Image and Vision Computing 30 (2012) 683–697

Contents lists available at SciVerse ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Static and dynamic 3D facial expression recognition: A comprehensive survey ☆

Georgia Sandbach [a,*], Stefanos Zafeiriou [a], Maja Pantic [a,b], Lijun Yin [c]

[a] Imperial College, Department of Computing, London, UK
[b] University of Twente, EEMCS, Twente, Netherlands
[c] Department of Computer Science, Binghamton University, Binghamton, New York

ARTICLE INFO

Article history:
Received 28 December 2011
Received in revised form 4 April 2012
Accepted 12 June 2012

Keywords:
Facial behaviour analysis
Facial expression recognition
3D facial surface
3D facial surface sequences (4D faces)

ABSTRACT

Automatic facial expression recognition constitutes an active research field due to the latest advances in computing technology that make the user's experience a clear priority. The majority of work conducted in this area involves 2D imagery, despite the problems this presents due to inherent pose and illumination variations. In order to deal with these problems, 3D and 4D (dynamic 3D) recordings are increasingly used in expression analysis research. In this paper we survey the recent advances in 3D and 4D facial expression recognition. We discuss developments in 3D facial data acquisition and tracking, and present currently available 3D/4D face databases suitable for 3D/4D facial expressions analysis as well as the existing facial expression recognition systems that exploit either 3D or 4D data in detail. Finally, challenges that have to be addressed if 3D facial expression recognition systems are to become a part of future applications are extensively discussed.

1. Introduction

Automatic human behaviour understanding has attracted a great deal of interest over the past two decades, mainly because of its many applications spanning various fields such as psychology, computer technology, medicine and security. It can be regarded as the essence of next-generation computing systems as it plays a crucial role in affective computing technologies (i.e. proactive and affective user interfaces), learner-adaptive tutoring systems, patient-profiled personal wellbeing technologies, etc. [1].

Facial expression is the most cogent, naturally preeminent means for humans to communicate emotions, to clarify and give emphasis, to signal comprehension disagreement, to express intentions and, more generally, to regulate interactions with the environment and other people [2]. These facts highlight the importance of automatic facial behaviour analysis, including facial expression of emotion and facial action unit (AU) recognition, and justify the interest this research area has attracted, in the past twenty years [3,4].

Until recently, most of the available data sets of expressive faces were of limited size containing only deliberately posed affective displays, mainly of the prototypical expressions of six basic emotions (i.e. anger, disgust, fear, happiness, sadness and surprise), recorded under highly controlled conditions. Recent efforts focus on the recognition of complex and spontaneous emotional phenomena (e.g. boredom or lack of attention, frustration, stress, etc.) rather than on the recognition of deliberately displayed prototypical expressions of emotions [5,4,6,7]. However, most of these systems are still highly sensitive to the recording conditions such as illumination, occlusions and other changes in facial appearance like makeup and facial hair. Furthermore, in most cases when 2D facial intensity images are used, it is necessary to maintain a consistent facial pose (preferably a frontal one) in order to achieve a good recognition performance, as even small changes in the facial pose can reduce the system's accuracy. Moreover, single-view 2D analysis is unable to fully exploit the information displayed by the face as 2D video recordings cannot capture out-of-plane changes of the facial surface, or difficult to see changes. Hence, many 2D views must be utilised simultaneously if the information in the face is to be fully captured. Alternatively, in order to tackle this problem, 3D data can be acquired and analysed. In the case of AU recognition, the subtle changes occurring in the depth of the facial surface are captured in detail when 3D data are used, with 2D data. For example, AU18 (Lip Pucker) is not easily distinguished from AU10 + AU17 + AU24 + AU24 (Upper Lip and Chin Raising and Lip Presser) in a 2D frontal view video. In a 3D capture the action is easily identified, as can be seen in Fig. 1. Similarly, AU 31 (Jaw Clencher), can be difficult to detect in a 2D view, but is easily captured by the full 3D data as can be seen in Fig. 2. Recent advances in structured light scanning, stereo photogrammetry and photometric stereo have made the acquisition of 3D facial structure and motion a feasible task.

In this survey we focus on the use of 3D and 4D data capture for automatic facial expression recognition and analysis. We first study the recent technological solutions that are available for acquiring static and dynamic 3D faces. We particularly focus on the difficulties encountered when applying these techniques in order to be able to capture naturalistic (spontaneous) expressions. We later examine the challenges existing in 3D face alignment, tracking and finding point correspondences

☆ This paper has been recommended for acceptance by Jan-Michael Frahm, Dr.-Ing.

* Corresponding author. Tel.: +44 7714768440.
E-mail addresses: gls09@imperial.ac.uk (G. Sandbach), s.zafeiriou@imperial.ac.uk (S. Zafeiriou), m.pantic@imperial.ac.uk (M. Pantic), lijun@cs.binghamton.edu (L. Yin).

0262-8856/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2012.06.005

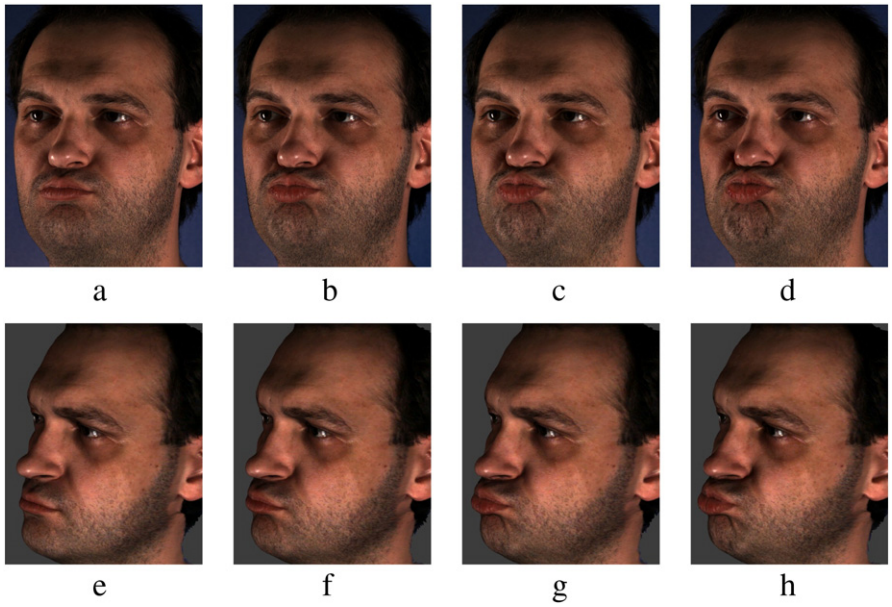

Fig. 1. AU 18 (Lip Pucker) captured in both 2D and 3D. (a)–(d) 2D nearly frontal view. (e)–(h) 3D reconstructed data.

and review existing methods. Furthermore, we survey the databases that have been created either for 3D and 4D facial expression analysis, or biometric applications but contain significant number of expressive examples. Next, we discuss the methods used for static and dynamic 3D facial expression recognition. Here, we mainly focus on feature extraction as this is what differentiates 3D methods from the corresponding 2D ones. Finally we examine the challenges that still remain and discuss the future research needed in tracking and recognition methodologies beyond the state of the art.

The rest of the paper is organised as follows. Section 2 reviews 3D acquisition, tracking and alignment methods. Section 3 presents in detail the available databases suitable for 3D facial expression analysis. Section 4 surveys the recognition systems that have been developed, both for static and dynamic analysis of 3D facial expressions. Section 5 discusses a number of open issues in the field. Finally, Section 6 concludes the paper.

2. Acquisition of 3D and 4D faces, dense correspondences, alignment and tracking

In the past decade the fields of capturing, reconstruction, alignment and tracking of static and dynamic 3D faces have witnessed tremendous development. This section focuses on the state-of-the-art methods in this field from the perspective of the kind of facial behaviour (posed or spontaneous) that the surveyed technology is able to capture. For acquisition, the focus is more on the actual process (i.e. how many cameras are needed, where they need to be placed, what kind of patterns should be projected onto the face of the subject, what lighting is required) and the

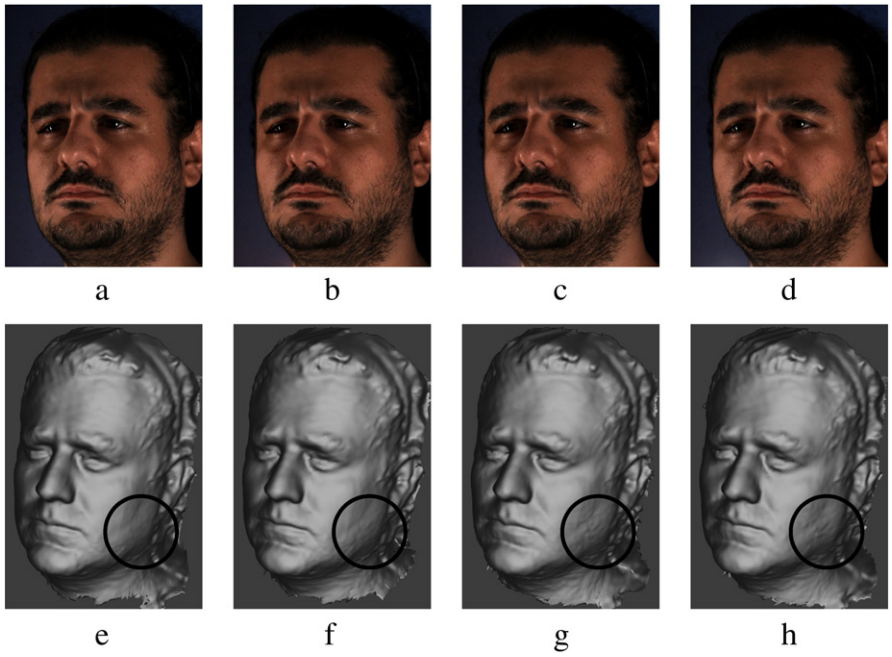

Fig. 2. AU 31 (Jaw Clencher) captured in both 2D and 3D. (a)–(d) 2D nearly frontal view. (e)–(h) 3D reconstructed data. The area of motion is shown in the circle.

quality and time needed for acquisition, rather than on the reconstruction methodology. We effectively show which setup can be used for capturing deliberately displayed (posed) facial behaviour and which can be used for capturing naturalistic and spontaneous facial behaviour in high resolution. We then describe the existing techniques for alignment and tracking of 3D facial surfaces, which is an important preprocessing stage for 3D facial expression recognition.

2.1. Devices and techniques for 3D face acquisition

The acquisition technique used for capturing 3D data is especially important when collecting facial expressions, as the equipment used can affect the level of imposition on the subject, thereby changing their behaviour significantly. A variety of devices and techniques have been employed previously for 3D facial expression data acquisition, including the use of single image reconstruction, structured light technologies, and two different methods for stereo reconstruction algorithms: photometric stereo and multiview stereo.

2.1.1. Single image reconstruction

Automatic 3D face reconstruction from a single facial image of low resolution is an emerging research topic in computer vision [8,9]. Such methods have great potential in facial behaviour research since recordings can be made in unconstrained environments with conventional 2D cameras, so that the subject under investigation has minimal awareness of the recording setup. However, these methods still result in errors in the reconstructed mesh which mean that the accuracy at high resolutions is not adequate for detection of subtle expressions and facial action units, and so far there are no extensive experiments on such data that have been reported.

One of the most prominent methodologies for reconstructing the 3D facial surface from 2D facial images captured in unconstrained environments is the 3D morphable model (3DMM) methodology [10–17] which also constitutes one of the most important recent developments in computational face modelling. Notably, the most well-known publicly available 3DMM is the one presented in [10,11] and has recently been made publicly available in [18]. An example of this model fitted to an image can be seen in Fig. 3. This 3DMM is built from 3D laser scans of human faces that are put into dense correspondence using their pixel intensities and 3D shape information. A 3DMM uses a statistical representation of both the 3D shape and texture of the human face. In order to reconstruct the 3D facial surface of a 2D intensity image the 3DMM is fitted and a set of parameters is retrieved which govern both the 3D facial surface and the shapeless texture. It has been demonstrated that the 3DMM can be efficiently used for extracting 3D facial surface and texture from static images. However, even the recently proposed extensions for 3DMM fitting algorithms require good initialisation of certain parameters (including the light directions) and are sensitive to outliers and partial face occlusion. In addition, since the fitting procedure is time consuming, the application of 3DMMs for capturing sequences of 3D facial surfaces is limited. Finally, 3DMMs are capable of capturing the general characteristics of the 3D surface of a face and hence are well suited for analysis and recognition of pronounced facial expressions (such as smiles). However, as of yet, there are no reported experiments demonstrating their capability of reconstructing subtle facial details such as wrinkles or furrows.

