LEARNING STRUCTURALLY DISCRIMINANT FEATURES IN 3D FACES

Sreenivas R. Sukumar[1], *Hamparsum Bozdogan*[2], *David L. Page*[1], *Andreas F. Koschan*[1], *Mongi A. Abidi*[1]
Email: {ssrangan, bozdogan, dpage, akoschan, abidi@ utk.edu}
[1] Imaging Robotics and Intelligent Systems Lab, Department of Electrical and Computer Engineering,
[2] Department of Statistics
The University of Tennessee, Knoxville, USA.

ABSTRACT

In this paper, we derive a data mining framework to analyze 3D features on human faces. The framework leverages kernel density estimators, genetic algorithm and an information complexity criterion to identify discriminant feature-clusters of lower dimensionality. We apply this framework on human face anthropometry data of 32 features collected from each of the 300 3D face mesh models. The feature-subsets that we infer as the output establishes domain knowledge for the challenging problem of 3D face recognition with dense 3D gallery models and sparse or low resolution probes.

Index Terms— 3D face recognition, feature learning, dimensionality reduction, informative-discrimant face features.

1. INTRODUCTION

Most of the success reported with 3D face recognition has been using alignment based methods [1]. A detailed study of these methods [2] reveals that alignment-based methods require dense models (few thousand points) of the face acquired for the probe and the gallery. But, in real world situations, we often cannot expect co-operation for acquiring dense 3D probe models. The question then that arises in applications where sparse 3D points and feature probes are extracted from surveillance videos [3] as shown in Figure 1 or by other forensic clinical means, is what geometric features in a human face should be extracted and matched with the dense gallery data for reliable face identification.

In such a scenario, curvature-based methods or moment-based methods are not suitable because of the sparse nature of the probe point cloud and the fact that curvature is extremely sensitive to noise and resolution. The alternative pattern recognition approach to 3D face recognition using geometric feature measurements would have been to use ideas like PCA [9] or LDA [8] on face features and search for the nearest match in a transformed multi-dimensional feature space. Though PCA and LDA are very efficient methods to perform recognition, the methods require that all features considered initially in the learning phase be present in the probe also. However, we might not be able to compute all the features from a sparsely/partially reconstructed probe. In such cases, PCA and LDA will not be able to tell us which of these features is actually contributing to discriminative separation in the component space for recognition.

Dense gallery (a) (b) Sparse probe (c)

Figure 1: Exploratory investigation to learn what facial features in a dense model can be matched with what is available in a sparse probe. (a) A dense 3D model of a human face used in the gallery. (b) 2D features extracted from a surveillance video frame. (c) Features tracked in several frames like (b) to generate a sparse 3D probe face [3].

Hence, we turn our attention to variable selection methods [10] to provide an understanding of the facial feature data. We realize that face features are an inherent multi-class dataset, where people from different ethnic backgrounds exhibit variation within the same feature and current variable selection methods are not able to accommodate within-feature multi-modal behavior. Hence, we design a feature selection framework for this purpose, explain the implementation details and draw significant statistical inference on 3D anthropometry data of human faces. With our results and conclusions, the two contributions we claim with this paper are as follows:

- The design of a generic variable sub-setting method leveraging kernel density estimators (KDE), genetic search algorithm, and an information theoretic fitness criterion as a three way hybrid dimensionality reduction tool.
- The application of the framework to learn structurally diverse, discriminant and informative 3D facial features extracted from a database of 300 faces consisting of people from different gender and ethnic backgrounds to aid face recognition with partial data.

978-1-4244-1764-3/08/$25.00 ©2008 IEEE

We have organized the paper as follows. In Section 2, we present related work on the face feature extraction and the analysis conducted in the literature. We describe the new variable selection approach in Section 3 before presenting experimental results in Section 4 and drawing conclusions in Section 5.

2. RELATED WORK

Our goal is to learn a minimal set of facial features, to assist recognition in applications where completeness of the probe data is not always guaranteed and only certain facial measurements are available. Hence, we have to consider features that can be extracted from both sparse and dense reconstructions of the human face. We consider easily extractable distance and angle features inspired by anthropometric and cognitive studies [4, 5]. We also considered the landmarks and features with 2D images from the computer vision community [6] and from the 3D graphics/gaming community [7]. The observations in a recent effort proving anthropometric 3D landmarks and feature proportions [8] performing better than arbitrarily chosen facial features helped us build a comprehensive list of 32 features tabulated in Table 1. The list intentionally includes as many expression invariant features as possible into our study.

Table 1: The features we extract automatically from anthropometric points.

