cavities. Clearly, shapes of this region (made up of cavities) are not reliable and can be excluded from the analysis.

To summarize, our method allows us to compare the shapes of nose regions. By carefully selecting the set of stable curves, we are able to increase identification and verification rates. However, the identification scenario costs a

performance until the seventh curve and, after that, results get worse.

lot in terms of time computation. In our case, to retrieve a shape in a dataset of 209 nasal surfaces, we need about 4 minutes. A hierarchical database organization can be performed in order to accelerate the identification process. We study an approach in order to cluster the gallery dataset using our framework.

4. Hierarchical Gallery Organization

One of the main goals for studying shapes of nose region is to conduct biometric searches where query is often compared to a set of gallery shapes. This comparison can be made more efficient if we can organize the gallery elements in form of a hierarchical database, i.e. a tree, where the comparisons are performed only at the nodes. To construct such a shape tree we need to be able to cluster similar shapes, and that is the problem we study next.

4.1. Clustering Algorithm

Consider the problem of clustering n noses (in $C^{[0,\lambda_0]}$) into k clusters. A general approach is to form clusters in such a way that they minimize total "within-cluster" variance [15]. Let a configuration C consists of clusters denoted by $C_1, C_2, \ldots, C_k$, and let μ_i be the mean shapes in C_is and n_is be the sizes of C_is. There are several cost functions used for clustering, e.g the sum of traces of covariances within clusters. However, the computation of means μ_is of large shape clusters, and therefore their variances, is computationally expensive, especially when they are updated at every iteration. As a solution, one often uses a variation, where a cluster is replaced by a scaled sum of distances (squared) between its elements :

$$Q(C) = \sum_{i=1}^{k} \frac{2}{n_i} \left(\sum_{N^a \in C_i, \, b<a, N^b \in C_i} d(N^a, N^b)^2 \right). \quad (5)$$

We seek configurations that minimize Q, i.e., $C^* = \arg\min Q(C)$. Notice that the metric used is the arithmetic mean d_a. We will minimize the clustering cost using a Markov chain search process on the configuration space. The basic idea is to start with a configuration of k clusters and reduce Q by re-arranging shapes amongst the clusters. The re-arrangement is performed in a stochastic fashion using two kinds of moves. These moves are performed with probability proportional to the negative exponential of the Q-value of the resulting configuration. The two types of moves are following : (1) **Move a shape :** Here we select a shape N_i randomly and re-assign it to another cluster. Let $Q_j^{(i)}$ be the clustering cost when a shape N_i is re-assigned to cluster C_j, keeping all other clusters fixed. If N_j is not a singleton, i.e. not the only element in its cluster, then the transfer of N_j to cluster C_j is performed with probability :

$P_M(j,i;T) = \frac{\exp(-Q_j^{(i)}/T)}{\sum_{l=1}^{k} \exp(-Q_l^{(i)}/T)}$, $i = 1, 2, \ldots, k$. Here T plays a role similar to temperature in simulated annealing. If N_j is a singleton, then moving it is not allowed in order to fix the number of clusters at k. (2) **Swap two shapes :** Here we select two shapes randomly from two different clusters and swap them. Let $Q^{(1)}$ and $Q^{(2)}$ be the Q-values of the original configuration (before swapping) and the new configuration (after swapping), respectively. Then, swapping is performed with probability : $P_S(T) = \frac{\exp(-Q^{(2)}/T)}{\sum_{l=1}^{2} \exp(-Q^{(l)}/T)}$.

In order to seek global optimization, we have adopted a simulated annealing approach. Although simulated annealing and the random nature of the search help in avoiding local minima, the convergence to a global minimum is difficult to establish. The main steps of the algorithm is given by Algorithm 2.

Algorithm 2 Stochastic Clustering Algorithm

For n shapes and k clusters, initialize by randomly distributing n shapes among k clusters. Set a high initial temperature T.

1- Compute pairwise geodesic distances between all n shapes. This requires $n(n-1)/2$ geodesic computations.