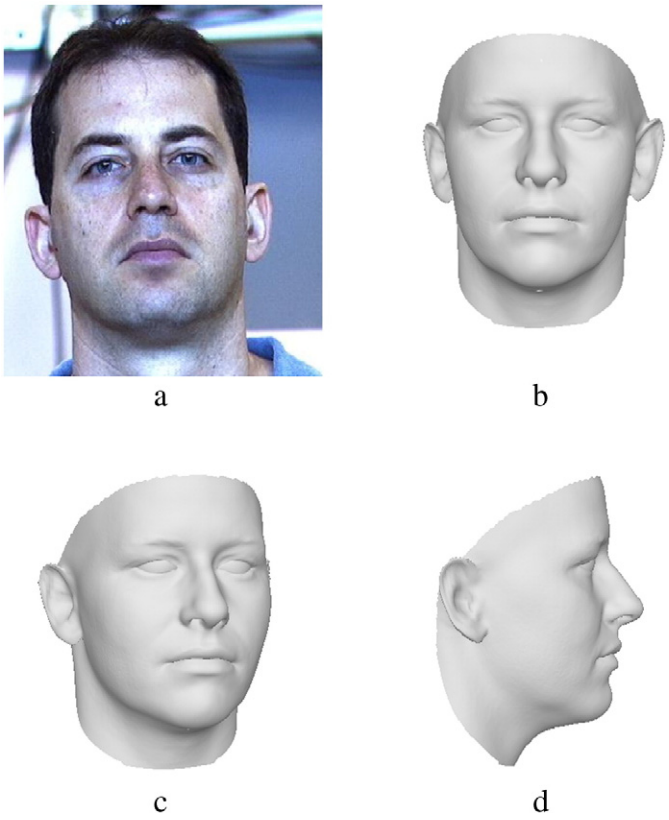

Fig. 3. Example of the morphable model fitted to a single image. (a) Single image of subject. (b)–(d) The fitted model in frontal, nearly frontal and profile poses.

2.1.2. Structured light

Among the most widely used technologies for acquisition of 3D facial surface are structured light methods [19–27]. The basic principle behind this technique is to project one or more encoded light patterns onto the scene and then measure the deformation on the objects' surfaces in order to extract shape information. An example of a pattern used in structured light, and this pattern projected onto a face, is depicted in Fig. 4. By switching rapidly between a coloured pattern and white light it is also possible to capture a colour image in addition to the depth image, with both images approximately synchronised. Unfortunately, the acquired range images can contain holes where points are missing, as well as small artefacts, mainly in areas that cannot be reached by the projected light or surfaces that are either highly refractive (e.g. eye-glasses) or have low reflective (e.g. hair, beard).

In most cases real-time structured light 3D face acquisitions systems use a single pattern, typically a colour pattern [28–30]. These methods sacrifice accuracy for improved acquisition speeds. The other structured light approach for real-time shape acquisition is to use multiple binary-coded patterns but switch them rapidly so that they can be captured in a short period of time [31–33]. The problem with this approach is that for binary-coding methods, the spatial resolution that can be achieved is relatively low because the stripe width must be larger than one pixel. Moreover, switching the patterns by repeatedly loading patterns to the projector limits the switching speed of the patterns and therefore the speed of shape acquisition. Recently, high-speed structure light based systems were proposed [24–27,34] which use projected fringe patterns. These can be either phase-shifted (coloured) sinusoidal, which allow simultaneous acquisition of the 2D intensity images, or trapezoidal fringe patterns for faster decoding. Sinusoidal patterns produce unique features which are easily extracted, which is a useful advantage of the phase-shifting method. However, when using fringe pattern systems, current systems are able to record at a higher frame rate of 40 Hz for 3D shape acquisition, and this rate could theoretically be extended to 120 Hz with the addition of new high speed cameras.

Structured light technology has several advantages for capturing 3D facial surfaces. The cost over normal 2D video capture is in most cases limited to a projector, its slides and a high-speed camera, as in most cases one camera suffices to capture the 3D facial surface. In addition, it can be used for real-time (or even high-speed) simultaneous acquisition of sequences of 3D facial surface and 2D intensity. But the main advantage is that the visible projected pattern in many systems does not distract the user as with high-speed systems the projected channels will appear as a full colour image, and it is also possible to use infrared light instead of visible

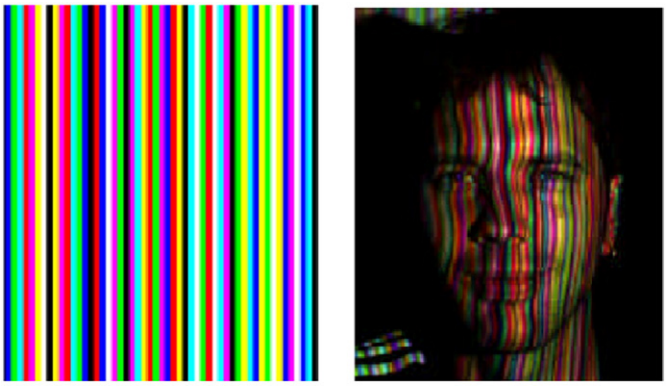

Fig. 4. Example of structured light. (a) The light structure used. (b) This light structure projected onto a face.
Images taken from [23].

light for capturing spontaneous facial motions without distractions for the subject.

However, there are some disadvantages. This method only allows a limited amount of movement of the face in the scene as it is restricted by the area simultaneously covered by the structured pattern and visible by the camera. The acquired range images may also contain holes and artefacts due to this restriction. In addition, the acquisition systems that use patterns within the visible spectrum, such as coloured patterns, cannot be easily applied for the acquisition of spontaneous facial behaviour, or for scenarios like interviews and conversations, as the lights will be visible for the subjects and cause distraction.

Many widely used static 3D acquisition systems are based on structured light technologies such as the Minolta Vivid 900/910 series [35], which was used for the capturing many popular face databases as will be described in Section 3, the Inspeck Mega Capturor II 3D [36] and notably the most widely used — the Kinect camera [37]. Furthermore, a custom made structured light system was used in one of the first studies of 3D facial expression recognition [21].

2.1.3. Photometric stereo

Another popular family of techniques for acquisition of 3D structure is photometric stereo. Photometric stereo, first proposed in [38], is a method for estimating the orientation field (normals) of a 3D surface of objects by capturing a set of images of the object under different illuminations. Fig. 5 shows some reconstructions produced by standard four lights photometric stereo and the multi-spectral photometric stereo in [39].

Photometric stereo is sensitive to the presence of projected (cast) shadows, highlights, and non-uniform lighting. Furthermore, the data computed directly in photometric stereo methods consists of the 3D normals rather than the mesh, and so to retrieve the surface an integration must then be performed. This procedure adds to the computation time required, and introduces additional errors [40–44]. A very recently deployed system for capturing faces in a real life face recognition scenario was presented in [45]. The setup is depicted in Fig. 6a and works as follows: individuals walk through the archway towards the camera located on the back panel and exit through the side. The presence of an individual is detected by an ultrasound proximity sensor placed before the archway. This can be seen in Fig. 6a on the horizontal beam towards the left-hand side of it. The device captures one image of the face for each light source in a total time of approximately 20 ms. This time was regarded as an adequately short period in which the inter-frame motion is no greater than a few pixels. Since the lights burst in only 20 ms the subject experiences only one total light source. An implementation of the booth with infrared lights was recently presented in [46]. One of the difficulties when capturing 3D sequences using photometric stereo is that the method requires visible/infrared lights to burst continually which can be quite unpleasant for the subject.

One way to overcome the distraction caused by bursting lights is to use continuously projected multispectral photometric stereo as proposed in [39]. This method uses red, green and blue lights at different positions to simultaneously capture different illuminations of the same scene, hence not requiring the flashing lights needed for traditional photometric stereo. However, this methodology also has disadvantages. Firstly it requires the subject to have their eyes closed for the duration of the expression capture as the visible coloured lights can be distracting for the subject, and also because this helps to avoid deformation artefacts in the data. This puts severe limitations on the ability to capture the natural behaviour of the subject. The second disadvantage is that the setup must be calibrated for every subject by first moving the head around with the same expression before expression capture can be performed.

2.1.4. Multi-view stereo

Multi-view stereo acquisition is another widely used technique for 3D facial reconstruction [47]. This family of methods employ multiple cameras placed at various known viewpoints from the subject. The different images of the scene then allow corresponding points to be found, subject to various constraints, and these can then be used for reconstruction. One recent method for high quality 3D face acquisition based on multi-view stereo was proposed in [48]. Here 3D capturing is performed both with high-quality equipment in order to provide very detailed face geometries, and also with consumer stereo cameras. Examples of commercial systems available that employ multi-view stereo techniques are the DI3D (Dimensional Imaging [49]) dynamic face capturing system, and the 3DMD dynamic 3D stereo system [50], each of which have been used for 3D database acquisition ([51] and [52,53] respectively). An example of the setup required for the DI3D system as employed in [51] can be seen in Fig. 6b. Here two stereo cameras can be seen along with an additional texture camera which captures the 2D image sequence. Each pair of stereo images is then processed using a passive stereo photogrammetry method to produce a range map. The advantage of this method is that it does not require flashing lights, as all cameras can record the same scene simultaneously, with constant light sources. This can allow more natural behaviour from the subjects being recorded. However, multiple cameras are required which may make the equipment more expensive than photometric stereo. In addition, accurate reconstruction of smooth surfaces such as faces can be difficult with this method, and the range of head movement that can be captured is very limited. Finally, 3D reconstruction is performed offline due to the computational complexity that this process can involve. This means that real-time systems using this method would currently be unfeasible. Some algorithms in use, such as that employed in [48] can take around 20 min to construct the 3D model from the input image, though there are other algorithms available, for example as used in the DI4D system, that take only around 15 s/frame.

2.2. 3D face tracking, 3D dense correspondence and alignment

Accurate alignment and tracking methods are very important for facial expression systems, as the features extracted can often rely on areas of the face falling in the same location between subjects, or finding the movement of particular points on the face. Dense correspondence between face meshes can also be required in order to track the full motion of the face mesh between subjects or frames. Many approaches have been proposed in order to tackle these problems. Rigid alignment between two similar meshes without large transformation can be achieved with algorithms based on the traditional iterative close point (ICP) algorithm [64]. However, non-rigid alignment is required to allow full alignment, dense correspondence and tracking, and this requires the use of more complex algorithms. The methods applied to this problem include the use of non-rigid ICP-based algorithms [54], free-form deformations (FFDs) [55–57], harmonic maps [55], conformal mappings [65,59,66],

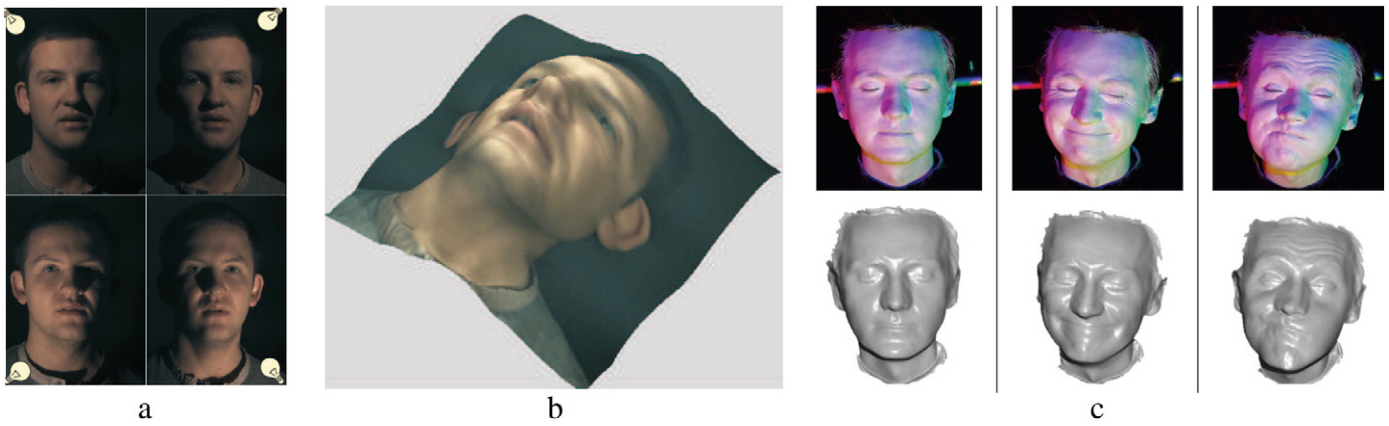

Fig. 5. Comparison of the photometric stereo and multispectral photometric stereo. (a) Photometric stereo illumination method. (b) 3D facial data captured by photometric stereo. (c) Multispectral photometric stereo capturing the subject with different coloured light in different direction.
Image (c) is taken from [39].

covariance matrix pyramids [58], active shape models (ASMs) [60], morphable models [62], and the annotated deformable model (ADM) [63].