Global features: 1.Surface area 2.Volume 3.Radius of inscribing sphere 4.Radius of circumscribing sphere **Geodesic distance features:** 5.Height of forehead 6.Width of forehead 7.Length of nose 8.Distance between lip & nose tip 9.Distance between chin to lip 10.Distance between temple and nearest eye tip 11.Chin to Neck 12.Distance Jaw chin Jaw (Perimeter) 13.Horizontal profile (Perimeter) 14.Vertical profile (Perimeter) 15.Length of eyebrow (Perimeter)	**Euclidean distance features:** 16.Length of eye 17.Width of eye 18.Width of nose 19.Depth of nose 20.Length of ear 21.Width of ear 22.Nose tip and eye ball 23.Length of lips 24.Width of lips 25.Distance between nose tip and jaw 26.Depth of eye 27.Distance between eye bone and right cheek 28.Distance between cheek bone and jaw 29.Distance between eyes **Angle features:** 30.Angle at eye brow (Angle) 31.Angle of nose along length (Angle) 32.Angle of nose along width (Angle)

Our goal is to present an extension to [6] moving from 2D pixel unit measurements to 3D metric units also including geodesic features. Our work differs from the 3D curvature-driven feature region analysis [11] in a human face for better recognition performance in the context that our method is unsupervised and is a top down approach of learning what are the discriminative features in a human face, listing their significance towards recognition rather than looking at what features are best suited for image to image matching. In the following section, we detail the variable selection framework that will help us learn clusters of useful features.

3. FEATURE LEARNING FRAMEWORK

From a pattern recognition perspective, domain knowledge is very important in the choice of features used for classification and recognition. With human faces, the literature on feature-based recognition is so vast one has to deal with a very high dimensional space. While some of these features might be redundant or even irrelevant, we realize that the problem is even more complicated when we use these features for recognition when presented with limited data. Hence, our goal is to learn and establish the domain knowledge of optimal feature combinations of lower dimensionality to help identify a face.

Let us start our learning process with the n samples of p features in data matrix D. For the face recognition motivation, this translates to using n human face models in our database and measuring p features $x_i = \{f_1, f_2, f_3 ..., f_p\}$ from the 3D face. Our objective is to choose k of the p features ($k \le p$) that is minimal, informative and discriminant. We begin by building statistical models of competing subsets of variables $\{f_1, f_2, f_3 ..., f_k\}$ using kernel density estimators in Equation 1.

$$\hat{f}_H(x) = \frac{1}{n(2\pi)^{(k/2)}} \sum_{i=1}^{n} | H_i |^{-1/2} \; K\{(x - x_i)' H_i^{-1} (x - x_i)\} \quad (1)$$

where K refers to a Gaussian kernel, H the bandwidth matrix, that is selected by likelihood cross validation of several competing bandwidth matrices H_j shown in equations below. L refers to the likelihood of the parameters given the data and $\hat{\Sigma}$ is the covariance matrix.

$$H_j = n^{-1/(j+4)} \hat{\Sigma}^{1/2}; \; j = 1,2,3..l \quad (2)$$

$$\hat{f}_{H(-i)}(x_i) = \frac{1}{n-1} \sum_{\substack{i=1 \\ j \ne i}}^{n} | H |^{-1/2} \; K\{(x_i - x_j) H^{-1} (x_i - x_j)\} \quad (3)$$

$$\log L(x_1, ..., x_n | H) = \sum_{i=1}^{n} \log \hat{f}_{H_{(-i)}}(x_i) \quad (4)$$

$$H_{opt} = \max_{\hat{H}} \frac{1}{n} \log L(x|H) \quad (5)$$

We use KDE for building our statistical model in order to preserve the separability of the features and at the same time avoid the drawbacks in the state-of-the-art methods that assume a functional form for the distribution [10]. Such an assumption of model structure particularly with human face feature data does not utilize the inherent classification information present within the dataset. KDE's on the other hand are data driven and encapsulate multi-modal feature variation. The choice of the optimal bandwidth also takes care of inter-feature correlation.

After constructing the KDE of each subset of feature variables, our next task is to choose the most informative of those subset(s). For large values of p, this becomes a combinatorial search problem. Hence, we encode each variable-subset as a chromosome and use the genetic algorithm (GA) as a systematic procedure to span the subset-space. A typical chromosome considering only f_1 and f_2 as informative features as a (n by 2) dataset would be encoded as shown in Figure 2. With the mutation aspect of the genetic algorithm introducing unexplored subsets and the crossover procedure producing healthy subsets from what was learnt in the previous iterations of the algorithm, we fasten the learning process and avoid exorbitant combinatorial evaluations. We would like to direct the reader to [12] for implementation details of the genetic algorithm in a similar scenario.

f_1	f_2	f_3	………………………………….						f_p
1	1	0	0	0	0	0	0	0	0

Figure 2: Encoding feature subsets as chromosomes to leverage genetic algorithm as a search tool over the subset space.

