Robust face recognition based on three dimensional data

Di Huang

ÉCOLE CENTRALE LYON

LIRIS

THESE

pour obtenir le grade de
DOCTEUR DE L'ECOLE CENTRALE DE LYON
Spécialité: Informatique

Robust Face Recognition based on Three Dimensional Data

dans le cadre de l'Ecole Doctorale InfoMaths
présentée et soutenue publiquement par

DI HUANG

Septembre 2011

Directeur de thèse: Prof. Liming CHEN
Co-directeur de thèse: Prof. Yunhong WANG
Co-directeur de thèse: Dr. Mohsen ARDABILIAN

JURY

Prof. Jean-Luc DUGELAY	Eurecom	Rapporteur
Prof. Abdenour HADID	Université Oulu	Rapporteur
Prof. Bill TRIGGS	Laboratoire Jean Kuntzmann	Examinateur
Prof. Christophe GARCIA	INSA de Lyon	Examinateur
Prof. Liming CHEN	Ecole Centrale de Lyon	Directeur de thèse
Prof. Yunhong WANG	Université Beihang	Co-directeur de thèse
Dr. Mohsen ARDABILIAN	Ecole Centrale de Lyon	Co-directeur de thèse

Contents

List of Tables

List of Figures

Abstract

The face is one of the best biometrics for person identification and verification related applications, because it is natural, non-intrusive, and socially well accepted. Unfortunately, all human faces are similar to each other and hence offer low distinctiveness as compared with other biometrics, e.g., fingerprints and irises. Furthermore, when employing facial texture images, intra-class variations due to factors as diverse as illumination and pose changes are usually greater than inter-class ones, making 2D face recognition far from reliable in the real condition. Recently, 3D face data have been extensively investigated by the research community to deal with the unsolved issues in 2D face recognition, i.e., illumination and pose changes.

This Ph.D thesis is dedicated to robust face recognition based on three dimensional data, including only 3D shape based face recognition, textured 3D face recognition as well as asymmetric 3D-2D face recognition.

In only 3D shape-based face recognition, since 3D face data, such as facial pointclouds and facial scans, are theoretically insensitive to lighting variations and generally allow easy pose correction using an ICP-based registration step, the key problem mainly lies in how to represent 3D facial surfaces accurately and achieve matching that is robust to facial expression changes. In this thesis, we design an effective and efficient approach in only 3D shape based face recognition. For facial description, we propose a novel geometric representation based on extended Local Binary Pattern (eLBP) depth maps, and it can comprehensively describe local geometry changes of 3D facial surfaces; while a SIFT-based local matching process further improved by facial component and configuration constraints is proposed to associate keypoints between corresponding facial representations of different facial scans belonging to the same subject. Evaluated on the FRGC v2.0 and Gavab databases, the proposed approach proves its effectiveness. Furthermore, due to the use of local matching, it does not require registration for nearly frontal facial scans and only needs a coarse alignment for the ones with severe pose variations, in contrast to most of the related tasks that are based on a time-consuming fine registration step.

Considering that most of the current 3D imaging systems deliver 3D face models along with their aligned texture counterpart, a major trend in the literature is to adopt both the 3D shape and 2D texture based modalities, arguing that the joint use of both clues can generally provides more accurate and robust performance than utilizing only either of the single modality. Two important factors in this issue are facial representation on both types of data as well as result fusion. In this thesis, we propose a biological vision-based facial representation, named Oriented Gradient Maps (OGMs), which can be applied to both facial range and texture images. The OGMs simulate the response of complex neurons to gradient information within a given neighborhood and have properties of being highly distinctive and robust to affine illumination and geometric transformations. The previously proposed matching process is then adopted to calculate similarity measurements between probe and

gallery faces. Because the biological vision-based facial representation produces an OGM for each quantized orientation of facial range and texture images, we finally use a score level fusion strategy that optimizes weights by a genetic algorithm in a learning process. The experimental results achieved on the FRGC v2.0 and 3DTEC datasets display the effectiveness of the proposed biological vision-based facial description and the optimized weighted sum fusion.

Indeed, (textured) 3D face recognition techniques also have their own downsides, and are currently limited by their high expense in data acquisition and computation. In this thesis, we present a novel framework, asymmetric 3D-2D face recognition, enrolling in textured 3D face models while performing identification only using 2D facial texture images. The motivation is to limit the use of 3D data where they really help to improve face recognition accuracy. The proposed method consists of a new preprocessing pipeline to enhance robustness to illumination and pose changes, an OGM-based facial representation to describe both local shape and texton variations of range and texture faces, as well as a twofold classification step which combines the matching between two facial texture images and the one between a facial range and texture image. The experiments carried out on the FRGC v2.0 database illustrate that the proposed method outperforms 2D intensity image based ones, and achieves comparable results as 3D data based ones do. Furthermore, it avoids the cost and inconvenience of facial data acquisition and computation in 3D based approaches.

Keywords: 2D, 3D and multi-modal face recognition, asymmetric face recognition, facial representation.

Résumé

La reconnaissance faciale est l'une des meilleures modalités biométriques pour des applications liées à l'identification ou l'authentification de personnes. En effet, c'est la modalité utilisée par les humains; elle est non intrusive, et socialement bien acceptée. Malheureusement, les visages humains sont semblables et offrent par conséquent une faible distinctivité par rapport à d'autres modalités biométriques, comme par exemple, les empreintes digitales et l'iris. Par ailleurs, lorsqu'il s'agit d'images de texture de visages, les variations intra-classe, dues à des facteurs aussi divers que les changements des conditions d'éclairage mais aussi de pose, sont généralement supérieures aux variations inter-classe, ce qui rend la reconnaissance faciale 2D peu fiable dans des conditions réelles. Récemment, les représentations 3D de visages ont été largement étudiées par la communauté scientifique pour palier les problèmes non résolus dans la reconnaissance faciale 2D, qui sont notamment causés par les changements d'illumination et de pose.

Cette thèse est consacrée à la reconnaissance faciale robuste utilisant les données faciales 3D, incluant la reconnaissance de visage 3D, la reconnaissance de visage 3D texturé ainsi que la reconnaissance faciale asymétrique 3D-2D.

La reconnaissance faciale 3D, utilisant l'information géométrique 3D représentée sous forme de nuage de points 3D ou d'image de profondeur, est théoriquement non affectée par les changements dans les conditions d'illumination et peut facilement corriger, par l'application d'une approche de recalage rigide comme ICP, les changements de pose. Le principal défi réside dans la représentation, avec précision, de la surface faciale 3D, mais aussi dans le recalage robuste aux changements d'expression faciale. Dans cette thèse, nous concevons une approche efficace et performante pour la reconnaissance de visage 3D. Concernant la description du visage, nous proposons une représentation géométrique basée sur les cartes extended Local Binary Pattern (eLBP), qui décrivent de manière précise les variations de la géométrie locale de la surface faciale 3D; tandis qu'une étape combinant l'appariement local, basé SIFT, aux informations compositionnelles du visage et aux contraintes de configuration permet d'apparier des points caractéristiques, d'un même individu, entre les différentes représentations de son visage. Évaluée sur les bases de données FRGC v2.0 et Gavab DB, l'approche proposée prouve son efficacité. Par ailleurs, contrairement à la plupart des approches nécessitant une étape d'alignement précise et couteuse, notre approche, en raison de l'utilisation de l'appariement local, ne nécessite pas d'enrôlement dans des conditions de pose frontale précise et se contente seulement d'un alignement grossier.

Considérant que la plupart des systèmes actuels d'imagerie 3D permettent la capture simultanée de modèles 3D du visage ainsi que de leur texture, une tendance majeure dans la littérature scientifique est d'adopter à la fois la modalité 3D et celle de texture 2D. On fait valoir que l'utilisation conjointe de ces deux types d'informations aboutit généralement à des résultats plus précis et plus robustes

que ceux obtenus par l'un des deux séparément. Néanmoins, les deux facteurs clés de la réussite sont la représentation bimodale du visage ainsi que la fusion des résultats obtenus selon chaque modalité. Dans cette thèse, nous proposons une représentation bio-inspirée du visage, appelée Cartes de Gradients Orientés (Oriented Gradient Maps: OGMs), qui peut être appliqué à la fois à la modalité 3D et à celle de texture 2D. Les OGMs simulent la réponse des neurones complexes, à l'information de gradient dans un voisinage donné et ont la propriété d'être très distinctifs et robustes aux transformations affines d'illumination et géométriques. Le processus d'appariement proposé précédemment est alors adopté pour calculer les mesures de similarité entre le visage probe et les visages de galerie. Etant donné que la représentation bio-inspirée du visage produit une carte OGM pour chaque orientation quantifiée de l'image de profondeur du visage et aussi celle texture, nous avons finalement utilisé une stratégie de fusion des scores, basée sur l'algorithme génétique, qui optimise les poids associés à travers un processus d'apprentissage. Les résultats expérimentaux obtenus sur les ensembles de données FRGC v2.0 et 3DTEC illustrent l'efficacité de la description bio-inspirée et de la fusion basée sur la somme pondérée optimisée.

Finalement, les techniques de reconnaissance de visage 3D ou visage 3D texturé ont aussi leurs propres inconvénients, et sont actuellement limitées par leur coût élevé d'acquisition de données et de calcul. Aussi, dans cette thèse, nous présentons un nouveau framework, asymétrique, de reconnaissance faciale 3D-2D. Dans celui-ci l'enrôlement se fait avec la modalité 3D texture alors que la reconnaissance nécessite uniquement les images de texture 2D du visage. La motivation est de limiter l'utilisation des données 3D là où elles permettent d'améliorer la précision de la reconnaissance faciale. La méthode proposée consiste en un nouveau pipeline de prétraitement, afin d'améliorer la robustesse aux changements d'illumination et de pose, une représentation de visage basée sur OGM permettant de décrire à la fois les variations locales de forme 3D et de texton des images de profondeur, et de texture faciales, ainsi que d'une double étape de classification qui combine l'appariement entre deux images de texture, et celui entre une image de profondeur et une image de texture. Les expérimentations menées sur la base de données FRGC v2.0 montrent que la méthode proposée surclasse celles utilisant seules la texture, et obtient des résultats comparables à ceux obtenus par les approches basées 3D. Par ailleurs, il permet d'éviter, entre autres inconvénients, le coût d'acquisition et de traitement des données 3D.

Keywords: reconnaissance de visage en 2D, 3D et multi-modale, reconnaissance faciale asymétrique, représentation faciale.

Chapter 1

Introduction

1.1 Face Recognition

In our daily life, many practical applications such as withdrawing money from cash dispensers, security checking at train stations or airports, logging in computer systems etc., require people to provide their passwords or documents (ID cards or passports), for identification. Unfortunately, common passwords are usually difficult to remember; and documents can be easily stolen or counterfeited, which causes plenty of inconvenience and potential safety hazards. Compared with the traditional password or document based identification approaches, biometrics, which determines or verifies the identity of an individual based on his or her physiological specificities (e.g. the fingerprint, iris, face, retina, palm print, hand vein, voice, ear, etc.) and biological behaviours (e.g. signature and gait), also offers good uniqueness and permanence for recognition, and meanwhile it reduces the risk of being lost or faked as well.

Fingerprint or iris recognition techniques grow mature gradually, and can achieve reliable performance. However, they demand much explicit cooperation from users. Specifically, fingerprint requires that the subject cooperate in making physical contact with the sensor surface, while iris imaging currently requires that the subject cooperate to careful position their eye relative to the sensor. Both facts raise problems of how to keep the surface clean and germ-free in a high-throughput application [Bowyer *et al.* 2006]. There hence is significant application driven demand for improved performance in face recognition, since the face is natural, non-intrusive, and socially well accepted. Hietmeyer [Hietmeyer 2000] analyzed and compared six biometric attributes containing the face, finger, hand, voice, eye, as well as signature, and concluded that facial feature scored the highest compatibility in a Machine Readable Travel Documents (MRTD) system based on a few evaluation factors, such as enrollment, renewal, machine requirements, and public perception, as shown in Fig. 1.1.

Face recognition is a primary activity in everyday life, and almost all human beings distinguish from each other through it. In the past decades, machine-based face recognition has received substantial attention in the community of biometrics, pattern recognition and computer vision. This common interest among researchers working in different fields is motivated not only by the fundamental and challenging problems in this domain, but by its numerous potential and practical applications as well. During the development process of face recognition, various databases (e.g. FERET [Phillips *et al.* 2000] and FRGC [Phillips *et al.* 2005]) have been released for special research goals and several milestone algorithms have been proposed, like Principal Component Analysis (PCA) [Turk & Pentland 1991], Linear Discriminant

Figure 1.1: A scenario of using biometric MRTD systems for passport control (left), and a comparison of diverse biometric features based on MRTD compatibility (the left picture from [Li & Jain 2005] and the right one from [Hietmeyer 2000].

Analysis (LDA) [Belhumeur *et al.* 1997], Elastic Bunch Graph Matching (EBGM) [Wiskott *et al.* 1997], Local Binary Patterns (LBP) [Ahonen *et al.* 2006], etc. Their applied ranges have been extended in security, forensic, commercial and entertainment fields, including mugshot identification (e.g. issuing driver licenses, face sketch recognition), access control, video surveillance, human computer interface design as well as content-based image and video dataset management. Many commercial face recognition are already available and competent to meet special requirements and contribute to our society.

The general term "face recognition" refers to two main scenarios, i.e. (i) verification or authentication; (ii) recognition or identification. In either scenario, facial images with known subjects are initially enrolled into the system. This set of persons is named as "gallery", and later facial images of these or other persons are used as "probe" or "query" to match against images in the gallery based on their specific similarity measurements. The verification scenario is an one-to-one matching problem, and it compares a probe face against the one in the gallery set whose identity is being claimed. The claimed identity is taken to be authenticated if the match quality exceeds some threshold. The identification scenario involves closed-set recognition problem as well as open-set recognition problem. Closed-set recognition is a one-to-many matching problem, and it compares a probe face with all the gallery ones to determine its identity. The identification of the probe facial image is assigned by locating the facial image in the gallery with which it has the extreme similarity (highest or lowest value). While open-set recognition is more challenging, since it embeds the verification and identification scenario into the same framework which determines whether the encountered person is known or not and finds out the identity if the person is known. In order to compare with the state of the art, this thesis only discusses verification and closed set face identification. The Cumulative Match Characteristic (CMC) curve and Receiver Operating Characteristic (ROC) curve can be both used to evaluate the performance of face recognition. The former one is designed for identification experiments, and it summarizes the percentage of

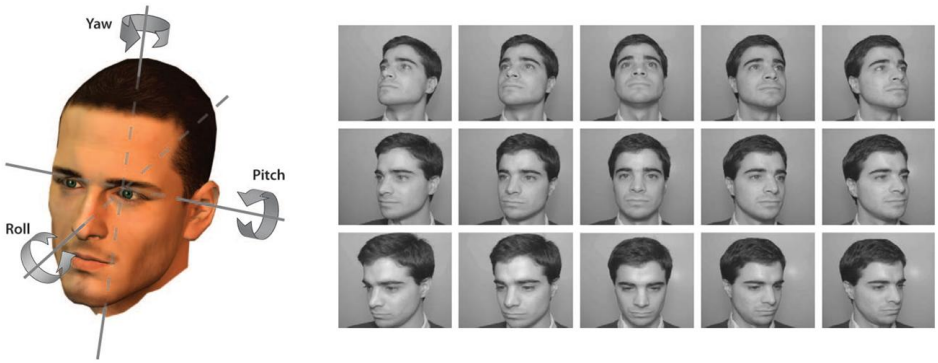


Figure 1.2: The left image: the three DOF of a human head can be described by the egocentric rotation angles pitch, roll, and yaw [Murphy-Chutorian & Trivedi 2009]; the right one: some face examples under pose variations [Beymer 1994].

a set of probes that is considered to be correctly matched as a function of the match rank that is counted as a right match; the rank-one recognition rate is the most commonly stated single number from the CMC curve. The latter one is schemed for verification experiments, and it depicts the True Acceptance Rate (TAR, the percentage of a set probe faces that is correctly accepted) as a trade-off against the False Acceptance Rate (FAR, the percentage that is falsely accepted). The verification rate (TAR) at a certain value of FAR (usually pre-defined at 0.1% in face recognition) is most commonly stated single number from the ROC curve. The academic results can be presented in the context of either recognition or verification, the core representation and matching issues are essentially the same. Actually, the raw matching similarities of the CMC curve for a recognition experiment can be readily tabulated in a different manner to generate the ROC curve fort a verification experiment [Bowyer *et al.* 2006].

1.2 Challenges in 2D Face Recognition

The face possesses obvious advantages over other biometrics, however, it has its own problems as well. All human faces are similar and thereby offer low distinctiveness as compared with other biometrics, e.g., fingerprint and iris [Jain *et al.* 2004]. Moreover, when utilizing 2D facial images, intra-class variations, due to factors as diverse as illumination and pose changes are often greater than inter-class ones. Although a great deal of efforts has been devoted to 2D intensity based face recognition, successful experimental results reported so far still require controlled situations, making 2D face recognition challenging and far from reliable in real conditions [Zhao *et al.* 2003].

Generally, a 2D face recognition system is expected to be robust to the variations of illumination, pose, facial expression, age as well as occlusion.

Pose variations occur in a full 3D orientation and position, while incorporating additional Degrees of Freedom (DOF) including the movement of facial muscles and jaw. Pose variations convey rich and interpersonal information, unfortunately, to

Figure 1.3: Facial appearances of one subject under different lighting conditions, and the face data are taken from CAS-PEAL dataset [Gao *et al.* 2008].

face recognition, different head poses result in distinct changes in facial appearances, which degrades system performance dramatically. Even though several facial features are claimed to be roll-invariant, they fail when pitch or yaw appears. See Fig. 1.2 for an illustration of three DOF and some face examples under pose variations.

Illumination variations have enormously complex influences on the image of a face appearance, because varying illumination directions leads to shifts in the location and shape of shadows, changes in highlights, and reversal of contrast gradients. Figure 1.3 shows an example of one face illuminated by different lighting conditions. To deal with illumination variations, thermal infrared [Socolinsky *et al.* 2003] [Yamaguchi *et al.* 1998] and near infrared (NIR) image [Li *et al.* 2007] based face recognition have been investigated, but the thermal image depends on the temperature of the subject reflected by his or her metabolic state, and it can be regarded as a new unstable factor; while the near infrared image needs to be captured under an active NIR source with an constrained intensity, which indicates that this problem has not been fundamentally solved yet.

Figure 1.4: Facial appearances of one subject with different expressions (The image is collected from Internet).

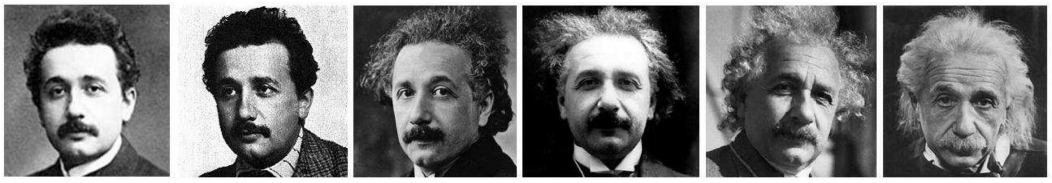


Figure 1.5: Face aging of Albert Einstein [Fu *et al.* 2010].

Facial expression variations are mainly caused by the movement of facial muscles in response to a person's internal emotional state, intentions, or social communications. Facial expression contains much important information, and facial expression recognition has thus been an active research topic, in which exaggerated expressions are generally easy to recognize accurately. In contrast to this, because severe movements of facial muscles influence the face appearance, facial expression is a negative factor to face recognition as well. Some facial images with various expressions are listed in Fig. 1.4.

Face aging is generally a slow and irreversible process. During the early growth and development of the face, from birth to adulthood, the greatest change is the craniofacial growth (shape change). During adult aging, from adulthood to old age, the most perceptible change becomes skin aging (texture change) [Fu *et al.* 2010]. Although this factor only slightly impacts the facial appearance in a short period, the long term influence still makes face recognition problematic. For example, we can easily figure out the aging process on the faces of Albert Einstein, as shown in Fig. 1.5.

Occlusions bring another challenge into this domain, because only partial faces are available as the input for recognition. System accuracies are likely to deteriorate when the percentage of occluded facial part become larger. Common occlusions can be caused by beard, hair, hands, glasses, hats, scarfs, or even cosmetic. Figure 1.6 displays two occlusions due to sunglasses and scarf respectively.

Figure 1.6: Occlusion examples.

1.3 Face Recognition based on 3D Data

The majority of face recognition research has focused on 2D facial images which are photometric appearances in typical photographs. However, several previous studies motivated possible advantages in 3D face recognition over 2D face recognition. For example, Medioni and Waupotitsch stated in [Medioni & Waupotitsch 2003] - "because we are working in 3D, we overcome limitations due to viewpoint and lighting variations"; Hesher et al. argued - "Range images have the advantage of capturing shape variation irrespective of illumination variabilities" [Hesher *et al.* 2003]; Gordon claimed - "Depth and curvature features have several advantages over traditional intensity-based features. Specifically, curvature descriptors: (1) have the potential for higher accuracy in representing surface-based events, (2) are better suited to describe properties of the face in the area such as the cheeks, forehead, and chin, and (3) are viewpoint invariant" [Gordon 1992]; etc. All theses works illustrate the general conclusion that using 3D data is a promising solution to deal with some key challenges in 2D face recognition, and the reasons are as follows:

(a) 3D face data conveys the exact shape information of facial surfaces defined (and can supposedly be acquired) independent of illumination, which is unlike photometric appearances;

(b) 3D face data allows more convenient pose correction than 2D face data;

(c) 3D face data tends to change less with variations in cosmetic use, skin coloration, similar surface reflectance factors than 2D face appearance [Faltemier 2007].

Nevertheless, the further development in 3D face recognition is being impeded by two main factors, i.e. data acquisition and facial expression variations.

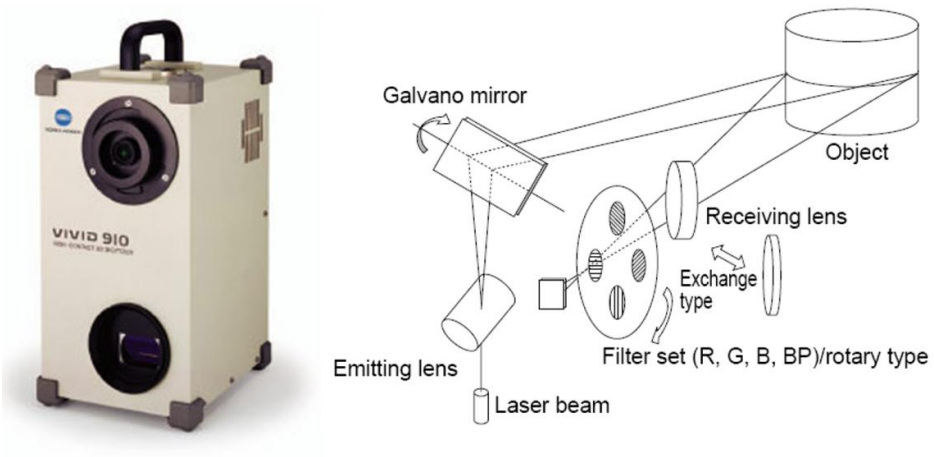


Figure 1.7: Production appearance of Minolta VIVID 910 and its working principle.

Currently, the sensing technologies of 3D imaging systems fall into three basic categories: active, passive, and hybrid [Bowyer *et al.* 2006]. The active type consists of a camera and a light projector with a pre-defined geometric relationship. A light pattern is projected onto the scene, detected in an image acquired by the camera, and the 3D location of points can then be computed. The sensors of VIVID 910

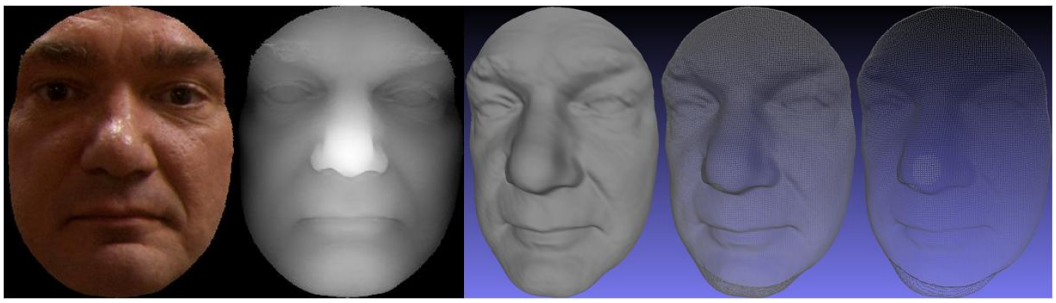


Figure 1.8: Different formats of a 3D face model, and from left to right is texture map, range image, shaded model, wireframe and point-cloud respectively.

serial production of Minolta used in [Lu & Jain 2005a, Chang *et al.* 2003a] would be a straightforward illustration of this type. The passive type used two cameras with a known geometric relationship to image the subject, corresponding points are found in the two images, and the 3D location of the points can be calculated. The Geometrix system [Medioni & Waupotitsch 2003] is a representative of this method. The hybrid type combines the way of the active and passive one. In this technique, a pattern is projected onto the scene and then imaged by a stereo camera rig. The projected pattern simplifies the selection of corresponding points, and can improve their density in the multiple images. The 3Q "Qlonerator" system made by 3DMD exploited this type of sensor. In general, the active sensor can provide a high depth accuracy, but it is limited by the acquisition time; while the passive sensor works at a high speed, but it lacks the ability of producing dense and accurate sampling of 3D points in regions where there is not much texture. The hybrid sensor combines the advantages of the active and passive ones, and is thus considered to be the most promising way. Figure 1.7 depicts the production appearance of Minolta VIVID 910 and its working principle.

In 2006, Bowyer et al. [Bowyer *et al.* 2006] concluded that an ideal 3D sensor for face recognition applications would combine at least the following properties: (i) image acquisition time similar to that of a typical 2D camera,(ii) a large depth of field; e.g, a meter or more in which there is essentially no loss in accuracy of depth resolution, (iii) robust operation under a range of "normal" lighting conditions, (iv) no eye safety issues arising from projected light, (v) dense sampling of depth values; perhaps 1000×1000, and (vi) depth resolution of better than 1 mm. Although the 3D imaging techniques have been greatly improved since then, evaluated by these criteria, existing 3D sensors are still only competent to support advanced research in this area, but cannot meet the requirements of practical application yet.

The 3D face shape is usually sensed in combination with a 2D intensity image, and 3D face can be represented as texture map (the 2D intensity image overlays on a 3D shape); range image (depth image), shaded model, wireframe and point-clound. Figure 1.8 shows an example of different formats of a 3D face model. Except texture map, all the other formats convey the exact shape information of 3D facial surfaces, which can drastically deform under facial expressions in a complex way,

since they are driven by highly complex muscle mechanism, leading to deterioration in accuracy of 3D face recognition. Our previous studies [Ben Soltana *et al.* 2010, Huang *et al.* 2011b] show that facial expression even has a greater effect on 3D face recognition compared with the 2D based one. Nevertheless, expression deformations and the 3D shape of a face are more accurately captured in 3D scans, as a result, this challenge probably can be better addressed in the 3D domain than in 2D domain.

To handle facial expression changes in 3D face recognition, significant effort has been made in the last few years. We can roughly classify them into four streams. The first stream focuses on the feature points or regions within facial areas where the shape changes the least with varying facial expression, and these areas are thus named as "(relatively) rigid ones". For example, the nose region is often emphasized, since it varies relatively little with expressions, while the mouth region is usually ignored, because its shape changes greatly with facial expression. However, up till now, there has been no study to show a large subset of the face that is perfectly shape invariant across all facial expression variations. The second stream concentrates on the improvement of the gallery set by enrolling different expressive 3D face models, aiming to reduce the dissimilarity between a probe and its enrolled samples. The problem of this approach is that the probe may still present an expression which is different from those contained in its face models in the gallery set, and the requirements of a great diversity of enrolled samples is unrealistic as well. The third stream relies on a generic model of 3D facial expressions, which can be applied to any 3D face models. The similarity between a probe and gallery face can be computed according to their parameters controlled by the pre-defined model. However, the same expression probably results in diverse facial shapes of different persons or even the same person at different times and cultural contexts, which indicates there likely is no general model to predict how facial expressions transfer into each other. The last stream tends to search for the facial features that are invariant or at least robust to facial expression changes, but really effective descriptors are seldom discovered. Moreover, this trend always leads to more sophisticated systems with the training process to learn the distribution of facial expression variations and/or the information fusion of different face modalities, i.e. 2D intensity and 3D shape. All the facts demonstrate that more effort should be put into this issue.

Considering both the challenges in the respect of data acquisition and facial expression, it is still premature to conclude that 3D face recognition has solved the key problems in 2D face recognition.

1.4 Landscape of 3D Face Recognition

Human faces are 3D objects containing shape (3D surface) and texture (2D intensity) information, and current 3D imaging system generally deliver 3D face models along with their aligned texture counterparts, as a result, to some extent, 2D intensity image based face recognition can be regarded as a part of 3D face recognition. According to which modality is adopted in the gallery and probe set, 3D face recognition can be carried out in several scenarios: solely 3D shape based face recognition, textured 3D (multi-modal 2D+3D) face recognition, and asymmetric (3D vs. 2D or

Figure 1.9: 3D face recognition scenarios.

vice versa) face recognition. Figure 1.9 shows an illustration of different scenarios in 3D face recognition.