Iterative close point (ICP) [64] is an algorithm that has been widely used for 3D rigid alignment problems in 3D facial expression analysis [67–69]. The algorithm takes a source and target mesh, and then works by selecting either a subset, or all, of the points in each mesh. Then for each point in the target, it finds the closest point in the source, and aims to minimise the error between these points by applying a rigid transformation between the two meshes. This process is repeated until a threshold error is reached. Many variants of ICP have also been proposed, and the interested reader may refer to the following [70–72].

However, this form of ICP only allows a rigid transformation, which does not find full correspondence of points between meshes of different individuals or when expression changes occur. Additional methods are therefore required to perform a mapping or non-rigid transformation that produce full dense correspondence. In [54], a non-rigid version of ICP was proposed. This algorithm worked by introducing a stiffness value which controlled the rigidity of the transformation that could be applied at each iteration. At the start this stiffness is given a high value, to force nearly rigid transformations, and then it is gradually reduced in order to allow progressively more non-rigid transformations to be applied as the iterations progress.

FFDs are another family of techniques used for non-rigid registration, first proposed in [73]. The idea is to deform an object by manipulating an underlying lattice of control points. The lattice is regular in the source mesh, and then deformed through an optimisation process in order to allow registration in the target mesh. B-spline interpolation of the deformation then models the motion of corresponding points between the two 3D meshes. This method was employed in [55] in order to fit a coarse face mesh model to the first frame in the sequence and in [56,57] in order to track the motion of the face meshes through 3D expression image sequences.

The method employed in [58] uses a simple 2D representation, purely the interpolated depth value, along with texture values at each x, y coordinate as input to a correspondence algorithm. This algorithm then exploits a pyramidal approach based on the covariance matrix of a region, with the Particle Swarm Optimisation (PSO) algorithm at each level to search for the corresponding point in the neighbourhood.

In [55], in order to find dense correspondences, harmonic maps were applied. The method embeds a mesh from a manifold with disc topology into a planar graph through minimisation of the harmonic energy. This method is beneficial as it does not suffer from local minima, folding or clustering of the mesh, and is not affected by the resolution, smoothness or pose of the original 3D data. The full tracking methodology uses FFDs in order to prepare meshes by fitting a generic face mesh model to the first frame. Harmonic maps are then used to fit the 3D point data onto a 2D disc. Feature point correspondence constraints are introduced by detecting specific features through corner detection and similar techniques, and these constraints are applied to the harmonic maps. The maps are then iteratively refined by using optical flow methods to update the feature correspondences.

Conformal mappings is a technique that has been widely exploited in 3D alignment and tracking. A conformal mapping is a function that maps points in the mesh into a new domain, whilst preserving angles between edges in the mesh. This idea is used in order to produce 2D representations of the 3D data in [65,74]. Circle pattern conformal mappings are employed to convert the data into a 2D planar mesh. A generic model, also mapped to 2D using the same algorithm, is used for first coarse, and then fine, alignment and vertex correspondence. An alternative conformal mapping, least squares conformal mapping (LSCM), is applied in [66] for a similar purpose. Here active appearance models (AAMs) are exploited to find features which allow a rough correspondence to be computed. LSCM is then applied to produce 2D planar meshes which were employed for dense correspondence. Harmonic maps, conformal mapping and LSCM are also used in [59].

An active shape model (ASM) was employed in [60,61] in order to perform tracking of facial features. The shape of the face is represented as a sequence of 81 points which correspond to salient facial features. These points are formed from basis shapes computed from the principal components of a set of training faces, added to the mean of this training data. The local appearance model for each landmark is computed from the image gradient information along a line, perpendicular to the facial contour which the landmark lies on, in the 2D training images. A local model of the gradient changes associated with each landmark is then built using a Gaussian distribution. The landmark positions in a new pair of images (2D + 3D) can then be estimated via an optimisation algorithm that aims to minimise the fitting error of the images onto the model. This global method is combined with two local detectors, that focus specifically on difficult areas — the eyebrows and the mouth. The landmarks around these features are found separately using ASMs that are constrained to these shapes, and in the case of the mouth have been predetermined to be one of two possible shapes — open or closed. The estimates from these local methods replace those generated by the global method in the optimisation algorithm in each iteration.

Morphable models, as described in the previous section, is an alternative method that is used in [62] to track 3D objects through image sequences. The model allows rigid motion in the form of translation and rotation away from the original mesh, plus non-rigid motion which is defined as a linear combination of basis vectors. The difference between one frame and the next was thus defined as being dependant on the

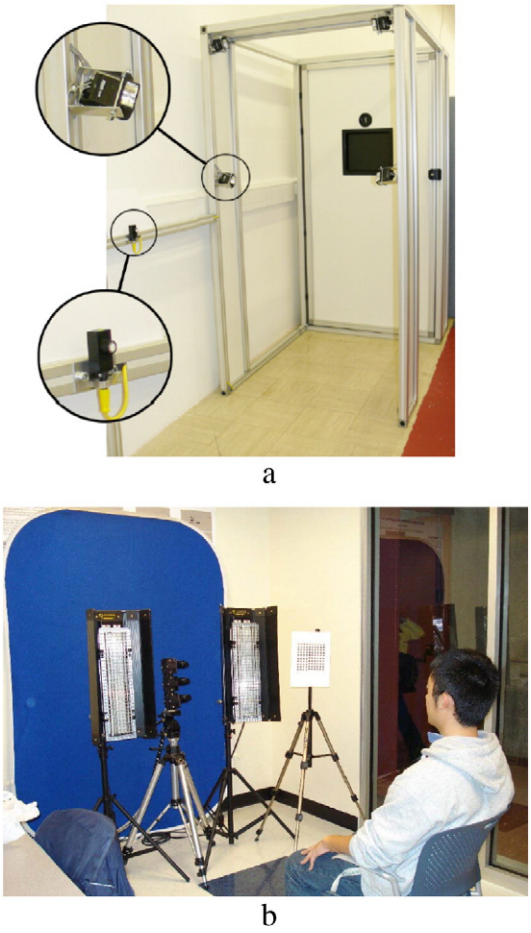

Fig. 6. The different setups for two types of stero acquisition. (a) Photometric stereo acquisition setup. (b) Multi-view stereo acquisition setup.

change in motion parameters that allowed the target to be aligned with the current image. A matrix factorisation was found, allowing a large constant structure matrix to be precomputed off-line. A small time-varying motion matrix can be efficiently computed online and then used to update the motion parameters.

Finally, an alignment method which can be applied for finding dense correspondences is the annotated deformable model (ADM) fitting proposed in [63]. The method fits the generic ADM in a novel 3D image and has been used for face recognition in the presence of facial expressions. As a preprocessing step an alignment method that uses spin images was applied in order to extract an initial correspondence between the data and the ADM. ICP is then employed, followed by refinement through the comparison of the z-buffer images for the model and data. Finally fitting is completed through iteratively deforming the generic model.

The advantages and disadvantages of all of the methods described in this section are summarised in Table 1.

3. Databases

During the past two decades a number of 3D face databases have been created in order to be used for face modelling and recognition. In this Section we review the existing 3D databases, including not only those that have been especially created for expression recognition, but also those that contain expressive faces despite having been recorded for other purposes (e.g. face recognition), as long as they contain enough available samples for training and testing 3D static and dynamic facial expression or action unit recognition systems.

The first 3D facial expression dataset created [21] consists of six subjects expressing the six basic facial expressions. It was collected using of-the-shelf NTSC video equipment and a custom-built system consisting of a camera/projector pair and active stereo using structured light projections, as described in Section 2. The database is not publicly available.

The first systematic effort to collect 3D facial data for facial expression recognition resulted in the creation of BU-3DFE dataset [75], examples of which can be seen in Fig. 7a. Static 3D expressive faces of 100 subjects, displaying the six prototypical expressions at four different intensity levels, were captured using the 3DMD acquisition setup [87]. The models created were of resolution in the range of 20,000 to 35,000 polygons, depending on the size of the subject's face. The database was accompanied by a set of metadata including the position of 83 facial feature points on each facial model, as depicted in Fig. 8.

The same institution continued the effort and recorded BU-4DFE [51], the first database consisting of 4D faces (sequences of 3D faces). The database includes 101 subjects and was created using the DI3D (Dimensional Imaging [49]) dynamic face capturing system. It contains sequences of the six prototypical facial expressions with their temporal segments (onset, apex and offset) with each sequence lasting approximately 4 s (examples can be seen in Fig. 7b). The temporal and spatial resolution are 25 frames/s and 35,000 vertices, respectively. Unfortunately, the database provides no AU annotation.

Another publicly available dataset consisting of static 3D facial models is the Bosphorus database [76]. The database was captured using Inspeck Mega Capturor II 3D [36], which is a commercial structured-light based 3D digitiser device. The database consists of 105 subjects (60 men and 45 women, with the majority of the subjects being Caucasian), 27 of whom were professional actors, in various poses, expressions and occlusion conditions. The subjects expressed the 6 six prototypical facial expressions (examples can be seen in Fig. 7c), and up to 24 AUs. The database is fully annotated with regards to 25 AUs, split as lower (18) AUs and upper (7) AUs. The texture images are of resolution 1600×1200 pixels while the 3D faces consist of approximately 35,000 vertices. The database is accompanied by a set of available metadata consisting of 24 manually labelled facial landmarks such as nose tip, inner eye corners, etc.

One of the most recently created facial expression databases is the ICT-3DRFE database [77]. The database consists of 3D data of very high resolution recorded under varying illumination conditions, in order to test the performance of automatic 2D facial expression recognition systems. The database contains 3D models for 23 subjects (17 male and 6 female) and 15 expressions: the six prototypical expressions, two neutral states (eyes closed and open), two eyebrow expressions, scrunched face expression, and four eye gaze expressions. Each model in the dataset contains up to 1,200,000 vertices with reflectance maps of 1296×1944 pixels, resolution that corresponds to a detail level of sub-millimetre skin pores. The ability to relight the data is ensured by the reflectance information provided with every 3D model. This information allows the faces to be rendered realistically under any given illumination. The database also includes photometric information that allows photorealistic rendering. The database is fully annotated with regards to AUs. AUs are also assigned scores between 0 and 1 depending on the degree of muscle activity.

Another available facial expression database that is not widely available is the one presented in [61]. The invisible near infrared spectrum is capable of quasi-synchronous acquisition of 3D and grayscale images. The database consists of 832 sequences of 52 participants, 12 female and 40 male. In each sequence, the human subject displays a single AU (11 in total) or mimics a facial expression (happy, sad, disgust, surprise, neutral) 2–4 times. Facial action periods are of approximate duration of 5–10 s.

A database created in order to assess the individuality of facial motion for person verification was presented in [52]. This dataset was collected using the 3DMD Face Dynamic System [50], and consists of 94 participants uttering a word. Smiles were recorded from about 50 subjects. Even though the authors started collecting data of various AUs the attempt was not continued, as the authors found during the recording

Table 1
Comparative review of the tracking and alignment methods employed in facial expression recognition systems.