The most important aspect of the genetic algorithm is the fitness criterion that evaluates the goodness of the KDE model for all possible clusters of k features. This being a model selection problem, we use the recently proposed extension to Akaike's information criterion called information complexity (ICOMP) [13] as our fitness function. ICOMP is computed using Equation 6. ICOMP brings together the likelihood of a subset model as the goodness of fit term at the same time penalizing for the feature parsimony using the Fisher information matrix.

This formulation of information complexity evaluates the information gain in entropic sense with the inclusion of a new feature into a cluster evaluating if the increase in dimensionality of the feature subset is a valuable addition of significant information. The parameter we will use for penalizing the dimensionality is the optimized bandwidth matrix H_{opt} that encapsulates within feature variation as well as inter-feature correlation.

$$ICOMP = -2\log(L(x \mid H_{opt}; \hat{f}_H(x))) + 2C_1(F^{-1}(H_{opt})) \tag{6}$$

where F^{-1} is the inverse Fisher information matrix.

$$C_1(F^{-1}(H_{opt})) = \frac{s}{2}\log\left[\frac{tr(F^{-1}(H_{opt})}{s}\right] - \frac{1}{2}\log\left|F^{-1}(H_{opt})\right| \tag{7}$$

with s being the rank of F^{-1}, |.| refers to the determinant and tr refers to the trace of the matrix. F^{-1} is computed as shown in Equation 8.

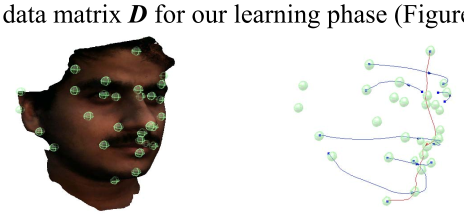
$$F^{-1}(H_{opt}) = \begin{bmatrix} H_{opt} & 0 \\ 0 & D^{+}{}_{p}(H_{opt} \otimes H_{opt})D^{+}{}_{p}' \end{bmatrix} \tag{8}$$

with $D^{+}{}_{p}$ being the Moore-Penrose inverse of vectorized H_{opt}, $\otimes$ represents the Kronecker product. The details behind the derivation of this formulation is available in [13].

The minimum value of ICOMP reveals the feature variable-subset that is optimal in dimensionality and information content. After every iteration of GA, we retain 20% of the least ICOMP value chromosomes as the healthy population for generating future generations towards the convergence. This variable-clustering or feature selection framework is generic for any dataset and quantitatively outperforms traditional methods at the cost of extra storage and computations in building the KDE for each cluster. In the following section, we apply this generic framework to learning 3D face feature clusters for recognition.

4. EXPERIMENTAL RESULTS

For experiments in the learning phase, we use 3D triangle mesh models of human faces acquired using structured light scanners both in our lab using the Genex 3D Face Cam and also the XM2VTSDB database from the University of Surrey. 300 models were used in the learning phase, with 56% males and 44% females. 20 % of the database belonged to Asian-Indian ethnicity, 19% were Asian-Chinese, 18% had African origin while the rest were Caucasian. This diversity in the database was purposefully introduced into the study with the hope that some features listed in Table 1 will easily differentiate people of different ethnicities and gender and that we will be able to find them using our feature clustering algorithm. To the best of our knowledge, our effort is the first to consider such diversity in the database for 3D face recognition. The anthropometric points were manually selected in the 3D model but the feature measurements were automatically computed by marching through the 3D mesh for all faces in the database to create data matrix $\boldsymbol{D}$ for our learning phase (Figure 3).



(a) (b)

Figure 3: Datasets used in the experiment. (a) Anthropometric points selected manually in a 3D face. The error in point localization was negligible compared to the size of the actual features. (b) Automatically computed geodesic and Euclidean features (Only some features are shown).

We conducted two experiments with the face data and our feature learning scheme. The first one was to learn feature clusters (combinations) that are minimal and informative by executing the GA to convergence. The GA pruned 200,000 feature clusters and the top 3 clusters with the lowest ICOMP values in the converged population were: (FC-1) width of nose, depth of nose, depth at the eye, angle at nose along its length (FC-2) depth at the eye, lip to chin distance, nose to lip distance, jaw to jaw distance (FC-3) vertical profile, length of eyebrow, length of nose. With these feature clusters we have essentially learnt what set of features to try and measure in situations where not all measurements of a sparse probe face are possible.

The second experiment that we conducted was by reinitializing the population for the GA after a fixed number of iterations and observing the convergence over several trials. We saved the optimal result after every 1000 iterations and ranked features based on their frequency of occurrence in the converged set over 100 such trials. We observed that the vertical profile, jaw to jaw distance, depth of nose, depth at the eye and chin to neck distance features repeated with a high frequency greater than 60% while other features repeated less than 30% of the time.

