

Chapter 1

3D Face Recognition

Berk Gökberk, Albert Ali Salah, Neşe Alyüz, Lale Akarun

1.1 Introduction

Face is the natural assertion of identity: We show our face as proof of who we are. Due to this widely accepted cultural convention, face is the most widely accepted biometric modality.

Face recognition has been a specialty of human vision: Something humans are so good at that even a days-old baby can track and recognize faces. Computer vision has long strived to imitate the success of human vision and in most cases, has come nowhere near its performance. However, the recent Face Recognition Vendor Test (FRVT06), has shown that automatic algorithms have caught up with the performance of humans in face recognition [73].

How has this increase in performance come about? This can partly be attributed to the advances in 3D face recognition in the last decade. 3D face recognition has important advantages over 2D; it makes use of shape and texture channels simultaneously, where the texture channel carries 2D image information. However, it is registered with the shape channel, and intensity can now be associated with shape attributes such as the surface normal. The shape channel does not suffer from certain problems that the texture suffers from such as poor illumination, or pose changes. Recent research in 3D face recognition has shown that shape carries significant information about identity. At the same time, the shape information makes it easier to eliminate the effects of illumination and pose from the texture. Processed together, the

Berk Gökberk, e-mail: berk.gokberk@philips.com
Philips Research, Eindhoven, The Netherlands

Albert Ali Salah, e-mail: A.A.Salah@cwi.nl
CWI, Amsterdam, The Netherlands

Neşe Alyüz, e-mail: nese.alyuz@boun.edu.tr, Lale Akarun, e-mail: akarun@boun.edu.tr
Boğaziçi University, Computer Engineering Dept. Bebek, TR-34342, Istanbul, Turkey

shape and the texture make it possible to achieve high performances under different illumination and pose conditions.

Although 3D offers additional information that can be exploited to infer the identity of the subject, this is still not a trivial task: *External factors* such as illumination and camera pose have been cited as complicating factors. However, there are *internal factors* as well: Faces are highly deformable objects, changing shape and appearance with speech and expressions. Humans use the mouth and the vocal tract to produce speech; and the whole set of facial muscles to produce facial expressions. Human vision can deal with face recognition under these conditions. Automatic systems are still trying to devise strategies to tackle expressions. A third dimension complicating face recognition is the *time dimension*. Human faces change primarily due to two factors. The first factor is ageing: All humans naturally age. This happens very fast at childhood, somewhat slower once adulthood is reached. The other factor is intentional: Humans try to change the appearance of their faces through hair style, make-up and accessories. Although the intention is usually to enhance the beauty of the individual, the detrimental effects for automatic face recognition are obvious.

This chapter will discuss advances in 3D face recognition together with open challenges and ongoing research to overcome these. In Section 2, we discuss real-world scenarios and acquisition technologies. In Section 3, we overview and compare 3D face recognition algorithms. In Section 4, we outline outstanding challenges and present a case study from our own work; and in chapter 5, we present conclusions and suggestions for research directions. A number of questions touching on the important points of the chapter can be found at the end.

1.2 Technology and Applications

1.2.1 Acquisition Technology

Among biometric alternatives, facial images offer a good trade-off between acceptability and reliability. Even though iris and fingerprint biometrics provide accurate authentication, and are more established as biometric technologies, the acceptability of face as a biometric makes it more convenient. 3D face recognition aims at bolstering the accuracy of the face modality, thereby creating a reliable and non-intrusive biometric.

There exist a wide range of 3D acquisition technologies, with different cost and operation characteristics. The most cost-effective solution is to use several calibrated 2D cameras to acquire images simultaneously, and to reconstruct a 3D surface. This method is called *stereo acquisition*, even though the number of cameras can be more than two. An advantage of these type

of systems is that the acquisition is fast, and the distance to the cameras can be adjusted via calibration settings, but these systems require good and constant illumination conditions.

The reconstruction process for stereo acquisition can be made easier by projecting a structured light pattern on the facial surface during acquisition. The structured light methods can work with a single camera, but require a projection apparatus. This usually entails a larger cost when compared to stereo systems, but a higher scan accuracy. The potential drawbacks of structured light systems is their sensitivity to external lighting conditions and the requirement of a specific acquisition distance for which the system is calibrated. Another problem associated with structured light is that the projected light interferes with the color image, and needs to be turned off to generate it. Some sensors avoid this problem by using near infrared structured light.

Yet a third category of scanners relies on active sensing: A laser beam reflected from the surface indicates the distance, producing a range image. These types of laser sensors, used in combination with a high resolution color camera, give high accuracies, but sensing takes time.

The typical acquisition distance for 3D scanners varies between 50 cm and 150 cm, and laser scanners are usually able to work with longer distances (up to 250 cm) when compared to stereo and structured light systems. Structured light and laser scanners require the subject to be motionless for a short duration (0.8 to 2.5 seconds in the currently available systems), and the effect of motion artifacts can be much more detrimental for 3D in comparison to 2D. Laser scanners are able to provide 20-100 μm accuracy in the acquired points. The presence of strong motion artifacts would make a strong smoothing necessary, which will dispel the benefits of having such a great accuracy. Simultaneous acquisition of a 2D image is an asset, as it enables fusion of 2D and 3D methods to potentially greater accuracy. The amount of collected data affects scan times, but also the time of transfer to the host computer, which can be significant. For instance a Minolta 910 scanner requires 0.3 seconds to scan the target in the fast mode (about 76K points), and about 1 second to transfer it to the computer. Longer scan times also result in motion-related problems, including poor 2D-3D correspondence. Table 1.1 lists properties of some commercial sensors.

Table 1.1 3D scanners used for face recognition.

Scanner	Scanning Technology	Scan Time	Range	Field of View	Accuracy	Databases	Website
3DMD	Stereo structured light	1.5 msec	0.5 m	not specified	0.5mm	BU-3DFE, ASU PRISM	www.3dmd.com
Konica Minolta 300/700/9i/910	Laser scanner	2.5 sec	0.6 to 1.2 m	463x347x500	0.16mm	FRGC, IV^2 , GavabDB	www.konicaminolta-3d.com
Inspeck Mega-Capturor II	Structured light	0.7 sec	1.1 m	435x350x450	0.3mm	Bosphorus	www.inspeck.com
Geometrix Face-vision (ALIVE Tech)	Stereo camera	1 sec	1 m	not specified	not specified	IDENT	www.geometrix.com
Bioscrypt VisionAccess (formerly A4)	Near infrared light	< 1 sec	0.9-1.8 m	not specified	not specified	N/A	www.bioscrypt.com
Cyberware PX	Laser scanner	16 sec	0.5 to 1 m	440x360	0.4mm	N/A	www.cyberware.com
Cyberware 3030	Laser scanner	30 sec	<0.5 m	300x340x300	0.35mm	BJUT-3D	www.cyberware.com
Genex 3D FaceCam	Stereo structured light	0.5 sec	1 m	510x400x300	0.6mm	N/A	www.genextech.com
Breuckmann FaceScan III	Structured light	2 sec	1 m	600x460	0.43mm	N/A	www.breuckmann.com

1.2.2 Application Scenarios

Scenario 1 - Border Control: Since 3D sensing technology is relatively costly, its primary application is the high-security, high-accuracy authentication setting, for instance the control point of an airport. In this scenario, the individual briefly stops in front of the scanner for acquisition. The full face scan can contain between 5.000 to 100.000 3D points, depending on the scanner technology. This data is processed to produce a biometric template, of the desired size for the given application. Template security considerations and the storage of the biometric templates are important issues. Biometric databases tend to grow as they are used; the FBI fingerprint database contains about 55 million templates.

In verification applications, the storage problem is not so vital since templates are stored in the cards such as e-passports. In verification, the biometric is used to verify that the scanned person is the person who supplied the biometric in the e-passport, but extra measures are necessary to ensure that the e-passport is not tampered with. With powerful hardware, it is possible to include a screening application to this setting, where the acquired image is compared to a small set of individuals. However, for a recognition setting where the individual is searched among a large set of templates, biometric templates should be compact.

Another challenge for civil ID applications that assume enrollment of the whole population is the deployment of biometric acquisition facilities, which can be very costly if the sensors are expensive. This cost is even greater if multiple biometrics are to be collected and used in conjunction.

Scenario 2 - Access Control: Another application scenario is the control of a building, or an office, with a manageable size of registered (and authorized) users. Depending on the technology, a few thousand users can be managed, and many commercial systems are scalable in terms of users with appropriate increase in hardware cost. In this scenario, the 3D face technology can be combined with RFID to have the template stored on a card together with the unique RFID tag. Here, the biometric is used to authenticate the card holder given his/her unique tag.

Scenario 3 - Criminal ID: In this scenario, face scans are acquired from registered criminals by a government-sanctioned entity, and suspects are searched in a database or in videos coming from surveillance cameras. This scenario would benefit most from advances in 2D-3D conversion methods. If 2D images can be reliably used to generate 3D models, the gallery can be enhanced with 3D models created from 2D images of criminals, and acquired 2D images from potential criminals can be used to initiate search in the gallery.