In *only 3D shape based face recognition*, face samples in the gallery and probe sets are range images, wireframes or point-clouds all of which only convey geometry information, and such a recognition scenario is thus theoretically invariant to illumination changes and allows easy pose correction. The core problem lies in how to represent facial surfaces accurately and achieve robust matching to facial expression variations.

In *textured 3D face recognition*, gallery and probe face models consist of both 2D intensity and 3D shape information, and this scenario computes the similarity measurement between 2D textures as well as the one between 3D surfaces, and both scores are then combined for final decision. Since it makes use of both clues and is thus expected to achieve more accurate and robust performance than only using either of the single modality. Textured 3D face recognition generally involves two critical issues, i.e., how to describe faces in both modalities and how to fuse their accuracies.

Recently, *Asymmetric face recognition* appeared as a novel topic in face recognition domain. The word "asymmetric" means that different types of face data are employed in the gallery and probe set respectively. For instance, 3D face models are used for enrollment, and only 2D intensity face images are taken as probes; otherwise, visa versa. Essentially, this scenario mainly concerns the matching between heterogenous face data, i.e. 3D shape vs. 2D intensity or textured 3D shape vs. 2D intensity. Indeed, when the face data in the gallery and probe set are asymmetric, we can reconstruct 3D face shapes using mutiple 2D intensity face images or produce 2D intensity face images by rendering 3D face shapes to achieve homologous matching, but both solutions fall into the above two scenarios again; therefore, we will not detailedly discussed them in this thesis.

1.5 3D Face Databases

Many 3D face datasets have been collected and released so far for public research activities, for example, FRGC v2.0 [Phillips *et al.* 2005], BU-3DFE [Yin *et al.* 2006b],

Table 1.1: A comprehensive introduction of public datasets.

Dataset	Device	Sub. Num.	Num./Sub.	Tex.	Changes
XM2VTS	-	295	1	Yes	-
3D_RMA	-	120	3	No	E, P, O
MPI	Cyberware	200	7	-	-
York	-	350	15	Yes	E, P, R
Norte Dame	Minolta Vivid 910	277	Total: 953	Yes	E, T
GavabDB	Minolta Vivid 700	61	9	No	P, E
BU-3DFE	3DMD	100	4	No	E
Bosohorus	-	105	Total: 4666	Yes	E, P, O
FRGC v2.0	Minolta Vivid 910	466	Total: 4007	Yes	E, O, I
FRAV3D	Minolta Vivid 700	106	16	Yes	E, P
3DTEC	Minolta Vivid 910	214	2	Yes	E, Twins
USF 3D	Cyberware 3030 PS	100	1	Yes	-
BJUT-3D	Cyberware 3030 PS	100	-	Yes	E

Variation labels: (P)Pose; (E)Expression; (O)Occlusion; (I)Illumination; (T)Time.

XM2VTS [Messer *et al.* 2003], 3D_RMA [Beumier & Acheroy 2001], Notre Dame, GavabDB [Moreno & Sanchez 2004], 3DTEC, USF 3D [Blanz & Vetter 1999], York [Heseltine 2005], FRAV3D [Conde *et al.* 2007], Bosphorus [Savran *et al.* 2008], MPI [Troje & Bulthoff 1996], etc. Please see Table 1.1 for a comprehensive introduction of public 3D face databases.

From Table 1.1, we can see that each 3D face dataset was designed for one or more specific research challenges in this domain, including:

- Models of a great number and demographic variety of subjects: FRGC v2.0;
- Models of a given person at repeated intervals of time: FRGC v2.0;
- Models with substantial variations of facial expressions: BU-3DFE, Bosphorus, FRGCv2.0, GavabDB, York, etc;
- Models of twins variations: 3DTEC.

In this thesis, three datasets are used for experimental evaluations and analysis on our proposed approaches, i.e. FRGC v2.0, GavabDB and 3DTEC.

1.5.1 The FRGC dataset

The FRGC database provides a number of experiments; however, since this thesis concentrates on face recognition based on 3D data, we mainly used the protocol of Experiment 3, that is, matching 3D faces (shape and texture) to 3D faces (shape and texture). The face data for Experiment 3 consist of 4950 3D face models along with their aligned texture counterparts acquired with the Minolta Vivid scanner 910 during 2003-2004 academic year. The spatial resolution of the scanner is 480×640, but the resolution of faces changes because they were scanned at different distances from the scanner. The data consists of frontal viewpoints of subjects mostly captured from the shoulder level up. Among all these subjects, 57% are male and 43% are

Figure 1.10: Samples of textured 3D face models in the FRGC dataset.

female, with the following age distribution: 65% 18-22 years old, 18% 23-27 years old and 17% 28 years old or over. Some of the subjects have facial hair, but none of them is wearing glasses. There are minor pose variations, and major illumination as well as facial expression variations. More detailed statics on the database are given by Phillips et al. [Phillips *et al.* 2005].

The 3D faces (shapes) are available in the form of four matrices, each of size 480×640. The first matrix is a binary mask indicating the valid pixels (or points) in the remaining three matrices that respectively contain the x, y, and z-coordinates of the pixels. The 2D faces (texture maps) are 480×640 color images having a one-to-one correspondence to their respective 3D face. The texture maps are correctly registered to the 3D faces in most cases, however, to a few examples, the registration between texture and shape of the same face model is incorrect. In this dataset, the original 3D face models contain noises, spikes, and holes.

All the face data are divided into three sets (based on acquisition time), namely, Spring2003, Fall2003, and Spring2004. The first subset, Spring2003, made up of 943 3D face models with a neutral expression is also called FRGC v1.0, and the other two compose FRGC v2.0 which contains 4007 3D face models of 466 subjects. The FRGC explicitly specifies that FRGC v1.0 be utilized for training while FRGC v2.0 be used for validation. Figure 1.10 lists some samples of textured 3D face models in the FRGC dataset.

1.5.2 The GavabDB dataset

To the best of our knowledge, GavabDB is the most noise-prone dataset currently available to the public. This database consists of 549 3D facial range scans of 61 different subjects captured by a Minolta Vivid 700 scanner. Texture information of each vertex was eliminated to reduce the size of all these face models. The subjects, of whom 45 are male and 16 are female, are all Caucasian, and most of them are aged between 18 and 40. Each subject was scanned 9 times for different poses and facial expressions. The scans with pose variations contain one facial scan while looking up (+35 degree), one while looking down (-35 degree), one for the right profile (+90 degree), one for the left profile (-90 degree), as well as one with a random pose and an arbitrary facial expression. The facial scans without pose changes include four different nearly frontal facial scans: two of them are with a neutral expression, one with a smile, and one with an accentuated laugh. Some examples of one subject in the GavabDB database are shown in fig. 1.11.

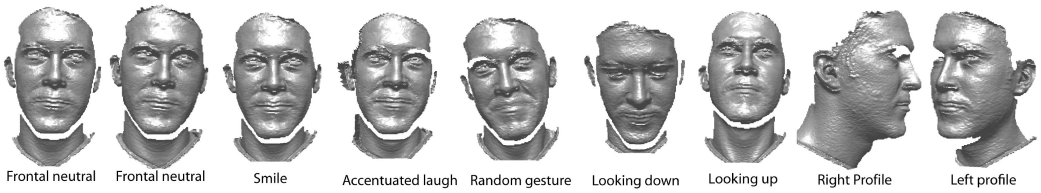


Figure 1.11: Samples of 3D face models in the GavabDB dataset [Drira *et al.* 2010].

1.5.3 The 3DTEC dataset

The Twins Days 2010 dataset was acquired at the Twins Days Festival in Twinsburg, Ohio, USA. It contains 266 subjects and each of them has two 3D facial scans taken using a range scanner: one with a neutral expression and another with a smiling expression. There are 106 sets of identical twins, one set of triplets, and the remainder are non-twins. Three pairs of twins come in for two recording sessions and everyone else only has a single session. Phillips et al. [Phillips *et al.* 2011] provides more details about the dataset.

The 3DTEC dataset is a subset of the Twins Days dataset, which consists of 3D face scans of 107 pairs of twins (two of the triplets were included as the 107th set of twins), and where only the first session for each person was used. To our knowledge, this is the only dataset of 3D face scans in existence that has more than a single pair of twins. The facial scans were taken with a Minolta VIVID 910 3D scanner in a controlled light setting, with the subjects posing in front of a black background. For each pair of twins, their neutral and smile images were taken in a 5 to 10 minute window of time. The 3DTEC only uses one session for each subject. Please see fig. 1.12 for an illustration of two identical twins.

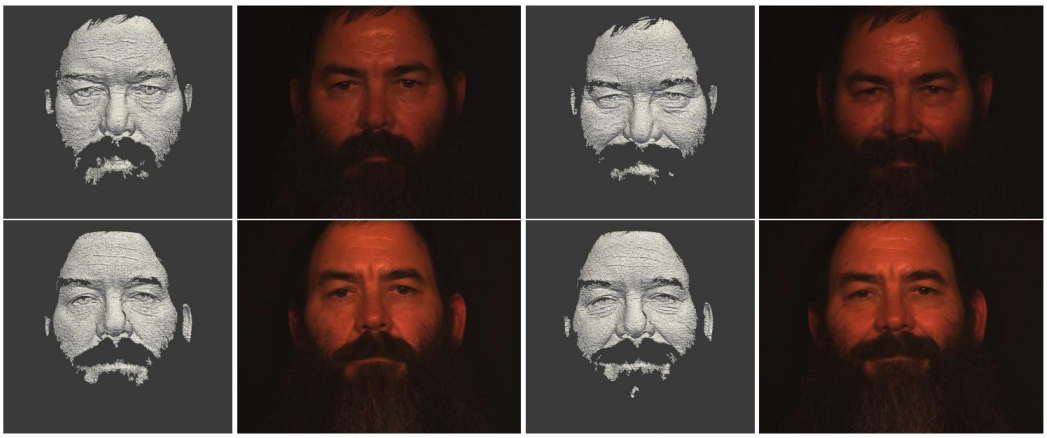


Figure 1.12: The images of two twins taken in a session: the top row shows the first twin and the bottom row, the second; the texture images are brightened to increase visibility in this figure.

1.6 Thesis Contributions

This Ph.D. thesis involves three main contributions in three scenarios respectively, i.e. only 3D shape based face recognition, textured 3D face recognition, as well as asymmetric 3D-2D face recognition.

In only 3D shape-based face recognition, we present an effective and efficient approach using Multi-Scale extended Local Binary Pattern Depth Maps (MS-eLBP-DMs) along with a SIFT-based hybrid matching step. The MS-eLBP-DMs comprehensively describe local shape changes of 3D facial surfaces; while the matching process further robustly associates keypoints between corresponding facial representations of different facial scans belonging to the same subject. Evaluated on the FRGC v2.0 dataset, it achieves a 97.6% rank-one recognition rate and a 98.4% verification rate with a FAR at 0.001 using the neutral vs. all protocol, and displays its insensitiveness to several key factors, such as facial expression variations, noisy data, data missing, and data decimation. Moreover, since the proposed approach possesses certain tolerance to moderate pose changes due to the utilization of local matching, it does not require the generally costly registration step in data preprocessing for nearly frontal faces as the ones in FRGC. When dealing with the face models with severe pose variations, e.g. the left and right profiles, in the GavabDB dataset, it only needs a coarse alignment to compute rotation and translation parameters. Both the facts are in contrast to most of the related tasks that are based on time-consuming fine registration.

In textured 3D face recognition, we propose a biological vision-based facial representation, named Oriented Gradient Maps (OGMs), which can be applied to both facial range and texture images. The OGMs simulate the response of complex neurons to gradient information within a given neighborhood and own properties of being highly distinctive and robust to affine lighting and geometric transformations. The previously proposed matching process is then adopted to calculate similarity measurements between probe and gallery faces. Because the biological vision-based facial representation produces an OGM for each quantized orientation of facial range and texture images, we finally introduced a score level fusion strategy that optimizes weights by a genetic algorithm in a learning step. The experimental results achieved on the FRGC v2.0 and 3DTEC databases clearly illustrate the effectiveness of the proposed biological vision-based facial description and the optimized weighted sum fusion.

Considering the face recognition approaches based on 3D data also have their own disadvantages, and are currently limited by their high cost of data acquisition and computation, in this thesis, we design a novel framework, namely asymmetric 3D-2D face recognition, enrolling in textured 3D face model while performing identification using only 2D facial images. The motivation is to limit the use of 3D data where it really helps to improve face recognition accuracy. The proposed method consists of a new preprocessing pipeline to enhance robustness to illumination and pose changes, an OGM-based facial representation to describe both local shape and texture variations, and a twofold classification step which combines the matching between two facial texture images and the one between a facial range and texture image. The experiments carried out on the FRGC v2.0 database demonstrate that

the proposed approach outperforms 2D facial image based methods, and achieves comparable results as 3D data based ones do. Furthermore, it avoids the cost and inconvenience of facial data acquisition and computation of 3D based approaches.

1.7 Thesis Organization

The remainder of this thesis is organized as follows.

- In Chapter 2, we review the development of face recognition algorithms from 2D domain to 3D domain.
- In Chapter 3, we introduce the proposed face recognition approach using solely 3D shape, including the multi-scale eLBP based geometric facial representation, SIFT based hybrid matching process, and experimental results on the FRGC v2.0 and GavabDB databases.
- In Chapter 4, we present the biological vision based facial description, namely Oriented Gradient Maps (OGMs), the fusion strategy using genetic algorithm, their application to textured 3D face recognition, as well as the evaluation on the FRGC v2.0 and 3DTEC databases.
- In Chapter 5, we describe the novel framework of asymmetric 3D-2D face recognition involving a new preprocessing pipeline, an OGM-based facial representation applied to both facial texture and range images, and a twofold classification step. Meanwhile, according to the performance comparison with traditional 2D and 3D based methods on the FRGC v2.0 database, we analyze its effectiveness and advantages.
- In Chapter 6, we summarize our contributions and propose future directions.

Chapter 2

Literature Review: From 2D to 3D

During last few decades, face recognition has been one of the most active topics in biometrics, pattern recognition and computer vision communities, and a great deal of progress has been made to improve its performance. The past years have witnessed the research attention on face recognition from 2D intensity images to (textured) 3D shape models. In this chapter, we first review several milestone algorithms in 2D domain, then introduce the recent development of face recognition using 3D data.

2.1 Intensity Image based 2D Face Recognition

Up to now, many techniques have been proposed for face identification or verification tasks using 2D intensity facial images. They can be generally divided into two main categories: i.e. holistic and local feature based [Zhao *et al.* 2003, Zou *et al.* 2007a]. The holistic approaches make use of the entire face region as the input data of the face recognition system, which are further projected and compared in a relatively low dimensional subspace in order to avoid the curse of dimensionality, including Principal Component Analysis (PCA) [Turk & Pentland 1991], Linear Discriminant Analysis (LDA) [Belhumeur *et al.* 1997], Independent Component Analysis (ICA) [Bartlett *et al.* 2002], etc. While the local based ones proceed first to locate a number of features or components from a facial image, and then classify them by combining and measuring with corresponding local statistics, such as Elastic Bunch Graph Matching (EBGM) [Wiskott *et al.* 1997], Scale-Invariant Feature Transform (SIFT) [Bicego *et al.* 2006], and Local Binary Patterns (LBP) [Ahonen *et al.* 2006]. There also exist some methods that do not belong to either of the two classes, like Active Appearance Model (AAM) [Edwards *et al.* 1998]. In the subsequent subsections, we introduce the representatives of each category in detail.

2.1.1 Holistic Approaches

Face recognition using holistic methods operates directly on the whole 2D facial images. These facial images can be represented as vectors, i.e., as points in a high dimensional vector space. Specifically, a 2D facial image with a resolution of $p \times q$ can be mapped to a vector $x \in R^{p \times q}$ by a lexicographic ordering of the pixel elements, e.g., by concatenating each row or column of the given image. Despite this high-dimensional embedding, the natural constraints of the physical world (and the imaging process) dictate that the data will in fact lie in a lower-dimensional manifold. The primary goal of subspace analysis is to identify, represent and parameterize this manifold in accordance with some optimality criteria [Moghaddam 1999].

Let $X = (x_1, x_2, ..., x_i, ..., x_N)$ represent a $n \times N$ data matrix, where each x_i is a face vector of dimension n, concatenated from a $p \times q$ facial image; thus $n = p \times q$. Here n is the total number of pixels in the facial image and N is the number of face samples in the training set. The mean vector of the training images $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ is subtracted from each facial image vector for normalization.

The facial images with the resolution of $p \times q$, each pixel of which has 256 gray scale, construct a face space, and in this space, each image represented as a vector corresponds to a point. The facial images only reside a very small portion out of total $256^{p \times q}$ possible instances in this space. The manifold or the distribution of all faces accounts for changes in the facial appearance. To analyze this face manifold, three classical holistic approaches, PCA, LDA, and ICA can be applied. Each method has its own representations (basis vectors) of a high dimensional face space based on different statistical viewpoint. By projecting facial images to the basis vectors, their corresponding coefficients are regarded as the feature of each facial image. The similarity measurement between the test facial image and training prototype is computed (e.g. Euclidean distance or the cosine of the angle) in the feature space, and a larger cosine value of the angle or a smaller Euclidean distance between their coefficient vectors indicates a better matching.

All the three representations can be considered as a linear transformation from the original image vector to a projection feature vector, as defined in eq. 2.1.

$$Y = W^T X \tag{2.1}$$

where Y is a $d \times N$ feature vector matrix, d is the dimension of the feature space, and W is the transformation matrix. It should be noted that $d \ll n$.

(1) PCA

Principal Component Analysis (PCA) searches for Y, which is best accounts for the distribution of facial images, and Y defines the subspace of these facial images as the face space [Turk & Pentland 1991]. All faces in the training set are projected into this subspace to find a set of weights which describes the contribution of each vector in the face space. To identify a facial image, one needs to project the test image into the face space to achieve the corresponding set of weights. By comparing the weights of the test image with the ones of the faces in the training set, the face in the test image can be identified.

The main process of PCA algorithm is based on Karhunen-Loeve transformation [Kirby & Sirovich 1990]. If facial image elements are considered to be random variables, the image can be seen as a sample of a stochastic procedure. The PCA basis vectors are defined as the eigenvectors of the $n \times n$ total scatter matrix S_T.

$$S_T = \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T \tag{2.2}$$

The transformation matrix W_{PCA} is composed of the eigenvectors corresponding to the d largest eigenvalues. The eigenvectors are also known as eigenfaces. In order to better demonstrate eigenfaces, we derive the ones corresponding to the 12 largest

Figure 2.1: Face samples from the ATT (ORL) face dataset.

eigenvalues from ATT face database [Samaria & Harter 1994] (formerly the ORL database) consisting of 40 different subjects, each of which has 10 2D facial images. Some face samples in ATT face database are shown in Fig. 2.1; and the mean face of this dataset is given in Fig. 2.2. The achieved eigenfaces corresponding to the 12 largest eigenvalues are depicted in Fig. 2.3.

Figure 2.2: The mean face derived from the ATT (ORL) face dataset.

After applying the projection, the input facial image represented as an n dimensional vector is reduced to a feature vector in a d dimensional subspace. For most applications, the eigenvectors corresponding to very small eigenvalues are regarded as noise, and hence not taken into account during the recognition step. Figure 2.4 illustrates the eigenfaces corresponding to the smallest 12 eigenvalues.

PCA is the most descriptive representation in terms of the least square reconstruction error. Moreover it operates efficiently and is easy to implement. Therefore, PCA is usually used as the baseline algorithms for face recognition. However, it is not the most discriminative in this domain.

(2) ICA

Independent Component Analysis (ICA) [Hyvarinen & Oja 2000] is quite similar to PCA, and their only difference lies in that the distribution of the components are designed to be non-Gaussian. In general, minimizing non-Gaussianity promotes statistical independence [Hyvarinen & Oja 2000].

Bartlett et al. [Bartlett *et al.* 2002] show that first- and second- order statistics

Figure 2.3: Eigenfaces (eigenvectors) of the 12 largest eigenvalues are shown as $p \times q$ resolution, where $p \times q = n$, derived from the ATT (ORL) face dataset.

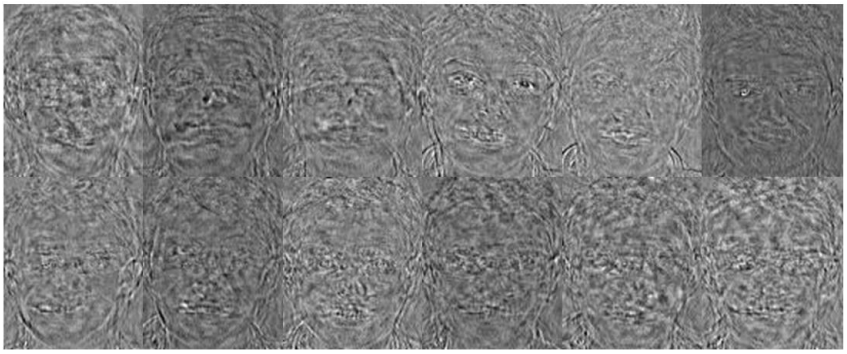


Figure 2.4: Eigenfaces (eigenvectors) of the 12 smallest eigenvalues are shown as resolution $p \times q$, where $p \times q = n$, derived from the ATT (ORL) face dataset.

hold information only about the amplitude spectrum of an image while discard the phase spectrum, however, some experiments then brought out that the human capability in recognizing objects is mainly driven by the phase spectrum. This is also the reason that they investigated ICA as a more powerful algorithm for face recognition. Since ICA separates the high-order moments of the input image in addition to the second-order moments utilized by PCA, it can be regarded as a generalization of PCA. Furthermore, ICA allows a better characterization of data in an n dimensional space and the basis vectors found by ICA are not necessarily orthogonals so that they also reduce the reconstruction error. Nevertheless, ICA has no general closed-form solution, and needs iterative methods to obtain its representation.

Bartlett et al. provided two architectures based on ICA, statistically independent basis images and a factorial code representation, for the face recognition tasks. Both the architectures display similar accuracies. The basis vectors based on fast fixed-point algorithm [Hyvarinen 1999] for ICA factorial code representation are shown in Fig. 2.5.

(3) LDA

PCA and ICA are unsupervised approaches since they construct the face space without using the face class information. Unlike them, Linear Discriminant Analysis

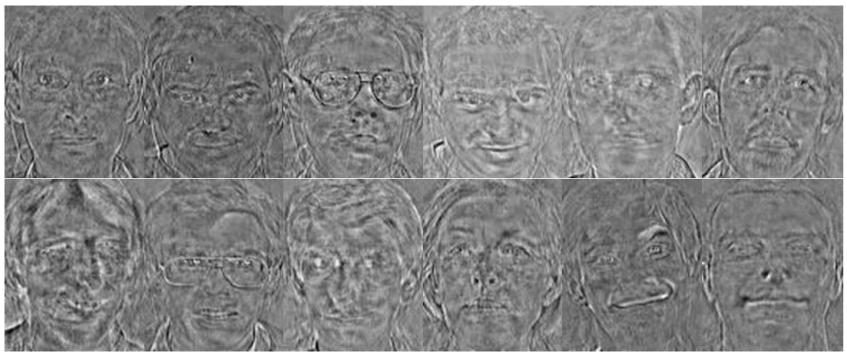


Figure 2.5: ICA basis vectors shown as resolution $p \times q$, where $p \times q = n$, derived from the ATT (ORL) face dataset based on the second architecture [Bartlett *et al.* 2002].

(LDA) aims to find an "optimal" way to represent the face vector space to maximize the discrimination between different subjects. Exploring class information can improve the performance of identification tasks [Belhumeur *et al.* 1997].

LDA seeks a transformation W_{LDA} defined as in eq. 2.3:

$$W_{LDA} = \arg\max_{W} \frac{W^T S_B W}{W^T S_W W} \tag{2.3}$$

where S_B and S_W are the between-class and within-class scatter matrix, and formally defined as in eq. 2.4 and eq. 2.5 respectively.

$$S_B = \sum_{i=1}^{c} N_i (x_i - \mu)(x_i - \mu)^T \tag{2.4}$$

$$S_W = \sum\nolimits_{i=1}^{c} \sum\nolimits_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \tag{2.5}$$

In the above equations, N_i is the number of training samples in class i; c is the total number of different classes; μ_i is the mean vector of the samples belonging to class i; and X_i represents the set of samples in class i. The LDA basis vectors are demonstrated in Fig. 2.6

Unfortunately, LDA is affected by "small sample size problem", especially common in face recognition, arising from the small number of available training samples compared to the dimensionality of the sample space. An effective solution is taking PCA as the preliminary step to reduce the dimensionality of the sample space and then applying LDA to the achieved space for real classification. The Fisherface algorithm [Belhumeur *et al.* 1997] is a right example, derived from the Fisher Linear Discriminant (FLD), which employs specific category information. By defining different classes with various statistics, the facial images in the training set are first divided into the corresponding classes. Then, techniques similar to the one applied in PCA approach are used. The Fisherface algorithm generally results in a

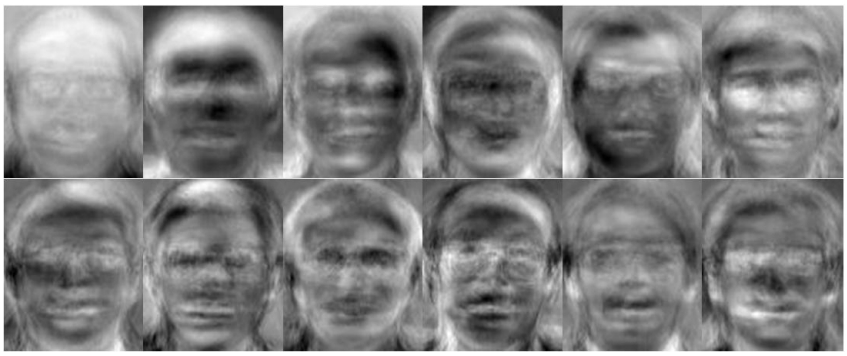


Figure 2.6: First 12 LDA basis vectors shown as resolution $p \times q$, where $p \times q = n$, derived from the ATT (ORL) face dataset.

higher accuracy rate for face recognition than PCA does. In contrast, arguing that combining PCA and LDA as the way of Fisherface, discriminant information is also discarded together with redundant one, Chen et al. [Chen *et al.* 2000], Yu and Yang [Yu & Yang 2001] suggest that in some cases LDA should be applied directly to the input space.

Many efforts have been made to ameliorate linear subspace analysis technique since PCA, ICA and LDA were proposed for face recognition. For example, Pentland et al. [Pentland *et al.* 1994] modified PCA to modular eigenfaces to deal with pose variations; Probabilistic subspaces [Moghaddam 2002] was introduced to derive a more meaningful similarity measurement under the probabilistic framework; Lu et al. [Lu *et al.* 2003] presented an hybrid between the D-LDA (Direct LDA) and the F-LDA (Fractional LDA), an LDA variant, in which weighted functions are used to avoid the misclassification caused by output classes that are too close; Vasilescu and Terzopoulos [Vasilescu & Terzopoulos 2002] proposed an approach based on multilinear tensor decomposition of image ensembles, namely TensorFace, to resolve the confusion of multiple factors constrained in the same face recognition system, such as illumination and pose; instead of representing the facial image as a vector, Yang et al. [Yang *et al.* 2004a] considered an image as a 2D matrix and developed a two dimensional PCA algorithm for face recognition; similarly, Kong et al. [Kong *et al.* 2005] generalized the conventional LDA into 2D Fisher discriminant analysis and applied it to face recognition; etc.

Although linear subspace analysis approaches, e.g. PCA, ICA, and LDA, have significantly advanced the development of face recognition, due to high non-linearity of the face manifolds [Li & Jain 2005], linear subspace analysis is an approximation of this non-manifold, and does not possess the modeling capacity which is powerful enough to preserve the variations of the face manifold and distinguish between individuals to achieve robust face recognition results. Recent studies in non-linear manifold analysis provide more flexibility and modeling power to analyze face manifold. Its main idea is that a pattern in the original input space is first mapped into a potentially much higher dimensional feature vector in the feature space; then perform dimensionality reduction techniques to learn the non-linear manifold, and the resulting projection coefficients are used as features for face classification. How-

ever, it is difficult to do so directly because it is computationally very intensive to compute the dot products in a high-dimensional feature space. Fortunately, kernel techniques can be introduced to avoid this difficulty. The algorithm can be actually implemented in the input space by virtue of kernel tricks satisfying the Mercer's condition [Vapnik 1995]

$$K(x_i, x_j) = \Psi(x_i) \cdot \Psi(x_j) \tag{2.6}$$

where the kernel function $K(x_i, x_j)$ in the input space corresponds to inner-product in the higher dimensional feature space. Common kernels include: polynomial (homogeneous), polynomial (inhomogeneous), Gaussian Radial Basis Function (RBF), hyperbolic tangent, etc. Therefore, the explicit mapping process is not required at all. For example, Yang [Yang 2002b] investigated KPCA (Kernel Principal Component Analysis) for face recognition. Unlike the traditional PCA and LDA, KPCA achieved a higher dimensionality than the input image, and provided better performance than Eigenface and Fisherface did. Liu et al. [Liu *et al.* 2002] proposed an face recognition approach using polynomial kernel based Fisher Discriminant Analysis (FDA, Fisherface), and claimed that it outperformed PCA and LDA. But in these kernel based methods, suitable kernels and their corresponding parameters can only be determined empirically. There also exist another two popular methods for non-linear manifold learning, i.e. ISOMAP [Tenenbaum *et al.* 2000] and Locally Linear Embedding (LLE) [Roweis & Saul 2000], which have been further applied to face recognition since they were originally proposed. Yang [Yang 2002a] introduced LDA to recognize faces using geodesic distance, which is the basis of the ISOMAP. He et al. [He *et al.* 2005] presented laplacianfaces based on the Locality Preserving Projections (LPP) for face subspace description. These manifold learning algorithms are quite interesting, but need to further prove their performance for robust face recognition systems.

