21

3D Face Recognition

Theodoros Papatheodorou and Daniel Rueckert
Department of Computing, Imperial College London
UK

1. Introduction

The survival of an individual in a socially complex world depends greatly on the ability to interpret visual information about the age, sex, race, identity and emotional state of another person based on that person's face. Despite a variety of different adverse conditions (varying facial expressions and facial poses, differences in illumination and appearance), humans can perform face identification with remarkable robustness without conscious effort.

Face recognition research using automatic or semi-automatic techniques emerged in the 1960s, and especially in the last two decades it has received significant attention. One reason for this growing interest is the wide range of possible applications for face recognition systems. Another reason is the emergence of affordable hardware, such as digital photography and video, which have made the acquisition of high-quality and high-resolution images much more ubiquitous. Despite this growing attention, the current state-of-the-art face recognition systems perform well when facial images are captured under uniform and controlled conditions. However, the development of face recognition systems that work robustly in uncontrolled situations is still an open research issue.

Even though there are various alternative biometric techniques that perform very well today, e.g. fingerprint analysis and iris scans, these methods require the cooperation of the subjects and follow a relatively strict data acquisition protocol. Face recognition is much more flexible since subjects are not necessarily required to cooperate or even be aware of being scanned and identified. This makes face recognition a less intrusive and potentially more effective identification technique. Finally, the public's perception of the face as a biometric modality is more positive compared to the other modalities (Hietmeyer, 2000).

1.1 Challenges for face recognition

The face is a three-dimensional (3D) object. Its appearance is determined by the shape as well as texture of the face. Broadly speaking, the obstacles that a face recognition systemmust overcome are differences in appearance due to variations in illumination, viewing angle, facial expressions, occlusion and changes over time.

Using 2D images for face recognition, the intensities or colours of pixels represent all the information that is available and therefore, any algorithm needs to cope with variation due to illumination explicitly. The human brain seems also to be affected by illumination in performing face recognition tasks (Hill et al., 1997). This is underlined by the difficulty of identifying familiar faces when lit from above (Johnston et al., 1992) or from different

Source: Face Recognition, Book edited by: Kresimir Delac and Mislav Grgic, ISBN 978-3-902613-03-5, pp.558, I-Tech, Vienna, Austria, June 2007

directions (Hill and Bruce, 1996). Similarly it has been shown that faces shown in photographic negatives had a detrimental effect on the identification of familiar faces (Bruce and Langton, 1994). Further studies have shown that the effect of lighting direction can be a determinant of the photographic negative effect (Liu et al., 1999). As a result, positive faces, which normally appear to be top-lit, may be difficult to recognize in negative partly because of the accompanying change in apparent lighting direction to bottom-lit. One explanation for these findings is that dramatic illumination or pigmentation changes interfere with the shape-from-shading processes involved in constructing representations of faces. If the brain reconstructs 3D shape from 2D images, it remains a question why face recognition by humans remains viewpointdependent to the extent it is.

One of the key challenges for face recognition is the fact that the difference between two images of the same subject photographed from different angles is greater than the differences between two images of different subjects photographed from the same angle. It has been reported that recognition rates for unfamiliar faces drop significantly when there are different viewpoints for the training and test set (Bruce, 1982). More recently, however, there has been debate about whether object recognition is viewpoint-dependent or not (Tarr and Bulthoff, 1995). It seems that the brain is good at generalizing from one viewpoint to another as long as the change in angle is not extreme. For example, matching a profile viewpoint to a frontal image is difficult, although the matching of a three-quarter view to a frontal seems to be less difficult (Hill et al., 1997). There have been suggestions that the brain might be storing a view-specific prototype abstraction of a face in order to deal with varying views (Bruce, 1994). Interpolation-based models (Poggio and Edelman, 1991), for example, support the idea that the brain identifies faces across different views by interpolating to the closest previously seen view of the face.

Another key challenge for face recognition is the effect of facial expressions on the appearance of the face. The face is a dynamic structure that changes its shape non-rigidly since muscles deform soft tissue and move bones. Neurophysiologic studies have suggested that facial expression recognition happens in parallel to face identification (Bruce, 1988). Some case studies in prosopagnostic patients show that they are able to recognize expressions even though identifying the actor remains a near-impossible task. Similarly, patients who suffer from *organic brain syndrome* perform very poorly in analyzing expressions but have no problems in performing face recognition. However, the appearance of the face also changes due to aging and people's different lifestyles. For example, skin becomes less elastic and more loose with age, the lip and hair-line often recedes, the skin color changes, people gain or lose weight, grow a beard, change hairstyle etc. This can lead to dramatic changes in the appearance of faces in images.

A final challenge for face recognition is related to the problem of occlusions. Such occlusions can happen for a number of reasons, e.g. part of the face maybe occluded and not visiblewhen images are taken from certain angles or because the subject grew a beard, is wearing glasses or a hat.

2. From 2D to 3D face recognition

2D face recognition is a much older research area than 3D face recognition research and broadly speaking, at the present, the former still outperforms the latter. However, the wealth of information available in 3D face data means that 3D face recognition techniques might in the near future overtake 2D techniques. In the following we examine some of the inherent differences between 2D and 3D face recognition.

2.1 Advantages and disadvantages of 3D face recognition

As previously discussed, face recognition using 2D images is sensitive to illumination changes. The light collected froma face is a function of the geometry of the face, the albedo of the face, the properties of the light source and the properties of the camera. Given this complexity, it is difficult to develop models that take all these variations into account. Training using different illumination scenarios as well as illumination normalization of 2D images has been used, but with limited success. In 3D images, variations in illumination only affect the texture of the face, yet the captured facial shape remains intact (Hesher et al., 2003).

Another differentiating factor between 2D and 3D face recognition is the effect of pose variation. In 2D images effort has been put into transforming an image into a canonical position (Kim and Kittler, 2005). However, this relies on accurate landmark placement and does not tackle the issue of occlusion. Moreover, in 2D this task is nearly impossible due to the projective nature of 2D images. To circumvent this problem it is possible to store different views of the face (Li et al., 2000). This, however, requires a large number of 2D images from many different views to be collected. An alternative approach to address the pose variation problem in 2D images is either based on statistical models for view interpolation (Lanitis et al., 1995; Cootes et al., 1998) or on the use of generative models (Prince and Elder, 2006). Other strategies including sampling the plenoptic function of a face using lightfield techniques (Gross et al., 2002). Using 3D images, this view interpolation can be simply solved by re-rendering the 3D face data with a new pose. This allows a 3D morphable model to estimate the 3D shape of unseen faces from non-frontal 2D input images and to generate 2D frontal views of the reconstructed faces by re-rendering (Blanz et al., 2005). Another pose-related problem is that the physical dimensions of the face in 2D images are unknown. The size of a face in 2D images is essentially a function of the distance of the subject from the sensor. However, in 3D images the physical dimensions of the face are known and are inherently encoded in the data.