Scenario 4 - Identification at a Distance: For scenarios 1 to 3, available commercial systems can be employed. A more challenging scenario is identification at a distance, when the subject is sensed in an arbitrary situation. In this scenario, people can be far away from the camera, unaware of

the sensors. In such cases, challenge stems from these types of un-cooperative users. Assuming that the template of the subject is acquired with a neutral expression, it is straightforward for a person who tries to avoid being detected to change parts of his or her facial surface by a smiling or open-mouthed expression. Similarly, growing a moustache or a beard, or wearing glasses can make the job of identifying the person difficult. A potential solution to this problem is to use only the rigid parts of the face, most notably the nose area, for recognition. However, restricting the input data to such a small area means that a lot of useful information will be lost, and the overall accuracy will decrease. Furthermore, certain facial expressions affect the nose and subsequently cause a drop in the recognition accuracy.

This scenario also includes consumer identification, where a commercial entity identifies a customer for personalized services. Since convenience is of utmost importance in this case, face biometrics are preferable to most alternatives.

Scenario 5 - Access to Consumer Applications: Finally, a host of potential applications are related with appliances and technological tools that can be proofed against theft with the help of biometrics. For this type of scenario, the overall system cost and user convenience are more important than recognition accuracy. Therefore, stereo camera based systems are more suited for these types of applications. Computers, or even cell phones with stereo cameras can be protected with this technology. Automatic identification has additional benefits that can increase the usefulness of such systems. For instance a driver authentication system using 3D facial characteristics may provide customization for multiple users in addition to ensuring security. Once the face acquisition and analysis tools are in place, this system can also be employed for opportunistic purposes, for instance to determine drowsiness of the driver by facial expression analysis.

1.3 3D Face Recognition Technology

A 3D face recognition system usually consists of the following stages: 1) preprocessing of raw 3D facial data, 2) registration of faces, 3) feature extraction, and 4) matching [2, 87, 19, 72, 30] Figure 1.1 illustrates the main components of a typical 3D face recognition system. Prior to these steps, the 3D face should be localized in a given 3D image. However, currently available 3D face acquisition devices have a very limited sensing range and the acquired image usually contains only the facial area. Under such circumstances, recognition systems do not need face detection modules. With the availability of more advanced 3D sensors that have large range of view, we foresee the development of highly accurate face detection systems that use 3D facial shape data together with the 2D texture information. For instance, in [25], a 3D face detector that can localize the upper facial part under occlusions is proposed.

The preprocessing stage usually involves simple but critical operations such as surface smoothing, noise removal, and hole filling. Depending on the type of the 3D sensor, the acquired facial data may contain significant amount of local surface perturbations and/or spikes. If the sensor relies on reflected light for 3D reconstruction, dark facial regions such as eyebrows and eye pupils do not produce 3D data, whereas specular surfaces scatter the light: As a result, these areas may contain holes. In addition, noise and spike removal algorithms also produce holes. These holes should be filled at the preprocessing phase.

After obtaining noise-free facial regions, the most important phases in the 3D face recognition pipeline are the registration and feature extraction phases. Since human faces are similar to each other, accurate registration is vital for extracting discriminative features. Face registration usually starts with acceptable initial conditions. For this purpose, facial landmarks are usually used to pre-align faces. However, facial feature localization is not an easy task under realistic conditions. Here, we survey methods that are proposed for 3D landmarking, registration, feature extraction, and matching.

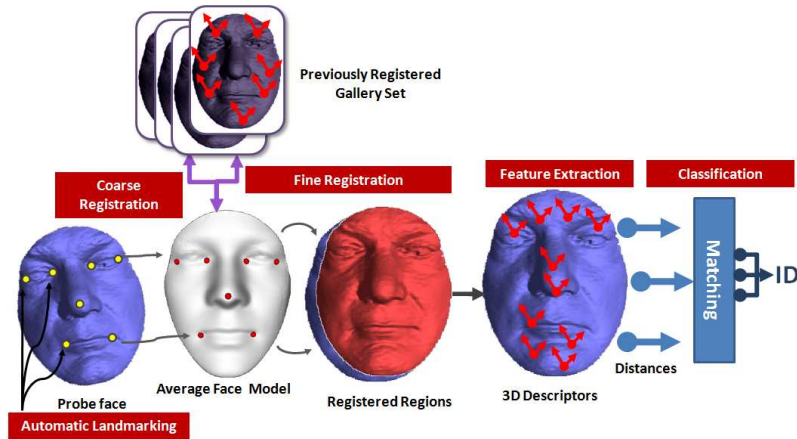


Fig. 1.1 Overall pipeline of a typical 3D face recognition system.

1.3.1 Automatic Landmarking

Robust localization of facial landmarks is an important step in 2D and 3D face recognition. When guided by accurately located landmarks, it is possible to coarsely register facial images, increasing the success of subsequent fine registration.

The most frequently used approach to facial landmark detection is to devise a number of heuristics that seem to work for the experimental conditions at hand [9, 16, 45, 53, 99, 101]. These can be simple rules, such as taking the point closest to the camera as the tip of the nose [24, 102], or using contrast differences to detect eye regions [54, 102]. For a particular dataset, these methods can produce very accurate results. However, for a new setting, these methods are not always applicable. Another typical approach in landmark localization is to detect the easiest landmark first, and to use it in constraining the location of the next landmark [9, 16, 24]. The problem with these methods is that one erroneously located landmark makes the localization of the next landmark more difficult, if not impossible.

The second popular approach avoids error accumulation by jointly optimizing structural relationships between landmark locations and local feature constraints [88, 98]. In [98], local features are modeled with Gabor jets, and a template library (called *the bunch*) is exhaustively searched for the best match at each feature location. A canonic graph serves as a template for the inter-feature distances, and deviations from this template are penalized by increases in internal energy. In [40], an attentive scheme is employed to constrain the detailed search to smaller areas. A feature graph is generated from the feature point candidates, and a simulated annealing scheme is used to find the distortion relative to the canonic graph that results in the best match. A large number of facial landmarks (typically 30-40) are used for these methods and the optimization is difficult as the matching function exhibits many local minima. Most of the landmarks used in this scenario do not have sufficiently discriminating local features associated with them. For instance landmarks along the face boundary produce very similar features.

The third approach is the adaptation of feature-based face detection algorithms to the problem of landmarking [12, 29]. Originally, these methods are aimed at finding a bounding box around the face. Their application to the problem of exact facial landmarking calls for fine-tuning steps.

The problem is no less formidable in 3D, although the prominence of the nose makes it a relatively easy candidate for fast, heuristic-based approaches. If the symmetry axis can be found, it is relatively easy to find the eye and mouth corners [45]. However, the search for the symmetry axis can be costly without the guiding landmarks. Curvature-based features seem to be promising in 3D due to their invariance to several transformations [53, 24]. Especially, Gaussian and mean curvatures are frequently used to locate and segment facial parts. For instance, in [5], multi-scale curvature features are used to localize several salient points such as eye pits and nose. However, curvature-based descriptors suffer from a number of problems. Reliable estimation of curvature requires a strong pre-processing that eliminates surface irregularities, especially near eye and mouth corners. Two problems are associated with this pre-processing: The computational cost is high, and the smoothing destroys local feature information to a great extent, producing many points with similar curvature values in each local neighbourhood. One

issue that makes consistent landmarking difficult is that the anatomical landmarks are defined in structural relations to each other, and the local feature information is sometimes not sufficient to determine them correctly. For flat-nosed persons, the “tip of the nose” is not a point, but a whole area of points with similar curvature. More elaborate 3D methods, like spin images, are very costly in practice [26].

It is possible to use 3D information in conjunction with 2D for landmark localization [16, 24, 6]. Although the illumination sensitivity of the 2D features will have a detrimental effect on the joint model, one can use features that are relatively robust to changes in illumination. In [24], 2D Harris corners are used together with 3D shape indices. In some cases 3D is just used to constrain the 2D search [16, 6]. Under consistent illumination conditions, 2D is richer in discriminative information, but 3D methods are found to be more robust under changing illumination conditions [81].

In [81] and [82] statistical feature models are used to detect each facial feature independently on 3D range images. The advantage of this method is that no heuristics are used to tailor the detection to each landmark separately. A structural analysis subsystem is used between coarse and fine detection. Separating structural analysis and local feature analysis avoids high computational load and local minima issues faced by joint optimization approaches [83]. Fig. 1.2 shows the amount of statistical information available in 2D and 3D face images for independent detection of different landmarks.

1.3.2 Automatic Registration

Registration of facial scans is guided by automatically detected landmarks, and greatly influences the subsequent recognition. 3D face recognition research is dominated by dense registration based methods, which establish point-to-point correspondences between two given faces. For recognition methods based on point cloud representation, this type of registration is the standard procedure, but even range image-based methods benefit from this type of registration.