On the other hand, the downside of the non-linear approaches is that their generalization capability is influenced by the sample size in real conditions, i.e., small number of face images available for training compared to the large variations of facial appearance in testing, leading to overfitting [Raudys & Jain 1991]. A possible solution is to use face synthesis methods to generate additional samples from available ones for training, proving helpful to enhance the accuracies of face recognition systems [Lu *et al.* 2004c, Vetter & Poggio 1997, Zhao & Chellappa 2000]. In addition, the techniques such as classifier combination [Lu & Jain 2003] and data re-sampling [Lu & Jain 2003] can also be employed for improved results.

2.1.2 Local Feature based Approaches

Face recognition using local feature based approaches usually consists of three major steps: i.e. alignment and partitioning, feature extraction, as well as classification and combination, as illustrated in Fig. 2.7.

As in the holistic methods, the preliminary step of most local feature based ones is to align the face images, while in contrast to the holistic ones, the aligned faces are then partitioned into local blocks for feature extraction. An common and efficient

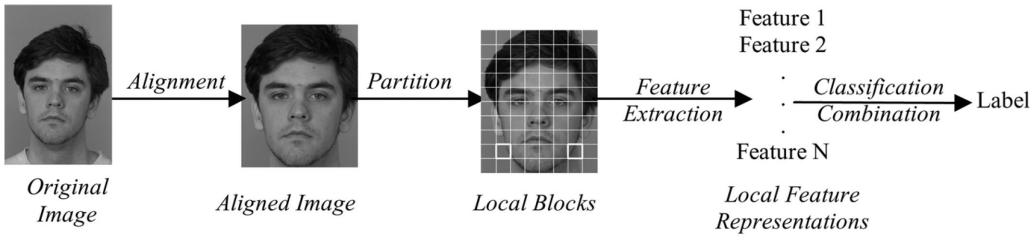


Figure 2.7: The framework of local feature based methods [Zou *et al.* 2007a].

way is to align facial images into the same coordinate system by a similarity transform (translation, rotation and scaling) based on a few fiducial landmark points, facial components (eyes, nose, and mouth) or even the entire face region provided by some detection techniques, e.g. AdaBoost [Viola & Jones 2004], and the face is further divided into local regions. Classification is finally carried out on the faces by comparing and combining each similarity of their local region in the given feature space. In this subsection, we discuss three typical local feature based approaches, namely, Gabor wavelets, LBP and SIFT.

(1) Gabor wavelets

Gabor wavelets were applied to image analysis due to their biological relevance and computational properties [Daugman 1985, Daugman 1988, Marcelja 1980]. The Gabor wavelets, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains as well

The Gabor wavelets (kernels, filters) can be defined as follows [Daugman 1980, Lades *et al.* 1993, Marcelja 1980]:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} \exp\left(\frac{-\|k_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2}\right) \left[\exp(ik_{\mu,\nu}z) - \exp(-\frac{\sigma^2}{2})\right] \tag{2.7}$$

where μ and ν denotes the orientation and scale of the Gabor kernels, $z = (x, y)$ respectively; $\|\cdot\|$ is the norm operator; and the wave vector $k_{\mu,\nu}$ is defined as:

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \tag{2.8}$$

where $k_\nu = k_{\max}/f^\nu$ and $\phi_\mu = \pi\mu/8$. k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain [Lades *et al.* 1993].

The Gabor kernels in eq. 2.7 are all self-similar because they can be generated from one filter, the mother wavelet, by scaling and rotation through the wave vector $k_{\mu,\nu}$. Each kernel is a product of a Gaussian envelope and a complex plane wave, while the first term in the square brackets in eq. 2.7 determines the oscillatory part of the kernel and the second term compensates for the DC value. The effect of the

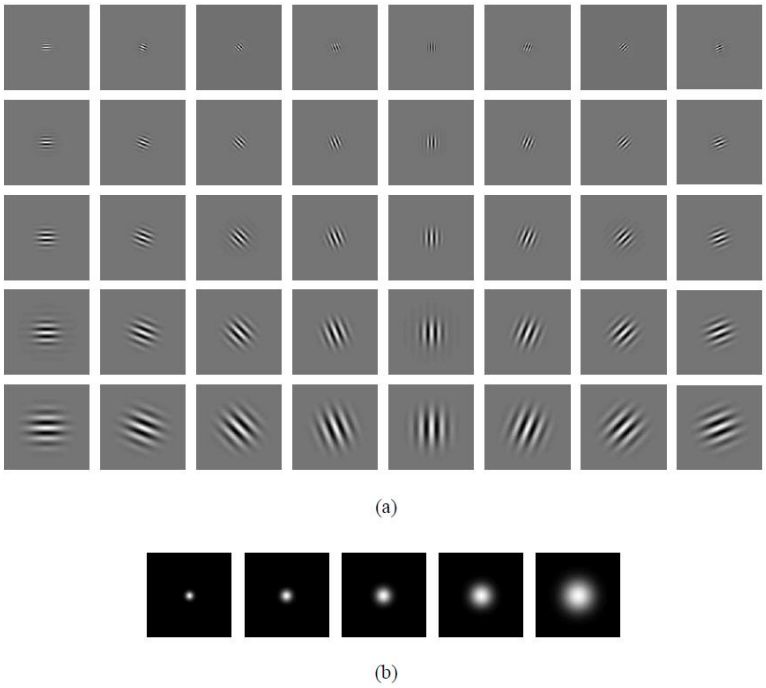


Figure 2.8: Gabor wavelets: (a) The real part of Gabor kernels at five scales and eight orientations with the following parameters: $\sigma = 2\pi$, $k_{\max} = \pi/2$, and $f = \sqrt{2}$; (b) The magnitude of Gabor kernels at five different scales. The kernels exhibit desirable characteristics of spatial frequency, spatial locality, and orientation selectivity [Liu & Wechsler 2002].

DC term becomes negligible when the parameter σ, which determines the ratio of the Gaussian window width to wavelength, has sufficiently large values.

In most cases, one would make use of the Gabor wavelets of five different scales, $\nu \in \{0,...,4\}$ and eight orientations, $\mu \in \{0,...,7\}$ [Burr *et al.* 1989, Field 1987, Jones & Palmer 1987]. Figure 2.8 shows the real part of the Gabor kernels at five scales and eight orientations and their magnitude, with the following parameters: $\sigma = 2\pi$, $k_{\max} = \pi/2$, and $f = \sqrt{2}$. The kernels exhibit desirable characteristics of spatial frequency, spatial locality, and orientation selectivity.

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor Kernels as defined by eq. 2.7. Let $I(x, y)$ be the gray level distribution of an image, the convolution of image I and a Gabor kernel $\psi_{\mu,\nu}(z)$ is defined as follow:

$$O_{\mu,\nu}(z) = I(z) * \psi_{\mu,\nu}(z) \tag{2.9}$$

where $*$ is the convolution operator, and $O_{\mu,\nu}(z)$ is the convolution result corresponding to the Gabor Kernel at different orientations and scales. As a result, the set $S = \{O_{\mu,\nu}(z) : \mu \in \{0,...,7\}, \nu \in \{0,...,4\}\}$ forms the Gabor wavelet representation of the image $I(z)$.

According to the convolution theorem, each $O_{\mu,\nu}(z)$ can be derived from eq. 2.9 via the Fast Fourier Transform (FFT)

$$\Im\{O_{\mu,\nu}(z)\} = \Im\{I(z)\}\Im\{\psi_{\mu,\nu}(z)\} \tag{2.10}$$

and

$$O_{\mu,\nu}(z) = \Im^{-1}\{\Im\{I(z)\}\Im\{\psi_{\mu,\nu}(z)\}\} \tag{2.11}$$

where $\Im$ and $\Im^{-1}$ denote the Fourier and inverse Fourier transform, respectively.

Figure 2.9 gives an example of the Gabor wavelet representation (the real part and the magnitude) of a sample image. These achieved representation results display scale, locality, and orientation properties corresponding to those displayed by the Gabor wavelets in Fig. 2.8. To encompass different spatial frequencies (scales), spatial localities, and orientation selectivities, all these representation results are concatenated to derive an augmented feature vector χ.

Since Gabor filters detect amplitude-invariant spatial frequencies of pixel gray values, they are known to be robust to illumination variations, which is an important contribution to 2D face recognition. On the other hand, for each pixel of a facial image, Gabor filters tend to generate a feature vector with dozens of dimensionality, and the final vector by concatenating the ones of each pixel will thus be classified in a rather high-dimensional feature space, leading to another difficulty in this domain. Therefore, many studies have discussed how to balance its performance and computational cost in the last decade.

Wiskott et al. [Wiskott *et al.* 1997] obtained good performance with the Elastic Bunch Graph Matching (EBGM) method in the FERET test [Phillips *et al.* 2000]. The elastic bunch graph is a graph-based face model with a set of jets (represented by Gabor wavelet components) attached to each node of the graph. The algorithm recognizes new faces by first locating a set of facial features (graph nodes) to build a graph, whose jets and topography are then used to compute the similarity. Because only the Gabor representations of a certain number of jets rather than each pixel are required to describe the variations of facial appearance, EBGM avoids the problem caused by high-dimensional space. Liu and Wechsler proposed to first downsample each $O_{\mu,\nu}(z)$ by a pre-defined factor to reduce the space dimension and then apply Enhanced Fisher linear discriminant model (EFM) [Liu & Wechsler 2000] to the augmented feature vector. Using the eigenface selectivity constraint of the EFM method, the dimensionality of the resulting vector space was further reduced to derive low-dimensional features with enhanced discriminative power for classification. To deal with the same issue that Gabor features adopted by most systems are redundant and too high dimensional, Yang et al. [Yang *et al.* 2004b] investigated AdaBoost to choose Gabor features, and the selected features proved not only low-dimensional but also discriminant.

(2) Local Binary Patterns (LBP)

As a non-parametric algorithm, LBP summarizes local structures of images efficiently by comparing each pixel with its neighbors. The most important properties

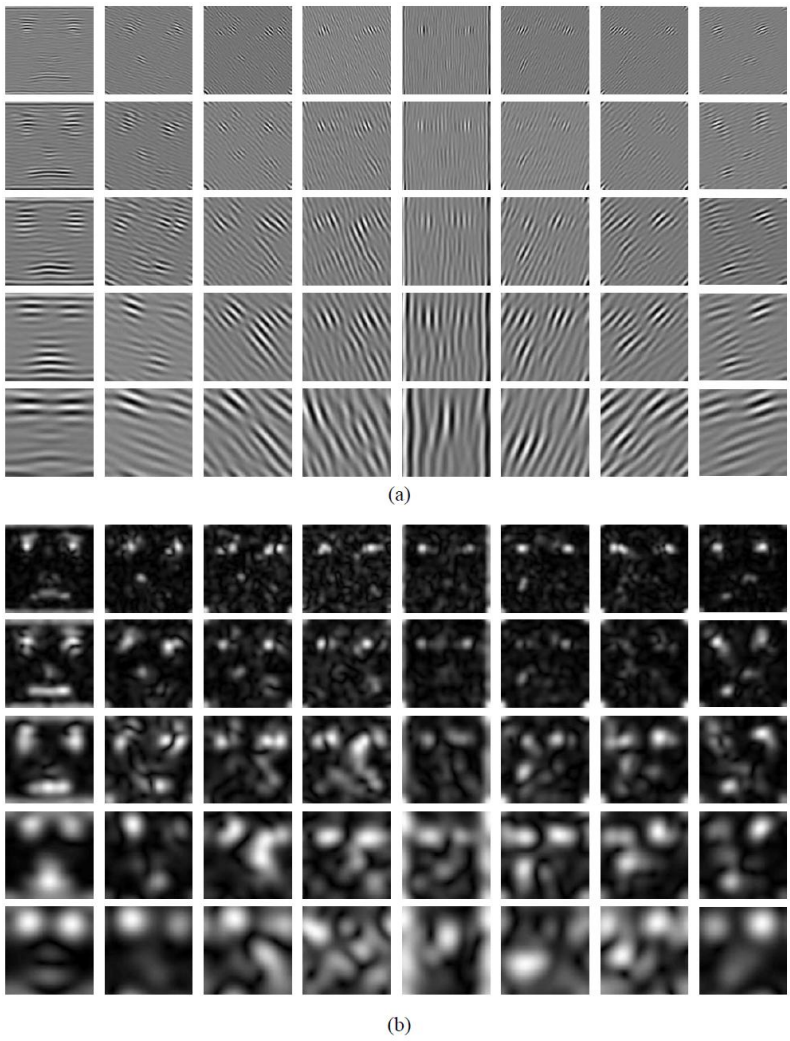


Figure 2.9: Gabor wavelet representation: (a) The real part of the representation; (b) The magnitude of the representation [Liu & Wechsler 2002].

of LBP are its tolerance to monotonic illumination changes and its computational simplicity. LBP was originally proposed for texture analysis [Ojala *et al.* 1996], and has proved a simple yet powerful approach to describe local structures. It has been extensively exploited in many applications, for instance, face image analysis, image and video retrieval, environment modeling, visual inspection, motion analysis, biomedical and aerial image analysis, remote sensing, so forth.

The original LBP operator labels the pixels of an image with decimal numbers, called *Local Binary Patterns* or *LBP codes*, which encode the local structure around each pixel. It proceeds thus, as illustrated in Fig.2.10: each pixel is compared with its eight neighbors in a 3×3 patch; the resulting strictly negative values are encoded with 0 and the others with 1; a binary number is obtained by concatenating all these binary codes in a clockwise direction starting from the top-left one and its

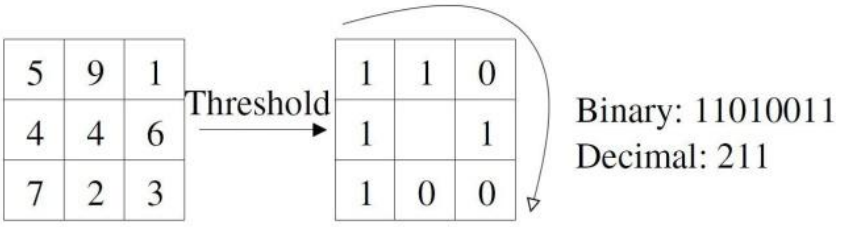


Figure 2.10: An example of the basic LBP operator [Ahonen *et al.* 2004].

corresponding decimal value is used for labeling.

One limitation of the basic LBP operator is that its small 3×3 neighborhood cannot capture the dominant features with large scale structures. The original LBP operator was later generalized to deal with different neighborhoods [Ojala *et al.* 2002]. A local neighborhood is defined as a set of sampling points evenly spaced on a circle which is centered at the pixel to be labeled, and the sampling points that do not fall within the pixels are interpolated using bilinear interpolation, thereby allowing for any radius and any number of sampling points in the neighborhood. Fig.2.11 shows some examples of the extended LBP operator, where the notation (P, R) denotes a neighborhood of P sampling points on a circle of radius of R.

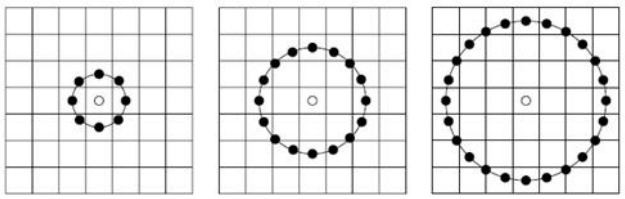


Figure 2.11: Some examples of the extended LBP operator [Ahonen *et al.* 2006]: the circular $(8, 1)$, $(16, 2)$, and $(24, 3)$ neighborhoods.

Formally, given a pixel at (x_c, y_c), the resulting LBP can be expressed in decimal form as:

$$LBP_{P,\,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(i_P - i_c)2^p \tag{2.12}$$

where i_c and i_P are respectively gray-level values of the central pixel and P surrounding pixels in the circle neighborhood with a radius R, and function $s(x)$ is defined as:

$$s(x) = \begin{cases} 1 & if \;\; x \geq 0 \\ 0 & if \;\; x < 0 \end{cases} \tag{2.13}$$

By the definition above, the basic LBP operator is invariant to monotonic gray-scale transformations preserving pixel intensity order in the local neighborhoods.

The histogram of LBP labels calculated over a region can be exploited as a texture descriptor.

The operator $LBP_{(P,R)}$ produces 2^p different output values, corresponding to 2^p different binary patterns formed by P pixels in the neighborhood. If the image is rotated, these surrounding pixels in each neighborhood will move correspondingly along the perimeter of the circle, resulting in a different LBP value, except patterns with only 1s and 0s. In order to remove rotation effect, a rotation-invariant LBP is proposed in [Ojala *et al.* 2002]:

$$LBP^{ri}_{(P,R)} = \min\{ROR(LBP_{(P,R)}, i)|i = 0,\ 1,\ ..., P-1\} \tag{2.14}$$

where $ROR(x, i)$ performs a circular bit-wise right shift on the P-bit number x i times. The $LBP^{ri}_{(P,R)}$ operator quantifies occurrence statistics of individual rotation invariant patterns corresponding to certain micro-features in the image; hence, the patterns can be considered as a feature detector [Ojala *et al.* 2002]. However, in [Pietikainen *et al.* 2000], it was shown that such a rotation-invariant LBP operator does not necessarily provide discriminative information, since the occurrence frequencies of the individual patterns incorporated in $LBP^{ri}_{(P,R)}$ vary greatly and the crude quantization of the angular spaces at 45° intervals.

It has been shown that certain patterns contain more information than others [Ojala *et al.* 2002]. It is possible to use only a subset of 2^p binary patterns to describe the texture of images. Ojala et al. named these patterns uniform patterns, denoted $LBP^{U2}_{(P,R)}$. A local binary pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the corresponding bit string is considered circular. For instance, 00000000 (0 transitions) and 01110000 (2 transitions) are both uniform whereas 11001001 (4 transitions) and 01010011 (6 transitions) are not. It is observed that the uniform patterns account for around 90% of all the patterns in a $(8, 1)$ neighborhood and around 70% in a $(16, 2)$ neighborhood in texture images [Ojala *et al.* 2002]. A similar experiment was conducted on the FERET face image database, and it was found that 90.6% of the patterns in a $(8, 1)$ neighborhood and 85.2% in a $(8, 2)$ neighborhood are uniform [Ahonen *et al.* 2004]. More recently, Shan and Gritti [Shan & Gritti 2008] verified validity of uniform patterns for representing faces from the viewpoint of machine learning. Specifically, they applied AdaBoost to select the discriminative patterns for facial expression recognition, and their experiments demonstrated that, using $LBP_{(8,2)}$ operator, 91.1% of these selected patterns are uniform. Accumulating the non-uniform patterns into a single bin yields an LBP operator with less than 2^p labels. For example, the number of labels with the neighborhood of 8 pixels is 256 for the standard LBP but only 59 for LBP^{U2}.

Ahonen et al. [Ahonen *et al.* 2004] introduced LBP for face recognition in 2004, and unlike its application to texture classification, in order to make use of the spatial information of facial physical components, they proposed to first divide face images into several local regions from which local LBP histograms can be extracted, and then to concatenate them into a single, spatially enhanced feature histogram (as illustrated in Fig.2.12). The resulting histogram encodes both the local texture and global shape information and hence describes a face comprehensively. Since

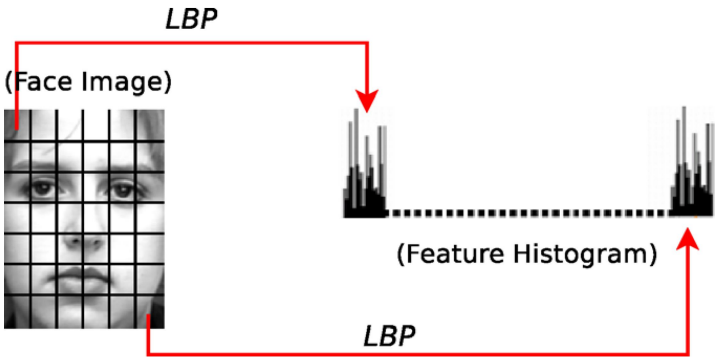


Figure 2.12: LBP based facial description [Shan *et al.* 2009].

then, LBP and its variants have been extensively studied and exploited in 2D face recognition. Zhang et al. [Zhang *et al.* 2005] combined Gabor and LBP in a serial strategy which consists in first applying Gabor filters and then LBP to the raw image. The Multiple Gabor feature maps (GFM) were computed by convolving input images with multi-scale and multi-orientation Gabor filters. Each GFM was divided into small non-overlapped regions from which LBP histograms were extracted and finally concatenated into a single facial feature histogram. They reported the best result on the FERET dataset at that moment. Tan and Triggs [Tan & Triggs 2007] modified LBP to a version with 3-value codes, called Local Ternary Patterns (LTP), by replacing zero with a user-specified threshold to measure the difference between the central pixel and its neighbors, claiming it is more resistant to noise and thus can improve performance of face recognition under difficult lighting conditions. In order to capture not only micro-structures but also macro-structures to enhance the discriminative power of LBP, Liao et al. [Liao *et al.* 2007] proposed Multi-Block LBP (MB-LBP) which, instead of comparing pixels, compares average intensities of neighboring sub-regions, and a similar scheme was introduced in [Wolf *et al.* 2008]: Three-Patch LBP (TP-LBP) and Four-Patch LBP (FP-LBP) were proposed to compare distances between the whole blocks (patches) concerned for face recognition in wild. More development details of LBP on face recognition is in [Huang *et al.* 2011c].

(3) Scale-Invariant Feature Transform (SIFT)

SIFT features have many important properties, such as invariant to image scaling and rotation, (partial) occlusion and to a certain extent also to changes in illumination and 3D camera viewpoint, which make them suitable for matching different images of an object or a scene. According to [Lowe 2004], the basics of the SIFT algorithm consists of four computational stages: (i) scale-space extrema detection, (ii) removal of unreliable keypoints, (iii) orientation assignment, (iv) keypoint descriptor calculation, and (v) matching.

In the **first** stage, interest points (or called keypoints), are detected in the scale space searching for image locations that represent the maxima or minima of the difference-of-Gaussian function. The scale space of an image is defined as a function $L(x, y, \sigma)$, that is produced from the convolution of a variable-scale Gaussian, $G(x, y, \sigma)$, with the input image, $I(x, y)$:

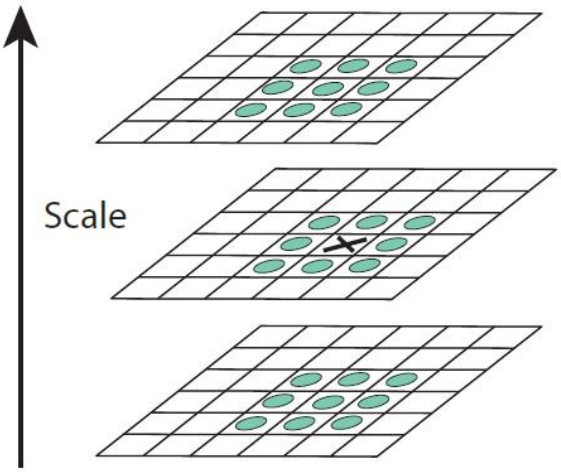


Figure 2.13: Maxima and minima of the difference-of-Gaussian images are detected by comparing a pixel (marked with '×') to its 26 neighbors in 3×3 regions at the current and adjacent scales (marked with circles) [Lowe 2004].

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{2.15}$$

with

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{2.16}$$

where σ denotes the standard deviation of the Gaussian $G(x, y, \sigma)$.

The difference-of-Gaussian function $D(x, y, \sigma)$ can be computed from the difference of Gaussians of two scales that are separated by a factor k:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{2.17}$$

Local maxima and minima of $D(x, y, \sigma)$ are computed based on the comparison of the sample point and its eight neighbors in the current image as well as the nine neighbors in the scale above and below (See Fig.2.13 for an illustration). If the pixel represents a local maximum or minimum, it is selected as a candidate keypoint.

The final qualified keypoints are selected based on measurements of their stability. During the **second** stage low contrast points (sensitive to noise) and poorly localized points along edges (unstable) are discarded. Two criteria are utilized for the detection of unreliable keypoints. One criterion evaluates the value of $|D(x, y, \sigma)|$ at each candidate keypoint. If the value is below some threshold, which means that the structure has low contrast, the keypoint is removed. Another criterion evaluates the ratio of principal curvatures of each candidate keypoint to find out poorly

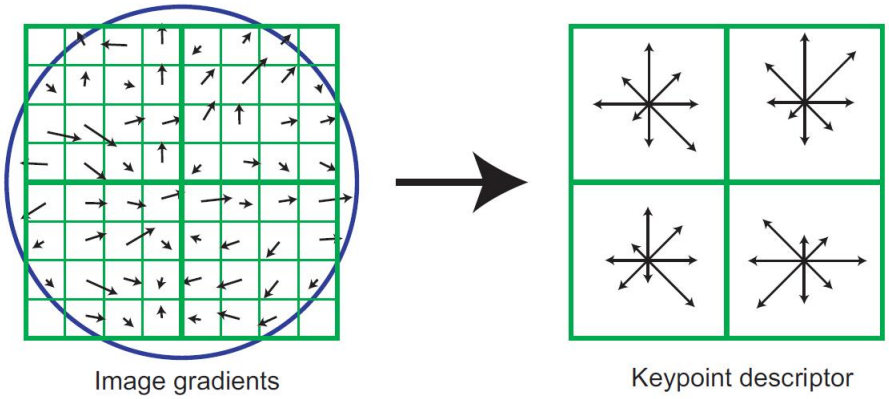


Figure 2.14: A keypoint descriptor is created by first computing the gradient magnitude and orientation at each image sample point in a region around the keypoint location, as shown on the left. These are weighted by a Gaussian window, indicated by the overlaid circle. These samples are then accumulated into orientation histograms summarizing the contents over 4×4 subregions, as shown on the right, with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region. This figure shows a 2×2 descriptor array computed from an 8×8 set of samples, whereas the experiments in this paper use 4×4 descriptors computed from a 16×16 sample array.[Lowe 2004].

defined peaks in the Difference-of-Gaussian function. For keypoints with high edge responses, the principal curvature across the edge will be much larger than the principal curvature along it. As a result, to remove unstable edge keypoints based on the second criterion, the ratio of principal curvatures of each candidate keypoint is checked. If the ratio is below some threshold, the keypoint is kept, otherwise it is removed.

The **third** stage assigns an orientation to each keypoint by building a histogram of gradient orientations $\theta(x, y)$ weighted by the gradient magnitudes $m(x, y)$ from the neighborhood of keypoint:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \quad (2.18)$$

$$\theta(x,y) = \tanh(L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)) \quad (2.19)$$

where L is a Gaussian smoothed image with a closest scale to that of a keypoint. By assigning a consistent orientation to each keypoint, the keypoint descriptor can be represented relative to this orientation and, therefore, invariance to image rotation is achieved.

In the **fourth** step, the keypoint descriptor is created by primarily computing the gradient magnitude and orientation at each image point of the 16×16 keypoint neighborhood (left side of Fig.2.14). This neighborhood is weighted by a Gaussian

Figure 2.15: An example of SIFT matching on 2D facial images.

window and then accumulated into orientation histograms summarizing the contents over subregions of the neighborhood of size 4×4 (see the right side of Fig.2.14), with the length of each arrow in Fig. 2.14 (right) corresponding to the sum of the gradient magnitudes near that direction within the region [Lowe 2004]. Each histogram contains 8 bins, therefore each keypoint descriptor features 4×4×8=128 elements. The coordinates of the descriptor and the gradient orientations are rotated relative to the keypoint orientation to achieve orientation invariance and the descriptor is normalized to enhance invariance to changes in illumination.

The previously retained descriptors are matched in the **last** step. When applying the SIFT algorithm to object recognition, each keypoint descriptor extracted from the query (or test) image is matched independently to the set of descriptors extracted from every image in the training set. The best match for each descriptor is found by identifying its nearest neighbor (closest descriptor) in the set of keypoint descriptors from the training image. Generally, many features from a test image do not have any correct match in the training database, because they are either not detected in the training image or they arose from background clutter. To discard keypoints whose descriptors do not have any good match to a training image, a subsequent threshold is used, which rejects matches that are too ambiguous. If the distance ratio between the closest neighbor and the second-closest neighbor is below some threshold, the match is kept; otherwise the match is rejected. The object in the database with the largest number of matching points is considered as the matched object.