In contrast to 2D images, 3D images are better at capturing the surface geometry of the face. Traditional 2D image-based face recognition focuses on high-contrast areas of the face such as eyes, mouth, nose and face boundary because low contrast areas such as the jaw boundary and cheeks are difficult to describe from intensity images (Gordon, 1992). 3D images, on the other hand, make no distinction between high- and low-contrast areas. 3D face recognition, however, is not without its problems. Illumination, for example, may not be an issue during the processing of 3D data, but it is still a problem during capturing. Depending on the sensor technology used, oily parts of the face with high reflectance may introduce artifacts under certain lighting on the surface. The overall quality of 3D image data collected using a range camera is perhaps not as reliable as 2D image data, because 3D sensor technology is currently not as mature as 2D sensors. Another disadvantage of 3D face recognition techniques is the cost of the hardware. 3D capturing equipment is getting cheaper and more widely available but its price is significantly higher compared to a high-resolution digital camera. Moreover, the current computational cost of processing 3D data is higher than for 2D data.

Finally, one of the most important disadvantages of 3D face recognition is the fact that 3D capturing technology requires cooperation from a subject. As mentioned above, lens or laserbased scanners require the subject to be at a certain distance from the sensor. Furthermore, a laser scanner requires a few seconds of complete immobility, while a traditional camera can capture images from far away with no cooperation from the subjects. In addition, there are currently very few high-quality 3D face databases available for testing and evaluation purposes. Those databases that are available are of very small size compared to 2D face databases used for benchmarking.

3. An overview of 3D face recognition

Despite some early work in 3D face recognition in the late 1980s (Cartoux et al., 1989) relatively few researchers have focused on this area during the 1990s. By the end of the last decade interest in 3D face recognition was revived and has increased rapidly since then. In the following we will review the current state-of-the-art in 3D face recognition. We have divided 3D face recognition techniques broadly into three categories: surface-based, statistical and model-based approaches.

3.1 Surface-based approaches

Surface-based approaches use directly the surface geometry that describes the face. These approaches can be classified into those that extract either local and global features of the surface (e.g. curvature), those that are based on profile lines, and those which use distance-based metrics between surfaces for 3D face recognition.

3.1.1 Local methods

One approach for 3D face recognition uses a description of local facial characteristics based on *Extended Gaussian Images* (EGI) (Lee and Milios, 1990). Alternatively the surface curvature can be used to segment the facial surfaces into features that can be used for matching (Gordon, 1992). Another approach is based on 3D descriptors of the facial surface in terms of their mean and Gaussian curvatures (Moreno et al., 2003) or in terms of distances and the ratios between feature points and the angles between feature points (Lee et al., 2005).

Another locally-oriented technique is based on using *point signatures*, an attempt to describe complex free-form surfaces, such as the face (Chua and Jarvis, 1997). The idea is to form a representation of the neighbourhood of a surface point. These point signatures can be used for surface comparisons by matching the signatures of data points of a "sensed" surface to the signatures of data points representing the model's surface (Chua et al., 2000). To improve the robustness towards facial expressions, those parts of the face that deform non-rigidly (mouth and chin) can be discarded and only other rigid regions (e.g. forehead, eyes, nose) are used for face recognition. In a similar approach this approach has been extended by fusing extracted 3D shape and 2D texture features (Wang et al., 2002).

Finally, hybrid techniques that use both local and global geometric surface information can be employed. In one such approach local shape information, in the form of *Gaussian-Hermite moments*, is used to describe an individual face along with a 3D mesh representing the whole facial surface. Both global and local shape information are encoded as a combined vector in a low-dimensional PCA space, and matching is based on minimum distance in that space (Xu et al., 2004).

3.1.2 Global methods

Global surface-based methods are methods that use the whole face as the input to a recognition system. One of the earliest systems is based on locating the face's plane of bilateral symmetry and to use this for aligning faces (Cartoux et al., 1989). The facial profiles along this plane are then extracted and compared. Faces can also be represented based on the analysis of maximum and minimum principal curvatures and their directions (Tanaka et al., 1998). In these approaches the entire face is represented as an EGI. Another approach uses EGIs to summarize the surface normal orientation statistics across the facial surface (Wong et al., 2004).

A different type of approach is based on distance-based techniques for face matching. For example, the *Hausdorff distance* has been used extensively for measuring the similarity between 3D faces (Ackermann, B. and Bunke, H., 2000; Pan et al., 2003). In addition, several modi- fied versions of the Hausdorff distance metric have been proposed (Lee and Shim, 2004; Russ et al., 2005). Several other authors have proposed to perform face alignment using rigid registration algorithms such as *iterative closest point algorithm* (ICP) Besl and McKay (1992). After registration the residual distances between faces can be measured and used to define a similarity metric (Medioni and Waupotitsch, 2003). In addition, surface geometry and texture can be used jointly for registration and similarity measurement in the registration process, and measures not only distances between surfaces but also between texture (Papatheodorou and Rueckert, 2004). In this case each point on the facial surface is described by its position and texture. An alternative strategy is to use a fusion approach for shape and texture (Maurer et al., 2005). In addition to texture, other surface characteristics such as the shape index can be integrated into the similarity measure (Lu et al., 2004). An important limitation of these approaches is the assumption that the face does not deform and therefore a rigid registration is sufficient to align faces. This assumption can be relaxed by allowing some non-rigid registration, e.g. using thin-plate splines (TPS) (Lu and Jain, 2005a).

Another common approach is based on the registration and analysis of 3D profiles and contours extracted from the face (Nagamine et al., 1992; Beumier and Acheroy, 2000; Wu et al., 2003). The techniques can also be used in combination with texture information (Beumier and Acheroy, 2001).