2- With equal probabilities pick one of the two moves :
– **Move a shape :** Pick a shape N_i randomly. If it is not a singleton in its cluster, then compute $Q_j^{(i)}$ for all $i = 1, 2, \ldots, k$. Compute the probability $P_M(j,i;T)$ for all $i = 1, \ldots, k$ and re-assign N_j to a cluster chosen according to that probability.
– **Swap two shapes :** Select two clusters randomly, and select a shape from each. Compute the probability $P_S(T)$ and swap the two shapes according to that probability.

3- Update the temperature using $T = T/\beta$ and return to Step 2.

4- We have used $\beta = 1.0001$.

It is important to note that once the pairwise distances are computed, they are not computed again in the iterations. Secondly, unlike k-mean clustering, the mean shapes are never calculated in this clustering.

The algorithms for computing Karcher mean and clustering can be applied repeatedly for organizing a large database of human noses into a hierarchy that allows efficient searches during identification process. As an illustration of this idea, we consider 500 nose scans corresponding to 50 distinct subjects. These noses form the bottom layer of the hierarchy, called level E in Figure 8. Then, we compute Karcher mean shapes (representative shapes) for each person to obtain shapes at level D. These shapes are further clustered together and a Karcher mean is computed for each cluster. These mean shapes form the level C of the hierarchy. Repeating this idea a few times, we reach the top of the tree which only one shape. We obtain so the final tree shown in Figure 8. If we follow a path from top to bottom of the tree, we see the shapes getting more particularized to groups and then to individuals as illustrated in Figure 8.
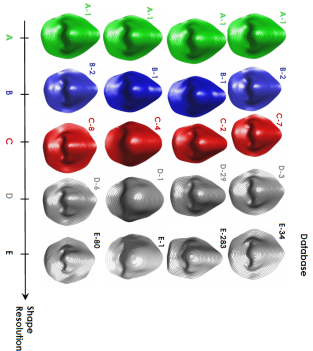
5. Conclusions

In this paper, we have proposed a geometric analysis of 3D nasal shapes for the use in partial human biometrics.

4.2. Hierarchical Shape Retrieval

Once the tree is formed, one can use this representation of data to conduct biometric search in order to reduce computation time, especially in identification scenario, which needs a comparison of the query shape to the whole gallery dataset. In view of this structure, a natural way is to start at the top, compare the query shape with the shapes at each level, and proceed down the branch that leads to the closest shape. At any level of the tree, there is a number, say h, of possible shapes and our goal is to find the shape that matches the query best. Figure 9 illustrates 2 examples of retrieval using the hierarchical organization of the dataset. For Q1 for example, the query nasal shape (at the top) is first compared to the shapes of level B. As it is closer to shape B_1, we proceed down the corresponding branch. There, the query proceeds down the branch of the shape C_2 as this shape is the closest one in this level to the query. The decision of the retrieval is given after comparison with shapes at level D. In this case the query is matched to the shape D_2. Actually, the shape D_2 is the mean shape of nasal shapes of the same person of the query. Notice that nasal shapes of each person at level E are represented by their mean at level D. The last match decides in which cluster the query belongs. According to our tree, time computation for shape retrieval is approximately 3–10 times faster than exhaustive comparison as given by table 1 for 11 examples of queries. Results on a larger experimental set will be presented in the final versions.

FIGURE 8. Paths from top to bottom in the tree show increasing shape resolutions

TABLE I. Average time computation and retrieval results

Query	Exhaustive	Hierarchical	Result
$Q_1, i \in [1..9]$	75 s	16.05 s	correct
Q_{10}, Q_{11}	75 s	16.5 s	incorrect

The main tool presented in this paper is the construction of geodesic paths between arbitrary two two nasal surfaces. The length of a geodesic between any two nasal surfaces is computed as the geodesic length between a set of their nasal curves. In particular, we have presented results for computing geodesics, computing statistical means and stochastic clustering to perform hierarchical clustering. We have demonstrated these ideas in two application contexts : the authentication and the identification biometric scenarios using nasal shapes on a large dataset involving 2000 scans, and hierarchical organization of noses gallery to allow efficient searches.