Method	Advantages	Disadvantages
Non-rigid ICP [54]	Able to handle variations in initial pose and occlusions well and gives dense correspondence.	Vulnerable to noisy data as will fit to all points.
FFDs [55–57]	Fast and efficient to compute, and gives dense correspondence between meshes.	Vulnerable to errors in noisy data and variations in pose.
Covariance Pyramids [58]	Able to handle varying pose and provides correspondence between individuals.	Performed on a point-by-point basis, so difficult to scale to dense correspondence.
Harmonic maps [55]	Robust to noisy data, does not suffer from local minima and gives dense correspondence.	Large differences between data and model may result in ambiguities in the correspondences.
Conformal maps [59]	Able to handle occlusions and noisy data, and gives dense correspondence.	Computationally expensive.
ASMs [60,61]	Very fast fitting process and robustness to noise.	Cannot give dense correspondence of mesh as restricted only to salient facial features.
Morphable Models [62]	Fitting process is robust to noise in the raw input and dense correspondence achieved.	Variations allowed by model are restricted by range of data used to create it.
ADMs [63]	Robust model fitting that achieves good dense correspondence.	Fitting process is computationally expensive.

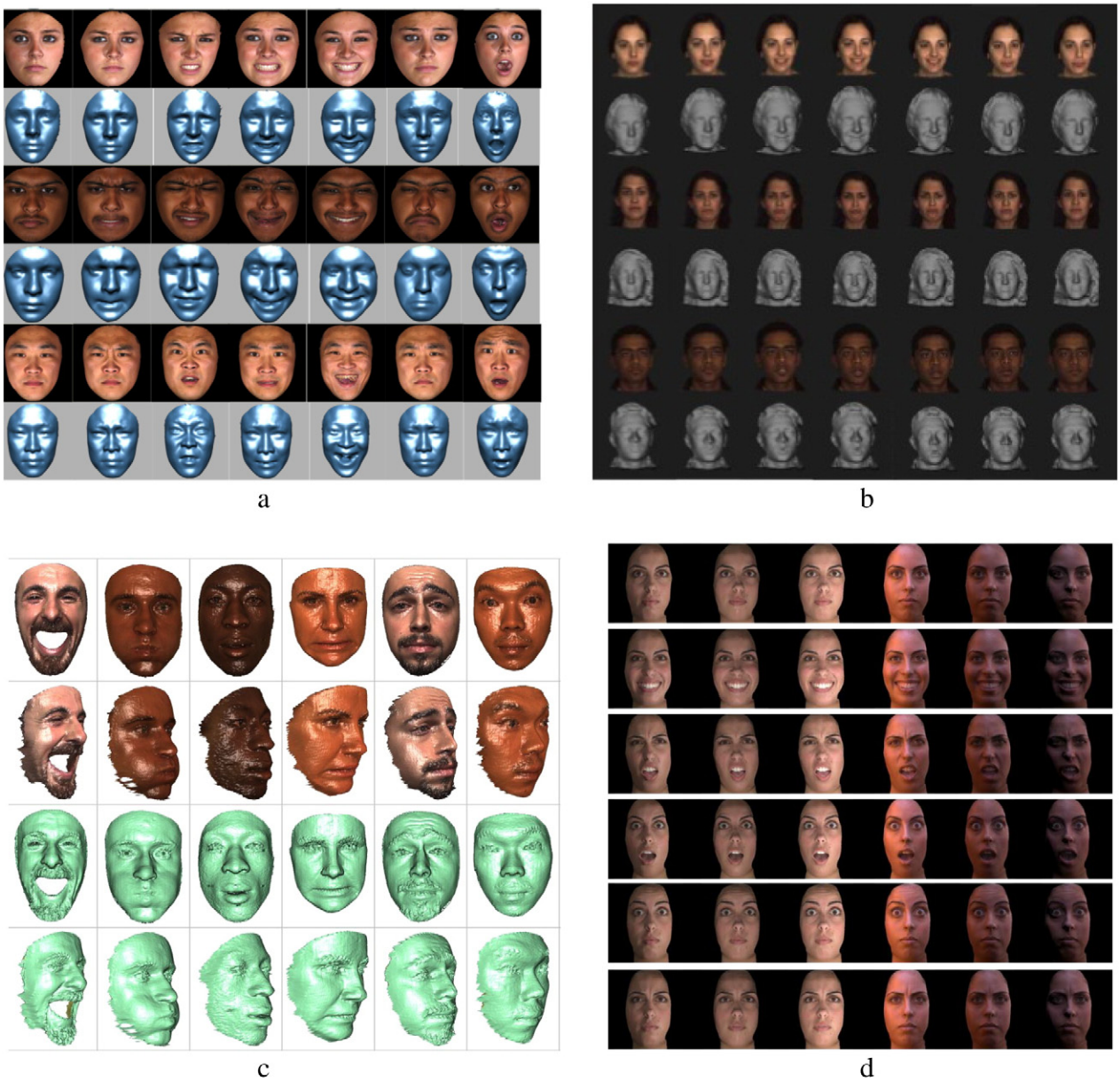

Fig. 7. Examples from the main 3D facial expression databases currently publicly available. (a) The BU-3DFE database. (b) The BU-4DFE database. (c) The Bosphorus database (d) The ICT-3DRFE database.

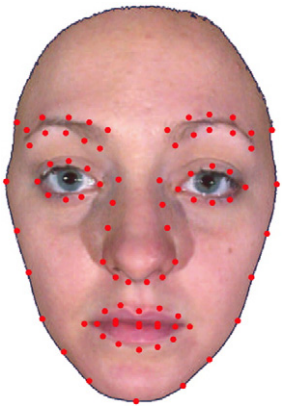

Fig. 8. The 83 facial points given in the BU-3DFE database.

sessions that the procedure of producing accurate AUs, let alone of repeating exactly the same performance several times, was of increased difficulty for the non-experienced users.

The first database to contain coded examples of dynamic 3D AUs, namely the D3DFACS, was presented in [53]. It contains 10 subjects, including 4 FACS experts, performing posed examples of up to 38 AUs in various combinations. In total, 519 AUs sequences were captured at 60 frames/s, consisting of approximately 90 frames each. The peak of each sequence has been coded by a FACS expert. The database was captured using the 3DMD Face Dynamic System [50]. It is the first database that will allow research into dynamic 3D AU recognition and analysis.

One of the earliest large publicly available 3D face databases for face verification was presented in [88,89]. The database was published in its first form in [10]. The database consists of more than 100 people and was collected using structured light technology [20]. More precisely, it contains 100 men and 100 women aged between 18 and 45 years, all of whom were Caucasians. 3D faces of the database were used for building the morphable models, after applying careful alignment [10]. The database was recorded using a CyberwareTM 3030PS laser scanner [90]. The same database in the form of a statistical morphable model was recently made publicly available [18]. This dataset was mainly used for 3D facial expression analysis and synthesis.

There exist other databases that were designed for testing the effect of expression in 3D face recognition, consisting of expressive faces. These are the extension of the FRGC dataset, the ND-2006 database [78,79], the CASIA 3D Face dataset [80,81], the Gavadb database [82], the York 3D dataset [83,84] and the Texas 3D face recognition database [85,86]. The ND-2006 contains 888 subjects with multiple images per subject displaying posed happiness, disgust, sadness and surprise. The images were acquired with a Minolta Vivid 910 range scanner [35]. The resolution of the 3D faces provided is up to 112,000 vertices. The CASIA 3D Face database contains 123 people with 10 images per person with different expressions (smile, laugh, anger and surprise) and closed eyes. It was captured using a Minolta Vivid 910. The University of York 3D face database contains 350 faces which display smiles and anger and also includes images with closed eyes and raised eyebrows. Acquisition was performed using structured light. The Texas database consists of 105 subjects performing smiles or talking, and includes 25 facial fiducial points. Finally, GavabDB is a database of 61 individuals (all Caucasian) which contains smiles (open/closed mouth) and random expressions chosen by the individuals. The database was captured by the Minolta VI-700 digitizer. Details of all these databases are summarised in Table 2.

4. Static and dynamic 3D facial expression recognition

A wide range of 3D facial expression recognition methodologies have been developed in order to perform analysis on static faces and, more recently, dynamic facial image sequences. Methods for 3D facial expression recognition generally consist of two main stages: feature extraction, and selection and classification of features. Dynamic systems may also employ temporal modelling of the expression as a further step. Here, we focus more on the techniques employed for 3D feature extraction. This is in contrast to the later stages of feature selection, classification and temporal modelling, which can be conducted using approaches similar to those used in 2D systems.

4.1. 3D feature extraction and representation

The majority of systems developed have attempted recognition of expressions from static 3D facial expression data [91–97,68,98,69,99–101,67]. However, more recent works employ dynamic 3D facial expression data for this purpose [21,74,102–104,56,57]. The features extracted for static and dynamic systems can differ greatly, due to the nature of data. We examine both kinds of systems in detail. Unless otherwise stated, quoted performance measures concern cases in which testing was performed on all available expressions or AUs in the database examined.

4.1.1. Static analysis

Several methods have been developed for the analysis of static 3D facial expressions. They use a range of different features for distinguishing

Table 2
3D face databases containing expression data. S/D: Static or dynamic data. Size: Number of subjects. Content: Expressions or AUs available (H/Sa/A/D/Su — Happy/Sad/Angry/Disgust/Surprise). Landmarks: Available landmarks. Annotation: Available annotation. P: Publicly available.

Name	S/D	Size	Content	Landmarks	Annotation	P
Chang et al. [21]	D	6 adults	6 basic expressions	N/A	N/A	N
BU-3DFE [75]	S	100 adults	6 basic expressions at 4 intensity levels	83 facial points	N/A	Y
BU-4DFE [51]	D	101 adults	6 basic expressions	83 facial points for every frame	N/A	Y
Bosphorus [76]	S	105 adults inc. 27 actors	24 AUs, neutral, 6 basic exps, occlusions	24 facial points	25 AUs	Y
ICT-3DRFE [77]	S	23 adults	15 exps: 6 basic, 2 neutral, 2 eyebrow, 1 scrunched face, 4 eye gaze	N/A	AUS with intensity levels	Y
Tsalakanidou et al. [61]	S	52 adults	11 AUs and 6 basic expressions	N/A	N/A	N
Benedikt et al. [52]	S	94 adults	Smiles and word utterance	N/A	N/A	N
D3DFACS [53]	D	10 adults inc. 4 FACS experts	Up to 38 AUs per subject	N/A	AU peaks	Y
Blanz Vetter [10,18]	S	200 adults	Neutral faces	N/A	N/A	Y
ND-2006 [78,79]	S	888 adults	Neutral and 5 exps: H, D, Sa, Su, random	N/A	N/A	Y
CASIA [80,81]	S	123 adults	Neutral and 5 exps: smile, laugh, A, Su, eyes closed	N/A	N/A	Y
Gavdb [82]	S	61 adults	3 exps: open/closed smiling and random	N/A	N/A	Y
York 3D [83,84]	S	350 adults	Neutral and 4 exps: H, A, eyes closed, eyebrows raised	N/A	N/A	Y
Texas [85,86]	S	105 adults	Neutral and smiling, or talking with open/closed eyes	25 facial points	N/A	Y

between expressions or AUs, including characteristic distances, features from statistical models such as morphable model and active shape model parameters, analysis of 2D representations and motion-based feature methods.

4.1.1.1. Distance-based features. One of the most popular methods for feature extraction in 3D static faces is the use of characteristic distances between certain facial landmarks, and the calculated changes that occur in these due to facial deformations. This is comparable to the common geometric 2D methods that track fiducial points on the face. The BU-3DFE database provides the coordinates of 83 facial points in each mesh (as depicted in Fig. 8). These points, as well as their distances, have been widely employed for static facial expression analysis [91,105,92,93,106–108,94–96].

The method developed in [91] uses six characteristic distances that are extracted from the distribution of 11 facial feature points from the given points in the BU-3DFE, thus achieving an average expression recognition rate of 91.3%. In the method proposed in [105] a larger number of distances are extracted, corresponding to how open the eyes are, the height of the eyebrows, and several features that describe the position of the mouth. This approach achieves a mean rate of 87.8%. The work in [92] is another example of the use of facial points in the BU-3DFE. The distances between these points are normalised by Facial Animation Parameter Units (FAPUs). In addition, the authors use the slope of the lines joining these points, divided by their norms in order to produce unit vectors, as an additional set of features, thus achieving an average rate of 95.1%. Similarly, [93] uses six distances that are related to the movement of particular parts of the face, plus the angles of some slopes that relate to the shape of the eyes and mouth, thus achieving an average rate of 90.2%. In [106] a wider range of distances are calculated based on the given points in the BU-3DFE, achieving an average rate of 87.1%. The distances among all pairs of available 83 facial points were also used as features in [107,108,94]. The average expression recognition rates achieved on the BU-3DFE database with these methods were equal to 93.7% [107,108], and 88.2% [94].