3.2 Statistical approaches

Statistical techniques such as Principal Component Analysis (PCA) are widely used for 2D facial images. More recently, PCA-based techniques have also been applied to 3D face data (Mavridis et al., 2001; Hesher et al., 2003; Chang et al., 2003; Papatheodorou and Rueckert, 2005). This idea can be extended to include multiple features into the PCA such as colour, depth and a combination of colour and depth (Tsalakanidou et al., 2003). These PCA-based techniques can also be used in conjunction with other classification techniques, e.g. *embedded* hidden Markov models (EHMM) (Tsalakanidou et al., 2004). An alternative approach is based on the use of Linear Discriminant Analysis (LDA) (G¨okberk et al., 2005) or Independent Component Analysis (ICA) (Srivastava et al., 2003) for the analysis of 3D face data.

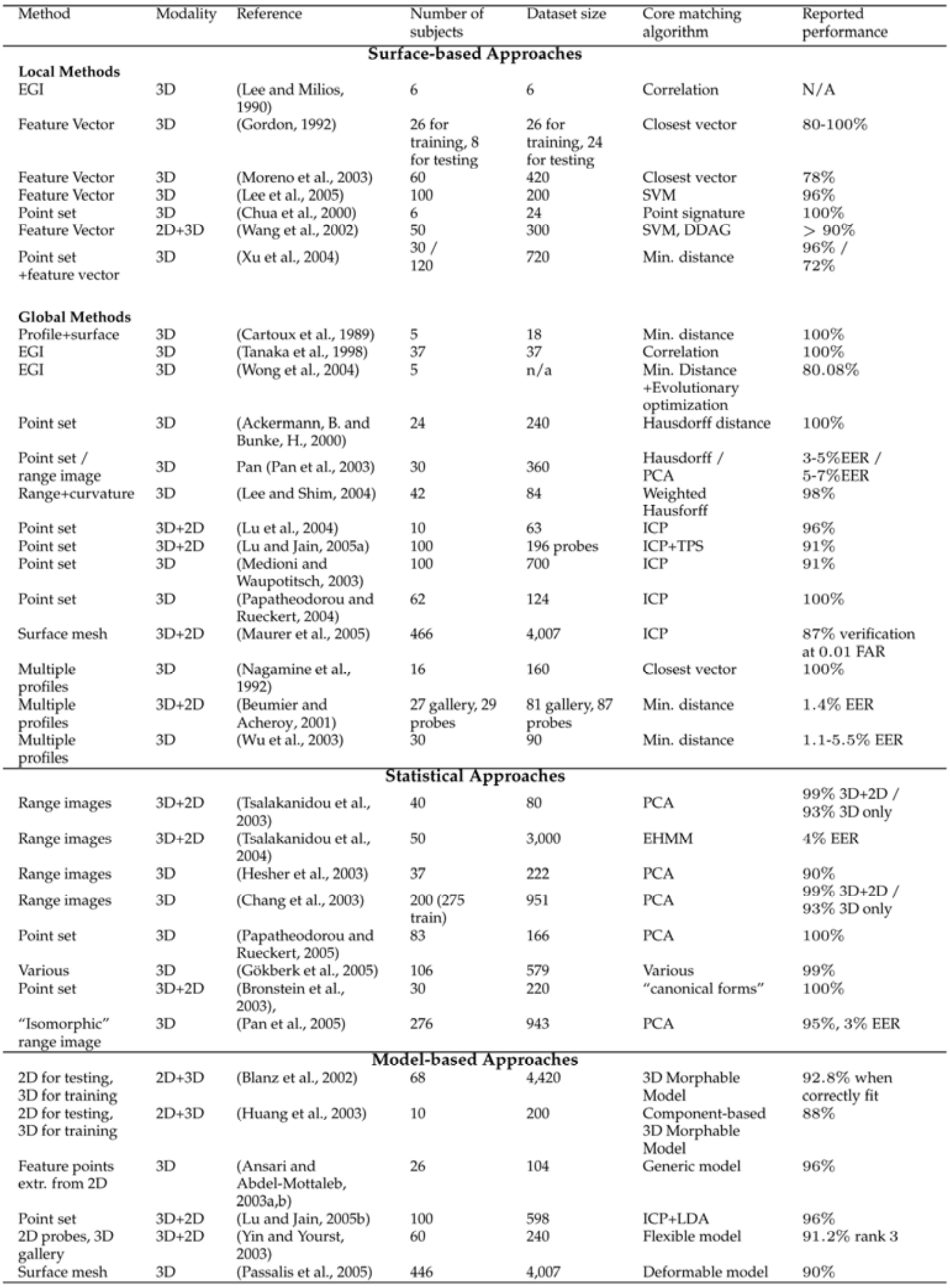
All of the statistical approaches discussed so far do not deal with the effects of facial expressions. In order to minimize these effects, several face representations have been developed which are invariant to isometric deformations, i.e. deformations which do not change the geodesic distance between points on the facial surface. One such approach is based on flattening the face onto a plane to form a canonical image which can be used for face recognition (Bronstein et al., 2003, 2005). These techniques rely on *multi-dimensional scaling* (MDS) to flatten complex surfaces onto a plane (Schwartz et al., 1989). Such an approach can be combined with techniques such as PCA for face recognition (Pan et al., 2005).

3.3 Model-based approaches

The key idea of model-based techniques for 3D face recognition is based on so-called 3D morphable models. In these approaches the appearance of the model is controlled by the model coefficients. These coefficients describe the 3D shape and surface colours (texture), based on the statistics observed in a training dataset. Since 3D shape and texture are independent of the viewing angle, the representation depends little on the specific imaging conditions (Blanz and Vetter, 1999). Such a model can then be fitted to 2D images and the model coefficients can be used to determine the identity of the person (Blanz et al., 2002). While this approach is fairly insensitive to the viewpoint, it relies on the correct matching of the 3D morphable model to a 2D image that is computationally expensive and sensitive to initialization. To tackle these diffi- culties, component-based morphable models have been proposed (Huang et al., 2003; Heisele et al., 2001).

Instead of using statistical 3D face models it is also possible to use generic 3D face models. These generic 3D face models can then be made subject-specific by deforming the generic face model using feature points extracted from frontal or profile face images (Ansari and Abdel- Mottaleb, 2003a,b). The resulting subject-specific 3D face model is then used for comparison with other 3D face models. A related approach is based on the use of an annotated face model (AFM) (Passalis et al., 2005). This model is based on an average 3D face mesh that is annotated using anatomical landmarks. This model is deformed non-rigidly to a new face, and the required deformation parameters are used as features for face recognition. A similar model has been used in combination with other physiological measurements such as visible spectrum maps (Kakadiaris et al., 2005).