Several tasks adapted SIFT to face recognition. Bicego et al. [Bicego *et al.* 2006] first investigated the performance of SIFT in face authentication, and designed a regular grid based matching step that was claimed to be more suitable to this topic. The proposed matching process takes into account the spatial information of the keypoints detected in the face area, emphasizing the importance of matches from corresponding local patches, and hence greatly eliminated wrong matches between spatially inconsistent SIFT descriptors. Luo et al. [Luo *et al.* 2007] further enhanced the SIFT-based face recognition system by assigning weights to different facial subregions using a K-means clustering method, highlighting the matches from the areas that are more critical to improve accuracies. Krizaj et al. [Krizaj *et al.* 2010] argued that even though SIFT is regarded as one of the state-of-the-art approaches to object

recognition, it has some deficiencies when applied to the problem of face recognition. Compared to general objects, there are less structures with high contrast or high-edge responses in facial images. Because keypoints along edges and low-contrast keypoints are removed by the original SIFT algorithm, interest points representing distinctive facial features can also be removed. Therefore, it is of paramount importance to properly adjust the thresholds governing the process of unstable keypoint removal, when applying the SIFT technique for the task of face recognition. They suggested to extract the SIFT descriptors at fixed predefined image locations learned during the training stage, and by fixing the keypoints to pre-defined spatial locations, they cancelled the need for threshold optimization and face image partitioning. In addition, they claimed their approach gained greater illumination invariance than other SIFT adaptations. See Fig.2.15 for an example of SIFT matching on 2D facial images.

2.1.3 Other Methods

Besides the two streams of face recognition mentioned above, another one, namely statistical model based methods, has been also widely explored. A statistical model is a formalization of relationships between variables in the form of mathematical equations, and it describes how one or more random variables are related to one or more random variables. The model is regarded to be statistical as the variables are not deterministically but stochastically related. To formulate face analysis with statistical model, we wish to build models of facial appearance and its variations and learn the ways in which the shape and texture of the human face vary across a range of images. Examples of the statistical model based approaches include Active Appearance Model (AAM) [Edwards *et al.* 1998], Gaussian Mixture Model (GMM) [Cardinaux *et al.* 2003, Lucey & Chen 2004], 1D Hidden Markov Model (HMM) [Samaria & Harter 1994], pseudo-2D HMM [Nefian & Hayes III 1999], etc. In this sub-section, as a typical representative of statistical model based techniques, AAM is introduced in detail not only due to its successful application in face recognition, but because of its fundamental contributions in facial image analysis as well.

Active Appearance Model (AAM)

An Active Appearance Model (AAM) is an integrated statistical model combining a model of shape variations with a model of appearance variations in a shape-normalized frame. It generally contains two steps: AAM building and AAM fitting. AAM can be used in many applications such as hand tracking, face recognition, facial expression analysis, etc. In this thesis, we take face as an example to illustrate the working process of AAM.

The AAM is constructed based on a training set of labeled images, where landmark points are marked on each example face at pre-defined key positions to outline the mean features (shown in Fig.2.16). To ensure the precise location of landmarks, manual labeling is required in the current model building scheme.

The shape of a face is represented by a vector consisting of the positions of the landmarks, $S = (x_1, y_1, x_2, y_2, ..., x_n, y_n)^T$, where (x_j, y_j) denotes the coordinates of the j^{th} landmark points in the 2D facial image. All the shape vectors of the faces in the training set are normalized to a common coordinate system. PCA is applied

Figure 2.16: A labeled training facial image gives a shape free patch and a set of points [Cootes *et al.* 2001].

to this set of shape vectors to construct the face shape model as:

$$S = \overline{S} + P_S B_S \tag{2.20}$$

where S is a shape vector; $\overline{S}$ is the mean shape; P_S is a set of orthogonal modes of shape variation; and B_S is a set of parameters of the shape model.

To construct the appearance model, example faces in the training set are warped to make the manually labeled control points match the mean shape. The warped region of the face image covered by the mean shape is sampled to extract the gray level intensity (texture) information. Similar to the way to build the shape model, a vector representation is generated, $G = (I_1, I_2, ..., I_m)^T$, where I_j denotes the intensity of the sampled pixel in the warped image. PCA is also exploited to construct a linear model:

$$G = \overline{G} + P_G B_G \tag{2.21}$$

where G is a texture vector; $\overline{G}$ is the mean texture; P_G is a set of orthogonal modes of gray-level variation; and B_G is a set of parameters of the gray-level model.

Thus, the shape and texture information of any training face can be summarized by the parameter vectors B_S and B_G, denoted as follows:

$$B = \begin{pmatrix} W_S B_S \\ B_G \end{pmatrix} = \begin{pmatrix} W_S P_S^T (S - \overline{S}) \\ P_G^T (G - \overline{G}) \end{pmatrix} \tag{2.22}$$

where W_S is a diagonal matrix of weights for each shape parameter, as a normalization factor, allowing for the difference in units between the shape and gray scale models. PCA is applied to vectors B, giving a further model:

$$B = QC \tag{2.23}$$

where Q is the eigenvector matrix, and C is a vector of appearance parameters controlling both the shape and gray levels of the model. By the nature of its construction, C has a zero mean across the training set. Figure 2.17 shows the first

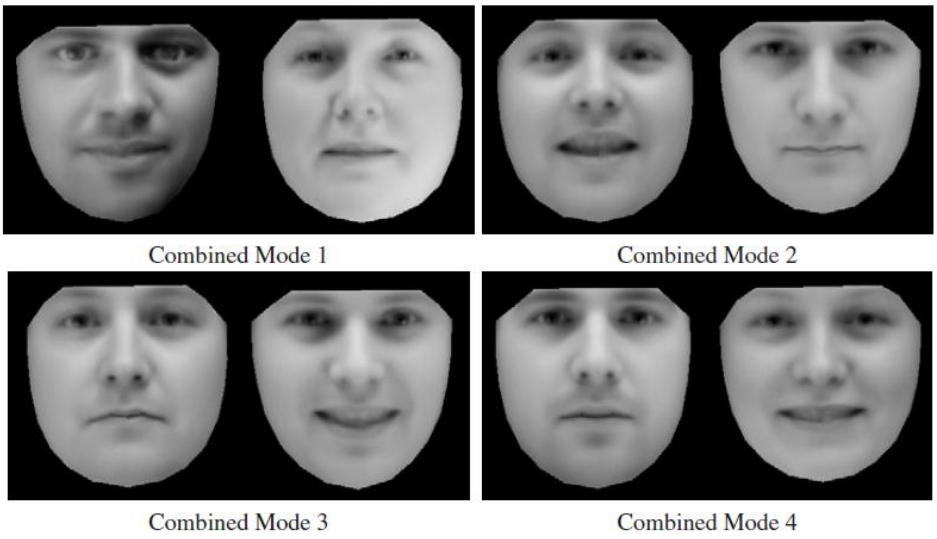


Figure 2.17: Four modes of combined shape and texture model (parameters varied by $\pm 2\sigma$ from the mean) [Li & Jain 2005].

four modes of a combined-appearance model built from 400 facial images of 100 individuals, and the model represents about 20,000 pixels. The modes combine the variation due to lighting, viewpoint, identity, and expression.

Given a test image and the face model generated in the previous step, the metric utilized to measure the match quality between the model and the image is $\Delta = |\delta I|^2$, where δI is the vector of intensity differences between the give facial image and the synthesized image generated by the model tuned by the model parameters, called the residual. The AAM fitting process seeks the optimal set of model parameters that best describes the given image. Cootes et al. [Cootes *et al.* 1998] observed that displacing each model parameter from the correct value induces a particular pattern in the residual. In the training phase, AAM learns a linear model which captures the relationship between parameter displacements and the corresponding residual. During the fitting process, it measures the residual and uses this model to correct the values of current parameters, leading to a better fitting result. After that many attempts were made to improve the performance of AAM [Matthews & Baker 2004] or overcome its limitation, i.e. sensitive to occlusions [Gross *et al.* 2006]. See Fig. 2.18 for an iterative AAM fitting process.

Given a facial image, its corresponding model parameter, C, can be estimated, which best match it with the model (those that synthesize a face as similar as possible to the target face). If the model is sufficiently complex, the parameters should summarize almost all the needed important information to describe the face and can thus be used for face interpretation. In particular, it is possible to adopt the parameters for face verification or identification. By comparing the vectors representing two images, c_1 and c_2, we can measure how similar they are. Experiments [Kang *et al.* 2002] suggest that an effective measure of difference is the normalized dot product as defined in eq. 2.24. The value of zero indicates a perfect match.

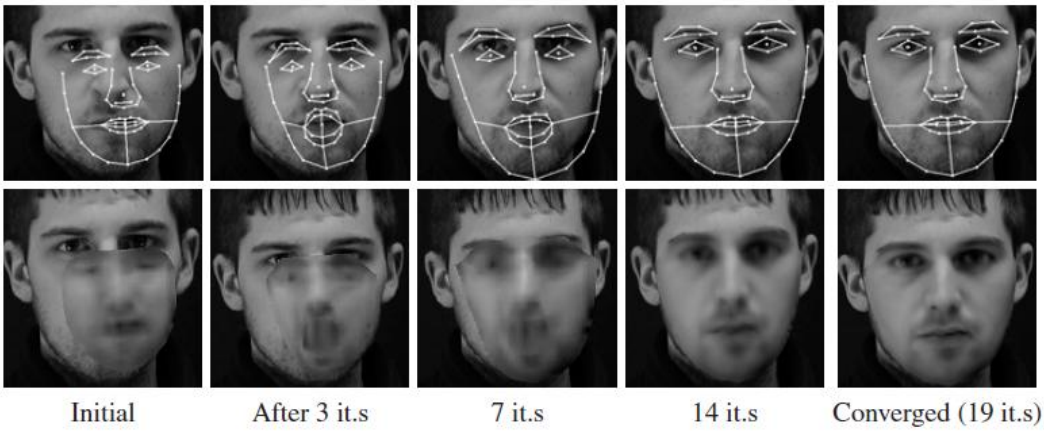


Figure 2.18: An example of AAM fitting iterations [Li & Jain 2005].

This method outperforms other simple metrics such as Euclidean distance or the dot product without normalization [Kang *et al.* 2002]. Meanwhile, LDA can also be exploited to construct the discriminant subspace using the model parameter for face recognition.

$$d = 1 - \frac{c_1}{|c_1|} \cdot \frac{c_2}{|c_2|} \tag{2.24}$$

2.1.4 Discussion

In this section, several representative algorithms belonging two main categories, i.e. holistic techniques as well as local feature based ones in intensity image based face recognition have been reviewed. Comparing these two categories, holistic methods tend to be less sensitive to noise, however, the local feature based approaches have shown their promising performance in recent years. It has been proved by Heisele et al. [Heisele *et al.* 2003] that the component-based face recognition approaches (local feature-based) perform better than global ones (holistic). The main reason is that holistic approaches always require facial images to be accurately normalized with regard to the factors of pose, illumination condition and scale. In addition, global features are also more sensitive to facial expression variations and occlusions. According to its theoretical nature, AAM can also be deemed as a special example of holistic techniques, and it thus possesses both their advantages and drawbacks in face recognition. Because local feature-based techniques extract features from local points or patches, there generally remain some invariant features even in the presence of facial expression variations or occlusions, and recognition can still be done by matching these invariant features. Therefore, local feature-based methods are potentially more robust than the holistic ones to facial expression changes and occlusions. Moreover, unlike the holistic approaches, they require fewer samples for enrollment, and can even achieve an excellent accuracy only with a single facial image registered in the gallery set [Tan *et al.* 2006]. There also exist several tasks,

combining holistic techniques with local feature-based ones and aiming to fuse their advantages, but at the same time they inherits their downsides from both types of categories as well.

One would further raise the question which is the best local descriptor for face recognition among LBP, SIFT and Gabor filters. Luo et al. [Luo *et al.* 2007] showed that SIFT is not as robust as LBP to illumination effects on the FERET dataset. Zou et al. [Zou *et al.* 2007a] compared Gabor wavelets and LBP using the same database, and concluded that Gabor wavelets are more insensitive to illumination changes because they detect amplitude-invariant spatial frequencies of gray values of image pixels, while LBP is greatly affected by non-monotonic gray value transformations. Ruiz-del-Solar et al. [del Solar *et al.* 2009] evaluated these three methods extensively not only on controlled datasets, e.g. FERET and FRGC, but also on the UCH FaceHRI database designed for Human-Robot Interaction and the LFW dataset captured in unconstrained environments. Concerning robustness to illumination variations, their study illustrated that Gabor wavelets achieve the best performance on the FERET database, LBP is not far behind; while SIFT is the last, thus supporting the previous conclusions. On the UCH FaceHRI database, the LBP approach gains the best results in all the specially-designed experiments with indoor and outdoor lighting, expression, scaling, and rotation, followed by Gabor wavelets and SIFT. On the LFW dataset, LBP and Gabor wavelets obtain a slightly better result than each other with aligned face and non-aligned face respectively, both of which surpassed that of SIFT. On the other hand, in their investigation on computation cost, LBP operates much faster than Gabor wavelet and SIFT. Based on the above facts, it is difficult to draw a general conclusion to rank them, since their performance is decided by a number of complicated factors. Which one should be selected for recognizing faces depends on the requirements of applications. Finally, their possible combinations probably contribute more to this community.

2.2 Geometric Information based 3D Face Recognition

Recently, adopting 3D face data has emerged as a major alternative to deal with the unsolved issues in 2D face recognition, i.e. lighting and pose variations, and various approaches have been investigated within this field. Table 2.1 lists a comprehensive overview of only 3D shape based face recognition methods investigated in the past twenty years. As demonstrated in Table 2.1, the early stage of development in 3D face recognition started from 1989 and last until 2000. During that period, there were only a few scientists and researchers who was dedicating in this domain, leading to very limited amount of academic studies which used simple techniques on small databases. The traditional algorithms originally proposed for 2D face recognition (e.g. PCA, HMM [Achermann *et al.* 1997], and Hausdorff Distance [Achermann & Bunke 2000]) were explored, and 3D points, profiles or local areas were also attempted for face matching. Meanwhile, we can notice that some features such as Extended Gaussian Image (EGI), curvature, and point signature, specific in 3D face recognition, came into use. Since 2003, scientific interest has been increased significantly, and the number of related papers in this topic remains over 10 every

year till 2010. Especially in 2005, due to the workshop on FRGC in conjunction with CVPR, this number reaches about 30. The databases exploited for technique evaluation have been greatly enlarged as well. After 2003, more and more research organizations begin to devote into this topic, and active ones include the biometrics group at MSU (USA) [Lu *et al.* 2004a, Lu *et al.* 2004b, Lu & Jain 2005a, Lu & Jain 2006, Lu & Jain 2008]; Bronstein brothers at Technion (Israel) [Bronstein *et al.* 2004b, Bronstein *et al.* 2005, Bronstein *et al.* 2007]; Gang Pan together with co-workers at Zhejiang University (China) [Pan *et al.* 2003c, Pan *et al.* 2003a, Pan *et al.* 2003b, Pan *et al.* 2005, Wang *et al.* 2007b]; the NLPR at CAS-IA (China) [Xu *et al.* 2004, Zhong *et al.* 2007, Huang *et al.* 2007, Zhong *et al.* 2008]; the CVRL at UND (USA) [Bowyer *et al.* 2006, Chang *et al.* 2005a, Chang *et al.* 2006, Faltemier *et al.* 2008a, Faltemier *et al.* 2008b]; Ajmal Mian's team at UWA (Australia) [Mian *et al.* 2005, Mian *et al.* 2006a, Mian *et al.* 2007, Mian *et al.* 2008]; and the CBL at UH (USA) [Passalis *et al.* 2005, Passalis *et al.* 2007, Kakadiaris *et al.* 2007]; etc. More classic algorithms renowned in 2D face recognition have been investigated for direct use to recognize 3D faces, such as LDA [Gokberk *et al.* 2005], LBP [Li *et al.* 2005b] and Gabor filters [Xu *et al.* 2009], and besides them, many methods have been designed diversely to employ 3D face data. The taxonomy (holistic and local feature based) previously exploited in 2D intensity image based face recognition can be expanded to classify the techniques of 3D geometric information based recognition. Holistic matching directly operates on 3D face data (range images, meshes or point-clouds) to compute similarity scores, including Iterative Closest Point (ICP) based matching [Lu *et al.* 2006], Hausdorff distance [Achermann & Bunke 2000] and so forth. While local feature-based matching compares the local descriptive points, curves, or regions of 3D facial scans in a certain feature space (coordinates, areas, distances, curvatures, shape index, as well as some more complex ones) and then combines them for decision. This category has been widely explored in several tasks in the literature, for instance, the point signature approach [Chua *et al.* 2000], 3D coordinate based descriptive pose invariant features [Mian *et al.* 2008], the region ensemble method [Faltemier *et al.* 2008a], iso-geodesic strips [Berretti *et al.* 2010], etc. In this section, we chronologically retrospect the techniques used in 3D face recognition, and then discuss them according to the aforementioned taxonomy.

Table 2.1: A comprehensive overview of face recognition algorithms solely using 3D shape information (Note: the dataset column uses the format A:B, where A denotes the number of individuals and B denotes the total number of 3D models. The sign $*$ means 3D model sequence).

Reference	Dataset	Key Feature	Matching
[Cartoux *et al.* 1989]	5:18	Profile	Minimum Dist.
[Lee & Milios 1990]	6:6	EGI	Correlation
[Gordon 1991]	8:24	Depth & Curvature	NN
[Gordon 1992]	8:24	Depth & Curvature	NN
[G. G. Gordon 1992]	8:24	Depth & Curvature	NN
[Nagamine *et al.* 1992]	16:160	Multiple profile	NN

Table 2.1

[Tanaka & Ikeda 1996]	37:37	EGI	Corr.
[Achermann *et al.* 1997]	24:240	Range Image	PCA & HMM
[Tanaka *et al.* 1998]	37:37	EGI	Correlation
[Achermann & Bunke 2000]	24:240	Point Coordinate	Hausdorff Dist.
[Chua *et al.* 2000]	6:24	Rigid region	Point Signature
[Beumier & Acheroy 2000]	30:180	Curvature of Profile	Minimum Dist.
[Hesher *et al.* 2003]	37:222	Range Image	PCA
[Lee *et al.* 2003]	35:70	Re-sampled Local Area	NN
[Medioni & Waupotitsch 2003]	100:700	Point Coordinate	ICP
[Moreno *et al.* 2003]	60:420	Region Shape Fea. Set	NN
[Pan *et al.* 2003c]	30:360	Point Coordinate	Hausdorff Dist.
[Pan *et al.* 2003a]	30:180	Profile	Hausdorff Dist.
[Pan *et al.* 2003b]	30:180	Profile & Depth Map	Weighted Hausdorff
[Wu *et al.* 2003]	30:180	Profile	Partial Hausdorff
[Bronstein *et al.* 2004b]	7:-	Point Coordinate	Geodesic Dist.
[Lee & Shim 2004]	42:84	Curvature	Weighted Hausdorff
[Lu *et al.* 2004a]	10:73	Point Coordinate	ICP
[Lu *et al.* 2004b]	18:131	Point Coordinate	ICP
[Russ *et al.* 2004]	200:468	Range Image	Hausdorff Dist.
[Xu *et al.* 2004]	120:720	Gaussian-Hermite	Minimum Dis.
[Wu *et al.* 2004]	6:31	Local Shape Map	Correlation
[Heseltine *et al.* 2004b]	100:330	Curva. & Grad. Map	PCA
[Heseltine *et al.* 2004a]	280:1770	Curva. & Grad. Map	LDA
[Irfanoglu *et al.* 2004]	30:90	Coordinate with TPS	Point Set Dist.
[Bronstein *et al.* 2005]	30:220	Geodesic Distance	Minimum Dist.
[Chang *et al.* 2005a]	466:4007	Multi-Region	ICP
[Chang *et al.* 2005b]	355:3205	Local Region	ICP
[Gokberk *et al.* 2005]	106:579	Multiple Shape Feature	Multi-Classifier
[Lee *et al.* 2005]	100:200	Curve, Dis., Angle etc	DP & SVM
[Lu & Jain 2005a]	100:296	Coordinate with TPS	ICP
[Pan *et al.* 2005]	276:943	Conformal Depth Map	PCA
[Passalis *et al.* 2005]	466:4007	AFM & Wavelet Coe.	Minimum Dis.
[Russ *et al.* 2005]	200:398	Range Image	Hausdorff Dist.
[Abate *et al.* 2005]	133:-	Normal Map	Difference Hist.
[Koudelka *et al.* 2005]	276:943	Curve and Region	Hausdorff Dist.
[Li *et al.* 2005a]	80:160	Profile and Contour	Mahalanobis Dist.
[Mian *et al.* 2005]	277:519	Dis & Angle Tensor	Correlation
[Moreno *et al.* 2005]	60:420	Geometric Attribute	PCA & SVM
[Samir *et al.* 2005]	12:144	MS. Reeb Graph	Graph Matching
[Uchida *et al.* 2005]	18:72	Point Coordinate	ICP
[Sun & Yin 2005]	25:72	GA Selected Curvature	Correlation
[Abate *et al.* 2006b]	120:1200	Normal & Fourier Coe.	Minimum Dist.
[Ben Amor *et al.* 2006a]	50:450	Rigid Region	Region ICP
[Ben Amor *et al.* 2006b]	50:450	Point Coordinate	ICP
[Antini *et al.* 2006]	61:427	Wrinkle Curve	Graph Matching
[Chang *et al.* 2006]	449:4000	Nose Point Coordinate	ICP
[Feng *et al.* 2006]	35:175	Curve	Jensen-Shannon Diverge.
[Levine & Rajwade 2006]	277:953	Point Coordinate	SVM-ICP
[Lin *et al.* 2006]	276:943	Curve	NN
[Lu & Jain 2006]	50:150	Deformation Synthesis	ICP
[Samir *et al.* 2006]	162:740	Level Curve Union	Differential Geodesic

Table 2.1

[Wang *et al.* 2006]	40:360	Point Coordinate	Partial ICP
[Zhang *et al.* 2006]	166:382	Profile Curvature	Minimum Dist.
[Huang *et al.* 2006]	466:4007	Multi. Feature Fusion	Minimum Dist.
[Cook *et al.* 2006a]	466:4007	Log-Gabor	Mahalinobis Cosine Dist.
[Bronstein *et al.* 2007]	4:104	Geodesic Distance	Spherical Embedding
[Cook *et al.* 2007]	466:4007	Weighted Range Patch	PCA
[Gupta *et al.* 2007a]	466:4007	Geodesic Dist.	LDA
[Gupta *et al.* 2007b]	12:360	Complex Wavelet	Stru. Similarity Metric
[Hwang *et al.* 2007]	97:970	Range Image	PCA
[Passalis *et al.* 2007]	466:4007	Wavelet on Depth Map	Minimum Distance
[Kakadiaris *et al.* 2007]	466:4007	Geometry & Normal Map	CW-SSIM
[Li & Zhang 2007]	90:480	Geometric Attribute	NN
[Wang *et al.* 2007b]	353:1891	Deformable Model	ICP
[Wang *et al.* 2007a]	10:100	Conformal Map	Correlation
[Wei *et al.* 2007]	100:600∗	GA selected Curvature	Correlation
[Zhong *et al.* 2007]	466:4007	Learned Gabor Filter	NN
[Zou *et al.* 2007b]	113:650	Warping Coe.	Mahalanobis Dist.
[Mahoor & Abdel-Mottaleb 2007]	61:183	Rigid Line	Hausdorff Dist.
[Lin *et al.* 2007]	466:4007	Patch based Curve	Weighted LDA
[Faltemier *et al.* 2007]	888:13450	Multiple Enrollment	ICP
[Huang *et al.* 2007]	466:4007	Depth Difference	Hist. Proportion
[McCool *et al.* 2008]	557:4950	2D-DCT of Depth Block	GMM
[Guan & Zhang 2008]	61:427	Angle & Region Map	NN
[Berretti *et al.* 2008]	61:427	Iso-Geodesic Stripes	Graph Matching
[Alyuz *et al.* 2008]	47:1507	Curvature & Depth Map	AFM-based Region ICP
[Faltemier *et al.* 2008a]	466:4007	Region Ensemble	ICP
[Amberg *et al.* 2008]	277:953	3D Morphable Model	Non-rigid ICP Fitting
[Castellani *et al.* 2008]	80:240	Depth Local Region	HMM
[Jahanbin *et al.* 2008]	119:1196	Iso-Dep. & Geo. Curve	NN & SVM
[Wang *et al.* 2008]	466:4007	Local Shape Difference	Boosting
[Zhong *et al.* 2008]	434:1357	Local Log-Gabor Hist.	NN
[Lu & Jain 2008]	100:977	Deformation Synthesis	ICP
[Llonch *et al.* 2008]	277:953	Spherical Coe.	PCA & LDA
[Queirolo *et al.* 2008]	466:4007	Multi-Region	SA-based SIM
[Faltemier *et al.* 2008b]	888:13450	Multiple Enrollment	ICP
[Mayo & Zhang 2009]	61:549	Multi-View Depth Map	SIFT Matching Voting
[Daniyal *et al.* 2009]	100:2500	Inter-Landmark Dist.	LDA
[Li *et al.* 2009]	120:600	Geometric Attribute	Sparse Representation
[Al-Osaimi *et al.* 2009]	466:4007	Expression Deformation	PCA
[Smeets *et al.* 2009]	100:900	Iso-Geodesic Dist.	Mahalanobis Dist.
[Ouji *et al.* 2009]	50:300	Point Coordinate	Geodesic ICP
[Mahoor & Abdel-Mottaleb 2009]	466:4007	Rigid Line	Hausdorff Dist. & ICP
[Drira *et al.* 2010]	61:549	Elastic Radial Curve	Geodesic Deformation
[Llonch *et al.* 2010]	277:953	Spherical Coe.	LDA
[Queirolo *et al.* 2010]	466:4007	Multi-Region	SA-based SIM
[Berretti *et al.* 2010]	466:4007	Iso-Geodesic Stripes	Graph Matching
[Wang *et al.* 2010]	466:4007	Local Shape Difference	Boosting
[Maes *et al.* 2010]	105:4666	Mesh-SIFT	Minimum Matched Points
[McCool *et al.* 2010]	557:4950	2D-DCT of Depth Block	HMM
[Alyuz *et al.* 2010]	466:4007	Multi-Region Curvature	LDA

2.2.1 Holistic Matching

Similar to the 2D intensity image based holistic approaches which exploit the entire 2D facial image as system input, 3D geometric information based holistic matching works directly on the whole range image, mesh or point-cloud of the face to calculate similarity measurement.

(1) Subspace (PCA and LDA)

The subspace based methods are generally applied to facial range images. See the last section for more of technical details. Achermann et al. [Achermann *et al.* 1997] apply the eigenface approach to 3D face recognition. They present the results for a dataset of 24 subjects each of which has 10 images, and report 100% recognition rate. Later, Hesher et al. [Hesher *et al.* 2003] explore PCA using different numbers of eigenvectors and image sizes. The image data used has 6 different facial expressions for each of 37 persons. The performance results from using multiple facial images per person in the gallery set, which effectively gives the probe image more chances to make a correct match, and is known to raise the recognition accuracy relative to having a single gallery sample per person. Heseltine et al. [Heseltine *et al.* 2004b] evaluate the eigenface approach with different distances as similarity measurement for face verification on a larger database containing 330 facial range images from 100 individuals and conclude that the lowest Equal Error Rate (EER) of 17.8% is achieved by Mahalanobis distance, better than Euclidean and Cosine distances. They [Heseltine *et al.* 2004a] further enlarge the test set to 1470 range images of 230 individuals and exploit LDA in the same framework, claiming that LDA performs better than PCA, and with Cosine distance, LDA obtains the lowest EER of 15.3%.