Registration is potentially the most expensive phase of the 3D face recognition process. We distinguish between *rigid* and *non-rigid* registration, where the former aligns facial scans by an affine transformation, and the latter applies deformations to align facial structures more closely. For any test scan, registration needs to be performed only once for an authentication setting. For the recognition setting, the two extreme approaches are registering the query face to all the faces in the gallery, or to a single average face model (AFM), which automatically establishes correspondence with all the gallery faces, which have been registered with the AFM prior to storage. In between these extremes, a few category-specific AFMs (for instance one AFM

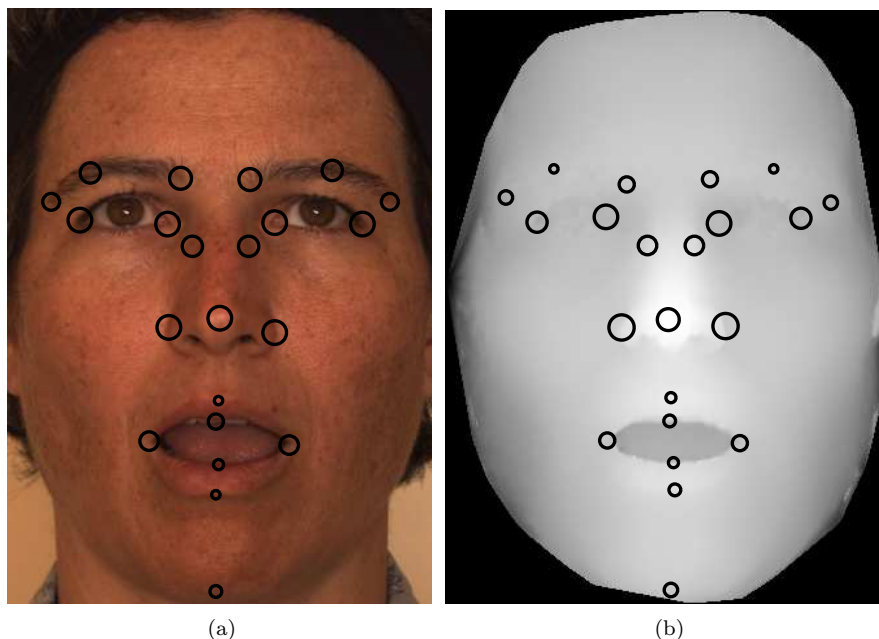


Fig. 1.2 The amount of statistical information available for independent detection of different landmarks in (a) 2D face images and (b) 3D face images, respectively. Marker sizes are proportional to localization accuracy (varies between 32-97 per cent). The 2D images are assumed to be acquired under controlled illumination. The method given in [81] is used on Bosphorus dataset [85].

for males, and one for females) can be beneficial to accuracy and still be computationally feasible [82].

For *rigid registration*, the standard technique is the iterative closest point (ICP) algorithm [11]. For registering shape S_1 to a coarsely aligned shape S_2 , the ICP procedure first finds the closest points in S_2 for all the points on S_1 , and computes the rotation and translation vectors that will minimize the total distance between these corresponding points. The procedure is applied iteratively, until a convergence criterion is met. Practical implementations follow a coarse-to-fine approach, where a subset of points in S_1 are used initially. Once two shapes are put into dense correspondence with ICP, it is straightforward to obtain the total distance of the shapes, as this is the value minimized by ICP. This value can be employed for both authentication and recognition purposes.

Previous work on ICP show that a good initialization is necessary for fast convergence and an accurate end-result. In [82], four approaches for coarse alignment to an AFM are contrasted:

1. Assume that the point with the greatest depth value is the tip of the nose, and find the translation to align it to the nose tip of the AFM. This heuristic is used in [24].
2. Use the manually annotated nose tip.
3. Use seven automatically located landmarks on the face (eye corners, nose tip, mouth corners), and use Procrustes analysis to align them to the AFM. Procrustes analysis finds a least squares alignment between two sets of landmark points, and can also be used to generate a mean shape from multiple sets of landmarks [39].
4. Use seven manually annotated landmarks with Procrustes analysis.

On the FRGC benchmark dataset, it was shown that the nose tip heuristic performed the worst (resulting in 82.85 per cent rank 1 recognition rate), followed by automatically located landmarks with Procrustes alignment (87.86 per cent), manually annotated nose tip (90.60 per cent) and manually annotated landmarks with Procrustes alignment (92.11 per cent) [82]. These results also confirmed that the nose tip is the most important landmark for 3D face registration.

Non-rigid registration techniques have been used for registration as well as for synthesis applications. Blanz and Vetter [15] have used deformable models to register faces to a 3D model which can then be used to synthesize faces with a specific expression and pose; which are subsequently used for recognition. A common feature of deformable algorithms is that they employ a common model to which all faces are registered to: This common face model, conveniently serves as an annotated face model (AFM), and serves to establish dense correspondence and method for annotating all faces. It has been shown that the construction of the AFM is critical for the success of the registration [82]. Many of the techniques for deformable registration employ the thin plate spline algorithm [17] to deform the surface so that a set of landmark points are brought in correspondence [44]. Most non-rigid registration techniques in the literature (such as [45, 44, 57, 91]) are derived from the work of Bookstein on thin-plate spline (TPS) models [18]. This method simulates the bending of a thin metal plate that is fixed by several anchor points. For a set of such points $P_i = (x_i, y_i), i = 1 \dots n$, the TPS interpolation is a vector-valued function $f(x, y) = [f_x(x, y), f_y(x, y)]$ that maps the anchor points to their specified homologues $P'_i = (x'_i, y'_i), i = 1 \dots n$, and specifies a surface which has the least possible bending, as measured by an integral bending norm. The mapping for the anchor points (i.e. specified landmarks on the facial surface) is exact, whereas the rest of the points are smoothly interpolated. This type of registration strongly depends on the number and accuracy of landmarks. If the landmark set is large, all surfaces will eventually resemble the AFM and lose their individuality. To avoid this problem, Mao et al. [57] deform the AFM rather than the individual facial surfaces. Tena et al. [91] optimize this algorithm with the help of facial symmetry and multiresolution analysis. A compromise between individuality and surface conformance is achieved through the minimization of an energy function combining inter-

nal and external forces [49]. The success of the fitting is highly dependent upon the initial pose alignment, the construction of the AFM, and the mesh optimization. Kakadiaris et. al [48] use an anthropomorphically correct AFM and optimize all steps very carefully and obtain very good performance.

1.3.3 Feature Extraction and Matching

The feature extraction technique essentially depends upon the previous step: the registration. As explained in the previous section, most registration approaches register facial surfaces onto a common model (the AFM), which serves to facilitate dense correspondence between facial points. The one-to-all ICP technique fails to do that: Surface pairs are registered to each other rather than to a common model. Therefore, the ICP error, which serves as a measure of how well the surfaces match serves as the feature extraction and matching technique. Many early systems for 3DFR use this convention [19], [60], [8]. Some representation techniques such as the point signatures [23, 95], spin images [96], or histogram based approaches [100] are special in that they do not require prior registration. In the rest of this subsection, we will assume that the surfaces are densely registered and a dense one-to-one mapping exists between facial surfaces.

The point cloud feature, which is simply the set of 3D coordinates of surface points of densely registered faces, is the simplest feature one can use, and the point cloud and point set difference, used directly as a feature, is analogous to the ICP error. PCA has been applied to the point cloud feature by [79, 70].

Geometrical features rely on differences between facial landmark points located on the facial surfaces [77, 52]. Authors have used as few as 19 [71] or as many as 73 [33] landmark points. Riccio and Dugelay [78] use 3D geometrical invariants derived from MPEG4 feature points.

Facial surfaces are often called 2.5D data since there exists only one z value for a given (x,y) pair. Therefore, a unique projection along the z axis, provides a unique depth image, sometimes called a range image, which can then be used to extract features. Common feature extraction techniques are subspace projection techniques such as PCA, LDA, ICA, DFT, DCT, or NMF [31]. Many researchers have applied these standard techniques [22, 42, 38], as well as propose other techniques such as optimal component analysis [105] or Discriminant Common vectors [105]. Other 2D face feature extraction techniques are also applicable [80, 28]. Most of the statistical feature extraction based methods treat faces globally. However, it is sometimes beneficial to perform local analysis, especially under adverse situations. For instance, in [32], authors perform local region-based DCT analysis on the depth images, and construct final biometric templates by concatenating local DCT features.

Similarly, DCT features derived from overlapping local windows placed over the upper facial region is employed in [59]

While depth images rely on a projection to obtain an image from a surface, one can intersect the surface with planes to generate a set of curves [13, 104]. 1D curve representation techniques can then be applied to represent the curves.

Curvature-based surface descriptors are among the most successful 3D surface representations. They have been used for facial surface segmentation [66] as well as representation [90]. Commonly used descriptors are maximum and minimum principal directions [90], and normal maps [3, 4]. Kakadiris et al. [47] have fused Haar and pyramid features of normal maps. Gokberk et al. [38] have used shape indices, principal directions, mean and Gaussian curvatures and have concluded that principal directions perform the best.

The combination of different representations has also attracted widespread interest. Some approaches use feature level fusion: A typical example is given in [69], where shape and texture information is merged at the point cloud level thus producing 4D point features. Wang and Chua [94] select 2D Gabor wavelet features as local descriptors for the texture modality, and use point signatures as local 3D shape descriptors, and use score level fusion using weighted sum rule. Osaimi et al. [7] fuse local and global fields in a histogram.

Score fusion is more commonly used to combine shape and texture information. Tsalakanidou et al. [93, 92] propose a classic approach where shape and texture images are coded using PCA and their scores are fused at the decision level. Malassiotis and Srinatzis [56] use an embedded hidden Markov model-based (EHMM) classifier which produces similarity scores and these scores are fused by a weighted sum rule. Chang et al. [22] use PCA-based matchers for shape (depth image) and texture modalities. The outputs of these matchers are fused by a weighted sum rule. BenAbdelkader and Griffin [10] concatenate depth image pixels with texture image pixels for data level fusion. Linear discriminant analysis (LDA) is then applied to the concatenated feature vectors to extract features.

A two-level sequential combination idea was used in [55] for 2D texture images, where the ICP-based surface matcher eliminates the unlikely classes at the first round, and at the second round, LDA analysis is performed on the texture information to finalize the identification at the second round.