A common problem of 3D face models is caused by the fact that 3D capture systems can only capture parts of the facial surface. This can be addressed by integrating multiple 3D surfaces or depth maps from different viewpoints into a more complete 3D face model which is less sensitive to changes in the viewpoint (Lu and Jain, 2005b). Instead of using 3D capture systems for the acquisition of 3D face data, it is also possible to construct 3D models from multiple frontal and profile views (Yin and Yourst, 2003).



Method	Modality	Reference	Number of subjects	Dataset size	Core matching algorithm	Reported performance
Surface-based Approaches						
Local Methods						
EGI	3D	(Lee and Milios, 1990)	6	6	Correlation	N/A
Feature Vector	3D	(Gordon, 1992)	26 for training, 8 for testing	26 for training, 24 for testing	Closest vector	80-100%
Feature Vector	3D	(Moreno et al., 2003)	60	420	Closest vector	78%
Feature Vector	3D	(Lee et al., 2005)	100	200	SVM	96%
Point set	3D	(Chua et al., 2000)	6	24	Point signature	100%
Feature Vector	2D+3D	(Wang et al., 2002)	50	300	SVM, DDAG	> 90%
Point set +feature vector	3D	(Xu et al., 2004)	30 / 120	720	Min. distance	96% / 72%
Global Methods						
Profile+surface	3D	(Cartoux et al., 1989)	5	18	Min. distance	100%
EGI	3D	(Tanaka et al., 1998)	37	37	Correlation	100%
EGI	3D	(Wong et al., 2004)	5	n/a	Min. Distance +Evolutionary optimization	80.08%
Point set	3D	(Ackermann, B. and Bunke, H., 2000)	24	240	Hausdorff distance	100%
Point set / range image	3D	Pan (Pan et al., 2003)	30	360	Hausdorff / PCA	3-5%EER / 5-7%EER
Range+curvature	3D	(Lee and Shim, 2004)	42	84	Weighted Hausforff	98%
Point set	3D+2D	(Lu et al., 2004)	10	63	ICP	96%
Point set	3D+2D	(Lu and Jain, 2005a)	100	196 probes	ICP+TPS	91%
Point set	3D	(Medioni and Waupotitsch, 2003)	100	700	ICP	91%
Point set	3D	(Papatheodorou and Rueckert, 2004)	62	124	ICP	100%
Surface mesh	3D+2D	(Maurer et al., 2005)	466	4,007	ICP	87% verification at 0.01 FAR
Multiple profiles	3D	(Nagamine et al., 1992)	16	160	Closest vector	100%
Multiple profiles	3D+2D	(Beumier and Acheroy, 2001)	27 gallery, 29 probes	81 gallery, 87 probes	Min. distance	1.4% EER
Multiple profiles	3D	(Wu et al., 2003)	30	90	Min. distance	1.1-5.5% EER
Statistical Approaches						
Range images	3D+2D	(Tsalakanidou et al., 2003)	40	80	PCA	99% 3D+2D / 93% 3D only
Range images	3D+2D	(Tsalakanidou et al., 2004)	50	3,000	EHMM	4% EER
Range images	3D	(Hesher et al., 2003)	37	222	PCA	90%
Range images	3D	(Chang et al., 2003)	200 (275 train)	951	PCA	99% 3D+2D / 93% 3D only
Point set	3D	(Papatheodorou and Rueckert, 2005)	83	166	PCA	100%
Various	3D	(Gökberk et al., 2005)	106	579	Various	99%
Point set	3D+2D	(Bronstein et al., 2003),	30	220	“canonical forms”	100%
“Isomorphic” range image	3D	(Pan et al., 2005)	276	943	PCA	95%, 3% EER
Model-based Approaches						
2D for testing, 3D for training	2D+3D	(Blanz et al., 2002)	68	4,420	3D Morphable Model	92.8% when correctly fit
2D for testing, 3D for training	2D+3D	(Huang et al., 2003)	10	200	Component-based 3D Morphable Model	88%
Feature points extr. from 2D	3D	(Ansari and Abdel-Mottaleb, 2003a,b)	26	104	Generic model	96%
Point set	3D+2D	(Lu and Jain, 2005b)	100	598	ICP+LDA	96%
2D probes, 3D gallery	3D+2D	(Yin and Yourst, 2003)	60	240	Flexible model	91.2% rank 3
Surface mesh	3D	(Passalis et al., 2005)	446	4,007	Deformable model	90%

Table 1. Overview Of Techniques

3.4 Summary

The comparison of different 3D face recognition techniques is very challenging for a number of reasons: Firstly, there are very few standardized 3D face databases which are used for benchmarking purposes. Thus, the size and type of 3D face datasets varies significantly across different publications. Secondly, there are differences in the experimental setup and in the metrics which are used to evaluate the performance of face recognition techniques. Table 3.4 gives an overview of the different methods discussed in the previous section, in terms of the data and algorithms used and the reported recognition performance.

Even though 3D face recognition is still a new and emerging area, there is a need to compare the strength of each technique in a controlled setting where they would be subjected to the same evaluation protocol on a large dataset. This need for objective evaluation prompted the design of the FRVT 2000 and FRVT 2002 evaluation studies aswell as the upcoming FRVT 2006 (http://www.frvt.org/). Both studies follow the principles of biometric evaluation laid down in the FERET evaluation strategy (Phillips et al., 2000). So far, these evaluation studies are limited to 2D face recognition techniques but will hopefully include 3D face recognition techniques in the near future.

4. 3D Face matching

As discussed before, statistical models of 3D faces have shown promising results in face recognition (Mavridis et al., 2001; Hesher et al., 2003; Chang et al., 2003; Papatheodorou and Rueckert, 2005) and also outside face recognition (Blanz and Vetter, 1999; Hutton, 2004). The basic premise of statistical face models is that given the structural regularity of the faces, one can exploit the redundancy in order to describe a face with fewer parameters. To exploit this redundancy, dimensionality reduction techniques such as PCA can be used. For 2D face images the dimensionality of the face space depends on the number of pixels in the input images (Cootes et al., 1998; Turk and Pentland, 1991). For 3D face images it depends on the number of points on the surface or on the resolution of the range images. Let us assume a set of 3D faces Γ_1, Γ_2, Γ_3,..., Γ_M can be described as surfaces with n surface points each. The average 3D face surface is then calculated by:

$$\overline{\Gamma} = \frac{1}{M} \sum_{i=1}^{M} \Gamma_i \tag{1}$$

and using the vector difference

$$\boldsymbol{\gamma}_i = \Gamma_i - \overline{\Gamma} \tag{2}$$

the covariance matrix ***C*** is computed by:

$$\boldsymbol{C} = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T \tag{3}$$