(2) Iterative Closest Point (ICP)

The ICP [Besl & McKay 1992] algorithm iteratively attempts to align two 3D surfaces represented as point-clouds or meshes. To accomplish this task, ICP first finds the closest point in the reference surface for each of the n points in the probe one. When implemented with a k-d tree [Bentley 1975] data structure, nearest neighbor searches can each be completed in $O(log(n))$ time. Beginning with a starting estimate, the algorithm calculates a sequence of rigid transformations T_i until there is no additional improvement in mean square distance between the two shapes. Specifically, given two point-clouds or meshes, the reference surface $M \triangleq \{m_i\}_{i=1}^{N_m}$ with N_m points and the query surface $P \triangleq \{p_i\}_{i=1}^{N_p}$ with N_p points, the transformation of the query shape to the reference one is assumed to be linear with a rotation matrix R and translation vector t. The goal of the ICP algorithm is to find the transformation parameters, for which the error between the transformed query shape points and the closest points of the reference shape achieves minimal. This characteristic is described by eq. 2.25

$$\min_{R,t,j\in\{1,2,\ldots,N_m\}} \left(\sum_{i=1}^{N_p} \|Rp_i + t - m_j\| \right) \tag{2.25}$$

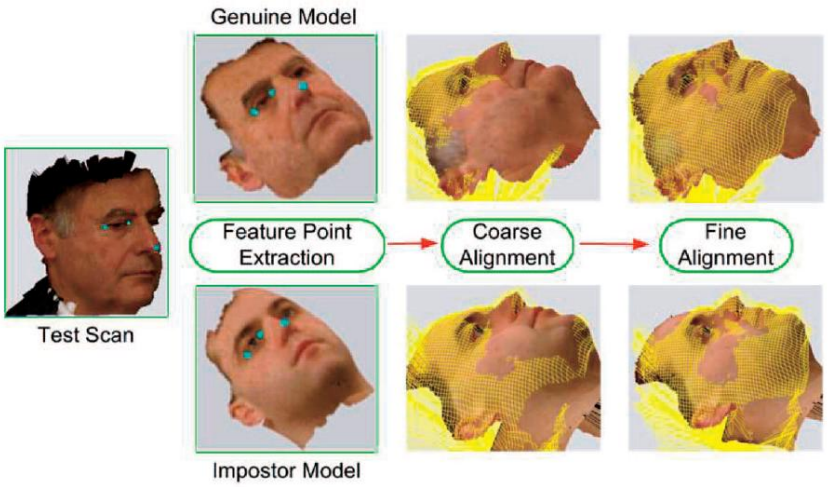


Figure 2.19: An ICP based coarse to fine registration process of 3D facial surfaces [Lu *et al.* 2006].

If the initial position is not good enough, ICP probably finds a local minima that corresponds to an improper registration. For this reason, the probe and gallery centroids of 3D facial surfaces are frequently aligned before ICP is performed to ensure an accurate final registration estimate.

ICP is widely utilized for face registration and matching in 3D face recognition. Medioni and Waupotitsch [Medioni & Waupotitsch 2003] perform 3D face recognition by using ICP to match facial surfaces. Whereas most of the works covered here use 3D shapes acquired through an active (structured-light) sensor, this work uses 3D shapes acquired by a passive stereo sensor. The database contains 100 subjects, each of which possesses 7 images sampling different poses. An EER of "better than 2%" is reported. Lu et al. [Lu *et al.* 2004b] present the results of face recognition based on an ICP-based approach. This approach assumes that the gallery face is a more complete 3D facial surface and the probe face is a frontal view that is likely a subset of the gallery face. In the experiments with models from 18 persons and multiple probe ones per person, incorporating some variation in pose and expression, a recognition rate of 97% is achieved. Wang et al. [Wang *et al.* 2006] modify ICP as Partial ICP, which selects a part of nearest point pairs to calculate dissimilarity measure during registration of facial surfaces, to reduce the negative effects caused by facial expression changes. The experiments on a dataset of 360 face models from 40 individuals show that the modified Partial ICP surpasses the original ICP. Besides used to compute similarity measurements, in most of the recent works, ICP is extensively introduced as a key step in preprocessing to correct 3D pose variations as in [Kakadiaris *et al.* 2007], [Li *et al.* 2009], etc. Figure 2.19 depicts an ICP based coarse to fine alignment process of 3D facial surfaces.

(3) Hausdorff Distance

The Hausdorff distance is a measure for the similarity of two sets of points [Huttenlocher *et al.* 1993]. It is very general and may be applied to a wide variety of problems.

The (undirected) Hausdorff distance between two sets of points, A and B, is defined as

$$H(A,B) = \max(h(A,B), h(B,A)) \tag{2.26}$$

where $h(A,B)$ denotes the directed Hausdorff distance,

$$h(A,B) = \max_{a \in A} \min_{b \in B} \|a - b\| \tag{2.27}$$

and $\|x\|$ is a norm. Usually, the well known Euclidean distance is utilized here.

The partial Hausdorff distance is defined as

$$H_{LK}(A,B) = \max(h_L(A,B), h_K(B,A)) \tag{2.28}$$

where the directed partial Hausdorff distance is

$$h_L(A,B) = L^{th}_{a \in A} \min_{b \in B} \|a - b\| \tag{2.29}$$

We notice that $h(A,B)$ is small, if every point in A is near some point in B. The same holds true for $h(B,A)$. Moreover, $H(A,B) = H(B,A)$ is small, if both $h(A,B)$ and $h(B,A)$ are small. In the partial Hausdorff distance, only the L and K best ranked points in sets A and B are taken into regard, respectively. Thus, it is possible to ignore outliers and handle cases where only part of the data fits a given model, or part of a model fits a given set of data points.

The Hausdorff distance was originally used for sets of points in the 2D plane [Huttenlocher *et al.* 1993]. However, eq. 2.26 to 2.29 can be applied to points in 3D space as well, and it is necessary to preregister the two sets under comparison, i.e., to find a translation and rotation that optimally aligns the two data sets.

Achermann and Bunke [Achermann & Bunke 2000] make use of an extension of Hausdorff distance based matching for 3D face recognition. They report on experiments using 240 range images, 10 images of each of 24 persons, and achieve 100% recognition for some instances of the algorithm. Pan et al. [Pan *et al.* 2003c] compare a Hausdorff distance approach and a PCA-based one in 3D face recognition. In experiments with images from the 3D_RMA database they report an EER in the range of 3% to 5% for the Hausdorff distance approach lower than that in the range of 5% to 7% for the PCA-based approach. Lee and Shim [Lee & Shim 2004] propose a "depth-weighted Hausdorff distance" to calculate the similarity score which is then fused with surface curvature information (the minimum, maximum, and Gaussian curvature) for 3D face recognition. They present the results of experiments with a dataset representing 42 persons, with two images for each person. A rank-one recognition rate is up to 98% for the best combination method investigated, whereas the plain Hausdorff distance achieves less than 90%. Russ et al. [Russ *et al.* 2004] also explore the Hausdorff matching on facial range images. In a verification experiment,

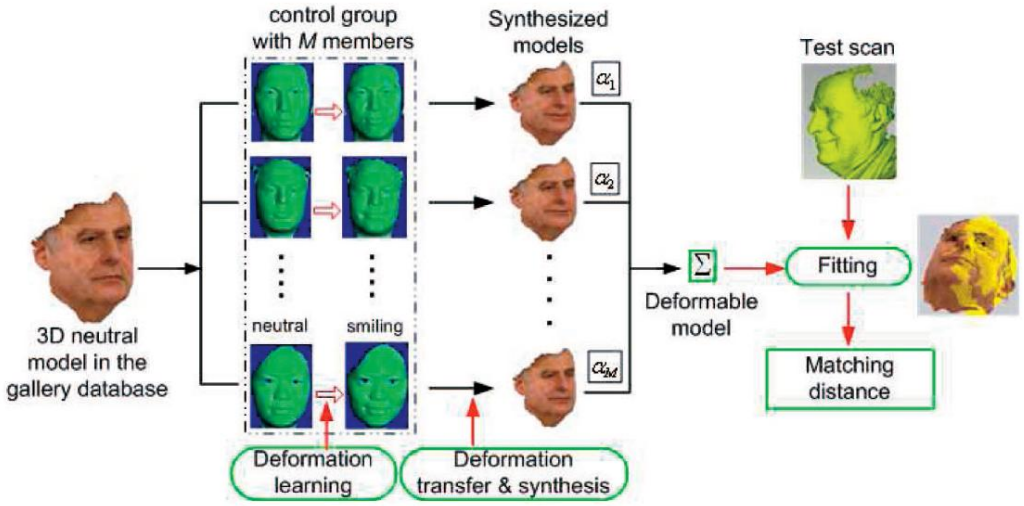


Figure 2.20: The deformation synthesis and transfer based 3D face recognition [Lu & Jain 2008].

200 persons are enrolled in the gallery, and the same 200 persons plus another 68 imposters are represented as probes. A probability of correct verification as high as 98% (of the 200) is achieved at a false alarm rate of 0 (of the 68). In a recognition experiment, 30 persons are enrolled as gallery samples and the same 30 persons imaged at a later time are used in the probe set. A 50% probability of recognition is achieved. The recognition experiment uses a subset of the available data "because of the computational cost of the current algorithm" [Russ *et al.* 2004].

(4) Deformable Model

Human faces are non-rigid, and facial expression variations can dramatically deform facial surfaces, which is quite challenging to holistic face matching mentioned above, i.e. Subspace, ICP as well as Hausdorff. Deformable model is proposed to be incorporated in holistic matching, aiming to better describe the non-rigid transformations of faces. Based on a training step using representative facial samples, the deformable model can be used for:

- learning the deformation quantity between various facial expressions and thus indicating whether it belongs to intra-class variations;
- learning the transformations between different facial expressions to enlarge the gallery set with more expressive face models or to convert an expressive probe face to a neutral one for recognition.

Lu and Jain [Lu & Jain 2005a] extend their previous work using an ICP based recognition approach [Lu *et al.* 2004b] to deal explicitly with variation in facial expression. The problem is solved as a rigid transformation of probe to gallery, done with ICP, along with a non-rigid deformation, done using Thin-Plate Spline (TPS) techniques. The proposed approach is evaluated using a 100-person dataset, with neutral-expression and smiling probes, matched to neutral-expression gallery images. The gallery entries are whole-head data structures, whereas all the probes are

frontal views. Most errors after the rigid transformation result from smiling probes, and these errors are reduced substantially after the non-rigid deformation stage. For the total 196 probes (98 neutral and 98 smiling), the rank-one recognition rate reaches 89%. Lu and Jain [Lu & Jain 2006, Lu & Jain 2008] also propose another 3D face recognition method which is robust to facial expression variations. Based on the information learned using the control group, to each subject, the approach can generate different expressive face models from the neutral one in the gallery set. A user-specific 3D deformable model is then built for each subject in the gallery with respect to the control group by combining the templates with synthesized deformations. By fitting this generative deformable model to a test facial scan, face recognition can be achieved. Figure 2.20 gives an illustration of the entire framework. A small control group consisting of 10 subjects, each with seven expressions (neutral, happy, angry, smile, surprise, deflated, and inflated), is collected from the MSU dataset. While the test set contains 877 scans of 100 individuals selected from FRGC v2.0, and the experimental results show that the proposed approach is able to handle expressions changes to some extent. Wei et al. [Wei *et al.* 2007] investigate to adopt 3D model sequences for identification. They conduct a fine adaptation procedure [Yin *et al.* 2006a] in order to deform the tracking model into a non-rigid facial surface area, which is realized using the energy minimization method based on the dissimilarity (error function) between the tracking model and the face scan. As a result, the 3D motion trajectories are estimated by vectors from the tracked points of the current frame to the corresponding points of the first frame with a neutral expression. The mean of the features of all deformed frames within one sequence is used as final representation for classification. This method is evaluated on 600 3D model sequences of 100 persons, the achieved accuracy is 90.7%. Wang et al. [Wang *et al.* 2007b] introduce a guidance-based constraint deformation model to cope with the shape distortion by expression. The basic idea is that, the face model with non-neutral expression is deformed toward its neutral one under certain constraint so that the distortion is reduced while inter-class discriminative information is preserved. This model exploits the neutral 3D face shape to guide the deformation, meanwhile brings in a rigid constraint on it. Both the steps are smoothly unified in the Poisson equation framework. The experimental results, calculated from an expressive subset of FRGC v2.0, demonstrate that this method outperforms ICP, and an EER of 6.2% is reported. Zou et al. [Zou *et al.* 2007b] employ a number of selected range images as example faces, and another range image is chosen as a "generic face". The generic face is then warped to match each of the example faces in the least mean square sense. Each warp is specified by a vector of displacement values. For feature extraction, when a target face image comes in, the generic face is warped to match it. The geometric transformation used in the warping is a linear combination of the example face warping vectors. The coefficients in the linear combination are adjusted to minimize the root mean square error, which are further used as facial features and fed to a Mahalanobis distance based classifier. The authors claim that this method is insensitive to holes, facial expression, and hair. Amberg et al. [Amberg *et al.* 2008] describe an expression-invariant method for face recognition by fitting an identity/expression separated 3D Morphable Model (3DMM) [Blanz & Vetter 1999] to probe facial surface. The fitting is performed with a ro-

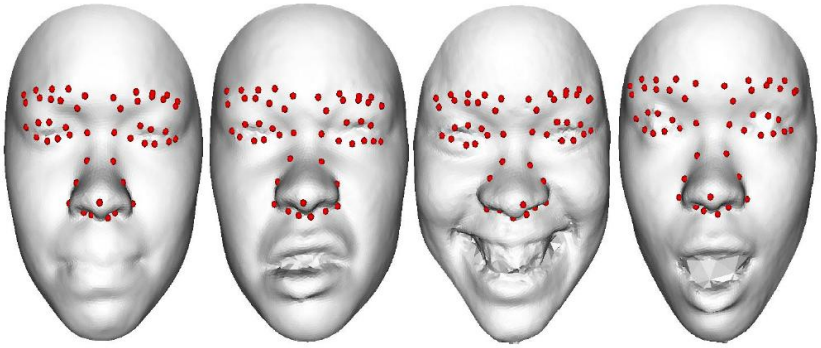


Figure 2.21: Examples of landmark fitting results on faces with different expressions [Daniyal *et al.* 2009].

bust non-rigid ICP algorithm [Amberg *et al.* 2007]. The experiments carried out on the Gavab and UND datasets illustrate that the proposed expression based 3DMM helps improve the ICP based face recognition approach in terms of the accuracy. Al-Osaimi et al. [Al-Osaimi *et al.* 2009] learn the patterns of expression deformations from training data in PCA eigenvectors, and these patterns are then used to morph out the expression deformations. PCA is performed in such a way it models only the facial expressions leaving out the interpersonal disparities. Similarity measures are extracted by matching the morphed 3D faces. The approach is applied on the FRGC v2.0 dataset and achieves the verification rates of 98.35% and 97.73% at 0.001 FAR for scans under neutral and non-neutral expressions, respectively.

2.2.2 Local Feature based Matching

The techniques of local feature based matching concentrate on local descriptive points, curves, or regions of 3D facial scans and compare them in a certain feature space (e.g. original coordinates, areas, distances, angles, curvatures, and even more complex ones) for final decision. This approach has been extensively investigated in 3D face recognition, since it generally achieves better results than holistic matching does. In this subsection, according to the utilized feature, we categorize local feature based matching into five classes: i.e. point based, curve based, region based, general descriptor based as well as multiple shape feature fusion based.

(1) Point based Methods

Point based local matching extracts distinctive information from a set of original points for matching, such as the original coordinate, the distance of an edge between two points, the area of a region composed of several points, the angle between two edges, or more complicated descriptors.

Wu et al. [Wu *et al.* 2004] propose Local Shape Map (LSM) to describe a point on the facial surface, and it is represented as a 2D histogram constructed from mapping 3D coordinates of surface's points within a sphere centralized at this point into a 2D space. The correlation coefficient is utilized as similarity metric to compare the LSM of every two points, and all the results are then incorporated into an ad-hoc voting method for classification. This method does not require to register two facial

surfaces. With a database of 31 facial range images from 6 subjects, the EER of 2.98% is obtained, and it also proves that LSM outperforms spin image in 3D face recognition tasks. Mian et al. [Mian *et al.* 2005] define the vertex pair according to a distance and angle constraint, and each valid pair is used to construct a third order tensor, which records the surface area of the face crossing each bin of the given cubic grid. The gallery is built by representing each facial image with multiple tensors, while in recognition, tensors are computed from the probe and then utilized to cast votes to the tensors in the gallery set. A similarity measurement is further calculated between each probe tensor and with only those gallery tensors which receive votes above a certain threshold. The top 10 faces with the highest similarities are registered to the probe face by aligning their matching pair of tensors, and this registration is refined with ICP to finally compute the similarity measurement. They report a rank one recognition rate of 86.4% on the UND dataset comprising 953 probe and gallery faces of 277 subjects. Gupta et al. [Gupta *et al.* 2007a] manually labeled 25 anthropometric fiducial points and calculate the Euclidean as well as geodesic distance as facial features, which are then fed into LDA for final decision. The experiments are carried out on a dataset made up of 1128 range images from 105 subjects, showing that geodesic based result (1.3% EER) is better than Euclidean based one (1.6% EER). The accuracies using arbitrary regularly spaced facial points illustrate that incorporating domain specific knowledge about the structural diversity of human faces significantly improves the performance of 3D face recognition. Castellani et al. [Castellani *et al.* 2008] introduce a generative learning method by adapting HMM to work on 3D meshes. The geometry of local area around face fiducial points is modeled by training HMMs providing a robust pose invariant point signature. Such description allows the matching by comparing the signature of corresponding points in a maximum-likelihood principle. The matching information of corresponding points are finally imported in a voting method to give the classification result. According to the preliminary performance on a small subset of FRGC v2.0, the authors claim that the proposed approach is robust to facial expression variations and works faster than the current ICP based 3D face recognition systems by maintaining a satisfactory recognition rate. Daniyal et al. [Daniyal *et al.* 2009] train a Point Distribution Model (PDM) using manual landmarks on 100 sample models, and automatically locate corresponding points on testing facial surfaces. All the Euclidean distances between every two points are exploited as features and classified by LDA for 3D recognition. The proposed approach is evaluated on the BU-3DFE database and achieves 96.5% recognition rate. Figure 2.21 gives an example of automatic landmarking results. Mayo et al. [Mayo & Zhang 2009] rotate each 3D point-cloud to represent a face around the x, y or z axes, iteratively and to project the 3D points onto multiple facial range images at each step of the rotation. Labelled keypoints are then extracted from the resulting collection of range images to describe the original facial scan and its projections in the face database. Unknown test faces are recognized by performing the same multiview keypoint extraction technique as well as weighted keypoint matching algorithm. Evaluated on the entire GavabDB database, the approach achieves up to 95% recognition rate for faces with a neutral expression only, and over 90% accuracy when expressions and random gestures appear. Maes et al. [Maes *et al.* 2010] extend the SIFT algorithm

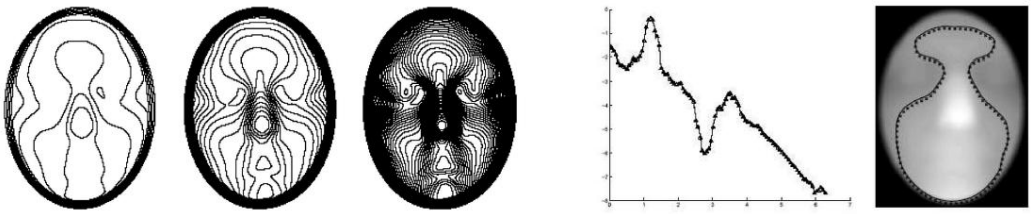


Figure 2.22: (a) Level curves of depth function for several levels; (b) Angle functions, observed (marked) and fitted (solid) for a level curve, and the corresponding curves on the facial range image [Samir *et al.* 2006].

to 3D mesh, namely mesh-SIFT. Unlike the gradient information extracted by the original SIFT operator, mesh-SIFT encodes the local histogram of shape index values. Tested on the whole Bosphorus dataset, it achieves a rank-one recognition rate of 93.7%.

(2) Curve based Methods

Curve based local matching compares corresponding curves from different 3D facial surfaces or computes similarity measurements using more robust features extracted from them.

Cartoux et al. [Cartoux *et al.* 1989] solve the problem of 3D face recognition by segmenting a facial range image based on principal curvature and finding a plane of bilateral symmetry through the face, which is used to correct pose variations. They consider approaches of matching the profile from the symmetry plane as well as of matching the facial surface, and report 100% recognition rate for either on a quite small dataset. Nagamine et al. [Nagamine *et al.* 1992] approach 3D face recognition by locating five feature points and using them to standardize face pose, matching is then operated based on various curves or profiles along the face data. Experiments are performed for 16 persons, with 10 images per person. The best recognition rates are achieved using vertical profile curves that pass through the central portion of the face. Wu et al. [Wu *et al.* 2003] introduce the method [Pan *et al.* 2003b] to extract the symmetry profile using along with two horizontal ones across forehead and nose respectively. The partial Hausdorff distance is employed to align and compare profiles. For each individual, a statistical model of facial surface is built to represent the distinct discriminative capability of the different parts in the facial surface, which is then incorporated into a weighted distance function to measure similarity of surfaces. An EER in the range of 1% to 5% is reported on the 3D_RMA database. Li et al. [Li *et al.* 2005a] recognize 3D face models based on the characterization of their contours and profiles. Experiments show that the central vertical profile and the contour are both very useful features for face recognition. When combined, better recognition rates can be obtained than just using any of them alone, meanwhile the performance of this method is better than that of PCA using a database of 80 subjects. Antini et al. [Antini *et al.* 2006] encode basic traits of a face by extracting

curves of salient ridges and ravines from the surface of a dense mesh. A compact graph representation is then constructed by these curves through an original modeling technique capable to quantitatively measure spatial relationships between curves in a 3D space for face matching. Experimental results on the Gavab database show that the proposed solution attains high recognition accuracy (91%) on neutral probes relative to the ones with expression and pose changes. Feng et al. [Feng *et al.* 2006] extract affine integral invariant from a set of manual landmark based facial curve, and a small subset is chosen for pose robust facial representation. The recognition procedure is based on the Discriminant Analysis and Jensen-Shannon Divergence analysis. Substantiating examples are provided with an achieved classification accuracy of 92.57% on a database of 175 facial range images from 35 individuals. Similarly, Lin et al. [Lin *et al.* 2006] exploit summation invariant generated from a number of curves for 3D face recognition, and report a verification rate of 97.2% with a 0.1% FAR on the FRGC v1.0 database. Samir et al. [Samir *et al.* 2006] represent facial surfaces by unions of level curves, called facial curves of the depth function (Fig. 2.22), which are further employed to compare shapes of surfaces using a differential geometric approach that computes geodesic lengths between closed curves on a shape manifold. With multiple gallery faces, a nearest-neighbor classifier achieves accuracies of 92% and 90.4% on the FSU and UND datasets respectively. Zhang et al. [Zhang *et al.* 2006] make use of 3D face models and their mirrors to extract central symmetry profiles, and based on the curvature values of the profiles, three essential points are located on each facial surface. These three essential points uniquely determine a Face Intrinsic Coordinate System (FICS), which is utilized to align different faces. The symmetry profile, together with two transverse profiles, composes a compact representation for face matching. The proposed method is tested on 382 face surfaces, coming from 166 individuals, and the EER values of face authentication for neutral and expressive scans are 0.8% and 10.8% respectively. Mahoor et al. [Mahoor & Abdel-Mottaleb 2007] introduce the principal curvature, kmax, to represent the face image as a 3D binary image called ridge image. The ridge image shows the locations of the ridge lines around the important facial regions on the face (i.e. the eyes, the nose, and the mouth). Hausdorff distance is then used to match the ridge image of a given probe to the created ridge images of the subjects in the gallery. Experiments on the Gavab dataset consisting of 61 subjects resulted in 93.5% ranked one recognition rate for a neutral expression and 82.0% for the faces with a smile expression. Berretti et al. [Berretti *et al.* 2008, Berretti *et al.* 2010] capture face shape information by iso-geodesic stripes. This information is encoded in an attributed relational graph. The similarity between two faces are computed by matching their graphs. The experiments on the Gavab dataset illustrate that this approach performs better than ICP [Berretti *et al.* 2008], and comprehensive results on the SHREC and FRGC 2.0 datasets are further given in [Berretti *et al.* 2010]. Jahanbin et al. [Jahanbin *et al.* 2008] describe 3D facial surfaces by iso-depth and iso-geodesic curves. The former is produced by intersecting a facial surface with parallel planes perpendicular to the direction of gaze, at different depths from the nose tip. The latter is defined to be the locus of all points on the facial surface with the same geodesic distance from a given facial landmark. Once the facial curves are extracted, their characteristics are encoded by several features like simple shape

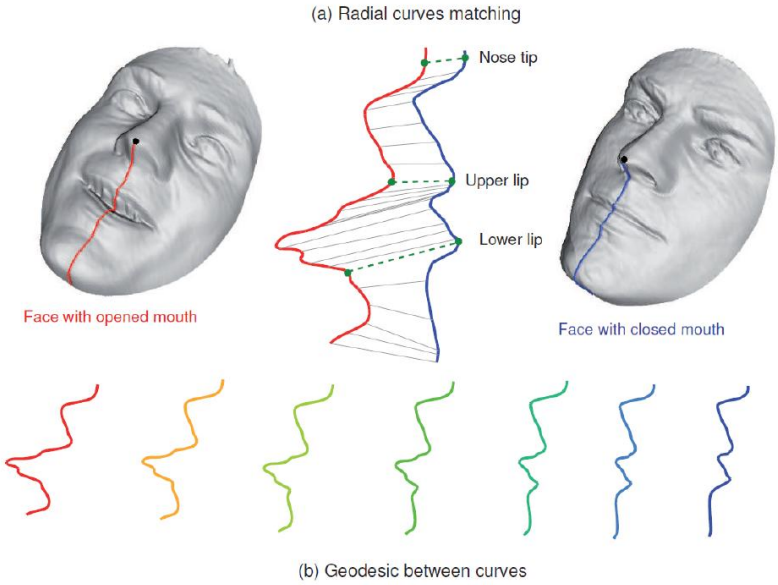


Figure 2.23: Matching and geodesic deforming radial curves [Drira *et al.* 2010].

descriptors (convexity, ratio of principal axes, compactness, circular variance and elliptic variance) or polar Euclidean distances from the origin. The final step is to verify or disapprove requests from users claiming the identity of registered individuals (gallery members) by comparing their features using Euclidean distance classifier or SVM. The experiments on a dataset containing 1196 models of 119 subjects prove that the performance of iso-geodesic is much better than that of iso-depth. Drira et al. [Drira *et al.* 2010] explore elastic radial curves to model 3D facial deformations caused by facial expression variations. They describe 3D facial surfaces by indexed collections of radial curves on them, emanating from the nose tips, and compare the facial shapes by measuring their corresponding curves. Geodesic distances between corresponding elastic curves are adopted as similarity measurements to compare facial surfaces. This technique achieves a more than 90% overall recognition rate on the complete Gavab dataset. Figure 2.23 demonstrates the matching and geodesic deforming process of radial curves.

(3) Region based Methods

The entire facial surface can be segmented into relative rigid and non-rigid parts. Region based local matching directly discards non-rigid facial areas or reduce their deformation effects by assigning lower weights. The selected facial regions are then compared in a certain feature space for face recognition.