To deal with degradation in the recognition performance due to facial expressions, part-based classification approaches are considered, where the similarity scores from individual classifiers are fused for the final classification.

In [51], the sign of mean and Gaussian curvatures are calculated at each point for a range image, and these values are used to segment a face into convex regions. EGIs corresponding to each region are created and for the regional classification correlation between the EGIs is utilized. In [64], Moreno *et al.* segment the 3D facial surface using mean and Gaussian curvatures and extract various descriptors for each segment. Cook *et al.* [27] use Log-Gabor

Templates (LGT) on range images and divide a range image into 147 regions. Classification is handled by fusing the scores of each individual classifier.

In [21], Chang *et al.* use multiple overlapping regions around the nose area. Individual regional surfaces are registered with ICP and the regional similarity measures are fused with sum, min or product rule. In [34], Faltemier *et al.* extend the use of multiple regions of [21] and utilize seven overlapping nose regions. ICP is used for individual alignment of facial segments. Threshold values determined for regions are utilized in committee voting fusion approach. In [35], the work of Faltemier *et al.* is expanded to utilize 38 regions segmented from the whole facial surface. The regional classifiers based on ICP alignment are fused with the modified borda count method.

In [47], a deformable facial model is used to describe a facial surface. The face model is segmented into regions and after the alignment, 3D geometry images and normal maps are constructed for regions of test images. The regional representations are analyzed with a wavelet transform and individual classifiers are fused with a weighted sum rule. Deformation information can also be used as a face descriptor. Instead of allowing deformations for better registration, that deformation field may uniquely represent a person. Zou *et al.* [107] follows this approach by selecting several prototype faces from the gallery set, and then learns the *warping space* from the training set. A given probe face is then warped to a generic face template where the warping parameters found at this stage are linear combinations of the previously learned warpings.

Deformation invariance can also be accomplished with the use of *geodesic distances* [20, 67]. It has been shown that the geodesic distance between two points over the facial surface does not change significantly when facial surface deforms slightly [20]. In [67], facial surface is represented using geodesic polar parametrization to cope with facial deformations. When a face is represented by geodesic polar coordinates, intrinsic properties are preserved and a deformation invariant representation is obtained. Using this representation, the face is assumed to contain 2D information embedded in 3D space. For recognition, 2D PCA classifiers in color and shape space are fused.

Mian *et al.* [62] extract inflection points around the nose tip and utilize these points for segmenting the face into eye-forehead and nose regions. The regions, that are less affected under expression variations, are separately matched with ICP and the similarity scores are fused at the metric level. In order to handle the time complexity problem of matching a probe face to every gallery face, authors propose a rejection classifier that eliminates unlikely classes prior to region-based ICP matching algorithm. The rejection classifier consists of two matchers: the first one uses spherical histogram of point cloud data for the 3D modality (spherical face representation, SFR) and the second one employs Scale-Invariant Feature Transform-based (SIFT) 2D texture features. By fusing each matchers similarity scores, the rejection classifier is able to eliminate 97% of the gallery faces which drastically speeds up the ICP-based matching complexity at the second phase.

1.3.3.1 Evaluation Campaigns for 3D Face Recognition

We have seen that there are many alternatives at each stage of a 3D face recognition system: The resulting combinations present abundant possibilities. Many 3D face recognition systems have been proposed over the years and performance has gradually increased to rival the performance of 2D face recognition techniques. Table 1.2 lists commonly used 3D face databases together with some statistics such as the number of subjects and the total number of 3D scans present in the databases. In the presence of literally hundreds of alternative systems, independent benchmarks are needed to evaluate alternative algorithms and to assess the viability of 3D face against other biometric modalities such as high-resolution 2D Faces, fingerprints and iris scans. Face Recognition Grand Challenge (FRGC) [76] and Face Recognition Vendor Test 2006 (FRVT'06) [73] are the two important evaluations where the 3D face modality is present.

Table 1.2 List of popular 3D face databases. Subj.: subject count, Samp.: number of samples per subject, Tot.: total number of scans. The UND database is a subset of the FRGC v.2. Pose labels: L: left, R:right, U: Up, and D: Down.

Database	Subject Count	Sample Count	Total Scans	Expressions	Pose
ND2006 [36]	888	1-63	13450	Neutral, happiness, sadness, surprise, disgust, other	-
York [41]	350	15	5250	Happy, angry, eyes closed, eyebrows raised	U,D
FRGC v.2 [74]	466	1-22	4007	Angry, happy, sad, surprised, disgusted, and puffy	-
BU-3DFE [103]	100	25	2500	Happiness, disgust, fear, angry, surprise, sadness (4 levels)	-
CASIA [106]	123	15	1845	Smile, laugh, anger, surprise and closed eyes	-
UND [68]	275	1-8	943	Smile	-
3DRMA [14]	120	6	720	-	L,R,U,D
GavabDB [65]	61	9	549	Smile, frontal laugh, frontal random gesture	L,R,U,D
Bosphorus [86, 85]	81	31-53	3396	34 expressions (28 different action units, 6 emotional expressions: happiness, surprise, fear, sadness, anger, disgust)	-
BJUT-3D [1]	500	1	500	-	-
Extended M2VTS Database [61]	295	4	1180	-	-
MIT-CBCL [97]	10	324	3240	-	Pose variations
ASU [89]	117	5-10	421	Smile, anger, surprise	-

Face Recognition Grand Challenge: FRGC is the first evaluation campaign that focuses expressly on face: 2D Face at different resolutions and illumination conditions and 3D face, alone or in combination with 2D [76, 75]. The FRGC data corpus contains 50,000 images where the 3D part is divided into two sets: *Development* (943 images) and *Evaluation set* (4007 images collected from 466 subjects). The evaluation set is composed of *target* and *query images*. Face images in the target set are to be used for enrollment, whereas face images in the query set represent the test images. Faces were acquired under controlled illumination conditions using a Minolta Vivid 900/910 sensor, a structured light sensor with a range resolution of 640×480 and a registered color image.

FRGC has three sets of 3D verification experiments: shape and texture together (Experiment 3), shape only (Experiment 3s), and texture only (Experiment 3t). The baseline algorithm for the 3D shape+texture experiment uses PCA applied to the shape and texture channels separately, the scores of which are fused to obtain the final scores. At the FAR rate of 0.1%, verification rate of the baseline system is found to be 54%. The best reported performance is 97% at FAR rate of 0.1% [76, 75]. Table 1.3 Summarizes the results of several published papers in the literature for a FAR rate of 0.001% using FRGC v.2 database. The FRGC 3D experiments have shown that the individual performance of the texture channel is better than the shape channel. However, fusing shape and texture channels together always results in better performance. Comparing 2D and 3D, high-resolution 2D images obtain slightly better verification rates than the 3D modality. However, at low resolution and extreme illumination conditions, 3D has a definite advantage.

Table 1.3 Verification rates in % of various algorithms at FAR rate of 0.001% on the FRGC v.2 dataset

System	Neutral vs All		Neutral vs Neutral		Neutral vs Non-neutral	
	3D	3D+2D	3D	3D+2D	3D	3D+2D
Mian et al. [63]	98.5	99.3	99.4	99.7	97.0	98.3
Kakadiaris et al. [46]	95.2	97.3	NA	99.0	NA	95.6
Husken et al. [43]	89.5	97.3	NA	NA	NA	NA
Maurer et al. [58]	86.5	95.8	97.8	99.2	NA	NA
FRGC baseline	45.0	54.0	NA	82.0	40.0	43.0

Face Recognition Vendor Test 2006: The FRVT 2006 is an independent large-scale evaluation campaign that aims to look at performance of high resolution 2D and 3D modalities [73] together with other modalities. The competition was open to academia and companies. The objectives of the FRVT 2006 tests were to compare face recognition performance to that of top-performing modalities. Another objective was to compare the performance to that of face recognition by humans.

Submitted algorithms were tested on sequestered data collected from 330 subjects (3,589 3D scans). The participants of the 3D part were *Cognitec*, *Viisage*, *Tsinghua University*, *Geometrics* and *University of Houston*. The best performers for 3D have a FRR interquartile range of 0.005 to 0.015 at a FAR of 0.001 for the *Viisage* normalization algorithm and a FRR interquartile range of 0.016 to 0.031 at a FAR of 0.001 for the *Viisage* 3D one-to-one algorithm. In FRVT 2006, it has been concluded that 1) 2D, 3D and iris biometrics are all comparable in terms of verification rates, 2) there is a decrease in the error rate by at least an order of magnitude over what was observed in the FRVT 2002. This decrease in error rate was achieved by still and by 3D face recognition algorithms, 3) At low false alarm rates for humans, automatic face recognition algorithms were comparable or better than humans in recognizing faces under different illumination conditions.

1.4 Challenges and A Case Study

1.4.1 Challenges

The scientific work of the last 20 years on 3D face recognition, the large evaluation campaigns organized, and the abundance of products available on the market all suggest that 3D face recognition is becoming available as a viable biometric identification technology. However, there are many technical challenges to be solved for 3D face recognition to be used widely in all application scenarios mentioned in the beginning of the chapter. The limitations can be grouped as follows:

Restrictions due to scanners: The first important restriction is cost: Reliable 3D face recognition still requires a high-cost, high-precision scanner; and that restricts its use to only very limited applications. A second limitation is the acquisition environment: current scanners require the object to stand at a fixed distance away; with controlled pose. Furthermore, most scanners require that the subject be motionless for a short time; since acquisition usually takes some time. As scanners get faster, not only will this requirement be relaxed, but other modes, such as 3D video will become available.