An eigenanalysis of ***C*** yields the eigenvectors ui and their associated eigenvalues λ_i sorted by decreasing eigenvalue. All surfaces are then projected on the facespace by:

$$\boldsymbol{\beta}_k = \boldsymbol{u}_k^T (\Gamma - \overline{\Gamma}) \tag{4}$$

where k = 1, ...,m. In analogy to active shape models in 2D (Cootes et al., 1995), every 3D surface can then be described by a vector of weights $\beta^T = [\beta_1, \beta_2, ..., \beta_m]$, which dictates how much each of the principal eigenfaces contributes to describing the input surface. The value of m is application and data-specific, but in general a value is used such that 98% of the population variation can be described. More formally (Cootes et al., 1995):

$$\frac{\sum_{k=1}^{m} \boldsymbol{\lambda}_k}{\sum_{j=1}^{M} \boldsymbol{\lambda}_j} \geq 0.98 \tag{5}$$

The similarity between two faces A and B can be assessed by comparing the weights β_A and β_B which are required to parameterize the faces. We will use two measurements for measuring the distance between the shape parameters of the two faces. The first one is the Euclidean distance which is defined as:

$$d_E(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B) = ||\boldsymbol{\beta}_A - \boldsymbol{\beta}_B|| = \sqrt{\sum_{i}^{m} (\beta_{A_i} - \beta_{B_i})^2} \tag{6}$$

In addition it is also possible calculated the distance of a face fromthe feature-space (Turk and Pentland, 1991). This effectively calculates how “face”-like the face is. Based on this, there are four distinct possibilities: (1) the face is near the feature-space and near a face class (the face is known), (2) the face is near the feature-space but not near a face class (face is unknown), (3) the face is distant from the feature-space and face class (image not a face) and finally (4) the face distant is from feature-space and near a face class (image not a face). This way images that are not faces can be detected. Typically case (3) leads to false positives in most recognition systems.

By computing the sample variance along each dimension one can use the Mahalanobis distance to calculate the similarity between faces (Yambor et al., 2000). In the Mahalanobis space, the variance along each dimension is normalized to one. In order to compare the shape parameters of two facial surfaces, the difference in shape parameters is divided by the corresponding standard deviation σ:

$$d_M(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B) = \sqrt{\frac{\sum_i^m (\beta_{A_i} - \beta_{B_i})^2}{\sigma_i^2}} \tag{7}$$

5. Construction of 3D statistical face models using registration

A fundamental problem when building statistical models is the fact that they require the determination of point correspondences between the different shapes. The manual identification of such correspondences is a time consuming and tedious task. This is particularly true in 3D where the amount of landmarks required to describe the shape accurately increases dramatically compared to 2D applications.

5.1 The correspondence problem

The key challenge of the correspondence problem is to find points on the facial surface that correspond, anatomically speaking, to the same surface points on other faces (Beymer and Poggio, 1996). It is interesting to note that early statistical approaches for describing faces did not address the correspondence problem explicitly (Turk and Pentland, 1991; Kirby and Sirovich, 1990).

Anatomical points landmarked	
Points	**Landmark Description**
Glabella	Area in the center of the forehead between the eyebrows, above the nose which is slightly protruding (1 landmark).
Eyes	Both the inner and outer corners of the eyelids are landmarked (4 landmarks).
Nasion	The intersection of the frontal and two nasal bones of the human skull where there is a clearly depressed area directly between the eyes above the bridge of the nose (1 landmark).
Nose tip	The most protruding part of the nose (1 landmark).
Subnasal	The middle point at the base of the nose (1 landmark).
Lips	Both left and right corners of the lips aswell as the top point of the upper lip and the lowest point of the lower lip (4 landmarks).
Gnathion	The lowest and most protruding point on the chin (1 landmark).

Table 2. The 13 manually selected landmarks chosen because of their anatomical distinctiveness

The gold standard to establish correspondence is by using manually placed landmarks to mark anatomically distinct points on a surface. As this can be a painstaking and error-prone process, several authors have proposed to automate this by using a template with annotated landmarks. This template can be then registered to other shapes and the landmarks can be propagated to these other shapes (Frangi et al., 2002; Rueckert et al., 2003). Similarly, techniques such as optical flow can be used for registration. For example, correspondences between 3D facial surfaces can be estimated by using optical flow on 2D textures to match anatomical features to each other Blanz and Vetter (1999). Some work has been done on combining registration techniqueswith a semi-automatic statistical technique, such as active shape models, in order to take advantage of the strengths of each (Hutton, 2004).

Yet another approach defines an objective function based on minimum description length (MDL) and thus treats the problem of correspondence estimation as an optimization problem (Davies, 2002). Another way of establishing correspondence between points on two surfaces is by analyzing their shape. For example, curvature information can be used to find similar areas on a surface in order to construct 3D shape models (Wang et al., 2000). Alternatively, the surfaces can be decimated in such a way that eliminates points from areas of low curvature. High curvature areas can then assumed to correspond to each other and are thus aligned (Brett and Taylor, 1998; Brett et al., 2000).

5.2 Landmark-based registration

One way of achieving correspondences is by using landmarks that are manually placed on 3D features of the face. The landmarks should be placed on anatomically distinct points of the face in order to ensure proper correspondence. However, parts of the face such as the cheeks are difficult to landmark because there are no uniquely distinguishable anatomical points across all faces. It is important to choose landmarks that contain both local feature information (eg. the size of the mouth and nose) as well as the overall size of the face (eg. the location of the eyebrows). Previous work on 3D face modelling for classification has shown that there is not much difference between the use of 11 and 59 landmarks (Hutton, 2004). In our experience 13 landmarks are sufficient to capture the shape and size variations of the face appropriately. Table 2 shows the landmarks that are used in the remainder of the chapter and Figure 1 shows an example of a face that was manually landmarked.

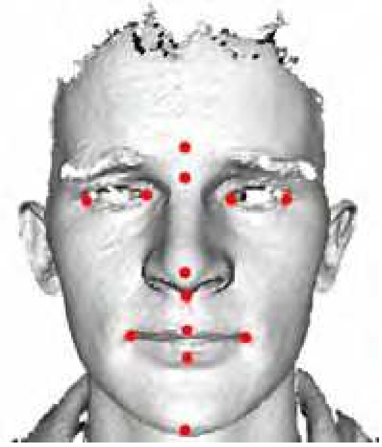


Figure 1. The 13 manually selected landmarks chosen because of their anatomical distinctiveness

5.2.1 Rigid registration

In order to perform rigid registration one face is chosen as a template face and all other faces are registered to this template face. Registration is achieved by minimizing the distance between corresponding landmarks in each face and the template face using the least square approach (Arun et al., 1987). Subsequently, a new landmark set is computed as the mean of all corresponding landmarks after rigid alignment. The registration process is then repeated using the mean landmark set as a template until the mean landmark set does not change anymore.