Lee and Milios [Lee & Milios 1990] claim that convex facial regions are asserted to change shape less than other regions in response to variations in facial expression. This gives some ability to cope with changes in facial expression. As a result, they segment convex regions in a range image based on the sign of the mean and Gaussian curvatures, and create an extended Gaussian image (EGI) for each convex region describing the shape by the distribution of surface normal over the facial surface. A match between a region in a probe image and in a gallery image is done by

correlating EGIs. Tanaka et al. [Tanaka *et al.* 1998] also perform curvature-based segmentation to distinguish convex and concave parts and represent the face using their EGI. Recognition is performed using a spherical correlation of the EGIs. Experiments are evaluated with a set of 37 images, and 100% recognition is reported. Chua et al. [Chua *et al.* 2000] propose point signatures to describe facial surfaces. To deal with facial expression variations, only the approximately rigid portion of the whole face from just below the nose up through the forehead is used in matching. Point signatures are applied to locate reference points in order to standardize the pose. Experiments are done using multiple facial images with different expressions from 6 subjects, and 100% recognition is achieved. Lee et al. [Lee *et al.* 2003] perform a set of contour lines to divide facial range image into several areas, which are further re-sampled and stored in consecutive location in feature vector. Euclidean distance is applied for classification, reporting 94% correct recognition at rank-five on 70 images from 35 individuals. Moreno et al. [Moreno *et al.* 2003] segment facial regions based on Gaussian curvature where they then create a facial feature vector including the information of angle, distance, area, and curvature. They obtain a recognition rate of 78% on a subset composed of frontal face meshes of the Gavab database. Chang et al. [Chang *et al.* 2005b, Chang *et al.* 2005a, Chang *et al.* 2006] present a type of classifier ensemble approach in which multiple overlapping subregions around the nose are independently matched using ICP, and the results of the multiple 3D matches are finally combined. The experimental evaluation uses essentially the FRGC v2.0 dataset, representing over 4000 images from over 400 persons. In a experiment in which one neutral-expression model is enrolled for each person, and all subsequent are as probes, performance of 92% rank-one recognition is reported. Ben Amor et al. [Ben Amor *et al.* 2006a] split the whole face into different regions according to their robustness to facial expression changes. Then, they apply ICP to calculate recognition score by a region-based similarity metric which takes into account labels of regions. The experiments prove that this scheme improves the accuracy when recognizing expressive facial surfaces. [Cook *et al.* 2007] assign learned weights to different patches of facial range images, and exploit PCA to compute similarity measurement between corresponding patches for fusion. They conclude that this strategy alleviates the performance issue related to facial expressions. Zhong et al. [Zhong *et al.* 2007] propose novel learned visual codebook (LVC) for 3D face recognition. They firstly extract intrinsic discriminative information embedded in 3D faces by Gabor filters, and the generated Gabor maps are divided into a certain number of blocks. K-means clustering is then adopted to learn the centers from the filter response vectors, and LVC is built by these learned centers. The mapping vector between each 3D probe face and LVC constructed by concatenating the corresponding histograms between each local patch and its corresponding sub-codebook is finally fed into a nearest neighbor (NN) classifier. This approach achieves the EER of 7.5% and 4.9% on the CASIA and FRGC v2.0 datasets respectively. Faltemier et al. [Faltemier *et al.* 2008a] introduce a 3D face recognition system based on the score level fusion of results from a committee of 28 small regions that have been independently matched by ICP. It achieves a rank one recognition of 97.2% and verification rates (Roc III) of 93.2% at a 0.1% false accept rate on the FRGC v2.0 dataset. Queirolo et al. [Queirolo *et al.* 2010] utilize Simulated An-

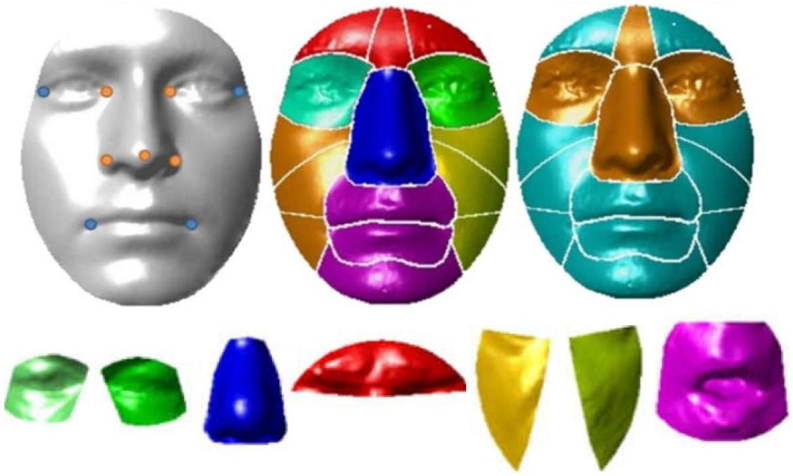


Figure 2.24: Landmark based facial segmentation [Alyuz *et al.* 2010].

nealing (SA)-based Surface Interpenetration Measure (SIM) to compute similarity measurements between corresponding pre-defined facial surface regions(circular and elliptical areas around the nose, forehead, and the entire face region), and combine them for final decision. Comprehensive experiments are performed on the FRGC v2.0 database, a verification rate (Roc III) of 96.5% percent at the FAR of 0.1% as well as a rank-one accuracy of 98.4% are reported. Alyuz et al. [Alyuz *et al.* 2010] cut the facial surface into different regions according to automatic landmarked points (see Fig. 2.24), then from each region, curvature based features are extracted and classified by LDA. It achieves a rank-one recognition rate of 97.5% and a verification rate (Roc III) of 86.1 with a 0.1% FAR in the FRGC v2.0 database. Moreover, a recognition accuracy of 98.2% is reported on the Bosphorus dataset. McCool et al. [McCool *et al.* 2008] introduce an approach to 3D face verification which divides the 3D face into separate parts. This method, termed 3D free-parts, considers each part of the face independently and consequently the spatial relationship is discarded for the purpose of obtaining many observations from each face. The distribution of 2D-DCT features of different blocks are modeled robustly using GMM. This approach demonstrates a significant improvement over the eigenfaces approach, lowering the false rejection rate (FRR) from 9.83% to 4.48% at a false acceptance rate of 0.1%, in tests conducted on 3D face data from FRGC. McCool et al. [McCool *et al.* 2010] further investigate HMM to replace GMM, and under the same framework, the EER is reduced from 0.88% for the GMM technique to 0.36% for the best HMM approach.

(4) General Descriptor based Methods

To represent each vertex of a facial 3D point-cloud (mesh) or each pixel of a facial range image, general descriptor based methods locally extract shape attributes within its neighborhood. The entire face can be described by a feature vector combining the information of all vertices or pixels, and finally fed into classifiers.

Gupta et al. [Gupta *et al.* 2007b] propose to use complex wavelet structural similarity metric (CW-SSIM) extracted from facial range images for 3D face recognition. CWSSIM is computationally efficient and is robust to small geometrical distortions.

Compared with Mean Squared Error (MSE) and Hausdorff distance, CW-SSIM provides better performance on a dataset that contains 360 3D face models of 12 subjects. Huang et al. [Huang *et al.* 2007] present a 3D face recognition method based on statistics of range image differences. Depth differences are computed from a neighbor district instead of direct subtraction to avoid the impact of non-precise registration, and the histogram proportion of depth differences is used to discriminate intra and inter personal differences for 3D face recognition. An EER of 12.4% is achieved in the experiment of ROC III on the FRGC v2.0 database. Wang et al. [Wang *et al.* 2008, Wang *et al.* 2010] employ Signed Shape Difference Map (SSDM) computed between two aligned 3D faces as a mediate representation for facial surface comparison. Based on the SSDMs, three kinds of features (Haar, MB-LBP and Gabor) are used to encode both the local similarity and the change characteristics between facial shapes. The most discriminative local features are selected optimally by boosting and trained as weak classifiers for assembling three collective strong classifiers. The reported rank-one recognition rate and verification rate (Roc III) at 0.1% FAR are both up to 98.0%.

(5) Multiple Shape Feature Fusion based Methods

All these local features mentioned above can be combined to improve the performance of 3D face recognition, and thus form the last category, namely multiple shape feature fusion based methods.

Gordon [Gordon 1992] begins with a curvature-based face segmentation. Then a set of features are extracted that describe both curvature and metric size properties of the face. Each face therefore becomes a point in a feature space, and Nearest Neighbor based matching is done. Experiments are carried out on a test set of three views of each of 8 faces and recognition rates as high as 100% are reported. Pan et al. [Pan *et al.* 2003b] present an approach for 3D face verification from range data. The method consists of profile and surface matching. The profile is extracted on the basis of symmetry of human face, and a global profile matching method based on k-th Hausdorff distance is used to align and compare profiles. For each individual, a statistical model of facial surface is built to represent the distinct discriminative capability of the different parts in the facial surface. Then the model is incorporated into a weighted distance function to measure similarity of surfaces. Finally two experts are combined to give a decision. An EER of 4.44% is obtained on the 3D_RMA database. A more comprehensive version can be found in [Pan & Wu 2005]. Xu et al. [Xu *et al.* 2004] develop a technique for 3D face recognition. The original 3D point-cloud is converted to a regular mesh. The nose region is found and used as an anchor to locate other local regions. Gaussian-Hermite (G-H) moments computed from the detected local regions along with holistic depth values are utilized as feature vectors. PCA is then applied for dimensionality reduction and Nearest Neighbor is finally employed for classification. The results achieved on the manual (30 subjects) and automatic (120 subjects) subsets of 3D_RMA are 96% and 72% respectively, but it should be noted that for each subject, 5 models are enrolled in the gallery. Gokberk et al. [Gokberk *et al.* 2005] compare five approaches to 3D face recognition. They compare methods based on EGI, ICP matching, range profile, PCA, and LDA. Using the dataset including 571 images from 106 people, they find that ICP and LDA offer the best results, although performance is relatively similar among

all approaches but PCA. They also explore methods of fusing the results of the five approaches and are able to achieve 99% rank-one recognition with a combination of recognizers. At that time, this work is relatively novel in comparing the performance of different 3D face recognition algorithms, and in documenting a performance increase by combining results of multiple algorithms. [Koudelka *et al.* 2005] introduce both radial symmetry and shape to extract five features on 3D facial range images. These facial features determine a very small subset of discriminating points which serve as the input to a prescreening method based on a Hausdorff fraction. Verified using the FRGC v1.0 dataset, and results show 97% of the extracted facial features are within 10 mm or less of manually marked ground truth, and the prescreener has a rank 6 recognition rate of 100%. Lee et al. [Lee *et al.* 2005] manually landmarked eight points, based on which two sets of facial features i.e., seven curves as well as some distance and angle values, are extracted. Dynamic programming is applied to classify the generated curves and obtains a recognition rate of 95% on a 5-individual test set; while SVM is used to classify the distance and angle values and achieves the accuracy of 96% on a 100-individual dataset. Based on pre-located eight landmarks, Moreno et al. [Moreno *et al.* 2005] calculate thirty local geometrical features containing distance, angles, area, curvatures, etc to model 3D facial surfaces. PCA and SVM are emplyed for classification, and on the Gavab dataset, they obtain results of 82.0% and 90.16% respectively when the faces are frontal views with a neutral expression, both of which are decreased to 76.2% and 77.9% under gesture and light face rotation. Li and Zhang [Li & Zhang 2007] explore the use of multiple intrinsic geometric attributes, including angles, geodesic distances, and curvatures, for 3D face recognition, where each face is represented by a triangle mesh, preprocessed to possess a uniform connectivity. They then train different weights to be applied to each individual attribute, as well as the weights used to combine the attributes, in order to adapt to expression variations. Using the eigenface approach based on the learned weights and a nearest neighbor classifier, the results on the Gavab database and the FRGC 3D database show that the fusion improves the performance. Li et al. [Li *et al.* 2009] further investigate Sparse Representation Classifier (SRC) for 3D face recognition using a similar feature set of geometric attribute features, and achieves more than 90% recognition rate on frontal samples using a dataset composed of 120 persons selected from the Gavab and FRGC v2.0 dataset. However, the comparison with the state of the art is not fair since they enroll five models of each person in the gallery set. Guan and Zhang [Guan & Zhang 2008] combine Facial Structural Angle (FSA) and Local Region Map (LRM) for 3D face recognition, claiming that this strategy is robust to facial expressions.

2.2.3 Other Matching Methods

In addition to holistic and local feature based 3D face matching techniques, there also exist some approaches aiming to find a point to point relationship between two 3D facial surfaces in a special mapping space and thus achieve their comparison for 3D face recognition. Considering this commonality, we categorize them as surface parameterization methods. Recently, following the research development in the respect of texture mapping, morphing, re-meshing, mesh compression etc, 3D surface

parameterization methods have been gradually driving to maturity stage. A parameterization of a surface can be deemed as a one-to-one mapping from the surface to a suitable domain. In general, the domain of parameters itself will be a surface and so constructing a parameterization means mapping one surface into another. This parameterization can be summarized as an optimization problem under some constraints, e.g. isometric, conformal, equiareal. Using surface parameterazition, 3D facial surfaces that are represented as meshes can be mapped onto a regular 2D image so that more sophisticated signal processing approaches can be applied or to an expression invariant standard surface for 3D face matching.

Abate et al. [Abate *et al.* 2005] map 3D meshes to a 2D plane so that they can compare any two faces represented as 3D polygonal surfaces through their corresponding normal maps, a bi-dimensional array which stores local curvature (mesh normals) as RGB components of pixels in a color image. The histogram of a difference map resulting from the comparison of normal maps, weighted by an expression mask, is utilized as facial feature for recognition.

Considering the influences caused by facial expression changes, Bronstein et al. [Bronstein *et al.* 2005] use isometric mapping to convert expressive facial surfaces to a defined standard model, namely canonical form. It is expected that face models with various expressions of the same subject should generally possess the same canonical form. Based on this idea, they carry out experiment on a test set containing 220 face models from 30 subjects, and report 100% recognition rate. Moreover, they also claim that this method can distinguish identical twins.

Pan et al. [Pan *et al.* 2005] explore a Mapped Depth Image based 3D face recognition method. Using the nose tip as the center point and an axis of symmetry for alignment, the face data are mapped onto a circular range image. The PCA based approach is finally introduced for classification. The performance is reported as 95% rank-one recognition rate and 2.8% EER in a verification scenario on the FRGC v1.0 dataset.

Passalis et al. [Passalis *et al.* 2005] firstly fit a pre-defined manually annotated deformable model to a given 3D facial surface so that it can be parameterized to a 2D plane to generate a color image, whose three channels contain the point coordinates of all the vertices. Wavelet transform is then applied to the achieved color image of certain individual, and corresponding coefficients are adopted as feature vector for L_1 distance based classification. Tested on the FRGC v2.0 database, the experiments show that this method performs much better than FRGC based line (PCA), and it is also insensitive to expression and gender variations.

Abate et al. [Abate *et al.* 2006b] project the vertices of 3D facial surfaces to a 3D sphere, and thus obtain their spherical coordinates. The angles between two normal vectors of adjacent basic facets are calculated and represented in a 2D image, to achieve the property of rotation invariant out of plane. While 2D Fourier based descriptor is used to extract the in-plane rotation invariant feature from the produced 2D image for classification. About 96% accuracy at the first rank is reported using a test set consisting of 120 individuals, each of which has 9 model.

Based on [Passalis *et al.* 2005], Kakadiaris et al. [Kakadiaris *et al.* 2007] further develop a more comprehensive 3D face recognition system, in which two types of wavelet transform are applied not only to geometry images, but also to normal maps.

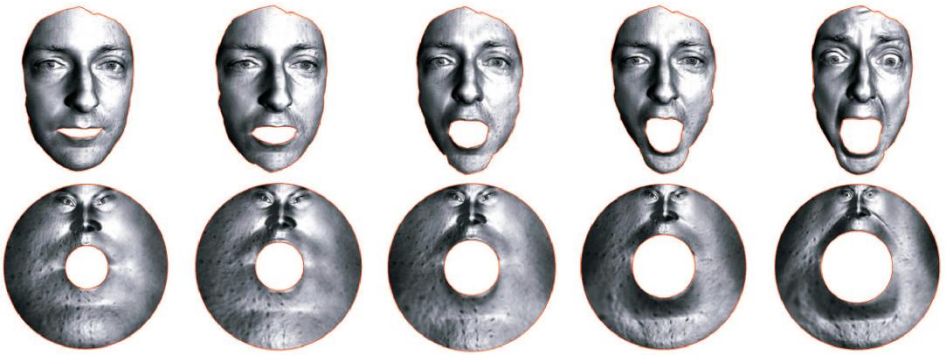


Figure 2.25: Dynamic human faces with expression variations and their conformal maps [Zeng *et al.* 2010].

By combination the similarity measurements of separate matching step, they report a 97.0% rank-one recognition rate and a 97.0% verification rate at 0.1% FAR on the FRGC v2.0 dataset.

Llonch et al. [Llonch *et al.* 2008] address such a problem of 3D face recognition using spherical sparse representations as well. After alignment, 3D facial surfaces are represented as signals on the 2D sphere, in order to take benefit of the geometry information. Simultaneous sparse approximations implement a dimensionality reduction process by subspace projection. Each face is typically described by a few spherical basis functions that are able to capture the salient facial characteristics. The dimensionality reduction step preserves the discriminant facial information and eventually permits an effective matching in the reduced space, where it can further be associated with LDA to improve recognition performance. They evaluate the algorithm on the FRGC v.1.0 database, and it outperforms applying PCA or LDA based approaches on depth facial images.

Wang et al. [Wang *et al.* 2007a] claim that conformal map is insensitive to facial expression variations and occlusions, therefore, they introduce conformal map to transform 3D facial surfaces to 2D images, whose pixel intensities are mean curvatures of vertices. Correlation is exploited to match faces, and it achieves a 97.3% recognition rate on 100 3D face models from a 10-person dataset. Several conformal maps are depicted in Fig. 2.25.

2.2.4 Discussion

After the development in the past two decades, current objective of 3D face recognition techniques is to meet the requirements of application in real condition, i.e. to provide high and robust accuracies (with one face model per subject for enrollment) while to operate efficiently.

As a typical topic in the pattern recognition community, a 3D face recognition system contains the step of feature extraction and classification. Based on the taxonomy in 3D face recognition discussed above, we can come into a conclusion, which is similar in the domain of 2D intensity image based face recognition, that local feature based matching generally outperforms holistic matching, especially to deal with

the problem caused by the variations in facial expression and pose. At this moment, many milestone 3D face recognition approaches based on holistic matching are usually seen as the baseline for performance comparison, such as PCA, ICP, Hausdorff distance and so on. Since facial expression variation is the main factor which largely degrades the accuracy of 3D system, in order to improve the performance of local feature based matching, a commonly adopted solution is to combine multiple features to represent local shapes of surfaces more comprehensively [Kakadiaris *et al.* 2007, Wang *et al.* 2010] or fuse similarity measurements computed from different regions [Faltemier *et al.* 2008a, Alyuz *et al.* 2010, Queirolo *et al.* 2010]. In addition, a training set to learn the distribution of expression changes is also useful to further enhance the results [Al-Osaimi *et al.* 2009, Wang *et al.* 2010].

Meanwhile, the preprocessing step plays an important role in face recognition. Since 3D data convey more useful information than 2D data, preprocessing using 3D data is more convenient and accurate. Generally, in 3D face analysis, the preprocessing step involves two main parts, i.e. fiducial landmark detection and facial surface registration. A few key fiducial landmark points are critical to estimate the initial positions of facial surfaces or locate more complex facial features (e.g. curves and regions etc.), particularly in the early studies. Curvature analysis is a dominant manner for this issue, and at present, considerable progress has been made to ameliorate landmarking techniques. Many tasks conclude that the detection accuracy of major points, including nose tip, nose corners, inner eye corners, on nearly frontal 3D facial surfaces is almost 100% within an affordable error range [Colbry *et al.* 2005, Szeptycki *et al.* 2009, Zhao *et al.* 2009, Alyuz *et al.* 2010], which is sufficient to support the latter steps. However, to precisely detect more landmarking points on the faces under large pose variations still remains a problem. Based on the generated landmarking points, ICP based surface alignment along with a coarse to fine strategy, is usually required in the literature to correct 3D facial pose changes, but it is the most time-consuming process in 3D face recognition. Therefore, as a part of 3D face recognition system, the performance and computational expense in preprocessing should be optimized as well. On the other hand, the 3D face recognition algorithms depending on less preprocessing work should be explored.

2.3 Summary

In the first half of this chapter, we have elaborately discussed several techniques which are regarded as milestones in 2D intensity image based face recognition, while in the latter half, we have extensively reviewed the development process formed by representative approaches in 3D model based face recognition.

In the 2D modality, almost all recent methods claim to be able to provide satisfactory recognition rates, but only when evaluated on standard databases. Actually, some fundamental challenges, i.e. illumination and pose variations, have not been solved completely. The 3D modality has the potential to boost recognition performance of 2D face recognition with respect to these challenges, and many techniques have already proved their competency in certain validation circumstances, however, several other difficult problems still require to be properly addressed, e.g. the pre-

cision and speed of data acquisition, the accuracy of landmarking on partial faces (right and left face profiles), the slowness of surface registration. Moreover, the 3D modality tends to be more insensitive to facial expression variations compared with the 2D based one.

As a result, besides investigating better solutions to improve the performance in each modality separately, we can also consider to make use of merits offered by both modalities so that the robustness of face recognition systems can be step up. A major trend is to achieve multi-modality face recognition (discussed in chapter 4) which combines 2D modality with 3D ones, since recent works confirm that fusing 2D and 3D approaches leads to results that overcome both 2D and 3D, in terms of recognition rate. Nevertheless, this category of systems inherits common problems of 3D processing, especially, concerning the acquisition process. Meanwhile, a smart and flexible integration between the two modalities should be explored as well aiming to reduce the expenditure in 3D data acquisition and computation, such as 3D aided 2D face recognition using 3D information to normalize pose and lighting changes in facial texture images based on which identification is performed, or asymmetric 3D-2D face recognition (described in chapter 5) exploiting (textured) 3D face models for enrollment while 2D facial images for test and computing similarity scores between two different types of data.

Additionally, even though many datasets have been released for research activities, each of which is designed and expected to give one or more difficulties, they cannot yet well simulate a more complex condition in the real world. More challenging factors including large population, diverse ethnicities, rich age distribution, low quality data, natural occlusions, make-up usage, etc. are attracting much greater scientific attention.

Chapter 3

3D Shape based Face Recognition

3.1 Introduction

As we discussed in the previous chapters, with the rapid development in 3D imaging systems, 2.5D and 3D facial scans have emerged as a major solution to deal with these unsolved issues in 2D intensity image based face recognition, i.e. illumination and pose changes [Bowyer *et al.* 2006, Scheenstra *et al.* 2005]. Meanwhile, even though 3D facial scans capture exact shape information of facial surfaces and 3D face recognition methods are thereby theoretically reputed to be robust to variations of illumination, they are likely to be more sensitive to facial expression variations. Besides, they generally require an accurate registration step before shape-based 3D matching.

Generally, similar as 2D intensity image based face recognition, how to describe facial surfaces is a core topic in the 3D domain as well. "Good" geometric features of 3D facial shapes should have the following properties [Hadid *et al.* 2004]: first, they can tolerate within-class variations while discriminating different classes well; second, they can easily be extracted from raw facial data to allow for fast processing; finally, they should lie in a space with moderate dimensionality in order to avoid unaffordable computational expense in the matching step. As a result, existing 3D face recognition techniques can be roughly classified according to the features they use: (1) *Original feature-based* techniques make use of the entire face region as input to compute similarity. Several works have explored PCA directly on facial range images [Achermann *et al.* 1997, Bronstein *et al.* 2003, Hesher *et al.* 2003]; while some have applied the ICP (Iterative Closest Point) algorithm [Besl & McKay 1992] or its modified version on facial point-clouds to match surfaces [Chang *et al.* 2005b, Lu & Jain 2006, Lu *et al.* 2006]. The Hausdorff distance has also been explored for matching facial surfaces [Lee & Shim 2004, Russ *et al.* 2005]. (2) *Region or point feature based* methods detect representative facial points or areas to construct corresponding feature spaces. The eye and nose areas are used in [Gordon 1992]; segmented facial regions and lines are utilized in [Moreno *et al.* 2003]; anthropometric facial fiducial keypoints are investigated in [Gupta *et al.* 2007a]. (3) *Curve feature based* approaches extract discriminative curves for facial surface representation. In [Nagamine *et al.* 1992], three facial curves were found to intersect the facial surface using horizontal, vertical planes and a cylinder; the central profile with maximal protrusion as well as two parallel profiles were searched in [Beumier & Acheroy 2000]; an union of the level curves of a depth function was proposed to describe 3D facial surfaces [Samir *et al.* 2006]; (4) *Shape feature based* algorithms focus on the geometric attributes of local surfaces, such as curvatures [Gordon 1992], point signature [Chua *et al.* 2000], Extended Gaussian Image (EGI) [Tanaka *et al.* 1998], Signed

Shape Difference Map (SSDM) [Wang *et al.* 2010] etc.

3D face recognition techniques can also be categorized according to the nature of their matching strategies, i.e. holistic or local face matching, even though this is highly dependent on the facial features used. We have extended this taxonomy to 3D face recognition in the second chapter. The holistic stream contains ICP-based matching [Lu *et al.* 2006], isometry invariant description [Bronstein *et al.* 2005] etc. This matching scheme is based on holistic facial features and hence generally requires an accurate normalization step with respect to pose and scale changes. Furthermore, it has proved sensitive to expression variations and partial occlusions. The local feature-based one utilizes local features of 3D facial scans and has been explored in several works in the literature, including point signature [Chua *et al.* 2000], Region-ICP [Ouji *et al.* 2009] and so on. Local feature-based matching has the potential advantage of being robust to facial expression, pose and lighting changes and even to partial occlusions. The downside of this scheme is the difficulty extracting sufficient informative feature points from similar or smooth 3D facial surfaces.

In this chapter, we propose a novel method for 3D face recognition making use of a geometric facial description along with a local feature matching scheme. Our main contributions can be summarized as follows.

(1) Since all facial surfaces are generally smooth and similar, in order to achieve accurate representation and enhance their distinctiveness, a shape based geometric facial description, consisting of a set of Multi-Scale extended Local Binary Pattern Depth Faces (MS-eLBP-DFs), is presented.

Because of its superior performance and computational simplicity, LBP is one of the most popular and successful local descriptors for 2D intensity image based face analysis. See [Huang *et al.* 2011d] for a review on its application to facial image analysis. In the literature, LBP has been also investigated for 3D face recognition [Li *et al.* 2005b, Huang *et al.* 2006]. However, LBP is not so discriminative for this purpose since it cannot distinguish similar local shapes well due to its thresholding strategy. To address such a problem, two solutions are considered. First, extended LBP (eLBP), generalized from our previous work on 3DLBP [Huang *et al.* 2006] and capable of handling different numbers of sampling points and various scales, is used. It not only extracts the relative gray value differences from the central pixel and its neighbors provided by LBP, but also focuses on their absolute differences that prove very critical to describe range faces as well. Secondly, a multi-scale strategy is introduced to represent local surfaces to different extents which are then combined for a comprehensive description. In addition, previous works simply repeated the method of 2D facial analysis that divides the face into a number of sub-regions, from which LBP based histograms are extracted; then concatenates all these local histograms into a global one to form a final facial feature. Unlike these works, we apply eLBP directly to the whole range facial area to generate Multi-Scale extended LBP Depth Faces (MS-eLBP-DFs) which retain all 2D spatial information on range faces. Finally, this method inherits the computational simplicity property from LBP and achieves fast processing.

(2) A local feature matching scheme is designed to measure similarities between gallery and probe facial scans once represented in terms of the proposed geometric facial description, i.e. MS-eLBP-DFs.

The recognition process works in a local manner by performing a feature-based matching using SIFT based features, which is further strengthened by facial component and configuration constraints. Since the proposed matching scheme possesses a certain tolerance to moderate pose changes, it enables the entire system to match nearly frontal 3D face models without generally costly registration.

(3) Most of the state of the art have proven their effectiveness over renowned databases like FRGC, but less effort has been made to analyze the robustness to quality degradations. In this chapter, we evaluate the performance of the proposed method with regard to data degradations such as Gaussian noise, decimation, and random holes etc. that are generated on a subset of FRGC.

The proposed method is evaluated on two public databases, namely FRGC v2.0 [Phillips *et al.* 2005] as well as Gavab DB [Moreno & Sanchez 2004]. Experimented on the FRGC v2.0 dataset for both tasks of 3D face recognition and verification, the proposed approach achieves a rank-one recognition rate of 97.6% and a 98.4% verification rate with an FAR of 0.1% respectively. Thanks to the local matching scheme, a generally costly face alignment step is not needed, because all 3D face models in the FRGC v2.0 dataset are nearly frontal, which is in contrast to most techniques so far in the literature. When it handles severe pose changes including left or right facial profiles leading to self-occlusions, coarse alignment based on a few landmark points is sufficient as a preprocessing step of this approach. This fact is demonstrated by the experiments on the Gavab database. Furthermore, additional results illustrate that the system is also robust to facial expression variations and data degradations such as Gaussian noise, decimation, random holes etc.

The remainder of this chapter is organized as follows. The proposed geometric facial description, MS-eLBP-DFs, is shown in section 3.2, and section 4.4 presents the local feature based matching step. Experimental results of both face recognition and verification tasks are described and analyzed in section 3.4. Section 4.7 concludes this chapter.

3.2 MS-eLBP based Facial Representation

Due to its ability to describe micro patterns and its computational simplicity, the Local Binary Patterns (LBP) technique has been widely explored in 2D face analysis [Ahonen *et al.* 2004, Yan *et al.* 2007, Chan *et al.* 2007, Shan & Gritti 2008], and recent investigation for 3D facial description [Li *et al.* 2005b, Huang *et al.* 2006]. However, using LBP directly on range faces does not allow accurate representation of local shape changes. In this section, we firstly analyze the descriptive ability of LBP for local facial surface representation. Then, we present extended LBP (eLBP) and the multi-scale scheme to generate a novel 3D geometric facial description, called MS-eLBP Depth Faces (MS-ELBP-DFs) that comprehensively encode local shape variations of range faces.

3.2.1 LBP and Its Descriptive Power for Local Shape Variations

In the second chapter, we have introduced the LBP methodology in detail. For the purpose of completeness, we briefly recall its basic conception.

The original LBP operator labels each pixel of a given 2D image by thresholding in a 3×3 neighborhood. If the values of the neighboring pixels are no lower than that of the central pixel, their corresponding binary bits are assigned to 1; otherwise they are assigned to 0. A binary number is hence constructed by concatenating all the eight binary bits in a clockwise direction starting from the top-left position, and the resulting decimal value (LBP code) is used for labeling the central pixel.

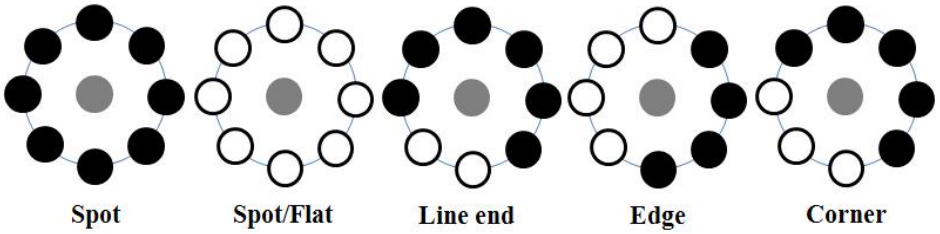


Figure 3.1: An example of texture patterns which can be encoded by LBP (white circles represent ones and black circles zeros).