Restrictions due to algorithms: Most studies have been conducted on datasets acquired in controlled environments, with controlled poses and expressions. Some datasets have incorporated illumination variances; and some have incorporated varying expressions. However, there is no database with joint pose, expression, and illumination differences and no studies on robust algorithms to withstand all these variations. There is almost no work on occlusions caused by glasses and hair; and surface irregularities caused by facial hair. In order for ubiquitous 3D face recognition scenarios to work, the recognition system should incorporate:

- 3D face detection in a cluttered environment
- 3D landmark detection and pose correction
- 3D face recognition under varying facial deformations
- 3D face recognition under occlusion

FRVT2006 has shown that significant progress has been made in dealing with external factors such as illumination and pose. The internal factors are now being addressed as well: In recent years, significant research effort has focused on expression-invariant 3D face recognition; and new databases that incorporate expression variations have become available. The time factor and deception attacks are yet to be addressed. Here, we point to the outstanding challenges and go over a case study for expression invariant face recognition.

Challenge 1: How to deal with changes in appearance in time The first factor that comes to mind when time is mentioned is naturally occurring ageing. A few attempts have been made to model ageing [50]. However, intentional or cosmetic attempts to change the appearance pose serious challenges as well: Beards, glasses, hairstyle and make-up all hamper the operation of face recognition algorithms. Assume the following case where we have a subject that grows a beard from time to time. Figure 1.3 shows sample 2D and 3D images of a person with or without beard. The gallery image of the subject can be bearded or not, but these cases are not symmetrical.

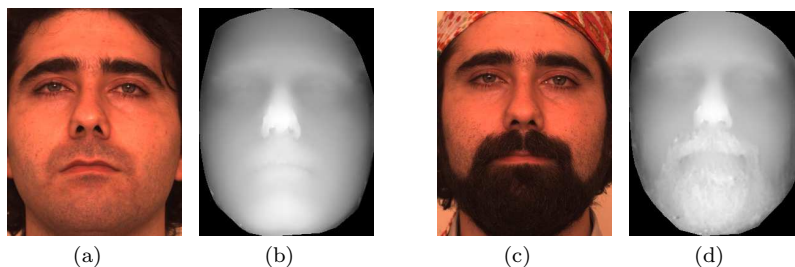


Fig. 1.3 2D and 3D images of a subject with and without beard and corresponding depth images.

Figure 1.4 shows the matching results for an experiment conducted with a 48-image gallery from the Bosphorus database, enhanced with such a subject. In the first case (the first three lines), the query is bearded, and the gallery image is not. Depending on the expression of the query (the first line), the correct image is mostly located in the gallery (the second line). The third line gives rank-1 match for a eye-region based registration and matching, and it is more successful. The second batch of experiments tell a different story. Now the gallery image is bearded, whereas the query (the fourth line) is not. This time, the full-scan registration fails to retrieve the correct gallery image for

the whole range of queries (the fifth line). The reason for this failure is the change in query image-gallery image distances. The total distance between the bearded and non-bearded images of the subject does not change, but it is large when compared to the distance between a non-bearded query image and a non-bearded gallery image belonging to a different subject. Thus, in the first experiment, the query produces large distances to all gallery images, from which the correct one can be retrieved, but in the second experiment, non-bearded false positives dominate because of their generally smaller distances. Subsequently, it is better to have the non-bearded face in the gallery. Alternatively, bearded subjects can be pre-identified and matched with region-based methods. The sixth line of Figure 1.4 shows that the eye-region based matcher correctly identifies the query in most of the cases.

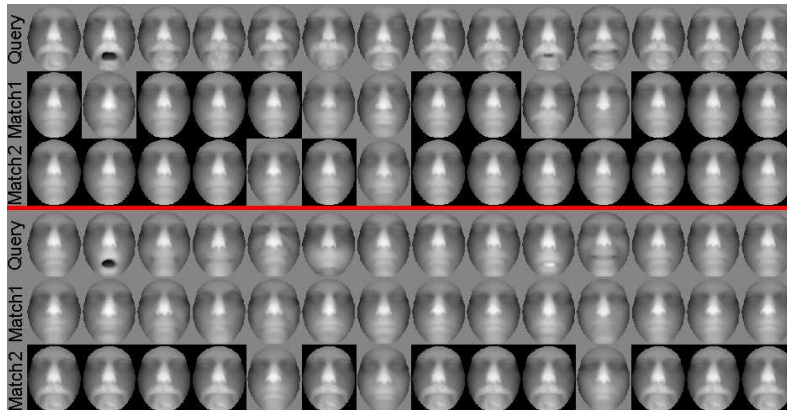


Fig. 1.4 The effect of beard for 3D face recognition. For each experiment, three lines of images are given: the query, rank-1 matched face with a complete facial matching, and rank-1 matched face with a eye-region based matching. The correct matches are shown with black borders. See text for more details.

Challenge 2: How to deal with Internal Factors

The facial surface is highly deformable. It has a single joint, the jaw; which is used to open the mouth; and sets of facial muscles that are used to open and close the eyes and the mouth, and to move the facial surface. The principal objective of mouth movements is speech; and the secondary objective of all facial deformations is to express emotions. Face recognition has largely ignored movement and assumed that the face is still. In recent years, many researchers have focused on expression invariant face recognition.

Here, we present a case study showing how expressions change the facial surface, on a database collected for this purpose, and an example system designed to deal with these variations.

1.4.2 A Case Study:

We have outlined two challenges above: Dealing with internal variations such as facial expressions and dealing with deception attacks, especially occlusions. To develop robust algorithms that can operate under these challenges, one needs to work with a special database that includes both a vast range of expression changes and occlusions. In this section, we will first introduce a database collected for these purposes, and then describe an example part-based system that is designed to deal with variations in the facial surface.

1.4.2.1 Bosphorus DB

The Bosphorus database is a 2D-3D face database including extreme and realistic expression, pose, and occlusion variations that may occur in real life [86, 84]. For facial data acquisition, a structured-light based 3D digitizer device, Inspeck Mega Capturor II 3D, is utilized. During acquisition, vertical straight lines are projected on the facial surface, and the reflections are used for information extraction. For 3D model reconstruction, a region of interest including the central facial region is manually selected, thus the background clutter is removed. The 3D sensor has $0.3mm$, $0.3mm$ and $0.4mm$ sensitivity in x , y , and z respectively, and a typical pre-processed scan consists of approximately 35K points. The texture images are high resolution (1600×1200) with perfect illumination conditions.

After the reconstruction and preprocessing phases, 22 fiducial points have been manually labeled on both 2D and 3D images, as shown in Fig. 1.5.

The Bosphorus database contains a total of 3396 facial scans acquired from 81 subjects, 51 men and 30 women. Majority of the subjects are Caucasian and aged between 25 and 35. The Bosphorus database has two parts: the first part, Bosphorus v.1, contains 34 subjects and each of these subjects has 31 scans: ten types of expressions, 13 different poses, four occlusions, and four neutral/frontal scans. The second part, Bosphorus v.2, has more expression variations. In the Bosphorus v.2, there are 47 subjects, each subject having 34 scans for different expressions including six emotional expressions and 28 facial action units, 13 scans for pose variations, four occlusions and one or two frontal/neutral faces. 30 of these 47 subjects are professional actors/actresses. Fig. 1.5 shows the total scan variations included in the Bosphorus v.2.

1.4.2.2 Example System

The rigid registration approaches are highly affected by facial expression diversities. To deal with deformations caused by expressions, we apply rigid registration in a regional manner. Registering all gallery faces to a common AFM off-line decreases run time cost. Motivated from this approach, we pro-

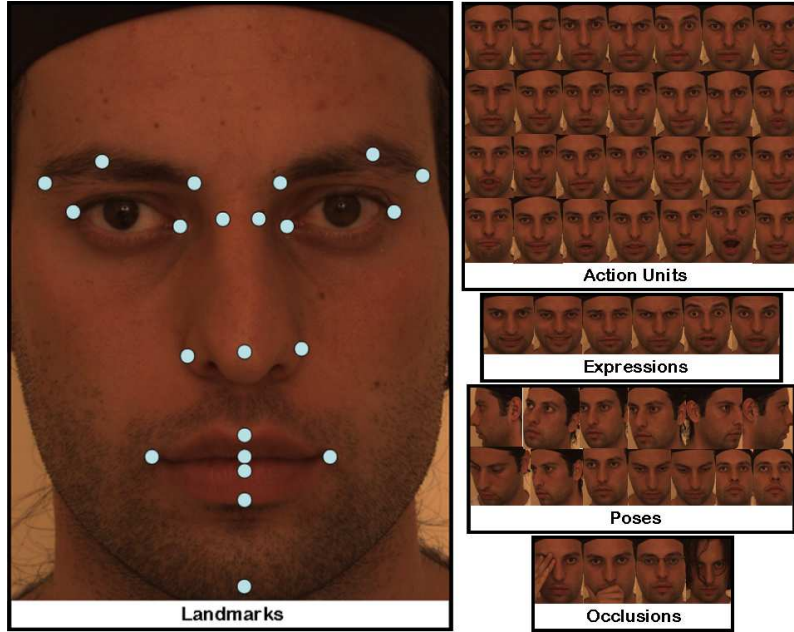


Fig. 1.5 a)Manually located landmark points and scan variations for the Bosphorus database.

posed to use regional models for regional dense registration. The Average Regional Models (ARMs) are constructed by manually segmenting an AFM. These ARMs are used for indexing the regions on gallery and probe faces. After regions are extracted, the facial segments can be used for recognition.