Figure 2 (top row) shows two faces aligned to the mean landmarks while the bottom row shows a frontal 2D projection of the outer landmarks of the same faces before and after rigid landmark registration. After registration it is possible to compute for each point in the template surface the closest surface point in each of the faces. This closest point is then assumed to be the corresponding surface point.

5.2.2 Non-rigid registration

The above rigid registration process assumes that the closest point between two faces after rigid registration establishes the correct anatomical correspondence between two faces. However, due to differences in the facial anatomy and facial expression across subjects this assumption is not valid and can lead to sub-optimal correspondences. To achieve better correspondences a non-rigid registration is required. A popular technique for non-rigid registration of landmarks are the so-called thin plate splines (Bookstein, 1989). Thin-plate splines use radial basis functions which have infinite support and therefore each landmark has a global effect on the entire transformation. Thus, their calculation is computationally inefficient. Nevertheless, thin-plate splines have been widely used in medical imaging as well as for the alignment of 3D faces using landmarks (Hutton, 2004).

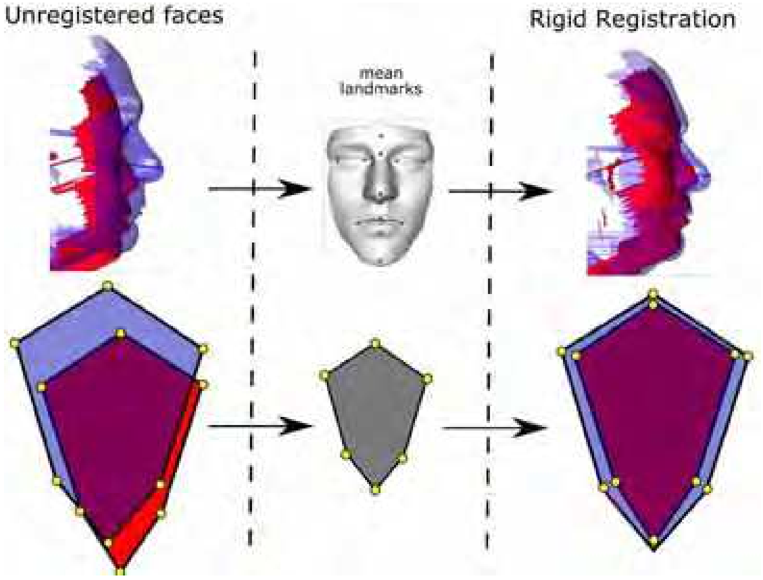


Figure 2. Rigid registration of faces using landmarks. The top rowshows the two faces aligned to the mean landmarks. The bottom row shows a frontal 2D projection of the outer landmarks of the same faces before and after registration

An alternative approach for the non-rigid registration of 3D faces is to use a so-called *free-form deformation* (FFD) (Sederberg and Parry, 1986) which can efficiently model local deformations. B-spline transformations, contrary to thin-plate splines, have local support, which means that each control point influences a limited region. Furthermore, the computational complexity of calculating a B-spline is significantly lower than a thin-plate spline. In the following, a nonrigid registration algorithm for landmarks based on multi-resolution B-splines is proposed.

Lee *et al.* described a fast algorithm for interpolating and approximating scattered data using a coarse-to-fine hierarchy of control lattices in order to generate a sequence of bicubic B-spline function whose sum approximates the desired interpolation function (Lee et al., 1997). We adopt this approach in order to calculate an optimal free-form deformation for two given sets of 3D landmarks. A rectangular grid of control points is initially defined (Figure 3) as a bounding box of all landmarks. The control points of the FFD are deformed in order to precisely align the facial landmarks. Between the facial landmarks the FFD provides a smooth interpolation of the deformation at the landmarks.

The transformation is defined by a $n_x \times n_y \times n_z$ grid Φ of control point vectors ϕ_{lmn} with uniform spacing δ:

$$\boldsymbol{T}(x,y,z) = \sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3} B_i(r)B_j(s)B_k(t)\boldsymbol{\phi}_{l+i,m+j,n+k} \tag{8}$$

where $l = \lfloor \frac{p_x}{\delta} \rfloor - 1, m = \lfloor \frac{p_y}{\delta} \rfloor - 1, n = \lfloor \frac{p_z}{\delta} \rfloor - 1, r = \frac{p_x}{\delta} - \lfloor \frac{p_x}{\delta} \rfloor, s = \frac{p_y}{\delta} - \lfloor \frac{p_y}{\delta} \rfloor$ and $t = \frac{p_z}{\delta} - \lfloor \frac{p_z}{\delta} \rfloor$ and where B_i, B_j , B_k represent the B-spline basis functions which define the contribution of each control point based on its distance from the landmark (Lee et al., 1996, 1997):

$$\begin{aligned} B_0(u) &= (1-u)^3/6 \\ B_1(u) &= (3u^3 - 6u^2 + 4)/6 \\ B_2(u) &= (-3u^3 + 3u^2 + 3u + 1)/6 \\ B_3(u) &= u^3/6 \end{aligned}$$

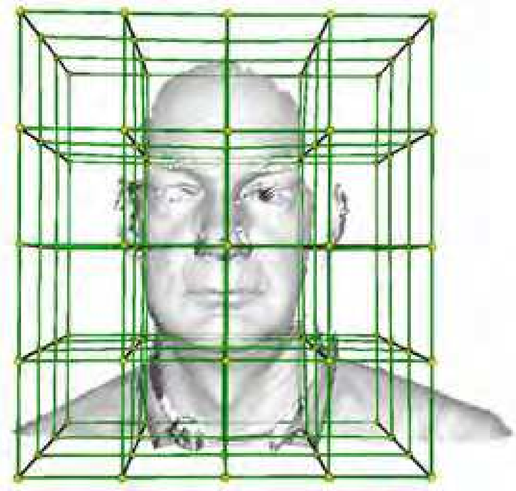


Figure 3. A free-formdeformation and the corresponding mesh of control points

Given a moving point set (source) $\boldsymbol{p} = \{(p_{e_x}, p_{e_y}, p_{e_z})\}$ and a fixed point set $\boldsymbol{q} = \{(q_{e_x}, q_{e_y}, q_{e_z})\}$, the algorithm estimates a set of displacement vectors $\boldsymbol{d} = \boldsymbol{p} - \boldsymbol{q}$ associated with the latter. The output is an array of displacement vectors Φ_{lmn} for the control points which provides a least squares approximation of the displacement vectors.