According to the definition of LBP, the LBP code is invariant to monotonic gray-scale transformations, preserving their pixel orders in local neighborhoods. When LBP operates on the images formed by light reflection, it can be used as a texture descriptor. Each of the 256 (2^8) LBP codes can be considered as a micro-texton. Local primitives codified by the bins of LBP using uniform pattern include different types of curved edges, spots, flat areas etc. Fig. 3.1 shows some examples. Similarly, as LBP works on range images which are based on depth information, it can also describe local shape structures, such as flat, concave, convex etc., as shown in Fig. 3.2.

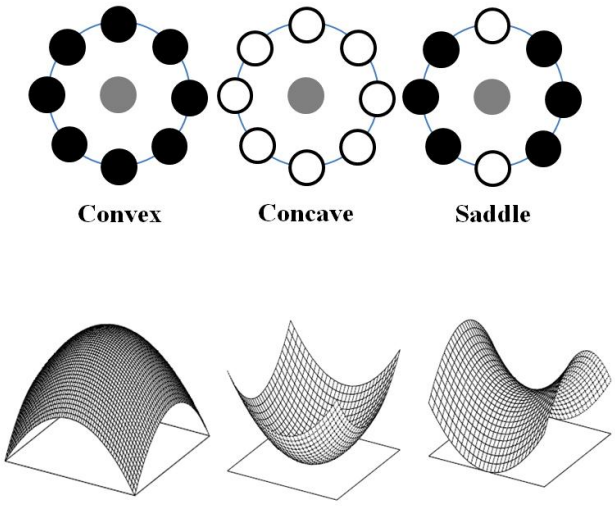


Figure 3.2: Examples of local shape patterns encoded by the original LBP operator (white circles represent ones and black circles zeros).

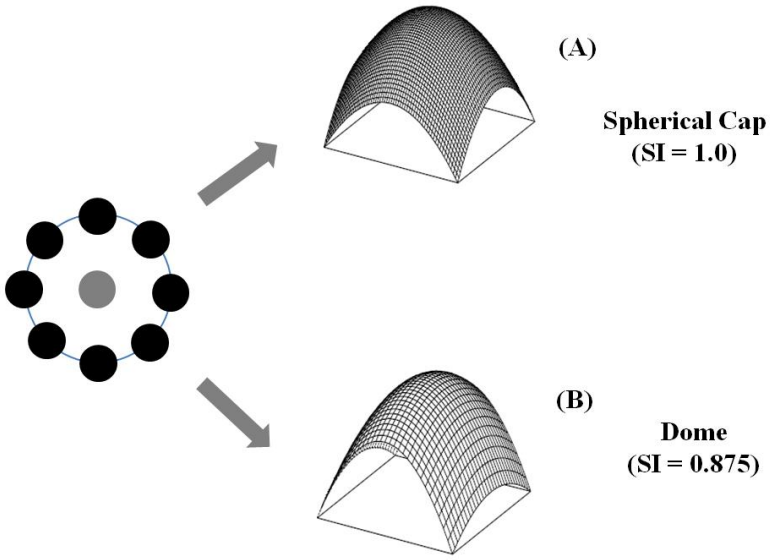


Figure 3.3: A confusion case of LBP as it describes similar but different local geometric shapes.

Unfortunately, the direct application of LBP to depict the shapes of 3D surfaces also leads to unexpected confusion when distinguishing similar yet different local shapes. Fig. 3.3 illustrates two similar shapes with different Shape Index (SI) values [Koenderink & Doorn 1992] while indeed sharing the same LBP code: shape (A) is a spherical cap; shape (B) is a dome. This lack of descriptive power is problematic when one needs to derive a facial description to enhance distinctiveness for the task of face recognition.

To address such a problem, we propose to adopt two complementary solutions. The first solution aims to improve the discriminative ability of LBP with the eLBP coding method, and the other one focuses on providing a more comprehensive geometric description of a given neighborhood by applying a multi-scale strategy. Each solution is discussed in the following two subsections respectively.

3.2.2 Extended Local Binary Patterns

The fact that LBP is not competent to distinguish between similar local shapes is due to its operation mode. It only encodes relative differences between a central pixel and its neighboring ones. In this section, we introduce eLBP to better describe local surface properties. Instead of the original LBP operator, eLBP not only extracts the relative gray value difference between the central pixel and its neighbors provided by LBP, but also focuses on their absolute differences which are also important to describe local shapes. eLBP is a generalized version of our previous work on 3DLBP [Huang *et al.* 2006] originally proposed for histogram-based 3D face recognition.

Specifically, the eLBP operator consists of several binary codes in multiple layers that encode the exact gray value differences (GD) between the central pixel and its neighboring ones. The first layer of eLBP is actually the original LBP code encoding

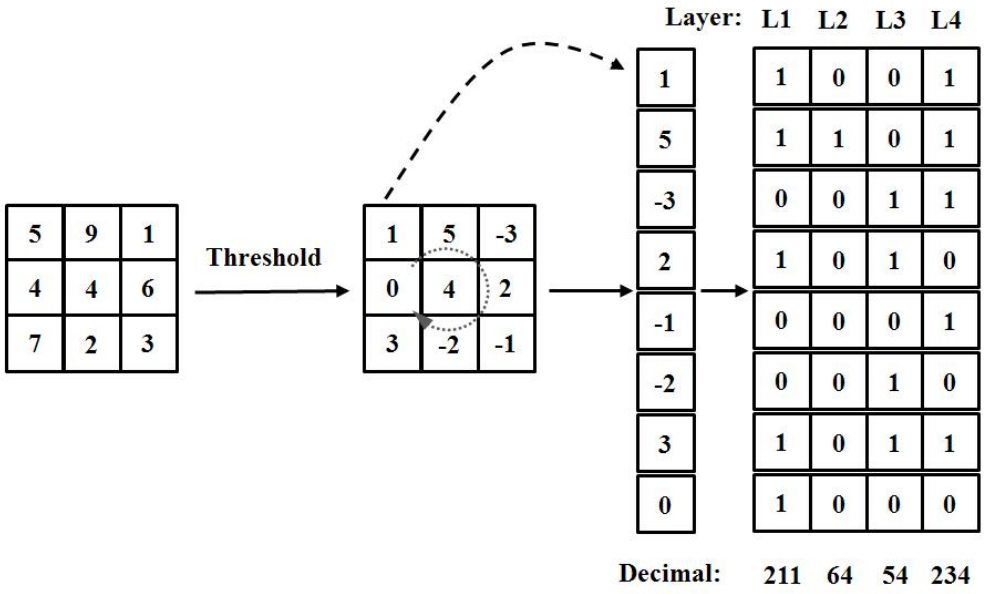


Figure 3.4: An example of the eLBP operator.

the GD sign. The following layers of eLBP then encode the absolute value of GD. Basically, each absolute GD value is first encoded in its binary representation and then all the binary values at a given layer result in an additional local binary pattern. The example of Fig. 2.10 can be expressed by eLBP as shown in Fig. 3.4. The first layer of eLBP code is simply the original LBP code that encodes the sign of GD, thus yielding a decimal number of 211 from its binary form $(11010011)_2$. The absolute values of GD, i.e. 1, 5, 3, 2, 1, 2, 3, 0, are preliminarily expressed as their binary numbers: $(001)_2$, $(101)_2$, $(011)_2$, $(010)_2$, ..., etc. Using the same weighting strategy of LBP on all the binary bits, we generate the eLBP code of its corresponding layer, e.g., L_2 is composed of $(01000000)_2$ and its decimal value is 64; L_3 is composed of $(00110110)_2$ and its decimal value is 54; finally L_4 is composed of $(11101010)_2$ and its decimal value is 234. As a result, as describing similar local shapes, although the first layer LBP is not discriminative enough, the information given by the additional layers can be used to distinguish them.

Theoretically, within one image, the maximum value of GDs is up to 255 (between 0 and 255), which means that 8 additional binary units are required to code GDs ($2^8 - 1 = 255$), and thus 7 additional layers should be produced. Nevertheless, we do not need so many layers in eLBP. Range faces are indeed very smooth; the GDs in a local surface generally do not vary dramatically. Some preliminary statistical work reveals that more than 85% GDs are smaller than 7 between points within eight pixels. Therefore, the number of added binary units, k, is determined by GD. Meanwhile k can also be exploited to control the trade-off between the description expressiveness of local shapes and the computational simplicity of eLBP. All the GDs which are larger than $2^k - 1$ can be assigned to $2^k - 1$ to decrease the com-

putational cost. In this study, three additional layers are extracted and analyzed to illustrate their contributions to the final accuracy.

3.2.3 Multi-Scale Strategy

As we know, the original LBP operator was extended later with different sizes of local neighborhood to deal with various scales [Ojala *et al.* 2002]. The local neighborhood of the LBP operator is defined as a set of sampling points evenly spaced on a circle centered on the pixel to be labeled. These sampling points which do not fall exactly on the pixels are expressed using bilinear interpolation, thus allowing any radius value and any number of points in the neighborhood. The notation (P, R) denotes the neighborhood possessing P sampling points on a circle of radius R. By adopting the same protocol, the eLBP operator can also handle different sampling points and scales.

Several LBP histogram-based tasks changed the neighborhood of the LBP operator in order to improve their performance. By varying the value of radius R, the LBP codes of different resolutions are obtained. The multi-scale scheme was originally used for texture classification [Ojala *et al.* 2002], and it was also introduced to 2D face recognition [Yan *et al.* 2007, Chan *et al.* 2007]. In [Shan & Gritti 2008], Shan and Gritti studied MS-LBP for facial expression recognition by firstly extracting MS-LBP histogram-based facial features and then using AdaBoost to learn the most discriminative bins. They pointed out that the boosted classifiers of MS-LBP consistently outperform those based on single-scale LBP, and the selected LBP bins distribute at all the scales. MS-LBP can hence be deemed as an efficient approach for facial representation. When considering it in 3D face analysis, this multi-scale technique can be applied to enhance the descriptive power of LBP.

3.2.4 Multi-Scale Extended LBP Depth Faces (MS-eLBP-DFs)

LBP facial representation can be achieved in two ways: one is by LBP histogram, while the other is by LBP face. The general idea of the former one is that one human face can be regarded as a composition of micro-patterns described by LBP. The images are divided into a certain number of local areas, from which LBP histograms are extracted. These histograms are concatenated and thus contain both local and global information of the faces. The latter manner is to generate LBP based maps, which regards the decimal number of the LBP code as the pixel values of an LBP map, and thus produces the corresponding LBP face. Due to its own strategy, an LBP histogram loses some 2D spatial information for representing faces. In this study, the second, eLBP face, is investigated.

Given a facial range image, we generate a set of MS-eLBP-DFs for facial representation. These MS-eLBP-DFs can be achieved by varying the neighborhood size of the eLBP operator, or by first down-sampling range faces and then adopting an eLBP operator with a fixed radius. Some face samples are illustrated in Fig. 3.5. In that figure, the number of sampling points is 8, and the radius value varies from 1 pixel to 8 pixels. As we can see, the original range face is very smooth, whilst the resulting MS-eLBP-DFs contain many more details of local shape variations.

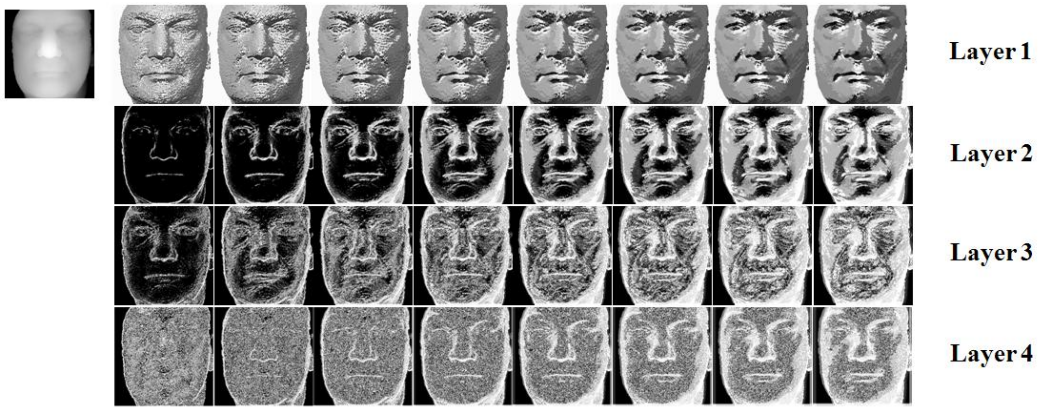


Figure 3.5: MS-eLBP-DFs of a facial range image with different radii from 1 to 8 (from left to right).

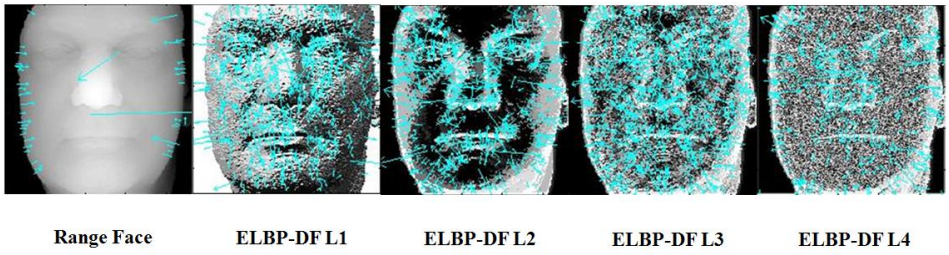


Figure 3.6: The SIFT-based keypoints detected from an original facial range image and its four associated eLBP-DFs.

3.3 Local Feature Matching

Once the proposed MS-eLBP-DFs have been produced, the widely-used SIFT features [Lowe 2004] are extracted from them. Then, a local matching process further improved by the facial component and configuration constraints is carried out to calculate a similarity score between two 3D facial scans.

3.3.1 Local Feature Extraction

The SIFT operator (more technique details are presented in the second chapter) works on each MS-eLBP-DF separately. Because MS-eLBP-DFs highlight the local shape characteristics of smooth facial range images, many more SIFT-based keypoints can be detected for the following matching step than in the original range faces. Some statistical work was done along with the experiments on the FRGC database. The average number of descriptors extracted from each of MS-eLBP-DFs is 553, while that of each original facial range image is limited to 41, and the detected keypoints are usually located on the edge of the face. Figure 3.6 shows the SIFT-based keypoints extracted from one range face image and its four associated eLBP-DFs respectively.

3.3.2 Local Matching Strategy

Given local facial features extracted from each MS-eLBP-DF pair of the gallery and the probe face scan respectively, two facial keypoint sets can be matched. Matching one keypoint to another is accepted only if the matching distance is less than a predefined threshold, t times the distance to the second closest match. In this research, t is empirically set at 0.6 as in [Lowe 2004]. Here, $N_{Li(P,R)}$ denotes the number of matched keypoints in the i_{th} layer of an eLBP-DF pair, generated by eLBP from range face images with a parameter setting of (P, R).

3.3.3 Facial Constraints

Unlike the samples used in object detection, all human faces have the same physical components and share a similar configuration. The local matching strategy is thus can be enhanced by holistically constraining the matched local features of the keypoints with respect to the facial components and configuration, .

(1) *Facial Component Constraint*: we propose to divide the entire facial range image into sub-regions, each of which contains roughly one component of nearly frontal faces, to restrict the matched keypoints of gallery and probe face scans only to those with similar physical meaning. That means the matched keypoints from the same facial part should be more important. Instead of the costly clustering process in [Luo *et al.* 2007] to automatically construct sub-regions based on keypoint locations from training samples, we simply use our common sense of facial component position, and divide the face area into 3×3 rectangle blocks of the same size. The similarity of the facial component constraint is defined from this facial composition scheme. A MS-eLBP-DF, I, is represented as $(m_1, m_2, ..., m_k)$; k is 9 in our case and m_i is the number of detected SIFT keypoints that fall within the i_{th} component. The local SIFT-based descriptors in all the k components can be denoted by:

$$I = (f_1^1, ..., f_1^{m_1}, f_2^1, ..., f_2^{m_2}, ..., f_k^1, ..., f_k^{m_k}) \tag{3.1}$$

where f_i^j means the j_{th} descriptor in the i_{th} facial component. Then the similarity between a gallery face I_g and a probe face I_p is computed by:

$$C(I_p, I_g) = \frac{1}{k} \sum_{i=1}^{k} (\max(\frac{< f_{p_i}^x, f_{g_i}^y >}{\|f_{p_i}^x\| \cdot \|f_{g_i}^y\|})) \tag{3.2}$$

where $x \in [1, ..., m_{pi}]$; $x \in [1, ..., m_{gi}]$. $\langle \cdot \rangle$ denotes the inner product of two vectors, and $\|\cdot\|$ denotes the norm of one vector. A bigger C value indicates the most similar attributes of the two faces represented by MS-eLBP-DFs. We thus obtain similarity values, $C_{Li(P,R)}$ for each MS-eLBP-DF.

(2) *Facial Configuration Constraint*: the former constraint on facial component emphasizes the importance of the matching score between local features of the same facial component-based area in the gallery and probe set, and we further improve the constraint by facial configuration.

All the facial range images are normalized to a certain size to build a public 2D coordinate system. For each 3D facial scan, the MS-eLBP-DFs are extracted from the range image, and there is a pixel-to-pixel correspondence between these images. Therefore, all the keypoints of the proposed facial surface representations share the same XY-plane with the range face image, and the pixel values of the corresponding facial range image can be regarded as the Z-axis values of these keypoints. Hence, each keypoint has its position in 3D space. After local feature-based matching, a 3D graph is formed for each MS-eLBP-DF of a probe F_p, by simply linking every two keypoints which have a matching relationship with keypoints detected on the corresponding MS-eLBP-DF of a gallery face F_g. The matched keypoints of F_g also construct a corresponding graph of F_p. Intuitively, if faces F_g and F_p are from the same subject, their corresponding graphs should have similar shapes and locations in 3D space.

One similarity measure between the two graphs is

$$d_e = \frac{1}{n_e} \sum_{i=1}^{n_e} |d_{p_i} - d_{g_i}| \tag{3.3}$$

where d_{pi} and d_{gi} are the lengths of corresponding edges in the probe and gallery graphs respectively. The value n_e is the total number of edges. If the number of matched keypoints is n_n, n_e will thus be expressed as $n_n * (n_n - 1)/2$. Equation 3.3 is an efficient way to measure the spatial error between the matched keypoint pairs of probe and gallery features.

Another similarity between two graphs is calculated as the mean Euclidean distance d_n between corresponding nodes:

$$d_n = \frac{1}{n_n} \sum_{i=1}^{n_n} |n_{p_i} - n_{g_i}| \tag{3.4}$$

where n_{pi} and n_{gi} are the coordinates of corresponding keypoints in a gallery and its probe graph respectively. n_n is the number of matched keypoints.

The final similarity measurement of the facial configuration constraint is:

$$D = w_e * d_e + w_n * d_n \tag{3.5}$$

where w_e and w_n are the corresponding weights of d_e and d_n, designated according to the scheme adopted in similarity fusion, and a smaller D value indicates the most similar attribute of two facial range images represented by the MS-eLBP-DFs. As in the facial component constraint, here, $D_{Li(P,R)}$ denotes the similarity score of each MS-eLBP-DF.

3.3.4 Similarity Fusion

In summary, the matching process of gallery and probe facial range images contains three types of similarity scores: the number of matched keypoint pairs N, similarity

C of the facial component constraint and similarity D of the facial configuration constraint. Except for D, all the other similarity measures have a positive polarity (a bigger value means a better matching relationship). A range face of the probe set is matched with every face in the gallery, resulting in three vectors S_N, S_C and S_D. The n_{th} element of each score vector corresponds to the similarity score between the probe and the n_{th} gallery face. Each vector is normalized to the interval of [0, 1] using the min-max rule. Elements of S_D are subtracted from 1 to reverse its polarity. The final similarity of the probe face with the faces in the gallery set is then calculated using a basic weighted sum rule:

$$S = w_N * S_N + w_C * S_C + w_D * (1 - S_D) \tag{3.6}$$

Their corresponding weights: w_N, w_C, and w_D are calculated dynamically during the online step using the scheme in [Mian *et al.* 2008]:

$$w_{S_i} = \frac{mean(S_i) - \min_1(S_i)}{mean(S_i) - \min_2(S_i)} \tag{3.7}$$

where i corresponds to the three similarities: N, C, and D, and operators $min_1(Si)$ and $min_2(Si)$ produce the first and second minimum value of the vector S_i. The gallery face image which has the maximum value in the vector S is declared as the identity of the probe face image when the decision is to be made on each MS-eLBP-DF independently.

3.4 Experimental Results

In order to demonstrate the effectiveness of the proposed method, we used two public databases for experiments, i.e. FRGC v2.0 [Phillips *et al.* 2005] as well as Gavab [Moreno & Sanchez 2004]. The first is for evaluating performance with large number of subjects and complicated conditions; the second is to analyze the accuracy on 3D face samples with extreme pose changes (left and right profiles).

3.4.1 Experiment Design

Most of the experiments were evaluated on the FRGC v2.0, one of the most comprehensive and popular datasets, containing 4007 3D face scans of 466 subjects. One facial range image was extracted from each 3D face model. A preprocessing step was applied to remove spikes with a median filter and fill holes using cubic interpolation. Thanks to the relative tolerance to moderate pose changes of hybrid matching, we did not perform any registration on these 3D face models, in contrast to most works such as [Kakadiaris *et al.* 2007, Mian *et al.* 2007, Faltemier *et al.* 2008a] etc. The facial range images were cropped by using a basic bounding box based on the mask provided by a 3D scanner indicating whether the point is valid or not in that position. Cropped faces have pose, expression changes, as well as partial occlusions caused by hair. All faces are normalized to facial range images with 150×150 pixels for computational simplicity.

The proposed approach was carried out on both the face recognition and verification tasks. The first 3D facial scan with a neutral expression from each subject makes up a gallery of 466 samples. The remaining face scans (4007-466=3541) were treated as probes. The 3D facial scans in the probe set were divided into two subsets according to their expression labels to evaluate its robustness to facial expression changes. The first subset is made up of facial scans with a neutral expression; whilst the other one with facial scans possessing only non-neutral facial expressions. Besides Neutral vs. All, two additional experiments: Neutral vs. Neutral and Neutral vs. Non-Neutral were also included to analyze its insensitiveness to facial expression variations. In the Neutral vs. Neutral and Neutral vs. Non-Neutral experiments, only the neutral and non-neutral subsets were used, respectively.

Furthermore, we tested the robustness of the proposed method to several degradations as they commonly occur in the acquisition process including Gaussian noise on depth information, decimation in terms of resolution, and random holes as data missing. In these experiments, we employed 410 persons each of which has more than one 3D model, and the other 56 with only one model per person were not considered. For each subject, the first neutral model was used to make up the gallery set, and another model randomly picked up from each subject was taken as a probe. To compare the performance with other approaches which require more precise preprocessing operations, in this experiment, besides removing spikes and filling holes, the nose tips of all the utilized faces were manually marked, and corresponding facial regions were cropped within a sphere of radius 100mm based on the nose-tip coordinates. Then, the probe set was altered to create degraded sets, according to the following rules (Fig. 3.7 shows examples of those degradations):

- Gaussian noise corresponds to the injection of an error with a Gaussian distribution on the Z coordinates in the depth image. This tends to emulate the behavior of electronic noise of acquisition devices, albeit a simplistic manner. We set the Root Mean Square (RMS) value of the error respectively to 0.2mm, 0.4mm and 0.8mm.
- Decimation aims at removing a number of vertices from the original facial surface. Vertices are picked up randomly and removed respectively from a ratio of $\times 2$, $\times 4$ and $\times 8$. For example, $\times 4$ means that the decimated model includes only a quarter of the vertices in the original ones.
- Holes are generated at random locations on facial surfaces. At first, we choose a random vertex on the surface of the face. Then, we crop the hole according to a sphere with a radius value of 1 centimeter centered at the selected vertex. For each level, we produce 1, 2, 3 holes on the entire face respectively.

To analyze the performance on severe pose changes and even partially occluded 3D facial scans, we also tested our approach on the Gavab database. This database consists of Minolta Vi-700 laser facial scans of 61 different individuals. Each subject was scanned 9 times for different poses and facial expressions.

In the experiments, a similar preprocessing pipeline was utilized as in the FRGC v2.0 dataset to remove spikes and fill holes. Since the Gavab dataset contains many severe pose changes, we performed a coarse alignment based on three landmarks for all facial scans. When the two inner corners of the eyes and the nose tip of one scan are available at the same time (all face scans excluding the extreme poses such as the right and left profiles), we used our previous technique [Szeptycki *et al.* 2009] to lo-

Figure 3.7: An example of degradations applied to one model: from left to right: the original facial surface, Gaussian noise applied, decimation applied, holes applied.

cate the three landmark points automatically and computed rotation and translation parameters; while for each of these left or right profiles, we manually landmarked four points, i.e., the inner and outer corner of one eye, the nose tip, and the corner of the nose, which are visible in that profile. After the coarse registration step, a facial range image is extracted from each facial scan, These range images hence only contain partial faces due to the self-occlusion caused by pose variations, and all the facial range images are further resized to 150×150 pixels. The first frontal facial scan of each subject was used as the gallery; while the others were treated as probes. We calculated rank-one face recognition rates.

3.4.2 Identification

The four sub-tables in Table 3.1 list the results based on depth faces of each eLBP layer with different parameters. Recall that P is the number of sampling points and R is the radius value.

In these sub-tables, the numbers displayed in the last row labeled as the "eLBP" performance are the fusion results according to the weighted sum rule as described in eq. 3.6 combining the similarity measurements provided by the first three layers (L_1, L_2 and L_3) using the same parameter setting; similarity scores at L_4 are omitted because of their low performance. As we can see from Table 3.1, using 8 sampling points achieves better results on L_1, and L_2 for almost all the radius values (except $R = 2, 3$), respectively; whilst the setting with 16 sampling points results in better performance on L_3 (except $R = 2$) and L_4, respectively. All the eLBP accuracies at different scales with different numbers of sampling points display recognition rates better than 90%.

Using the weighted sum rule described in eq. 3.6, we then fused the similarity scores of eLBP with 8 sampling points but different radius values from 2 to 8 pixels, and compared their rank-one face recognition rates with the state of the art as shown in Table 3.2. Except for ICP, all results are cited from the original papers.

In order to test the discriminative power of LBP and eLBP to characterize local geometric shapes, Shape Index (SI) face maps are produced and associated with the proposed local matching scheme for the task of 3D face recognition as well. With a rank-one recognition rate of 91.8% as indicated in Table 3.2, SI faces outperform any of the single scale LBP-DFs (i.e. the layer of eLBP L_1). Meanwhile, most of the

Table 3.1: Rank-one recognition rates based on depth faces of each eLBP layer with different parameters in the FRGC v2.0 dataset.

(A) $P = 4$	$R = 2$	$R = 3$	$R = 4$	$R = 5$	$R = 6$	$R = 7$	$R = 8$
eLBP L_1	81.6%	84.8%	86.9%	87.7%	87.6%	86.2%	85.9%
eLBP L_2	75.2%	83.3%	85.7%	87.1%	87.6%	87.3%	87.0%
eLBP L_3	76.9%	74.7%	71.6%	68.8%	67.4%	63.7%	61.9%
eLBP L_4	4.5%	8.0%	12.7%	16.0%	25.9%	33.2%	40.6%
eLBP	90.0%	90.9%	92.0%	92.6%	92.4%	92.3%	92.3%
(B) $P = 8$	$R = 2$	$R = 3$	$R = 4$	$R = 5$	$R = 6$	$R = 7$	$R = 8$
eLBP L_1	86.1%	87.8%	88.5%	88.3%	87.7%	86.6%	86.0%
eLBP L_2	73.6%	84.6%	88.6%	89.2%	89.2%	89.3%	89.9%
eLBP L_3	80.1%	78.3%	76.4%	76.3%	75.6%	76.6%	76.3%
eLBP L_4	6.6%	11.1%	17.8%	29.8%	40.3%	50.8%	55.6%
eLBP	91.3%	92.5%	93.5%	93.4%	93.0%	92.7%	92.6%
(C) $P = 12$	$R = 2$	$R = 3$	$R = 4$	$R = 5$	$R = 6$	$R = 7$	$R = 8$
eLBP L_1	85.3%	86.1%	86.2%	87.2%	85.8%	85.4%	84.6%
eLBP L_2	71.7%	84.4%	87.3%	88.6%	89.3%	88.9%	88.4%
eLBP L_3	81.9%	78.7%	78.1%	76.6%	78.5%	78.9%	79.6%
eLBP L_4	6.2%	12.3%	22.1%	35.6%	48.7%	57.4%	63.2%
eLBP	90.9%	92.1%	92.9%	93.3%	92.3%	92.3%	91.5%
(D) $P = 16$	$R = 2$	$R = 3$	$R = 4$	$R = 5$	$R = 6$	$R = 7$	$R = 8$
eLBP L_1	82.1%	82.9%	85.3%	84.2%	84.3%	83.5%	82.7%
eLBP L_2	73.7%	86.1%	87.9%	88.6%	88.2%	87.5%	87.7%
eLBP L_3	81.6%	80.0%	78.7%	78.4%	79.4%	79.1%	79.7%
eLBP L_4	7.2%	11.8%	27.7%	42.3%	52.3%	60.0%	66.1%
eLBP	90.6%	91.9%	92.4%	92.4%	91.8%	91.6%	91.6%

results based on single scale eLBP-DF surpass that of SI faces; furthermore, when fusing the matching scores of eLBP-DF at different scales to achieve MS-eLBP-DFs, the rank-one recognition rate is increased by more than 5 points, from 91.8% for SI face to 97.6% for MS-eLBP-DFs. These results clearly indicate how well Multi-Scale eLBP describes geometric shape variations.