ARM-based Registration: In regional registration, each region is considered separately when aligning two faces. For fast registration we have adapted the AFM-based registration for regions, where regional models act as index files. ARMs are obtained by manually segmenting a whole facial model. The average model is constructed using the gallery set. In this study, we have divided the face into four basic logical regions: forehead-eyes, nose, cheeks, mouth-chin. In Fig. 1.7, the AFM for the Bosphorus v.1 database and the constructed ARMs are given.

In ARM-based registration, a test face is registered individually to each regional model and the related part is labeled and cropped. Registering a test face to the whole gallery consists of four individual alignments, one specific for each region.

Part-based Recognition: After registering test faces to ARMs, the cropped 3D point clouds are regularly re-sampled, hence the point set difference calculation is reduced to a computation between only the depth vectors.

As a result of point set difference calculations, four dissimilarity measure sets are obtained to represent distance between gallery and test faces. Each

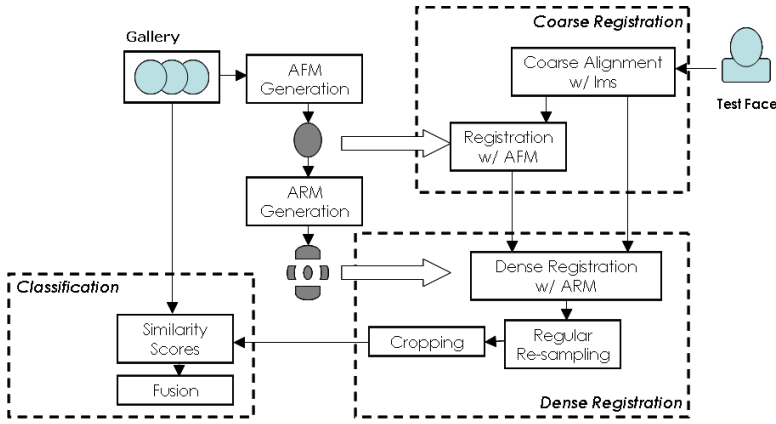


Fig. 1.6 The outline of the proposed system which consists of several steps: facial model construction, dense registration, coarse registration, classification.

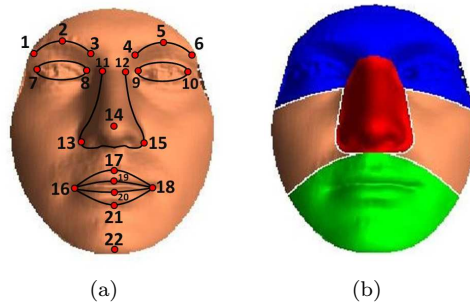


Fig. 1.7 (a) AFM for the Bosphorus v.1 gallery set, (b) four ARMs for forehead-eyes, nose, cheeks and mouth-chin regions.

regional registration is considered as an individual classifier and fusion techniques are applied to combine regional classification results. In this study, we have utilized various fusion approaches from different levels: plurality voting at the abstract-level; sum rule and product rule at the score-level and modified plurality voting at the abstract level [38].

In plurality voting (PLUR), each classifier votes for the nearest gallery identity and the identity with the highest vote is assigned as the final label. When there are ties among closest classes, the final label is assigned randomly among these class labels. In modified plurality voting, each classifier votes for the nearest gallery identity and the identity with the highest vote is assigned as the final label. A value to define the confidence of a classifier is also present and when there are ties, the label of the class with the highest confidence is

chosen as the final decision. More details on confidence-aided fusion methods can be found in [38], [37].

At the score-level, SUM and PRODUCT rules are tested, where similarity scores of individual classifiers are fused using simple arithmetic operations. For these approaches, the scores are normalized with the min-max normalization method prior to fusion.

Experiments: In our experiments, we have utilized both v.1 and v.2 of the Bosphorus database. For each version, we have grouped the first neutral scan of each subject into the gallery. The scans containing expression and AU variations are grouped into the probe set. The number of scans in each gallery and probe set are given in Table 1.4. It is observed that when expressions are present, the baseline AFM-based classifier’s performance drops by about 30%.

Table 1.4 Gallery and probe sets for Bosphorus Db. AFM-based ICP accuracies are also shown for each version.

Bosphorus		Gallery	Probe	AFM results
v.1	neutral scans	34	102	100.00%
	expression scans	-	339	71.39%
v.2	neutral scans	47	-	-
	expression scans	-	1508	67.67%

To deal with expression variations, we have proposed to use a part-based registration approach. The ICP registration is greatly affected by the accuracy of the coarse alignment of surfaces. To analyze the affect, we have proposed two different coarse alignment approaches for the ARM-based registration. The first method is referred to as *the one-pass registration*, where coarse alignment of facial surfaces are handled by Procrustes analysis of 22 manual landmarks. In the second approach, namely *the two-pass registration*, before registering with the regional models, dense alignment with the AFM is obtained.

In Table 1.5, recognition accuracies obtained for ARM-based registration using both coarse alignment approaches are given. As observed in the results, when expression diversity is large, as in v.2, better results are obtained by utilizing the two-pass registration.

As the results in Table 1.5 exhibit, the nose and forehead-eyes regions are less affected by deformations caused by facial expressions and therefore these regional classifiers yield better results. However, different expressions affect different facial regions, and fusing the results of all regions always yields better results than relying on a single region. Table 1.6 shows the results of fusion using different fusion rules on scores obtained by the two-pass registration. It is observed that the best performance is achieved by the product rule. Above

Table 1.5 Comparison of coarse alignment approaches.

ARM	One-pass		Two-pass	
	v.1	v.2	v.1	v.2
forehead-eyes	82.89	82.16	82.89	83.09
nose	85.55	82.23	85.84	83.95
cheeks	53.39	52.12	54.57	51.72
mouth-chin	42.48	34.55	45.72	34.95

95% for both datasets. It is observed that this performance is more than 10% above the performance of the best regional classifier, the nose.

The accuracy of the MOD-PLUR method, which utilizes the classifier confidences, follows the performance of the product rule. The second score-level fusion method we have used, the sum rule, does not perform as good as the product rule or the confidence-aided fusion schemes. The accuracy of the sum rule can be improved by weighting the effect of regional classifiers. For the weighted-sum rule, the weights are calculated from an independent set: We have used v.1 to calculate the weights for the v.2. The optimal weights calculated from the v.1 database are: $w_{nose} = 0.40$, $w_{eye} = 0.30$, $w_{cheek} = 0.10$, and $w_{chin} = 0.20$. Due to the weights chosen, nose and forehead-eyes regions have greater contribution to total recognition performance.

Table 1.6 Recognition rates (%) for fusion techniques.

Fusion Method	v.1	v.2
MOD-PLUR	94.40	94.03
SUM	88.79	91.78
Weighted SUM	93.51	93.50
PROD	95.87	95.29

1.5 Conclusions

3D face recognition has matured to match the performance of 2D face recognition. When used together with 2D, it makes face a very strong biometric: Face as a biometric modality is widely acceptable for the general public, and face recognition technology is able to meet the accuracy demands of a wide range of applications.

While the accuracy of algorithms have met requirements in controlled tests, 3D face recognition systems have yet to be tested under real application scenarios. For certain application scenarios such as airport screening and access control, systems are being tested in the field. The algorithms in these application scenarios will need to be improved to perform robustly under

time changes and uncooperative users. For other application scenarios, such as convenience and consumer applications, the technology is not yet appropriate: The sensors should get faster, cheaper and less intrusive; and algorithms should adapt to the new sensor technologies to yield good performance with coarser and noisier data.

One property of 3D face recognition sets it apart from other biometric modalities: It is inherently a multimodal biometric, comprising texture and shape. Therefore, a lot of research effort has gone into the fusion of 2D and 3D information. There are yet areas to be explored in the interplay of 2D and 3D: How to obtain one from the other; how to match one to the other, how to use one to constrain the other. In the future, with the widespread use of 3D video, the time dimension will open new possibilities for research, and it will be possible to combine 3D face with behavioral biometrics expressed in the time dimension.

1.6 Questions

- What are the advantages of 3D over 2D for face recognition, and vice versa? Would a 2D+3D system overcome the drawbacks of each of these systems, or suffer under all these drawbacks?
- Consider the five scenarios presented in the first section. What are the security vulnerabilities for each of these scenarios? How would you overcome these vulnerabilities?
- Propose a method for a 3D-face based biometric authentication system for banking applications. Which sensor technology is appropriate? How would the biometric templates be defined? Where would they be stored? What would be the processing requirements?
- Discuss the complexity of an airport security system, in terms of memory size and processing load, under different system alternatives.
- If the 3D data acquired from a sensor is noisy, what can be done?
- How many landmark points are needed for a 3D FR system? Where would they be chosen?
- What are the pros and cons of deformable registration vs. rigid registration?
- Propose an analysis-by-synthesis approach for 3D face recognition.
- Is feature extraction possible before registration? If yes, propose a method.
- Suppose a 3DFR system represents shape features by Mean and Gaussian curvatures, and texture features by Gabor features. Which fusion approach is appropriate? Data level or decision level fusion? Discuss and propose a fusion method.
- What would you do to deceive a 3D face recognizer? What would you add to the face recognition system to overcome your deception attacks?