Since B-splines have local support, each source point pe is affected by the closest 64 control points. The displacement vectors of the control points associated with this source point can be denoted as Φ_{ijk}:

$$\phi_{ijk} = \frac{w_{ijk}\boldsymbol{d}}{\sum_{a=0}^{3}\sum_{b=0}^{3}\sum_{c=0}^{3} w_{abc}^2} \tag{9}$$

where $w_{ijk} = B_i(r)\ B_j(s)\ B_k(t)$ and $i, j, k = 0, 1, 2, 3$. Because of the locality of B-splines, the spacing of control points has a significant impact on the quality of the least squares approximation and the smoothness of the deformation: Large control point spacings lead to poor approximations and high smoothness whereas small control point spacings lead to good approximations but less smoothness. To avoid these problems, a multilevel version of the B-spline approximation is used (Lee et al., 1997). In this approach an initial coarse grid is used initially and then iteratively subdivided to enable closer and closer approximation between two point sets. Before every subdivision of the grid the current transformation $\boldsymbol{T}$ is applied to points $\boldsymbol{p}$ and the displacement vectors $\boldsymbol{d}$ are recomputed.

5.3 Surface-based registration

A drawback of the registration techniques discussed in the previous section is the need for landmarks. The identification of landmarks is a tedious and time-consuming step which typically requires a human observer. This introduces inter- and intra-observer variability into the landmark identification process. In this section we will focus on surface-based registration techniques which do not require landmarks.

5.3.1 Rigid registration

The most popular approach for surface registration is based on the *iterative closest point* (ICP) algorithm (Besl and McKay, 1992): Given two facial surfaces, i.e. a moving face $A = \{a_i\}$ and a fixed (template) face $B = \{b_i\}$, the goal is to estimate the optimal rotation $\boldsymbol{R}$ and translation t that best aligns the faces. The function to be minimized is the mean square difference function between the corresponding points on the two faces:

$$f(\boldsymbol{T}_{rigid}) = \frac{1}{|A|}\sum_{i=1}^{|A|} ||\boldsymbol{b}_i - \boldsymbol{R}\boldsymbol{a}_i - \boldsymbol{t}||^2. \tag{10}$$

where pointswith the same index correspond to each other. The correspondence is established by looping over each point a on face A and finding the closest point, in Euclidean space, on face B:

$$d(\boldsymbol{a}, B) = \min_{\boldsymbol{b}\in B} ||\boldsymbol{b} - \boldsymbol{a}|| \tag{11}$$

This process is repeated until the optimal transformation is found. As before it is possible after this registration to compute for each point in the template surface the closest surface point in each of the faces. This closest point is then assumed to be the corresponding surface point.

5.3.2 Non-rigid registration

As before, rigid surface registration can only correct for difference in pose but not for differences across the facial anatomy and expression of different subjects. Thus, the correspondences obtained fromrigid surface registration are sub-optimal. This is especially pronounced in areas of high curvature where the faces might differ significantly, such as around the lips or nose. As a result the correspondence established between surface points tends to be incorrect. In this section we propose a technique for non-rigid surface registration which aims to improve correspondences between surfaces.

Given surfaces A and B, made up of two point sets $\boldsymbol{a}$ and $\boldsymbol{b}$, the similarity function that we want to minimize is:

$$f(\boldsymbol{T}_{nonrigid}) = \frac{1}{|A|}\sum_{i=1}^{|A|} ||\boldsymbol{b}_i - \boldsymbol{T}_{nonrigid}(\boldsymbol{a}_i)||^2. \tag{12}$$

where $\boldsymbol{T}_{nonrigid}$ is a non-rigid transformation. A convienient model for such a non-rigid transformation is the FFD model described in eq. (8). Once more one can assume that the correspondence between surface points is unknown. In order to pair points on two surfaces to each other, just as with ICP, one can assume that corresponding points will be closer to each other than non-corresponding ones. A distance metric d is defined between an individual source point a and a target shape B:

$$d(\boldsymbol{a}, B) = \min_{\boldsymbol{b} \in B} ||\boldsymbol{b} - \boldsymbol{a}|| \tag{13}$$

Using this distance metric the closest point in B from all points in A is located. Let Y denote the resulting set of closest points and $\boldsymbol{C}$ the closest point operator:

$$Y = \mathcal{C}(A, B) \tag{14}$$

After closest-point correspondence is established, the point-based non-rigid registration algorithm can be used to calculate the optimal non-rigid transformation $\boldsymbol{T}_{nonrigid}$. This is represented here by the operator $\mathcal{M}$. In order for the deformation of the surfaces to be smooth, a multi-resolution approach was adopted, where the control point grid of the transformation is subdivided iteratively to provide increasing levels of accuracy. The non-rigid surface registration algorithm is displayed in Listing 1.

Listing 1 The non-rigid surface registration algorithm.

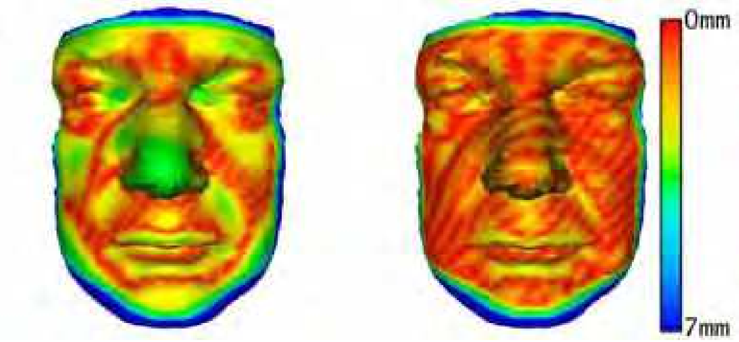
```
Start with surfaces $\boldsymbol{A}$ and a target point set $\boldsymbol{B}$.
Set subdivision counter $k = 0$, $A^{(0)} = A$ and reset $\boldsymbol{T}_{nonrigid}$.
repeat
Find the closest points between $A$ and $B$ by: $Y^{(k)} = \boldsymbol{C}(A^{(k)}, B)$
Compute the ideal non-rigid transformation to align $Y^{(k)}$ and $A^{(0)}$ by:
$\boldsymbol{T}^{(k)}_{nonrigid} = \mathcal{M}(A^{(0)}, Y^{(k)})$ (see section 5.2.2).
Apply the transformation: $A^{(k+1)} = \boldsymbol{T}^{(k)}_{nonrigid}(A^{(0)})$
until $k$ equals user-defined maximum subdivisions limit
```