Moreover, in [95] features were extracted by calculating the distances among all pairs of available face points. In addition, the surface curvature at each point in the mesh was classified as belonging to one of eight categories. The face was divided into triangles using a subset of the given facial points, and histograms were formed for each triangle of the surface curvature types. This approach resulted in an average expression recognition rate of 83.5%. In [96] the authors used residues, which give both the magnitude and direction of the displacement of the given points in the BU-3DFE database, as features. A feature matrix was then formed by concatenating the different matrices in each of the three spatial directions in order to form one 2D matrix. The average rate achieved with this method was 91.7%.

Fig. 9a and b show some of the distance based features used in the literature.

4.1.1.2. Patch-based features. Patches are another method that is widely employed for feature extraction in expression recognition systems. They are used to capture information about the shape of the face over a small local region around either every point in the mesh [110], or around landmarks or feature points [101,109].

The authors in [110] computed a set of parameters for a smooth polynomial patch fitted to the local surface at each point in the mesh, which were subsequently used as inputs to rules that allowed the labelling of the surface at each point with primitives defining the type of curvature feature. An average expression recognition rate of 83.6% was achieved on a custom built database containing the six basic expressions.

Alternatively, patches were found around landmarks in the 3D mesh in [101,109]. These patches were used to define curves circling the points which show the level of the patch at those points, and the square-root velocity function (SRVF), that captures the shape of a curve, was then calculated. The extracted values of the function were used to compute the necessary deformations between curves and hence find a geodesic distance that represents the dissimilarity. The dissimilarity values were then summed for all curves in a particular patch, in order to find one distance that represents the differences among patches. This method achieved average expression recognition rates of 96.1% [101] and 98.8% [109] when tested on the BU-3DFE database.

Finally, [111] also found patches around landmarks in the face through fitting of the Statistical Facial Feature Model (SFAM), which is expressed as linear combinations of components of three different variations: shape, intensity and range value. These patches were then compared to the equivalent region from the six prototypical facial expressions through attempting to align them with ICP, and the distance between the patches after this process was used as features for classification. This approach achieved an average recognition rate of 75.8% on the BU-3DFE database.

Fig. 9c shows some of the patch based features used in the literature.

4.1.1.3. Morphable models. An alternative approach followed for feature extraction is the use of morphable models. Different implementations of morphable models, as well as of their general case, deformable models, have been used in the literature, in order to model identity, expressions, or both kinds of variations.

The Morphable Expression Model (MEM) was used in [97], and was able to model a range of different expressions for a particular individual. First the corresponding points in the expressive faces of a given subject were identified by reducing an energy function between points. Then the MEM was created by taking the principal components of the expressive faces of a person along with the average face, and performing a weighted summation of these eigen-expressions to reconstruct a new face. Subsequently, the weights formed the representation of a new expression to be recognised. This method achieved an average expression recognition rate of 97.0% over a custom database containing neutral faces and three expressions: happy, sad and angry.

Another alternative was the Basic Facial Shape Component (BFSC) model, that was able to model different identities with neutral expressions [68]. It was created as a linear combination of neutral faces. After mesh alignment using (Iterative Closest Point (ICP), the BFSC was fitted to each mesh. Due to its nature BFSC is capable of modelling only neutral faces, therefore the subtraction of the depth map of BFSC from the depth map of the original aligned mesh provides the Expression Shape Component (ESC), that contains the expression information. This difference was then used to form the expression feature vector. This method resulted in an average expression recognition rate of 76.2% on the BU-3DFE database.

The SFAM was employed as an alternative type of morphable model in [98]. The model was fitted to the meshes under examination, and the parameters of the fitting were used to extract features. The intensity and range values were directly used, while the mean of the shape parameters was subtracted from this vector to extract a set of displacement features. In addition, the shape index was calculated from these parameters, and they were subsequently encoded via multi-scale local binary patterns (LBPs) to provide further descriptors. This approach achieved average expression recognition rates of 87.2% and 82.3%, on the BU-3DFE database using manually and automatically selected landmarks, respectively. In [112], the SFAM was also used to extract features. Landmarks in the face were selected from the SFAM, either manually or automatically and features were extracted. The features consisted of the coordinates of the landmarks, as well as the morphology, texture and range parameters from local grids centred at the landmarks. LBP operators were applied to both texture and range parameters in order to encode the local properties around the landmarks. The changes in distances between some pairs of the landmarks were also calculated. This method was tested for AU recognition in the Bosphorus database, and achieved an average recognition rate of 94.2% when using all features combined.

An elastically deformable bilinear 3D model was employed in [69]. This morphable model captures variations in both identity and

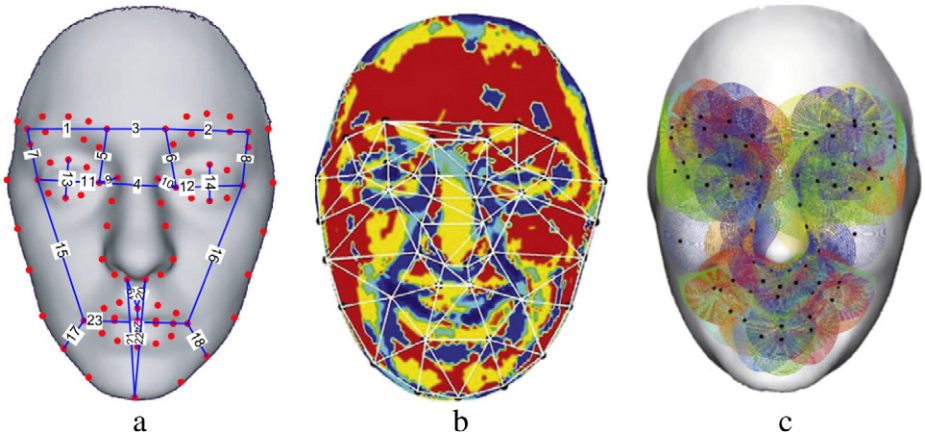

Fig. 9. Different features based on the 83 given facial points in the BU-3DFE database. (a) Distance between particular given facial points used in [106,92,107,108]. (b) Distance and curvature features used in [95]. (c) Circular patches around each facial point used in [101,109].
Image (a) taken from [106], (b) taken from [95] (c) taken from [101].

expressions (as shown in Fig. 10). A prototypic facial surface model with neutral expression and average identity was fitted at the original point cloud data, and was later used to establish correspondences between different faces. The model was fitted to the point cloud via landmarks that were identified on both the model and the cloud. A subdivision surface between these points was created by minimising an energy function using an optimisation process. The energy function was formed by taking various requirements into consideration, such as the distances of the vertices in the model from the points (and vice-versa), and the smoothness of the mesh. The mouth boundary was also detected prior to this process by first using a corner detector and then fitting a spline curve to the boundary. Once correspondence had been established, Principal Component Analysis (PCA) was applied to find the principal components of the base-mesh deformation, allowing any novel face to be written as a summation of these eigenmeshes. The face was then modelled via an asymmetric bilinear model based on these base meshes, thus allowing classification of both identity and expression via different methods. In the expression case, vector representations for each face model were formed during the fitting of the model. This approach achieved an average expression recognition rate of 90.5% on the BU-3DFE database. This feature extraction method was also used in [113,114], though the optimal parameters were in this case found by differentiating the energy function, setting it equal to zero and then performing SVD to solve the acquired linear equations. The feature extraction method was also used in [115], though the energy function now required only the distances between the points to be minimised, rather than a bidirectional pull of both sets of points to one another. These methods achieved average recognition rates of 92.3% [113], 89.5% [114], and 90.5% [115] on the BU-3DFE database, respectively.

4.1.1.4. 2D representations. An alternative approach to the problem of feature extraction from 3D image sequences includes mapping the 3D data into a 2D representation. This representation can then be used for alignment, for the division of the mesh area prior to the 3D features extraction, or for the direct application of traditional 2D techniques.

The depth map of the 3D facial meshes and the original z values at each x,y position were used as a 2D representation in [116]. The Scale-Invariant Feature Transform (SIFT) algorithm was then applied to extract features. Landmarks in the face were used as keypoints for the algorithms and local descriptors around each of these points were produced. This approach achieved an average expression recognition rate of 78.4% on the BU-3DFE database. The depth map was also employed in [117]. In this case the depth map is processed to achieve histogram equalisation over the image, and then Zernike moments [118] are taken to be used as features for classification. This method achieves average expression recognition rates of 73.0% and 60.5% for the BU-3DFE and Bosphorus databases, respectively.

Differential geometry-based features were used to convert the 3D face data into a 2D representation in [99]. The acquired 2D representation was then analysed using a traditional 2D AU detection method. The 3D data was preprocessed to smooth and remove spikes, before mapping them into 2D curvature images, as seen in Fig. 11a and b, respectively. The curvature images were subsequently used to extract various 2D features such as Gabor wavelets. This approach was tested for AU recognition in the Bosphorus database on examples with intensity C or higher, and achieved a 95.3% and 96.1% average area under the ROC curve when using either only 3D features or combined 2D and 3D features, respectively. Similarly, expressive maps were created in [119] through analysis of a variety of features such as the geometry, normals and local curvature. The maps described the discriminative nature of the points across the mesh, and could be used directly as features for classification purposes. This approach achieved a maximum average expression recognition rate of 90.4% when tested on the BU-3DFE database.

LSCM was used to form the 2D images from the 3D data in [120]. This was implemented through the method described in [121], which is an angle-preserving parameterisation method that produces consistent 2D images for 3D shapes. 2D elastic deformations were then used to estimate the correspondence between the image and a reference. Registration was performed using Gaussian image pyramids and multi-resolution meshes. Adaptive meshes were generated in order to smooth the contours and provide suitable point densities over the different parts of the face. These meshes were subsequently used for estimating the deformation via a non-rigid registration method that employs deformable triangular meshes that deform according to the stresses induced by the image matching errors. This method achieved an average area under the ROC curve of 96.2% when tested on the same data.

The authors in [74] used conformal mappings to convert the 3D meshes to 2D planar meshes and find correspondences, as described in Section 2. An example of the conformal mapping representation found for the face data in Fig. 11c can be seen in Fig. 11d. Facial surface features on the mapped mesh were then labelled according to twelve primitives to form a facial expression label map (FELM). The labels were applied through estimation of the surface principal curvatures and directions by fitting a local facial surface and using the characteristics of the resulting Hessian matrix. This method achieved an average expression recognition rate of 81.2% on the BU-3DFE database.

A combination of features derived from 2D texture, 2D and 3D surface, and 3D curvature were employed in [100]. These were determined from coefficients found from surface fitting using cubic functions and by computing the Gabor wavelet coefficients around landmarks in the face in order to compute moment invariants. During the validation stage of the

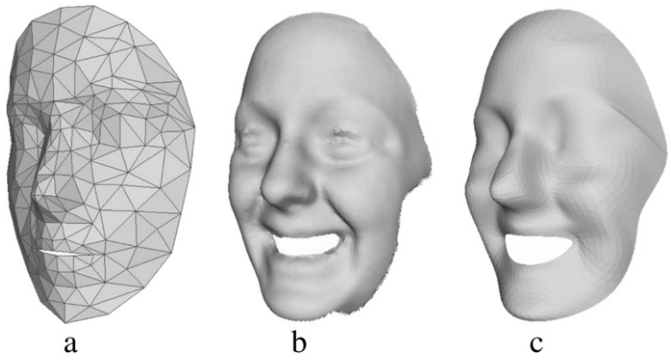

Fig. 10. Morphable model fitting in [69,113,114]. (a) The base mesh. (b) The original surface data. (c) The base mesh fitted to the surface.
Images taken from [69].

testing, the method was tested on a custom built database containing four expressions: happy, angry, fear and sadness, achieving an average expression recognition rate of 83.0%. 2D and 3D wavelet tranforms were employed in [122] in order to extract multiscale features from the 3D face data. These were used for classification purposes and resulted in an average expression recognition rate of 81.0% on a custom built dataset.