References

1. *The BJUT-3D Large-Scale Chinese Face Database, MISKL-TR-05-FMFR-001*, 2005.
2. D. Riccio G. Sabatino A. F. Abate, M. Nappi. 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, 28:1885–1906, 2007.
3. A. Abate, M. Nappi, S. Ricciardi, and G. Sabatino. Fast 3d face recognition based on normal map. In *IEEE International Conference on Image Processing*, pages 946–949, 2005.
4. A. Abate, M. Nappi, D. Riccio, and G. Sabatino. 3d face recognition using normal sphere and general fourier descriptor. In *ICPR*, 2006.
5. E. Akagunduz and I. Ulusoy. 3d object representation using transform and scale invariant 3d features. *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, 14–21 Oct. 2007.
6. H. Çınar Akakin, A.A. Salah, L. Akarun, and B. Sankur. 2d/3d facial feature extraction. In *Proc. SPIE Conference on Electronic Imaging*, 2006.
7. Bennamoun M. Al-Osaimi F.R. and Mian A. Integration of local and global geometrical cues for 3D face recognition. *Pattern Recognition*, 41(2):1030–1040, 2008.
8. B. B. Amor, M. Ardabilian, and L. Chen. New experiments on icp-based 3D face recognition and authentication. *ICPR 2006*, 2006.
9. S. Arca, P. Campadelli, and R. Lanzarotti. A face recognition system based on automatically determined facial fiducial points. *Pattern Recognition*, 39:432–443, 2006.
10. C. BenAbdelkader and P.A. Griffin. Comparing and combining depth and texture cues for face recognition. *Image and Vision Computing*, 23(3):339–352, 2005.
11. P. Besl and N. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
12. GM Beumer, Q. Tao, AM Bazen, and RNJ Veldhuis. A landmark paper in face recognition. *Proc. 7th Int. Conf. on Automatic Face and Gesture Recognition*, pages 73–78, 2006.
13. C. Beumier and M. Acheroy. Automatic 3D face authentication. *Image and Vision Computing*, 18(4):315–321, 2000.
14. C. Beumier and M. Acheroy. Face verification from 3d and grey level cues. *Pattern Recognition Letters*, 22:1321–1329, 2001.
15. V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.
16. C. Boehnen and T. Russ. A fast multi-modal approach to facial feature detection. In *Proc. 7th IEEE Workshop on Applications of Computer Vision*, pages 135–142, 2005.
17. F. Bookstein. Shape and the information in medical images: A decade of the morphometric synthesis. *Computer Vision and Image Understanding*, pages 99–118, 1997.
18. F. L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11:567–585, 1989.
19. K. Bowyer, Chang K., and P. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition. *Computer Vision and Image Understanding*, 101:1–15, 2006.
20. A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, 5(30), 2005.
21. K. I. Chang, K. W. Bowyer, and P. J. Flynn. Adaptive rigid multi-region selection for handling expression variation in 3D face recognition. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pages 157–164, 2005.

22. K. I. Chang, K. W. Bowyer, and P. J. Flynn. An evaluation of multi-modal 2D+3D face biometrics. *IEEE Trans. on PAMI*, 27(4):619–624, 2005.
23. C.S. Chua, F. Han, and Y.K. Ho. 3d human face recognition using point signature. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 233–238, 2000.
24. D. Colbry, G. Stockman, and A.K. Jain. Detection of anchor points for 3d face verification. In *Proc. IEEE Workshop on Advanced 3D Imaging for Safety and Security*, 2005.
25. Alessandro Colombo, Claudio Cusano, and Raimondo Schettini. 3d face detection using curvature analysis. *Pattern Recogn.*, 39(3):444–455, 2006.
26. C. Conde, A. Serrano, L.J. Rodríguez-Aragón, and E. Cabello. 3d facial normalization with spin images and influence of range data calculation over face verification. In *IEEE Conf. Computer Vision and Pattern Recognition*, 2005.
27. J. Cook, V. Chandran, and C. Fookes. 3D face recognition using log-gabor templates. In *British Machine Vision Conference*, pages 83–92, 2006.
28. J. Cook, V. Chandran, S. Sridharan, and C. Fookes. Gabor filter bank representation for 3D face recognition. *Proceedings of the Digital Imaging Computing: Techniques and Applications (DICTA)*, 2005.
29. D. Cristinacce and TF Cootes. Facial feature detection and tracking with automatic template selection. *Proc. 7th Int. Conf. on Automatic Face and Gesture Recognition*, pages 429–434, 2006.
30. Kresimir Delac and Mislav Grgic. *Face Recognition*. I-Tech Education and Publishing, Vienna, Austria, 2007.
31. HK Ekenel, H. Gao, and R. Stiefelhagen. 3-D Face Recognition Using Local Appearance-Based Models. *Information Forensics and Security, IEEE Transactions on*, 2(3 Part 2):630–636, 2007.
32. H.K. Ekenel, H. Gao, and R. Stiefelhagen. 3-d face recognition using local appearance-based models. *IEEE Transactions on Information Forensics and Security*, 2(3):630–636, 2007.
33. A.H. Eraslan. 3d universal face-identification technology: Knowledge-based composite-photogrammetry. In *Biometrics Consortium*, 2004.
34. T. Faltemier, K. W. Bowyer, and P. J. Flynn. 3D face recognition with region committee voting. In *Proc. 3DPVT*, pages 318–325, 2006.
35. T. Faltemier, K. W. Bowyer, and P. J. Flynn. A region ensemble for 3D face recognition. *IEEE Transactions on Information Forensics and Security*, 3(1):62–73, 2007.
36. Timothy C. Faltemier, Kevin W. Bowyer, and Patrick J. Flynn. Using a multi-instance enrollment representation to improve 3D face recognition. In *Proc. of Biometrics: Theory, Applications, and Systems, (BTAS)*, pages 1–6, 2007.
37. B. Gökberk and L. Akarun. Comparative analysis of decision-level fusion algorithms for 3D face recognition. In *Proc. ICPR*, pages 1018–1021, 2006.
38. B. Gökberk, H. Dutagaci, L. Akarun, and B. Sankur. Representation plurality and decision level fusion for 3D face recognition. *IEEE Trans. on Systems, Man, and Cybernetics*, in press.
39. C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society B*, 53(2):285–339, 1991.
40. R. Herpers and G. Sommer. An attentive processing strategy for the analysis of facial features. *NATO ASI series. Series F: computer and system sciences*, pages 457–468, 1998.
41. Thomas Heseltine, Nick Pears, and Jim Austin. Three-dimensional face recognition using combinations of surface feature map subspace components. *Image and Vision Computing*, 26:382–396, March 2008.
42. C. Heshner, A. Srivastava, and G. Erlebacher. A novel technique for face recognition using range imaging. *Proc. of the Seventh Int. Symposium on Signal Processing and Its Applications*, pages 201–204, 2003.

43. M. Hüskén, M. Brauckmann, S. Gehlen, and C. von der Malsburg. Strategies and benefits of fusion of 2d and 3d face recognition. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005.
44. T. Hutton, B. Buxton, and P. Hammond. Dense surface point distribution models of the human face. In *IEEE Workshop on Mathematical Methods in Biomedic Image Analysis*, pages 153–160, 2001.
45. M.O. İrfanoğlu, B. Gökberk, and L. Akarun. 3d shape-based face recognition using automatically registered facial surfaces. In *Proc. Int. Conf. on Pattern Recognition*, volume 4, pages 183–186, 2004.
46. I. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. 3d face recognition in the presence of facial expressions: an annotated deformable model approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(4):640–649, 2007.
47. I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. *IEEE Trans. on PAMI*, 29(4):640–649, 2004.
48. I.A. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, and N. Murtuza. Multimodal face recognition: combination of geometry with physiological information. In *Proc. Computer Vision and Pattern Recognition Conference*, pages 1022–1029, 2005.
49. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
50. A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):442–455, 2002.
51. J. C. Lee and E. Milios. Matching range images of human faces. In *International Conference on Computer Vision*, pages 722–726, 1990.
52. Y. Lee, H. Song, U. Yang, and H. Shin K. Sohn. Local feature based 3D face recognition. In *International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 2005)*, pages 909–918, 2005.
53. P. Li, B.D. Corner, and S. Paquette. Automatic landmark extraction from three-dimensional head scan data. *Proceedings of SPIE*, 4661:169, 2002.
54. CT Liao, YK Wu, and SH Lai. Locating facial feature points using support vector machines. *Proc. 9th Int. Workshop on Cellular Neural Networks and Their Applications*, pages 296–299, 2005.
55. X. Lu, A. Jain, and D. Colbry. Matching 2.5D face scans to 3D models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):31–43, 2006.
56. S. Malassiotis and M.G. Strintzis. Pose and illumination compensation for 3d face recognition. In *Proc. International Conference on Image Processing*, 2004.
57. Z. Mao, P. Siebert, P. Cockshott, and A. Ayoub. Constructing dense correspondences to analyze 3d facial change. In *International Conference on Pattern Recognition*, pages 144–148, 2004.
58. T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni. Performance of geometrix activeid 3d face recognition engine on the frgc data. In *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments*, 2005.
59. C. McCool, V. Chandran, S. Sridharan, and C. Fookes. 3d face verification using a free-parts approach. *Pattern Recognition Letters (In Press)*, 2008.
60. G. Medioni and R. Waupotitsch. Face recognition and modeling in 3d. In *IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, pages 232–233, 2003.
61. K. Messer, J. Matas, J. Kittler, J. Luetttin, and G. Maitre. Xm2vtsdb: The extended m2vts database. In *Proc. 2nd International Conference on Audio and Video-based Biometric Person Authentication*, 1999.