Figure 4 shows a colour map of the distance between two faces after rigid and non-rigid surface registration. It can be clearly seen that the non-rigid surface registration improves the alignment of the faces when compared to rigid surface registration. Similarly, non-rigid surface registration also better aligns the facial surfaces than non-rigid landmark registration:

Figure 5 (a) shows a color map of the distance between two paces after landmark-based registration. Notice that the areas near the landmarks (eyes, mouth, nose, chin) are much better aligned than other areas. Figure 5 (b) shows a colour map after surface-based registration. In this case the registration has reduced the distances between faces in all areas and provides a better alignment.

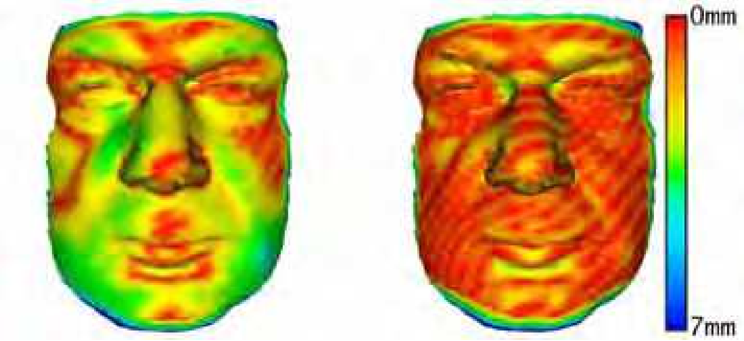
6. Evaluation of 3D statistical face models

To investigate the impact of different registration techniques for correspondence estimation on the quality of the 3D model for face recognition, we have constructed a 3D statistical face model using 150 datasets (University of Notre Dame, 2004). These datasets were acquired using a Minolta VIVID 910 camera which uses a structured light sensor to scan surfaces. A typical face consists of about 20,000 points. Figure 6 shows an example face.



(a) (b)

Figure 4. Two faces after (a) rigid and (b) non-rigid surface registration. The colour scale indicates the distance between the closest surface points



(a) (b)

Figure 5. Two faces after (a) rigid landmark registration and (b) rigid landmark registration followed by non-rigid surface registration. The colour scale indicates the distance between the closest surface points

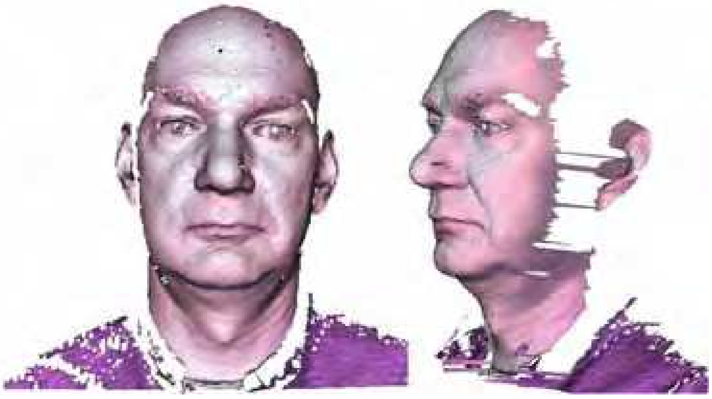


Figure 6. Example of a Notre Dame dataset

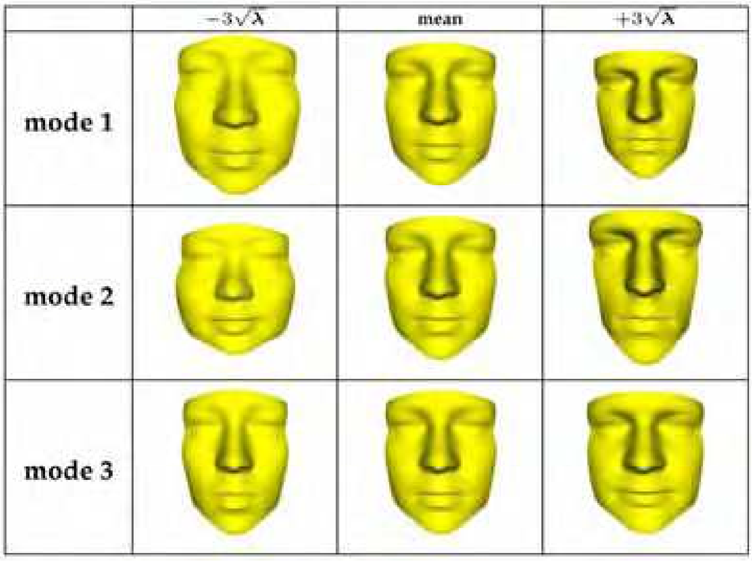


Table 3. The first three principal modes variation of the landmark registration-based model (frontal view)

6.1 Qualitative comparison

A visual comparison of the models generated shows some differences between them. Figure 7 shows two views of the landmark-based mean (left) and the surface-based mean (right). In both cases non-rigid registration has been used. The facial features on the model built using landmark-based registration are much sharper than the features of the model built using surface registration. Given that the features of the surfaces are aligned to each other using non-rigid registration, it is only natural that the resulting mean would be a surface with much more clearly defined features. For example, the lips of every face in the landmark-based model are always aligned to lips and therefore the points representing them would approximately be the same with only their location in space changing. On the other hand the lips in the surface-based model are not always represented by the same points. The upper lip on one face might match with the lower lip on the template face, which results in an average face model with less pronounced features. This is expected, as the faces are aligned using a global transformation and there is no effort made to align individual features together.

Another visual difference between the two models is the fact that facial size is encoded more explicitly in the landmark-based model. The first mode of variation in Table 3 clearly encodes the size of the face. On the other hand the surface-based model in Table 4 does not encode the size of the face explicitly. It is also interesting to observe that the first mode of the surfacebased model, at first sight, seems to encode the facial width. However, on closer inspection in can be seen that the geodesic distance from one side of the face to the other (i.e. left to right) changes very little. Figure 8 shows a schematic representation of a template mesh and a face as seen from the top. The geodesic distance between points x and y in the template mesh is the same as the geodesic distance between points p and q in the subject's