4.1.2. Dynamic analysis

Instead of employing single or multiple static images for 3D facial expression recognition, some work has begun to utilise 3D image sequences for analysis of facial expressions dynamics. Here we examine the methods that have been proposed for this purpose.

For example, in [21], feature points were tracked in order to capture the deformation of the 3D mesh during the expression. Although a small 3D database was created for expression recognition, no testing results have been reported.

One of the first works that employed 3D motion-based features for facial expression analysis was presented in [67]. A deformable model was used to track the changes between frames, thus calculating motion vectors. The acquired motion vectors were then classified via an extracted 3D facial expression label map which was produced for each expression. This resulted in an average expression recognition rate of 80.2% on the BU-3DFE database.

The method in [74], as described in the previous Section, was also applied to 3D dynamic data. This approach achieved an average expression recognition rate of 85.9% on BU-3DFE database. The tracking technique from [60] was used in both [102] and [61] to track the movement of landmarks in the face. The extracted information was subsequently used to determine the presence of different deformations in the face corresponding to particular AUs considering the change in measurements in different polygonal shapes represented by the landmarks. These methods were tested on a custom built database. The approach in [102] achieved an average recognition rate of 84.0% over neutral sequences and four expressions: happy, disgust, sadness and surprise, while the method in [61] resulted in an average expression recognition rate of 89.5% over the same four expressions, and an average AU detection rate of 89.5% over 11 AUs.

A motion-based approach was also followed in [123]. The 3D facial meshes were mapped onto a uniform 3D matrix before subtracting the matrix corresponding to the neutral state for that subject. In that way a flow matrix was produced, showing the movement appearing due to the expression evolvement through time. The Fourier Transform was subsequently applied to this matrix, and the rows of the resulting spectral matrix were concatenated to form a feature vector representing the expression. This method resulted in an average expression recognition rate of 85.6% when tested on the BU-3DFE database.

One of the first works to make use of the BU-4DFE database for the analysis of facial expression dynamics was [103], in which the deformable model presented in [67] was adapted to each frame in the image, and its changes were tracked in order to extract geometric features. This approach achieved an average expression recognition rate of 90.4% when tested on the BU-4DFE database. Facial level curves were used in [104] to extract spatio-temporal features which were subsequently used to analyse 3D dynamic expression sequences. These level curves were acquired from each frame by extracting the points that lay at a particular height on the face for different levels, after applying alignment and cropping. Features were then extracted by comparing the curves across frames using Chamfer distances. The Chamfer distances were applied to segments of the curve after partitioning by an arclength parameterised function. The features extracted from the previous, current and next frames were also considered for each frame in order to exploit temporal information. This method was tested on three expressions from the BU-4DFE database: happy, sadness and surprise, and achieved an average recognition rate of 92.2%.

Finally, motion-based features were employed in [56,57]. These were extracted using FFDs that modelled the motion between frames as the B-spline interpolation of a deformed lattice of control points. Vector projections were used to establish the spatial and temporal areas in the image sequences that contained the highest concentration of motion during the onset and offset segments of each expression. They were also used to perform a quad-tree decomposition in each pair of axes which divided the image into regions of difference sizes, with smaller regions covering the areas with the most motion. Features were then extracted from each region, consisting of measures of the amount of motion, divergence and curl of the vector field and the

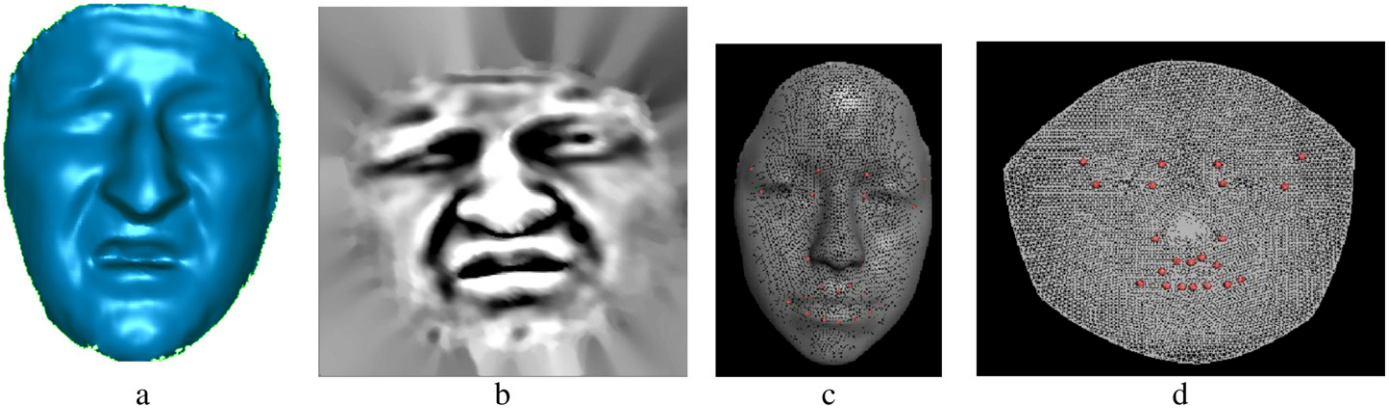

Fig. 11. 2D representations of 3D face data. (a) Original face range data. (b) 2D curvature representation used in [99]. (c) Original face range data. (d) 2D conformal map representation used in [74].
Images (a)–(b) taken from [99].

direction of the motion. This approach was tested on three expressions from the BU-4DFE database: happy, angry and surprise, and achieved an average F_1-measure of 83.0%.

4.2. Feature selection and classification

Feature selection and classification methods used for 3D facial expression analysis are generally similar to those used for the 2D cases. In this Section we briefly present existing techniques.

PCA is a widely used technique for dimensionality reduction which has been employed in several 3D facial expression recognition methods [107,108,95]. In some of these methods, linear discriminant analysis (LDA) was subsequently applied to create a discriminant subspace [107,108,67].

GentleBoost, a variation of AdaBoost was used for feature selection in [56,57]. Discriminant measures have also been employed to determine the features that should be considered for classification. The Fisher criterion was employed in [107,108], whereas the Kullback–Leibler divergence measure was used in [106] to determine the discriminative power of the feature vectors. Finally, the normalised cut-based filter (NCBF) algorithm, that aims to represent this discriminative ability with as few features as possible, was used in [95] prior to the application of PCA.

A wide range of classification techniques have been employed in 3D facial expression recognition systems. These include methods such as LDA [110,74,95] and linear classifiers [119], nearest neighbour classification [100,123], clustering algorithms [97] and Maximum Likelihood classifiers [69,114,115,120]. Rule-based classifiers, widely used for the 2D cases, have also been employed for 3D facial expression analysis [113,102,61]. In [113], the rules were discovered via ant colony and particle swarm optimisations (ACO and PCO). One of the main methods of classification that have been employed is Support Vector Machines (SVMs) [110,99,96,68,101,116,109,95,111], including multi-class SVMs [92,117]. Another technique that has been widely used is AdaBoost classification [106,99,101,109] with a selection of different weak classifiers such as linear regressors and LDA. A variation of this is GentleBoost classification, which was employed in [56,57]. Bayes classifiers have been also widely used [110,99]. Neural network classifiers constitute another popular approach [91]. Indeed, probabilistic neural networks are employed in [105,108,94,93]. Manifold learning has also been applied to the 3D feature classification problem [100].

4.3. Temporal modelling

So far, the majority of works regarding in 3D facial expression analysis have not used temporal modelling as a final stage in the classification process. This is due to the fact that existing approaches mainly employ static methods, or purely encode the temporal aspects of the data into the feature descriptors, as in [74,102,61,65]. However temporal models are widely employed in 2D dynamic facial expression analysis in order to model the dynamics as part of the classification process. This approach has been followed in a limited number of works in the literature.

Several methods use hidden Markov models (HMMs) for temporal modelling [104,56,57]. A variation of simple HMMs is the use of 2D spatio-temporal HMMs, which are employed in [103] to model both the spatial and temporal relationships in the features. An alternative method presented in [21] employed a manifold learning technique in order to embed image sequences as lines that can be traced through a low dimensional representation of the expression space.

5. Challenges and discussion

Research in 3D facial expression analysis is still in its infant stage, with a large number of works expected in the near future as the current technological advances allow the easy and affordable acquisition of high quality 3D data. However, there exist several issues that remain unsolved in this field.

There are many databases that can be used for static 3D facial expression analysis. However the current trend has shown a shift in interest of researchers towards the analysis of facial expression dynamics, as these allow the encoding of temporal cues that are indicative of more complex states and expressions. Currently, contrary to the 2D facial expression analysis, there exist only two publicly available datasets of dynamic 3D facial samples, namely BU-4DFE and D3DFACS. BU-4DFE was mainly used for facial expression recognition, and the recently published D3DFACS, which contains only 10 subjects, was designed for AU analysis. In order to design cross-database experiments, a standard procedure has to be followed in 2D facial expression and AU analysis [124,125], and in order to test the real generalisation capabilities of 4D facial expression and AU recognition/analysis algorithms, more databases of posed 4D facial expressions and AUs must be created.

Existing works in the field of facial expressions in 3D are all based on databases of acted, exaggerated expressions of the six basic emotions, although they rarely occur in our daily life. In addition, increasing evidence suggests that deliberate or acted behaviours differ in appearance and timing from spontaneous ones [126]. For instance, acted smiles have larger amplitude, shorter duration, and faster onset and offset velocity than naturally occurring smiles [127]. In turn, automatic approaches trained in laboratory settings on recordings of acted behaviour fail to generalise to the complexity of expressive behaviours found in real-world settings.

Furthermore, the human face is capable of micro-expressions which can last less than 0.04 s and be of very low intensity. A frame rate of at least 50–60 frames/s is therefore required to capture such micro expressions on at least 2–3 frames. In addition, the resolution of the recorded frames is expected to be very high in order to capture motion in very small parts of the face. Currently, a small number of solutions are available for the recording of such data. To the best of our knowledge two such commercial products are available [49,18] but both allow for restricted recording scenarios and none supports real time 3D face reconstruction. Therefore, another challenge would be the recording of 4D face databases consisting of spontaneous behaviour, captured in a range of contexts having both high spatial and temporal resolution in real-time. The equipment and setup required for the acquisition of such 3D data constitutes an even greater challenge as strategies will have to be developed in order to distract the subjects from the restrictions imposed from the environment and allow them to behave naturally, while inducing at the same time extremes of the universal expressions and other affective states. Finally, there is a demand of capturing 360 degree view of a face model. Due the limitation of current imaging systems, 4D capture is still limited to ear-to-ear frontal face, which is incapable of handling the very large pose variation issue. To handle the arbitrary head movement with large poses or postures, a full-range capture with multiple cameras setting around a head is needed. If so, a complete facial expression model can be captured under any circumstance of over 60° or 90° of rotation of a head. To this end, the issue of partial face capture or occlusion could be resolved.

A major challenge in 3D face tracking is that many existing 3D correspondence algorithms are computationally expensive. As the amount of data captured in 3D databases increases, being of high resolution and frame rates, the problem will become even bigger. Thus the area of research on optimisation methods for either model fitting or finding dense correspondences constitutes an open one. The ultimate goal for 3D facial expression systems will be real-time analysis, requiring real-time alignment and tracking, two operations that require low computational cost. Finally, another important challenge in 4D facial expression analysis occurs by the availability of both 2D texture and 3D facial surface. That is, schemes that fuse, in a dynamical manner, both sources (2D and 3D) should be developed. Source fusion constitutes a scientific field on its own , which combines elements of statistics, signal processing and machine learning. Current decision level and/or feature level

fusion schemes can be studied, but fusion algorithms that are specifically designed for the problem at hand should be preferred. The challenge is even greater in the case where information is extracted using other modalities (e.g., speech). For open problems and challenges in multimodal fusion for the task of automatic human behaviour analysis the interested reader may refer to [5].

6. Conclusions

Several approaches have been followed in the field of 3D facial expression analysis. The development of 3D data acquisition methods has allowed the creation of several databases containing 3D static faces and facial image sequences demonstrating expressions. The public availability of these databases has facilitated research in this area, particularly in static analysis. Many methods have been developed for the tracking and alignment of 3D facial meshes, a crucial step before feature extraction. Several promising approaches have been proposed for facial expression analysis. The developed systems generally share several similarities with 2D systems regarding the classification and temporal modelling techniques used. However, they differ greatly in the feature extraction methodologies used in order to exploit the benefits of the 3D facial geometries. In this survey paper we reviewed the state-of-the-art work in each of these areas, and highlighted possible directions towards which research should focus on, in order to progress beyond the state of the art.