62. A. S. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE Trans. on PAMI*, 29(11):1927–1943, 2007.
63. Ajmal S. Mian, Mohammed Bennamoun, and Robyn Owens. An efficient multimodal 2d-3d hybrid approach to automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11):1927–1943, 2007.
64. A. B. Moreno, A. Sanchez, J. F. Velez, and F. J. Diaz. Face recognition using 3D surface-extracted descriptors. In *Irish Machine Vision and Image Processing Conference (IMVIP 2003)*, 2003.
65. A.B. Moreno and Á. Sánchez. Gavabdb: A 3d face database. In *Proc. 2nd COST275 Workshop on Biometrics on the Internet*, 2004.
66. A.B. Moreno, A. Sanchez, J. F. Velez, and F. J. Diaz. Face recognition using 3d surface-extracted descriptors. In *Proc. of the Irish Machine Vision and Image Processing Conf.*, page 997, 2003.
67. I. Mpiperis, S. Malassiotis, and MG Strintzis. 3-D Face Recognition With the Geodesic Polar Representation. *Information Forensics and Security, IEEE Transactions on*, 2(3 Part 2):537–547, 2007.
68. University of Notre Dame (UND) Face Database. <http://www.nd.edu/~cvrl/>.
69. T. Papatheodorou and D. Reuckert. Evaluation of automatic 4D face recognition using surface and texture registration. *Sixth International Conference on Automated Face and Gesture Recognition*, pages 321–326, 2004.
70. T. Papatheodorou and D. Rueckert. Evaluation of 3D face recognition using registration and pca. *AVBPA, LNCS*, 3546:997–1009, 2005.
71. T. Papatheodorou and D. Rueckert. Evaluation of 3d face recognition using registration and pca. In *AVBPA05*, page 997, 2005.
72. Dijana Petrovska-Delacrtaç, Grard Chollet, and Bernadette Dorizzi. *Guide to Biometric Reference Systems and Performance Evaluation (in publication)*. Springer-Verlag, London, 2008.
73. P. Jonathon Phillips, W. Todd Scruggs, Alice J. OToole, Patrick J. Flynn, Kevin W. Bowyer, Cathy L. Schott, and Matthew Sharpe. *FRVT 2006 and ICE 2006 Large-Scale Results (NISTIR 7408)*, March 2007.
74. P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, Jin Chang, K. Hoffman, J. Marques, Jaesik Min, and W. Worek. Overview of the face recognition grand challenge. In *Proc. of Computer Vision and Pattern Recognition*, volume 1, pages 947–954, 2005.
75. P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, and W. Worek. Preliminary face recognition grand challenge results. In *Proceedings 7th International Conference on Automatic Face and Gesture Recognition*, pages 15–24, 2006.
76. P.J. Phillips, P.J. Flynn, W.T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W.J. Worek. Overview of the face recognition grand challenge. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 1, pages 947–954, 2005.
77. D. Riccio and J.L. Dugelay. Asymmetric 3d/2d processing: a novel approach for face recognition. In *13th Int. Conf. on Image Analysis and Processing LNCS*, volume 3617, pages 986–993, 2005.
78. Daniel Riccio and Jean-Luc Dugelay. Geometric invariants for 2d/3d face recognition. *Pattern Recogn. Lett.*, 28(14):1907–1914, 2007.
79. T. Russ, C. Boehnen, and T. Peters. 3D face recognition using 3D alignment for pca. *Proc. of the IEEE Computer Vision and Pattern Recognition (CVPR06)*, 2006.
80. T. Russ, M. Koch, and C. Little. A 2D range hausdorff approach for 3D face recognition. *IEEE Workshop on Face Recognition Grand Challenge Experiments*, 2005.
81. A. A. Salah and L. Akarun. 3d facial feature localization for registration. In *Proc. Int. Workshop on Multimedia Content Representation, Classification and Security LNCS*, volume 4105/2006, pages 338–345, 2006.

82. A.A. Salah, N. Alyüz, and L. Akarun. Registration of three-dimensional face scans with average face models. *Journal of Electronic Imaging*, 17:011006, 2008.
83. Albert Ali Salah, Hatice Cinar, Lale Akarun, and Bulent Sankur. Robust facial landmarking for registration. *Annals of Telecommunications*, 62(1-2):1608–1633, 2007.
84. A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun. Bosphorus database for 3d face analysis. In *European Workshop on Biometrics and Identity Management (accepted)*, 2008.
85. Arman Savran, Neşe Alyüz, Hamdi Dibeklioglu, Oya Çeliktutan, Berk Gökberk, Lale Akarun, and Bülent Sankur. Bosphorus database for 3D face analysis. In *Submitted to the First European Workshop on Biometrics and Identity Management Workshop (BioID 2008)*.
86. Arman Savran, Oya Çeliktutan, Aydın Akyol, Jana Trojanova, Hamdi Dibeklioglu, Semih Esenlik, Nesli Bozkurt, Cem Demirkır, Erdem Akagündüz, Kerem Çalıskan, Neşe Alyüz, Bülent Sankur, İlkey Ulusoy, Lale Akarun, and T. Metin Sezgin. 3D face recognition performance under adversarial conditions. In *Proc. eNTERFACE07 Workshop on Multimodal Interfaces*, 2007.
87. A. Scheenstra, A. Ruifrok, and R.C. Veltkamp. A survey of 3D face recognition methods. *Proceedings of the International Conference on Audioand Video-Based Biometric Person Authentication (AVBPA)*, 2005.
88. R. Senaratne and S. Halgamuge. Optimised Landmark Model Matching for Face Recognition. *Proc. 7th Int. Conf. on Automatic Face and Gesture Recognition*, pages 120–125, 2006.
89. Myung Soo-Bae, Anshuman Razdan, and Gerald Farin. Automated 3d face authentication and recognition. In *IEEE International Conference on Advanced Video and Signal based Surveillance*, 2007.
90. H. Tanaka, M. Ikeda, and H. Chiaki. Curvature-based face surface recognition using spherical correlation principal directions for curved object recognition. In *International Conference on Automated Face and Gesture Recognition*, pages 372–377, 1998.
91. J.R. Tena, M. Hamouz, A.Hilton, and J. Illingworth. A validated method for dense non-rigid 3d face registration. In *International Conference on Video and Signal Based Surveillance*, pages 81–81, 2006.
92. F. Tsalakanidou, S. Malassiotis, and M. Strinzis. Integration of 2d and 3d images for enhanced face authentication. In *Proc. AFGR*, pages 266–271, 2004.
93. F. Tsalakanidou, D. Tzovaras, and M. Strinzis. Use of depth and colour eigenfaces for face recognition. *Pattern Recognition Letters*, 24:1427–1435, 2003.
94. Y. Wang and C.-S. Chua. Face recognition from 2d and 3d images using 3d gabor filters. *Image and Vision Computing*, 23(11):1018–1028, 2005.
95. Y. Wang and C.S. Chua. Robust face recognition from 2D and 3D images using structural Hausdorff distance. *Image and Vision Computing*, 24(2):176–185, 2006.
96. Y. Wang, G. Pan, Z.Wu, and S. Han. Sphere-spin-image: A viewpoint-invariant surface representation for 3D face recognition. In *ICCS, LNCS 3037*, pages 427–434, 2004.
97. B. Weyrauch, J. Huang, B. Heisele, and V. Blanz. Component-based face recognition with 3d morphable models. In *Proc. First IEEE Workshop on Face Processing in Video*, 2004.
98. L. Wiskott, J.-M Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.
99. K. Wong, K. Lam, and W. Siu. An efficient algorithm for human face detection and facial feature extraction under different conditions. *Pattern Recognition*, 34:1993–2004, 2001.
100. Z. Wu, Y. Wang, and G. Pan. 3D face recognition using local shape map. In *Proceedings of the Int. Conf. on Image Processing*, pages 2003–2006, 2004.
101. C. Xu, T. Tan, Y. Wang, and L. Quan. Combining local features for robust nose location in 3D facial data. *Pattern Recognition Letters*, 27(13):1487–1494, 2006.

102. Y. Yan and K. Challapali. A system for the automatic extraction of 3-d facial feature points for face model calibration. *Proc. Int. Conf. on Image Processing*, 2:223–226, 2000.
103. Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and M.J. Rosato. A 3D facial expression database for facial behavior research. In *Proc of FGR*, pages 211–216, 2006.
104. Liyan Zhang, Anshuman Razdan, Gerald Farin, John Femiani, Myungsoo Bae, , and Charles Lockwood. 3d face authentication and recognition based on bilateral symmetry analysis. *The Visual Computer*, 22(1):43–55, 2006.
105. C. Zhong, T. Tan, C. Xu, and J. Li. Automatic 3D face recognition using discriminant common vectors. *International Conference on Biometrics, LNCS*, 3832:85–91, 2006.
106. Cheng Zhong, Zhenan Sun, and Tieniu Tan. Robust 3D face recognition using learned visual codebook. In *Proc of CVPR*, pages 1–6, 2007.
107. Le Zou, S. Cheng, Zixiang Xiong, Mi Lu, and K. R. Castleman. 3-d face recognition based on warped example faces. *Information Forensics and Security, IEEE Transactions on*, 2(3):513–528, Sept. 2007.