With the knowledge of minimal and informative features, our next experiment was to evaluate the recognition performance with these features. Our probes were a low resolution sparse point cloud of the face generated by decimating the dense 3D model to a mesh with only 100 3D vertices. This is the typical resolution of deformed generic meshes used in face modeling using video sequences. We used the Euclidean distance between features as our recognition metric to generate the results in Figure 4.

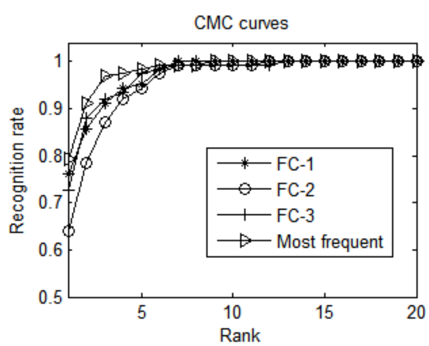


Figure 4: CMC curve of recognition using the structurally diverse feature clusters learnt using our formulation.

5. CONCLUSIONS AND FUTURE WORK

3D face recognition has faced criticism for its inability to handle expressions and for the co-operation required for success in real-world crime situations. Our study has helped us cluster facial features based on their discriminatory characteristics after studying a large-diverse database of human faces. Our conclusions explain the success of several heuristics experimented in the past for face recognition using only horizontal and vertical profiles and caricature inspired 3D face recognition.

Our future efforts are towards constructing an *n*-point anthropometric 3D face-graph of expression invariant geometric features, with anthropometric point features as nodes and informative-discriminative geometric features as weighted attributes in the graph. We expect the graph matching to provide us a reliable expression-invariant 3D face recognition method extending the previous effort with 2D face images [14] to 3D.

ACKNOWLEDGEMENTS

This work was supported by the DOE University Research program in Robotics under grant #DOE-DEFG02-86NE37968 and NSF-CITeR grant #01-598B-UT. The authors also thank Pushpalatha M.B for helping us with the experiments.

REFERENCES

[1] A. Koschan, M. Pollefeys, and M. Abidi (eds.), *3D Imaging for Safety and Security, Part I*, Springer, Dordrecht, Netherlands,2007.
[2] A. F. Abate, M. Nappi, D. Riccio and G. Sabatino, “2D and 3D face recognition: A survey”, *Pattern Recognition Letters*, Vol. 28(14), pp.1885-1906, 2007.
[3] Y. Yao, S. Sukumar, B. Abidi, D. Page, A. Koschan, and M. Abidi, "Automated Scene-Specific Selection of Feature Detectors for 3D Face Reconstruction," *Proc. 3rd International Symposium on Visual Computing Part I, LNCS 4841*, pp. 476-487, 2007.
[4] L.G. Farkas, *Anthropometry of the head and the face*, Second Edition, Raven Press, 1994.
[5] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, “Face Recognition by Humans: Nineteen results all computer vision researchers should know about”, *Proc. of the IEEE*, Vol. 94, No. 11, pp. 1948-1962, 2006.
[6] J. Shi, A. Samal and D. Marx, “How effective are landmarks and their geometry for face recognition?”, *Computer Vision and Image Understanding*, Vol. 102(2), pp. 117-133, 2006.
[7] D. DeCarlo, D. Metaxas and M. Stone, “An Anthropometric Face Model using Variational Techniques”, *In Proceedings of SIGGRAPH*, pp. 67-74, 1998.
[8] S. Gupta, J.K. Aggarwal, M.K. Markey, A.C. Bovik, “3D Face Recognition Founded on the Structural Diversity of Human Faces,” *in the Proc. of IEEE CVPR*, pp.1-7, 2007.
[9] C. Xu, Y. Wang, T. Tan and L. Quan, “Automatic 3D face recognition combining global geometric features with local shape variation information,” *in Proc. AFGR*, pp. 308-313, 2004.
[10] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection”, *Journal of Machine Learning Research*, 3, pp. 1157-1182, 2003.
[11] Y. Sun and L. Yin, “A Genetic Algorithm based feature selection approach for 3D face recognition”, *Proceedings of the Biometric Consortium*, 2005.
[12] H. Bozdogan, *Statistical data mining and knowledge discovery*, CRC Press, 2004.
[13] H. Bozdogan, “Akaike’s Information Criterion and Recent Developments in Information Complexity”, *Journal of Mathematical Psychology*, Vol. 44, pp. 62-91, 2000.
[14] L. Wiskott, J. M. Fellous, N. Kuiger, and C. von der Malsburg, “*Face recognition by elastic bunch graph matching*”, IEEE PAMI, 19, pp. 775-779, 1997.