3D facial expression analysis constitutes an open research field that is still at its infant stage. In order for the research to progress beyond the state of the art, additional databases of dynamic 3D facial expression data, plus some examples of spontaneous and natural behaviour captured in 3D are required. In addition, for the recognition of more complex affective states, capturing micro-expressions is essential, something that requires higher resolution 3D data. Real or near real-time tracking methods that are robust to occlusions and a wide range of contexts will have to be developed ensuring that important information is preserved through mapping or model fitting. As a consequence of all of the above expression analysis systems will need to become more robust and be able to adapt to spontaneous expressions and more complex states.

References

[1] M. Pantic, A. Nijholt, A. Pentland, T. Huang, Human-Centred Intelligent Human–Computer Interaction (HCI^2): how far are we from attaining it? Int. J. Auton. Adapt. Commun. Syst. 1 (2) (2008) 168–187.
[2] N. Ambady, R. Rosenthal, Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis, Psychol. Bull. 111 (2) (1992) 256.
[3] F. De la Torre, J.F. Cohn, Guide to visual analysis of humans: looking at people, Ch. Facial Expression Analysis, Springer, 2011.
[4] H. Gunes, M. Pantic, Automatic, dimensional and continuous emotion recognition, Int. J. Synthet. Emot. 1 (1) (2010) 68–99.
[5] Z. Zeng, M. Pantic, G. Roisman, T. Huang, A survey of affect recognition methods: audio, visual, and spontaneous expressions, IEEE Trans. Pattern Anal. Mach. Intell. 31 (1) (2009) 39–58.
[6] M. Nicolaou, H. Gunes, M. Pantic, Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space, IEEE Trans. Affective Comput 2 (2) (2011) 92–105 (April–June).
[7] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'ericco, M. Schroeder, Bridging the gap between social animal and unsocial machine: a survey of social signal processing, IEEE Trans, Affective Comput. IEEE Trans. Affective Comput. 3 (1) (2012) 69–87.
[8] I. Kemelmacher-Shlizerman, R. Basri, 3d face reconstruction from a single image using a single reference face shape, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2) (2011) 394–405.
[9] S. Wang, S. Lai, Reconstructing 3d face model with associated expression deformation from a single face image via constructing a low-dimensional expression deformation manifold, IEEE Trans. Pattern Anal. Mach. Intell. 3 (10) (2011) 2115–2121.
[10] V. Blanz, T. Vetter, A morphable model for the synthesis of 3D faces, In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, 1999, pp. 187–194.
[11] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Trans. Pattern Anal. Mach. Intell. (2003) 1063–1074.
[12] S. Romdhani, T. Vetter, Efficient, robust and accurate fitting of a 3D morphable model, In: Proceedings. Ninth IEEE International Conference on Computer Vision, 2003, 2003.
[13] P. Breuer, K.-I. Kim, W. Kienzle, B. Scholkopf, V. Blanz, Automatic 3D face reconstruction from single images or video, In: Automatic Face Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, 2008, pp. 1–8.
[14] A. Patel, W. Smith, 3D morphable face models revisited, In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009, pp. 1327–1334.
[15] N. Faggian, A. Paplinski, J. Sherrah, 3D morphable model fitting from multiple views, In: Automatic Face Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, 2008, pp. 1–6.
[16] D. Sibbing, M. Habbecke, L. Kobbelt, Markerless reconstruction of dynamic facial expressions, In: Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, IEEE, 2009, pp. 1778–1785.
[17] W. Brand, Morphable 3D models from video, In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001 (CVPR 2001), vol. 2, 2001, pp. II-456–II-463 vol.2.
[18] 3D morphable modelsURL, http://faces.cs.unibas.ch/bfm/main.phpMay 2011.
[19] R. Jarvis, Range sensing for computer vision, 1993.
[20] C. Beumier, M. Acheroy, 3D facial surface acquisition by structured light, In: In International Workshop on Synthetic–Natural Hybrid Coding and Three Dimensional Imaging, 1999.
[21] Y. Chang, M. Vieira, M. Turk, L. Velho, Automatic 3D facial expression analysis in videos, Anal. Model. Faces Gestures (2005) 293–307.
[22] M. Vieira, L. Velho, A. Sa, P. Carvalho, A camera–projector system for real-time 3d video, In: Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on, IEEE, 2005, p. 96.
[23] F. Tsalakanidou, F. Forster, S. Malassiotis, M. Strintzis, Real-time acquisition of depth and color images using structured light and its application to 3D face recognition, Real-Time Imaging 11 (5–6) (2005) 358–369.
[24] S. Zhang, P. Huang, High-resolution, real-time 3D shape acquisition, In: Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on, IEEE, 2004, p. 28.
[25] S. Zhang, S. Yau, High-resolution, real-time 3D absolute coordinate measurement based on a phase-shifting method, Opt. Express 14 (7) (2006) 2644–2649.
[26] Y. Wang, X. Huang, C. Lee, S. Zhang, Z. Li, D. Samaras, D. Metaxas, A. Elgammal, P. Huang, High resolution acquisition, learning and transfer of dynamic 3-D facial expressions, In: Computer Graphics Forum, vol. 23, Wiley Online Library, 2004, pp. 677–686.
[27] P. Huang, C. Zhang, F. Chiang, High-speed 3-D shape measurement based on digital fringe projection, Opt. Eng. 42 (2003) 163.
[28] Z. Geng, Rainbow three-dimensional camera: new concept of high-speed three-dimensional vision systems, Opt. Eng. 35 (1996) 376.
[29] C. Wust, D. Capson, Surface profile measurement using color fringe projection, Mach. Vis. Appl. 4 (3) (1991) 193–203.
[30] P. Huang, Q. Hu, F. Jin, F. Chiang, Color-encoded digital fringe projection technique for high-speed three-dimensional surface contouring, Opt. Eng. 38 (1999) 1065.
[31] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, H. Fuchs, The office of the future: a unified approach to image-based modeling and spatially immersive displays, In: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, ACM, 1998, pp. 179–188.
[32] S. Rusinkiewicz, O. Hall-Holt, M. Levoy, Real-time 3D model acquisition, In: ACM Transactions on Graphics (TOG), vol. 21, ACM, 2002, pp. 438–446.
[33] O. Hall-Holt, S. Rusinkiewicz, Stripe boundary codes for real-time structured-light range scanning of moving objects, In: Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, vol. 2, IEEE, 2001, pp. 359–366.
[34] T. Weise, B. Leibe, L. Van Gool, Fast 3D scanning with automatic motion compensation, In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007, pp. 1–8.
[35] Minolta Vivid 910URL, http://www.konicaminolta.com/instruments/products/3d/non-contact/vivid910/features.htmlMay 2011.
[36] Inspeck Mega Capturor II DigitizerURL, http://www.inspeck.comMay 2011.
[37] KinectURL, http://www.xbox.com/en-GB/kinectMay 2011.
[38] R. Woodham, Photometric method for determining surface orientation from multiple images, Opt. Eng. 19 (1) (1980) 139–144.
[39] G. Brostow, C. Hernandez, G. Vogiatzis, B. Stenger, R. Cipolla, Video normals from colored lights, IEEE Trans. Pattern Anal. Mach. Intell. 1 (2011).
[40] R. Frankot, R. Chellappa, A method for enforcing integrability in shape from shading algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 10 (4) (1988) 439–451.
[41] A. Agrawal, R. Raskar, R. Chellappa, What is the range of surface reconstructions from a gradient field? In: Lecture Notes in Computer Science (European Conference on Computer Vision (ECCV 2006)), 3951, 2006, p. 578.
[42] A. Agrawal, R. Chellappa, R. Raskar, An algebraic approach to surface reconstruction from gradient fields, In: Proc. Intl Conf. Computer Vision, vol. 1, 2005, pp. 174–181.
[43] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.
[44] T. Simchony, R. Chellappa, M. Shao, Direct analytical methods for solving Poisson equations in computer vision problems, IEEE Trans. Pattern Anal. Mach. Intell. 12 (5) (1990) 435–446.
[45] S. Zafeiriou, M. Hansen, G. Atkinson, V. Argyriou, M. Petrou, M. Smith, L. Smith, The photoface database, In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Workshop on Biometrics), 2011 (CVPRW 2011), 2011.
[46] M. Hansen, G. Atkinson, L. Smith, M. Smith, 3D face reconstructions from photometric stereo using near infrared and visible light, Comput. Vision Image Underst. 114 (8) (2010) 942–951.
[47] S. Seitz, B. Curless, J. Diebel, D. Scharstein, R. Szeliski, A comparison and evaluation of multi-view stereo reconstruction algorithms, In: Computer Vision and

Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 1, IEEE, 2006, pp. 519–528.
[48] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, M. Gross, High-quality single-shot capture of facial geometry, ACM Trans. on Graphics (Proc. SIGGRAPH) 29 (3) (2010).
[49] DI4D — 4D Capture SystemsURL, http://www.di3d.com/products/4dsystems/May 2011.
[50] 3DMD 4D CaptureURL, http://www.3dmd.com/3dmd4d-capture.htmlMay 2011.
[51] L. Yin, X. Chen, Y. Sun, T. Worm, M. Reale, A high-resolution 3D dynamic facial expression database, In: 8th International Conference on Automatic Face and Gesture Recognition, 2008 (FGR 2008), 2008, pp. 1–6.
[52] L. Benedikt, D. Cosker, P. Rosin, D. Marshall, Assessing the uniqueness and permanence of facial actions for use in biometric applications, IEEE Trans. Syst. Man Cybern. A Syst. Hum. 40 (3) (2010) 449–460.
[53] D. Cosker, E. Krumhuber, A. Hilton, A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling, In: Proc. of IEEE International Conference on Computer Vision (ICCV), 2011.
[54] B. Amberg, S. Romdhani, T. Vetter, Optimal step nonrigid ICP algorithms for surface registration, In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007, pp. 1–8.
[55] Y. Wang, M. Gupta, S. Zhang, S. Wang, X. Gu, D. Samaras, P. Huang, High resolution tracking of non-rigid motion of densely sampled 3D data using harmonic maps, Int. J. Comput. Vis. 76 (3) (2008) 283–300.
[56] G. Sandbach, S. Zafeiriou, M. Pantic, D. Rueckert, A dynamic approach to the recognition of 3D facial expressions and their temporal models, In: Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, 2011, pp. 406–413.
[57] G. Sandbach, S. Zafeiriou, M. Pantic, D. Rueckert, Recognition of 3D facial expression dynamics, Image Vision Comput. (0) (2012) http://dx.doi.org/10.1016/j.imavis.2012.01.006. URL http://www.sciencedirect.com/science/article/pii/S02628 85612000157.
[58] M. Kaiser, B. Kwolek, C. Staub, G. Rigoll, Registration of 3D facial surfaces using covariance matrix pyramids, In: Robotics and Automation (ICRA), 2010 IEEE International Conference on, 2010, pp. 1002–1007.
[59] S. Wang, Y. Wang, M. Jin, X. Gu, D. Samaras, Conformal geometry and its applications on 3D shape matching, recognition, and stitching, IEEE Trans. Pattern Anal. Mach. Intell. (2007) 1209–1220.
[60] F. Tsalakanidou, S. Malassiotis, Real-time facial feature tracking from 2D + 3D video streams, In: 3DTV Conference, 2010.
[61] F. Tsalakanidou, S. Malassiotis, Real-time 2D + 3D facial action and expression recognition, Pattern Recognit. 43 (2010) 1763–1775.
[62] E. Munandoz, J. Buenaposada, L. Baumela, A direct approach for efficiently tracking with 3D morphable models, In: Computer Vision, 2009 IEEE 12th International Conference on, 2009, pp. 1615–1622.
[63] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, T. Theoharis, Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 640–649 [doi:http://doi.ieeecomputersociety.org/10.1109/TPAMI.2007.1017].
[64] P. Besl, N. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. (1992) 239–256.
[65] Y. Sun, X. Chen, M. Rosato, L. Yin, Tracking vertex flow and model adaptation for three-dimensional spatiotemporal face analysis, IEEE Trans. Syst. Man Cybern. A Syst. Hum. 40 (3) (2010) 461–474.
[66] S. Wang, X. Gu, H. Qin, Automatic non-rigid registration of 3D dynamic data for facial expression synthesis and transfer, In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1–8.
[67] L. Yin, X. Wei, P. Longo, A. Bhuvanesh, Analyzing facial expressions using intensity-variant 3D data for human computer interaction, In: Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, vol. 1, 2006, pp. 1248–1251.
[68] 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 Seventeen ACM International Conference on Multimedia, 2009, pp. 569–572.
[69] I. Mpiperis, S. Malassiotis, M. Strintzis, Bilinear models for 3-D face and facial expression recognition, IEEE Trans. Inf. Forensics Secur. 3 (3) (2008) 498–511.
[70] D. Chetverikov, D. Svirko, D. Stepanov, P. Krsek, The trimmed iterative closest point algorithm, In: Pattern Recognition, 2002. Proceedings. 16th International Conference on, vol. 3, IEEE, 2002, pp. 545–548.
[71] C. Stewart, C. Tsai, B. Roysam, The dual-bootstrap iterative closest point algorithm with application to retinal image registration, IEEE Trans. Med. Imaging 22 (11) (2003) 1379–1394.
[72] S. Rusinkiewicz, M. Levoy, Efficient variants of the ICP algorithm, In: 3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on, IEEE, 2001, pp. 145–152.
[73] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, D. Hawkes, Nonrigid registration using free-form deformations: application to breast MR images, IEEE Trans. Med. Imaging 18 (8) (1999) 712–721.
[74] M. Rosato, X. Chen, L. Yin, Automatic registration of vertex correspondences for 3D facial expression analysis, In: Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on, 2008, pp. 1–7.
[75] L. Yin, X. Wei, Y. Sun, J. Wang, M. Rosato, A 3D facial expression database for facial behavior research, In: 7th International Conference on Automatic Face and Gesture Recognition, 2006 (FGR 2006), 2006, pp. 211–216.
[76] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, L. Akarun, Bosphorus database for 3D face analysis, In: Proc. First COST 2101 Workshop on Biometrics and Identity Management, Roskilde University, Denmark, 2008, pp. 47–56.
[77] G. Stratou, A. Ghosh, P. Debevec, L.-P. Morency, Effect of illumination on automatic expression recognition: a novel 3D relightable facial database, In: 9th International Conference on Automatic Face and Gesture Recognition, 2011 (FGR 2011), Santa Barbara, California, USA, 2011, pp. 1–6.
[78] T. Faltemier, K. Bowyer, P. Flynn, Using a multi-instance enrollment representation to improve 3D face recognition, IEEE BTAS 1. First IEEE International Conference on Biometrics: Theory, Applications, and Systems (2007) BTAS 2007.
[79] T. Faltemier, K. Bowyer, P. Flynn, Using multi-instance enrollment to improve performance of 3D face recognition, Comput. Vision Image Underst. 112 (2) (2008) 114–125.
[80] C. Zhong, Z. Sun, T. Tan, Robust 3D face recognition using learned visual codebook, In: IEEE Conference on Computer Vision and Pattern Recognition, 2007 (CVPR'07), 2007, pp. 1–6.
[81] Casia 3D face databaseURL, http://www.cbsr.ia.ac.cn/english/3DFaceDatabases.aspMay 2011.
[82] A. Moreno, A. Sanchez, Gavabdb: a 3D face database, In: Proc. 2nd COST275 Workshop on Biometrics on the Internet, Vigo (Spain), 2004, pp. 75–80.
[83] York 3D databaseURL, http://www-users.cs.york.ac.uk/nep/research/3Dface/tomh/3DFaceRecognition.htmlMay 2011.
[84] T. Heseltine, N. Pears, J. Austin, Three-dimensional face recognition using combinations of surface feature map subspace components, Image Vision Comput. 26 (3) (2008) 382–396.
[85] Texas 3D Face Recognition DatabaseURL, http://live.ece.utexas.edu/research/texas3dfrMay 2010.
[86] S. Gupta, K. Castleman, M. Markey, A. Bovik, Texas 3D face recognition database, In: Image Analysis & Interpretation (SSIAI), 2010 IEEE Southwest Symposium on, IEEE, 2010, pp. 97–100.
[87] 3DMD 3D Face CapturingURL, http://www.3dmd.com/3dmdface.htmlMay 2011.
[88] C. Beumier, M. Acheroy, Face verification from 3D and grey level clues, Pattern Recognit. Lett. 22 (12) (2001) 1321–1329.
[89] C. Beumier, M. Acheroy, SIC DB: multi-modal database for person authentication, In: Image Analysis and Processing, 1999. Proceedings. International Conference on, 1999, pp. 704–708.
[90] CyberwareTM 3030PSURL, http://www.cyberware.com/products/scanners/px.htmlMay 2011.
[91] H. Soyel, H. Demirel, Facial expression recognition using 3D facial feature distances, Image Anal. Recognit. (2007) 831–838.
[92] H. Tang, T. Huang, 3D facial expression recognition based on properties of line segments connecting facial feature points, In: Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, 2008, pp. 1–6.
[93] X. Li, Q. Ruan, Y. Ming, 3D facial expression recognition based on basic geometric features, In: Signal Processing (ICSP), 2010 IEEE 10th International Conference on, 2010, pp. 1366–1369.
[94] U. Tekguc, H. Soyel, H. Demirel, Feature selection for person-independent 3D facial expression recognition using NSGA — II, In: Computer and Information Sciences, 2009. ISCIS 2009. 24th International Symposium on, 2009, pp. 35–38.
[95] T. Sha, M. Song, J. Bu, C. Chen, D. Tao, Feature level analysis for 3D facial expression recognition, Neurocomputing (in press).
[96] R. Srivastava, S. Roy, 3D facial expression recognition using residues, In: TENCON 2009–2009 IEEE Region 10 Conference, 2009, pp. 1–5.
[97] S. Ramanathan, A. Kassim, Y. Venkatesh, W. Wah, Human facial expression recognition using a 3D morphable model, In: Image Processing, 2006 IEEE International Conference on, 2007, pp. 661–664.
[98] X. Zhao, E. Di Huang, L. Chen, Automatic 3D facial expression recognition based on a Bayesian belief net and a statistical facial feature model, In: 2010 International Conference on Pattern Recognition, 2010, pp. 3724–3727.
[99] A. Savran, B. Sankur, T. Bilge, Comparative evaluation of 3D versus 2D modality for automatic detection of facial action units, Pattern Recognition 45, (February) (2012), 767–782.
[100] P. Wang, C. Kohler, F. Barrett, R. Gur, R. Gur, R. Verma, Quantifying facial expression abnormality in schizophrenia by combining 2D and 3D features, In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[101] A. Maalej, B. Amor, M. Daoudi, A. Srivastava, S. Berretti, Local 3D shape analysis for facial expression recognition, In: 2010 International Conference on Pattern Recognition, 2010, pp. 4129–4132.
[102] F. Tsalakanidou, S. Malassiotis, Robust facial action recognition from real-time 3D streams, In: Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on, 2009, pp. 4–11.
[103] Y. Sun, L. Yin, Facial expression recognition based on 3D dynamic range model sequences, In: Computer Vision — ECCV 2008, 2008, pp. 58–71.
[104] V. Le, H. Tang, T. Huang, Expression recognition from 3D dynamic faces using robust spatio-temporal shape features, In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition (FG'11), Special Session: 3D Facial Behavior Analysis and Understanding, Santa Barbara, CA, USA, 2011.
[105] H. Soyel, H. Demirel, 3D facial expression recognition with geometrically localized facial features, In: Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on, 2008, pp. 1–4.
[106] H. Tang, T. Huang, 3D facial expression recognition based on automatically selected features, In: Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on, 2008, pp. 1–8.
[107] H. Soyel, H. Demirel, Optimal feature selection for 3D facial expression recognition with geometrically localized facial features, In: Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control, 2009. ICSCCW 2009. Fifth International Conference on, 2009, pp. 1–4.
[108] H. Soyel, H. Demirel, Optimal feature selection for 3D facial expression recognition using coarse-to-fine classification, Turk. J. Electr. Eng. Comput. Sci. 18 (6) (2010) 1031–1040.

[109] A. Maalej, B. Amor, M. Daoudi, A. Srivastava, S. Berretti, Shape analysis of local facial patches for 3D facial expression recognition, Pattern Recognit. 44 (8) (2011) 1581–1589.
[110] J. Wang, L. Yin, X. Wei, Y. Sun, 3D facial expression recognition based on primitive surface feature distribution, In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 2, 2006, pp. 1399–1406.
[111] P. Lemaire, B. Ben Amor, M. Ardabilian, L. Chen, M. Daoudi, Fully automatic 3D facial expression recognition using a region-based approach, In: Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding, J-HGBU'11, ACM, New York, NY, USA, 2011, pp. 53–58, http://dx.doi.org/10.1145/2072572.2072589. URL http://doi.acm.org/10.1145/2072572.2072589.
[112] X. Zhao, E. Dellandrea, L. Chen, D. Samaras, Au recognition on 3D faces based on an extended statistical facial feature model, In: Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, 2010, pp. 1–6.
[113] I. Mpiperis, S. Malassiotis, V. Petridis, M. Strintzis, 3D facial expression recognition using swarm intelligence, In: Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, 2008, pp. 2133–2136.
[114] I. Mpiperis, S. Malassiotis, M. Strintzis, Bilinear elastically deformable models with application to 3D face and facial expression recognition, In: Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, 2008, pp. 1–8.
[115] S. Mpiperis, S. Malassiotis, M.G. Strintzis, Bilinear decomposition of 3D face images: an application to facial expression recognition, In: 10th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2009), 2009.
[116] S. Berretti, B. Ben Amor, M. Daoudi, A. del Bimbo, 3D facial expression recognition using sift descriptors of automatically detected keypoints, Vis. Comput. 27 (2011) 1021–1036 [http://dx.doi.org/10.1007/s00371-011-0611-x URL http://dx.doi.org/10.1007/s00371-011-0611-x].
[117] N. Vretos, N. Nikolaidis, I. Pitas, 3D facial expression recognition using Zernike moments on depth images, In: Image Processing (ICIP), 2011 18th IEEE International Conference on, 2011, pp. 773–776, http://dx.doi.org/10.1109/ICIP.2011.6116669.
[118] F. Zernike, Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode, Physica 1 (7) (1934) 689–704.
[119] O. Ocegueda, T. Fang, S. Shah, I. Kakadiaris, Expressive maps for 3D facial expression recognition, In: Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, 2011, pp. 1270–1275, http://dx.doi.org/10.1109/ICCVW.2011.6130397.
[120] A. Savran, B. Sankur, Automatic detection of facial actions from 3D data, In: Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, 2009, pp. 1993–2000.
[121] A. Savran, B. Sankur, Non-rigid registration of 3D surfaces by deformable 2D triangular meshes, In: Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on, 2008, pp. 1–6.
[122] S. Pinto, J. Mena-Chalco, F. Lopes, L. Velho, R. Cesar, 3D facial expression analysis by using 2D and 3D wavelet transforms, In: Image Processing (ICIP), 2011 18th IEEE International Conference on, 2011, pp. 1281–1284, http://dx.doi.org/10.1109/ICIP.2011.6115668.
[123] Y. Venkatesh, A. Kassim, O. Murthy, Resampling approach to facial expression recognition using 3D meshes, In: 2010 International Conference on Pattern Recognition, 2010, pp. 3772–3775.
[124] S. Koelstra, M. Pantic, I. Patras, A dynamic texture-based approach to recognition of facial actions and their temporal models, IEEE Trans. Pattern Anal. Mach. Intell. (2010) 1940–1954.
[125] S. Valstar, M. Pantic, J. Cohn, Fully automatic recognition of the temporal phases of facial actions, Sys. Man and Cybern. Part B: Cybern. IEEE Trans (99) (2012) 1–16.
[126] Z. Ambadar, J. Schooler, J. Cohn, Deciphering the enigmatic face, Psychol. Sci. 16 (5) (2005) 403–410.
[127] M. Valstar, H. Gunes, M. Pantic, How to distinguish posed from spontaneous smiles using geometric features, In: Proceedings of ACM Int'l Conf. Multimodal Interfaces (ICMI'07), Nagoya, Japan, 2007, pp. 38–45.