References

[1] B. B. Amor, H. Drira, L. Ballihi, A. Srivastava, and M. Daoudi. An experimental illustration of 3d facial shape analysis under facial expressions. *Annals of telecommunications*, 64(5-6) :369–379, 2009.

[2] A. M. Bronstein, M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, 64(1) :5–30, 2005.

[3] H. Chen and B. Bhanu. Human ear recognition in 3d. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4) :718–737, 2007.

[4] H. Drira, B. B. Amor, M. Daoudi, and A. Srivastava. Nasal region contribution in 3d face biometrics using shape analysis framework. In M. Tistarelli and M. S. Nixon, editors, *Advances in Biometrics, Third International Conference, ICB 2009, Alghero, Italy, June 2-5, 2009, Proceedings*, volume 5558 of *Lecture Notes in Computer Science*, pages 357–366. Springer, 2009.

[5] T. Faltemier, K. Bowyer, and P. Flynn. A region ensemble for 3-d face recognition. *IEEE Transactions on Information Forensics and Security*, 3(1) :62–73, 2008.

[6] T. Hofmann and J. M. Buhmann. Pairwise data clustering by deterministic annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(1) :1–14, 1997.
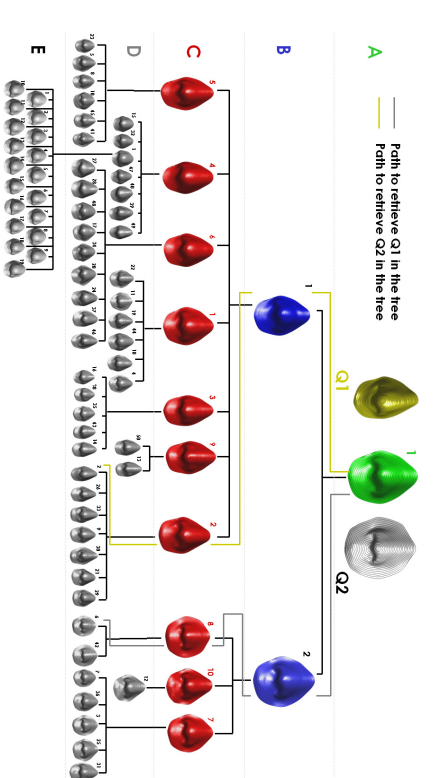
[7] S. H. Joshi, E. Klassen, A. Srivastava, and I. H. Jermyn. A novel representation for efficient computation of geodesics between n-dimensional curves. In *IEEE CVPR*, pages 1–7, 2007.

[8] S. H. Joshi, E. Klassen, A. Srivastava, and I. H. Jermyn. Removing shape-preserving transformations in square-root elastic (SRE) framework for shape analysis of curves. In A. Y. et al., editor, *EMMCVPR, LNCS 4679*, pages 387–398, 2007.

[9] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions : An annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4) :640-649, 2007.

[10] E. Klassen and A. Srivastava. Geodesics between 3D closed curves using path-straightening. In *Proceedings of ECCV, Lecture Notes in Computer Science*, pages 1 :95–106, 2006.

[11] E. Klassen, A. Srivastava, W. Mio, and S. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Pattern Analysis and Machine Intelligence*, 26(3) :372–383, March, 2004.

[12] P. W. Michor, D. Mumford, J. Shah, and L. Younes. A metric on shape space with explicit geodesics. *rend.lincei mat.appl.*, 9 :25, 2008.

[13] P.J. Phillips, P.J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE CVPR*, pages 947–954, 2005.

[14] C. Samir, A. Srivastava, M. Daoudi, and E. Klassen. An intrinsic framework for analysis of facial surfaces. *International Journal of Computer Vision*, 82(1) :80–95, 2009.

[15] A. Srivastava, S. H. Joshi, W. Mio, and X. Liu. Statistical shape analysis : Clustering, learning, and testing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4) :590–602, 2005.

[16] P. Yan and K. W. Bowyer. Biometric recognition using 3d ear shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8) :1297–1308, 2007.

FIGURE 9. Examples of paths parsed by two queries across the tree.