3.4.3 Verification

The proposed approach was evaluated for face verification as well; the verification rate with FAR at 0.1% using the experimental protocol Neutral vs. All are given in Table 3.3. Further experiments were carried out on ROC I, ROC II, and ROC III. These three ROC curves are based on three masks provided by the FRGC v2.0 dataset. They are defined over the square similarity matrix with a dimensionality of 4007×4007, and they are of increasing difficulty reflecting the time elapsed between

Table 3.2: Rank-one face recognition rate compared with the state of the art on the FRGC v2.0 dataset.

Different Approaches	Recognition Rate
(1) ICP	72.2%
(2) SI Faces	91.8%
(3) MS-LBP-DFs	94.1%
(4) Wang et al. [Wang *et al.* 2007b]	87.7%
(5) Chang et al. [Chang *et al.* 2005a]	91.9%
(6) Mian et al. [Mian *et al.* 2008]	93.5%
(7) Cook et al. [Cook *et al.* 2006a]	94.6%
(8) Huang et al. [Huang *et al.* 2010b]	96.1%
(9) Mian et al. [Mian *et al.* 2007]	96.2%
(10) Kakadiaris et al. [Kakadiaris *et al.* 2007]	97.0%
(11) Faltemier et al. [Faltemier *et al.* 2008a]	98.1%
(12) Wang et al. [Wang *et al.* 2010]	98.4%
(13) Alyuz et al. [Alyuz *et al.* 2010]	97.5%
(13) MS-eLBP-DFs	97.6%

probe and gallery data. Their comparisons are shown in Table 3.4.

Table 3.3: Face verification rates using the experimental protocol of Neutral vs. All at 0.1% FAR on the FRGC v2.0 dataset.

Different Approaches	Verification Rate
SI Faces	94.4%
MS-LBP-DFs	96.1%
MS-eLBP-DFs	98.4%
Maurer et al. [Maurer *et al.* 2005]	92.0%
Passalis et al. [Passalis *et al.* 2005]	85.1%
Husken et al. [Husken *et al.* 2005]	89.5%
Cook et al. [Cook *et al.* 2006a]	95.8%
Mian et al. [Mian *et al.* 2008]	97.4%
Mian et al. [Mian *et al.* 2007]	98.5%
Wang et al. [Wang *et al.* 2010]	98.6%

As we can see from Table 3.4, the performance of the proposed method on ROC I, ROC II and ROC III is slightly lower but still comparable to the best of ones in the literature. Meanwhile, it is noteworthy that the proposed approach does not require any registration for nearly frontal face scans as those in the FRGC v2.0 dataset. This is clearly in contrast to works [Wang *et al.* 2010, Kakadiaris *et al.* 2007]. In [Kakadiaris *et al.* 2007], Kakadiaris et al. used ICP-based alignment with a coarse

to fine strategy for preprocessing, and two kinds of features, i.e. Haar and Pyramid were extracted from normal and geometry maps. In [Wang *et al.* 2010], a self-dependent registration step was used, and a large training database with neutral and non-neutral facial scans was required to learn inter-class and intra-class distributions. Moreover, they fused multiple information i.e. Haar-, Gabor- and LBP-based facial features to make the final decision.

Table 3.4: Comparison of verification rates at 0.1% FAR using ROC I, ROC II, ROC III and All VS. All protocol on the FRGC v2.0 dataset.

Different Approaches	ROC I	ROC II	ROC III	All vs. All
Maurer et al. [Maurer *et al.* 2005]	NA	NA	92.0%	87.0%
Cook et al. [Cook *et al.* 2006a]	93.7%	92.9%	92.0%	92.3%
Husken et al. [Husken *et al.* 2005]	NA	NA	89.5%	NA
Faltemier et al. [Faltemier *et al.* 2008a]	NA	NA	94.8%	93.2%
Kakadiaris et al. [Kakadiaris *et al.* 2007]	97.3%	97.2%	97.0%	NA
Mian et al. [Mian *et al.* 2007]	NA	NA	NA	86.6%
Alyuz et al. [Alyuz *et al.* 2010]	85.4%	85.6%	85.6%	NA
Wang et al. [Wang *et al.* 2010]	98.0%	98.0%	98.0%	98.1%
MS-eLBP-DFs	95.1%	95.1%	95.0%	94.2%

Table 3.5: Rank-one face recognition rates using the facial expression protocols on the FRGC v2.0 dataset.

Different Approaches	Subset I	Subset II	Degradation
SI Faces	97.2%	84.1%	13.1%
MS-LBP-DFs	97.7%	88.9%	8.8%
MS-eLBP-DFs	99.2%	95.1%	4.1%
Huang et al. [Huang *et al.* 2010b]	99.1%	92.5%	6.6%
Mian et al. only 3D [Mian *et al.* 2008]	99.0%	86.7%	12.3%
Mian et al. 2D + 3D [Mian *et al.* 2008]	99.4%	92.1%	7.3%

Subset I: Neutral vs. Neutral
Subset II: Neutral vs. Non-neutral

3.4.4 Evaluation on Facial Expression Variations

According to the same experimental protocol, we compared the performance of the proposed approach with the one by Mian et al. [Mian *et al.* 2008] and Huang et al. [Huang *et al.* 2010b] for robustness analysis on facial expression changes (see Table 3.5). The results of our approach are 99.2% and 95.1% for the Neutral vs. Neutral

and Neutral vs. Non-Neutral experiment, respectively. The face recognition rate on Subset I is comparable to the best ones of the state-of-the-art, while significant progress was made on subset II, i.e. Neutral vs. Non-Neutral, displaying a 95.1% rank-one recognition rate. The degradation as the samples with non-neutral facial expressions are included drops by 4.1%, not only much lower than the figure of 12.3% in [Mian *et al.* 2008] when only 3D shape information was utilized, but also lower than the drop of 7.3% in [Mian *et al.* 2008] when both shape and texture clues were combined. Moreover, Table 3.5 also indicates that the MS-eLBP-DFs outperform SI face in both the additional experiments on Subset I and II, and the performance degradation of MS-ELBP-DFs is much lower than that of SI face. These accuracies hence suggest that our approach tends to be insensitive to facial expression changes. The achieved verification rates match those in face recognition, as illustrated in Table 3.6.

Table 3.6: Comparison of verification rates at 0.1% FAR using the facial expression protocols on the FRGC v2.0 dataset.

Different Approaches	Subset I	Subset II	Degradation
SI Faces	98.9%	87.5%	11.4%
MS-LBP-DFs	99.1%	91.9%	7.2%
MS-eLBP-DFs	99.6%	97.2%	2.4%
Passalis et al. [Passalis *et al.* 2005]	94.9%	79.4%	15.5%
Mian et al. only 3D [Mian *et al.* 2008]	99.9%	92.7%	7.2%
Mian et al. 2D + 3D [Mian *et al.* 2008]	99.9%	96.6%	3.3%

Subset I: Neutral vs. Neutral
Subset II: Neutral vs. Non-neutral

Figure 4.7 indicates the verification rates by the ROC curves in the three experiments: Neutral vs. All, Neutral vs. Neutral, as well as Neutral vs. Non-neutral.

3.4.5 Evaluation on Data Degradation

As discussed in the subsection 3.4.1, we tested the robustness of the proposed method to some types of face data degradation, i.e., Gaussian noise, decimation, and random holes. To show the effectiveness of our approach, we compared the performance with several other tasks, including ICP, TPS, and Elastic Radial Curve Matching [Drira *et al.* 2010]. All the techniques were benchmarked on the subset of 410 subjects as defined previously in subsection 3.4.1 as well as various degradations. In the Figure 3.9 to 3.11, the proposed method is marked in red; ICP is marked in violet; TPS is marked in green; and Elastic Radial Curve Matching is marked in blue.

Figures 3.9 to 3.11 present the comparisons of rank-one recognition rates under Gaussian noise, decimation and missing data respectively. As we can see from these figures, to some extent, all these methods record performance drops under degra-

Table 3.7: Comparisons of rank-one face recognition rates on the Gavab database: (A) without pose variations; (B) only with pose variations.

Different Approaches	I. Neutr.	II. Expr.	I + II
Li et al. [Li *et al.* 2009]	96.67%	93.33%	94.68%
Moreno et al. [Moreno *et al.* 2005]	90.16%	77.90%	NA
Mahoor et al. [Mahoor & Abdel-Mottaleb 2009]	95.00%	72.00%	78.00%
Berretti et al. [Berretti *et al.* 2006]	94.00%	81.00%	84.25%
Mousavi et al. [Mousavi *et al.* 2008]	NA	NA	91.00%
Drira et al. [Drira *et al.* 2010]	100.00%	NA	94.67%
MS-eLBP-DFs	100.00%	93.99%	95.49%

(A)

Different Approaches	(a)	(b)	(c)	(d)	(e)
Mahoor et al. [Mahoor & Abdel-Mottaleb 2009]	85.30%	88.60%	NA	NA	NA
Berretti et al. [Berretti *et al.* 2006]	80.00%	79.00%	NA	NA	NA
Drira et al. [Drira *et al.* 2010]	100.00%	98.36%	70.49%	86.89%	88.94%
MS-eLBP-DFs	96.72%	96.72%	78.69%	93.44%	91.39%

(B)

(a): Looking down
(b): Looking up
(c): Right profile
(d): Left profile
(e): Overall

dations in face model quality. Specifically, they resist relatively well to decimation and missing data, their performance drops drastically when the noise is increased to 0.8mm. However, compared with other three methods, the proposed approach shows a better tolerance to these data degradations.

3.4.6 Evaluation on Occlusion

To analyze its tolerance to occlusions, we also launched experiments on the Gavab database and calculated the rank-one recognition rates. Table 3.7 shows the matching accuracies for different categories of probe faces: (A) displays the results without pose variations; while (B) lists those only with the facial scans with pose changes. In (A), the neutral subset contains one frontal facial scan of each subject, and the expressive subset includes a smile; accentuated laugh and random gesture (random facial expression), three facial scans of each subject. To the best of our knowledge, the work [Drira *et al.* 2010] is the only one that has carried out experiments on the entire Gavab database before this work. Therefore, we compared our results with theirs on the subset of four severe pose variations as well as the overall performance. It is worth noting that the difference between their work and ours is that Drira et al.

manually landmarked nose tips on all the face scans in the dataset for an ICP-based fine registration, while we only manually landmarked facial scans of right and left profiles ((c) and (d) in Table 3.7 (B)), and for all the faces, only a coarse alignment step was utilized to rotate and translate them.

From Table 3.7 (A), we can see that for frontal neutral probes, the rank-one recognition rate is 100% as in [Drira *et al.* 2010]; while regarding expressive faces, our approach surpasses all the others. Moreover, when evaluating the robustness to severe pose variations (Table 3.7 (B)), we achieved an overall accuracy of 91.4% on these four subsets; whilst that reported by [Drira *et al.* 2010] is 88.9%.

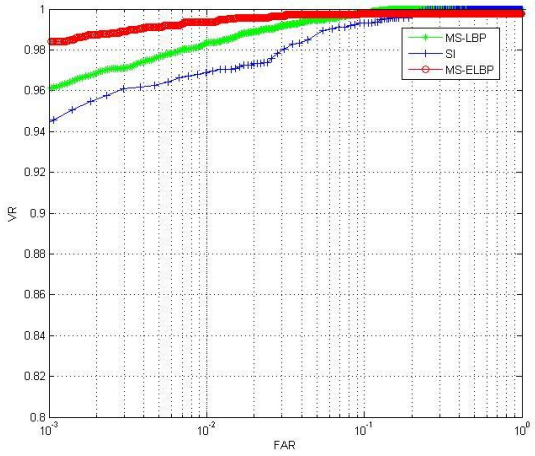
To sum up, the experimental performance on the Gavab database further confirms that our approach possesses a good tolerance to large pose changes and even partial occlusions.

3.4.7 Computation Cost

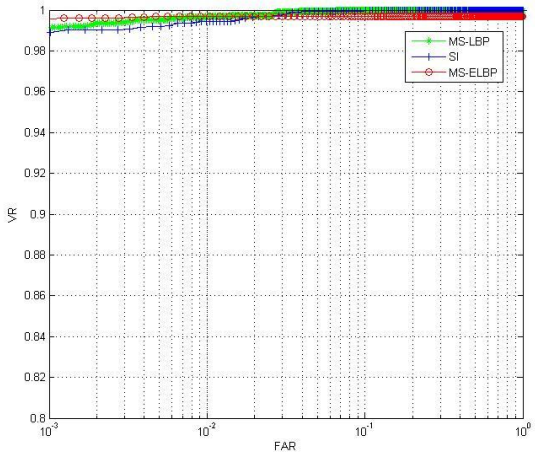
Currently, an unoptimized implementation of our approach with MATLAB (R2010a) can perform about 330 matches between the gallery and probe faces per second using a machine using Intel(R)Core(TM) i7 CPU (3.07 GHz) and 8 GB RAM.

3.5 Summary

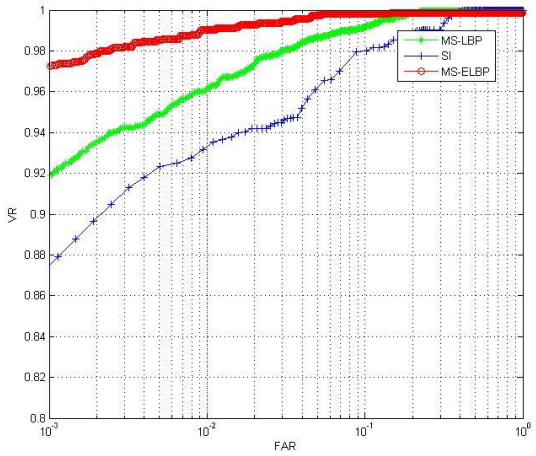
We have presented an effective method for 3D face recognition using a novel geometric facial representation and a local feature matching scheme. The proposed facial representation is based on MS-eLBP and allows for accurate and fast description of local shape variations, thereby enhancing the distinctiveness of generally smooth facial range images. SIFT-based local matching further improved by facial component and configuration constraints robustly associates keypoints between two faces of the same subject. The proposed method was evaluated in 3D face recognition and verification, achieving a recognition rate of 97.6% and a 98.4% verification rate with a 0.001 FAR respectively on the FRGC v2.0 database which consists of nearly frontal facial scans with rich expression changes. Experimental results on the Gavab database containing severe pose changes clearly illustrate that the entire system is also robust to partially occluded faces. Moreover, generally costly registration was not needed thanks to the relative tolerance of the proposed local matching strategy to moderate pose variations like the ones in the FRGC v2.0 dataset. When dealing with extreme poses, i.e. left or right profiles, a coarse alignment step based on a few manually landmarked points was sufficient in preprocessing as indicated by the experiments on the Gavab database. Finally, additional experiments demonstrate that the proposed method is not sensitive to facial expression variations and has a good tolerance to data degradations.



(A)



(B)



(C)

Figure 3.8: The ROC curves are depicted based on the SI Faces, MS-LBP-DFs, and MS-eLBP-DFs respectively in three experiments with neutral face models enrolled: (A) Neutral vs. All; (B) Neutral vs. Neutral; (C) Neutral vs. Non-neutral.

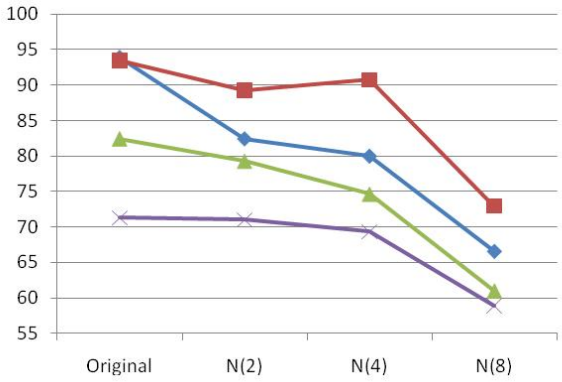


Figure 3.9: The comparison of rank-one recognition rates under Gaussian noise.

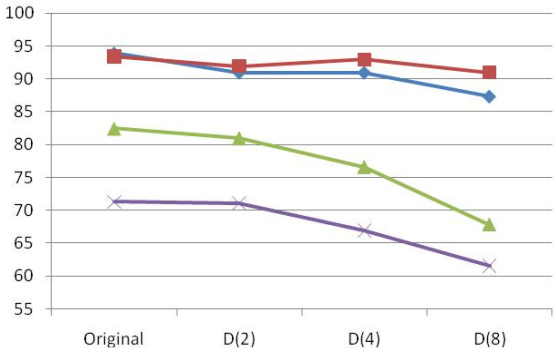


Figure 3.10: The comparison of rank-one recognition rates under decimation.

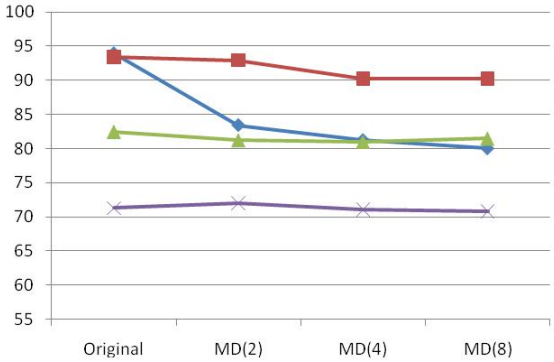


Figure 3.11: The comparison of rank-one recognition rates under missing data.

Chapter 4

Textured 3D Face Recognition

4.1 Introduction

In the previous chapter, we introduced a 3D face recognition method in order to cope with the problems usually encountered in 2D face recognition, caused by illumination and pose changes. However, even if it is so, the challenge of facial expression variations is even more difficult than in 2D modality, as 3D face models provide exact shape information of facial surfaces.

Since current 3D imaging systems generally deliver 3D face models along with their aligned texture counterpart, a major trend in the literature of face recognition is to adopt both the 3D shape and 2D texture based modalities, arguing that the joint use of these two clues can generally achieves more accurate and robust performance than using only either of the single modality. See section 4.2 for more information about the state of the art.

A majority of techniques in the literature for face recognition do not operate directly on original input facial images because all the human faces are similar and undergo the variations of illumination and pose. Instead, they try to search for an intermediate facial representation, for instance, eigenface [Turk & Pentland 1991], fisher face [Belhumeur *et al.* 1997], LBP face [Ahonen *et al.* 2004] etc., aiming to highlight intra-class similarity and inter-class dissimilarity.

In this chapter, we propose a novel biological vision-based facial representation, namely Oriented Gradient Maps (OGMs), which is applied to both facial range and texture images for textured 3D face recognition. These OGMs simulate the response of complex neurons to gradient information in a neighborhood and own properties of being highly distinctive and robust to affine lighting and geometric transformations. Based on this intermediate facial representation, SIFT-based local feature matching is then used to calculate similarity scores between probe and gallery faces. Because the facial representation generates an OGM for each of quantized orientations of facial range as well as texture images, we further propose a score level fusion scheme that optimizes weights by a genetic algorithm in a learning step. Evaluated on the complete FRGC v2.0 dataset, the rank-one recognition rate using only 3D or 2D modality is 95.5% and 95.9% respectively; while combining these two modalities, i.e. range- and texture-based OGMs, the final accuracy is 98.0%, demonstrating the effectiveness of the proposed biological vision-inspired facial description and the optimized weighted sum fusion. In addition, based on the 3DTEC dataset, we also explore the capacity of the proposed method for differential identical twins.

The remainder of this chapter is organized as: Section 4.2 depicts an extensively review on the related works. Section 4.3 introduces the biological vision-based facial representation that is applied on both facial range and texture images. The following

SIFT-based local feature matching step is shown in section 4.4. Section 4.5 presents a weighted sum rule based score fusion approach optimized by a genetic algorithm from a learning database. Section 4.6 analyzes and discusses the experimental accuracies achieved on the FRGC 2.0 and 3DTEC datasets. Section 4.7 concludes the paper.

4.2 Related Works

Although the first research work in the domain of 3D face recognition dates back to before 1990, multi-modal face recognition did not appear until the beginning of this century. Early efforts in this area investigate relatively simplistic ways to fuse the results obtained independently from 3D and 2D face data. For example, the most common approach is to apply an eigenface method to each of the 2D and 3D facial images separately, and then combine two matching measurements. More recently, a great number of studies appear to take a variety of quite different facial features as well as fusion schemes, and several commercial companies already have the ability to develop the software for multi-modal 2D and 3D face recognition.

Beumier and Acheroy [Beumier & Acheroy 2001] perform the multi-modal face recognition using a weighted sum rule (score level fusion). They exploit a central profile and a lateral profile, each in both 3D and 2D. They thus have a total of four classifiers, and an overall decision is made using a weighted sum of the similarity metrics. In this paper, results are reported on a subset of 3D_RMA, which consists of 27-person gallery and a 29-person probe faces. An EER of 1.4% is reported for multi-modal recognition, lower than either 3D or 2D modality alone.

Wang et al. [Wang *et al.* 2002] make use of Gabor filter coefficients in 2D and "point signatures" in 3D to achieve multi-modal face recognition. PCA is applied to each type of feature separately for dimensionality reduction. The generated 2D and 3D features together form a feature vector (feature level fusion). Classification is done by SVM. Experiments are carried out with images from 50 subjects, six images per subject, with pose and expression variations. Rank one recognition rate exceeding 90% is presented.

Chang et al. [Chang *et al.* 2003b] employ an eigenface based recognition to 3D and 2D images independently, their similarity scores are further combined with a weighted sum rule. The experiments tested on a 116-individual dataset show that the rank one recognition rates based on single modality are almost the same, but the fusion accuracy is much better than both of them. Chang et al. [Chang *et al.* 2003a] soon report another performance using the same method on a test set containing 676 probes from 200 individuals. The fusion result is up to 98.8% rank-one recognition for multi-modal; 92.8% for 3D alone; 89.5% for 2D alone. They [Chang *et al.* 2005c] further design a contrast test to take two 2D facial images in the gallery set, and calculate their similarity measurements with a given probe face respectively for a final fusion, concluding that multi-modal face recognition outperforms that based on multiple 2D enrollment .

Tsalakanidou et al. [Tsalakanidou *et al.* 2003] also explore a PCA based method for 3D and 2D image based face recognition. Unlike other tasks, they utilize YUV components of color images instead of gray-scale ones in 2D domain and a product

fusion rule to replace the sum rule for the combination of two similarity measurements. The recognition rate using images of 40 persons from the XM2VTS dataset are as high as 99% for the multi-modal algorithm, which is found to be higher than for either 3D or 2D clue alone.

Bronstein et al. [Bronstein *et al.* 2003] investigate an isometric transformation to map 3D facial surfaces into 2D canonical images using MDS, trying to better handle changes due to facial expression. Based on the correspondence between 2D facial image pixels and 3D surface vertices, facial texture images can be projected onto the aligned canonical surface in the canonical form space as well, and flattened texture is thus produced. Eigen-decomposition is applied to flattened textures and canonical images for multi-modal recognition. They show examples of correct and incorrect recognition by different algorithms, but do not mention any overall quantitative results. Later, Bronstein et al. [Bronstein *et al.* 2004a] further enhance this work. Using multiple 2D facial images, they firstly reconstruct 3D facial surfaces and their albedos. With re-lighted facial texture images and generated canonical images, multi-modal face recognition can be achieved.

Papatheodorou and Rueckert [Papatheodorou & Rueckert 2004] propose an ICP variant based on point distances in a 4D space (x, y, z, intensity) for multi-modal face recognition. This method integrates shape and texture information at an early stage, rather than fuse them at score level or decision level. They compute results from experiments with frontal neutral-expression models of 62 subjects in the gallery, and probe sets of different pose and facial expressions. They report 98% to 100% correct recognition in matching frontal neutral-expression probes. Recognition drops when the expression and pose of the probe faces is varying from those of the gallery images.

Tsalakanidou et al. [Tsalakanidou *et al.* 2004] present a method to multi-modal face recognition based on an embedded hidden Markov model (EHMM) for each modality. Meanwhile, based on 3D face models, they synthesize facial range and texture images to enlarge the gallery set. Their experimental results prove that combining multiple modalities and increasing enrolled images are both effective solutions to improve face recognition performance. A longer version of this work appears in [Tsalakanidou *et al.* 2005].

Malassiotis and Strintzis [Malassiotis & Strintzis 2004] describe a face recognition system using a combination of color and depth images. To cope with lighting and pose variations, 3D information is used for correcting pose and recovering illumination to generate the samples with upright orientation and frontal lighting. Computed on a dataset containing 2200 faces from 20 individuals, the experimental results show that multi-modal performance improves the accuracy of each single modality.

Zhang et al. [Li *et al.* 2005b] implement a systematic framework, which extracts LBP based features from 2D and 3D modalities for a fusion at both feature and decision levels, by exploring synergies of the two modalities at these levels. A Boosting statistical learning procedure is used for feature selection and classifier. Experiments are carried out on a database consisting of 252 individuals, and achieved results show that the fusion at both levels yields significantly better performance than fusion at the decision level. Moreover, LBP features prove more discriminative than eigenface

based ones.

Godil et al. [Godil *et al.* 2005] make a comprehensive evaluation on the PCA based method for 3D + 2D multi-modal face recognition using the CAESAR anthropometric database, which includes 200 persons. They compare the performance of feature and score level fusion, and analyze the impact on results of two normalization methods (Max-Min and Z-Sore) as well as four fusion schemes (Sum, Max, Min, and Product Rule). Their best rank-one recognition rate reported in this paper is 82%.

Colombo et al. [Colombo *et al.* 2005] exploit a curvature based method to detect 3D faces and locate a few initial landmarks for pose correction. A PCA based holistic matching is then applied on both facial texture and rang images extracted from aligned 3D face models to achieve multi-modal face recognition. They compare the experimental results using three different fusion strategies, i.e. feature level fusion, score level fusion as well as decision level fusion, and point out all these three fusion accuracies are better than the ones only based on single modality. Moreover, in their case, feature level fusion outperform the other two in the scenario of verification, while score level fusion obtains the best result in identification.

Husken et al. [Husken *et al.* 2005] describe a system (from Viisage company) to multi-modal recognition. The 3D matching follows the style of Hierarchical Graph Matching (HGM) already implemented for 2D face recognition, which allows faster matching in comparison to the ICP method or similar iterative techniques. Combining the results from the two clues is done at the score level, and they reported a 93% verification at 0.01 FAR as multi-modal performance on the FRGC v2.0 database, which is better than either of single modality based result. In addition, they also claim that the fusion accuracy is closely related to statistical independence between the two clues. The more independent the matching scores of the algorithms are, the higher the benefit of the fusion, as errors in the one algorithm can be counterbalanced by the other one.

Maurer et al. [Maurer *et al.* 2005] develop the Geometrix system for textured 3D face recognition. The 3D face matching builds on the method proposed by Medioni and Waupotitsch [Medioni & Waupotitsch 2003], whereas the 2D matching adopts the method of Neven Vision [Okada *et al.* 1998]. A weighted sum rule is then used to fuse the two similarity measurements. The system is evaluated on the FRGC v2.0 dataset. The facial expression changes are categorized into "neutral", "small", as well as "large", and results are presented separately for these three categories. Multi-modal performance for the "all versus all" matching protocol of the 4007 facial images reaches approximately 87% verification rate at 0.01 FAR. They also conclude that using both texture and shape information outperforms 3D alone by a noticeable increment, and that the verification rates for 2D alone are below those for 3D alone.

Lu and Jain [Lu & Jain 2005b] extend their earlier study for 3D facial surface matching [Lu *et al.* 2004a] to multi-modal 3D + 2D face recognition. In contrast to other related tasks, two modalities are operated in a serial way instead of a parallel one. The surface matching component is firstly carried out based on ICP with a coarse-to-fine strategy, and the candidate list used for 2D matching is dynamically generated based on the output of surface matching, in order to reduce the complexity of the 2D matching stage. The 3D model in the gallery is employed to synthesize

new 2D facial images with pose and illumination variations that are then imported into an LDA classifier. The weighted sum rule is finally applied to combine the two matching components. Experimental results are given on a database of 100 3D face models with 598 2.5D test scans acquired in different pose and lighting conditions, and with some smiling expression, illustrating the effectiveness of the fusion.

Cook et al. [Cook *et al.* 2005] introduce Gabor filter based features to represent both facial texture and range images. Then, the PCA approach is applied to reduce their dimensionality and de-correlate the data. The scores of both modalities are finally combined for multi-modal face recognition. The UND database containing 943 individual captured from 277 subjects is taken for evaluation. The lowest EER reported in this work is 0.66%, showing that frequency bands are more discriminative than original face data and combining both clues achieves better results. Meanwhile, they also analyze the performance using some facial parts as well as that based on different components of the Gabor filters, such as real, mag, imag, phase and some possible combinations. In the next year, Cook et al. [Cook *et al.* 2006b] strengthen this method by bringing in a local matching strategy and SVM classifier. Tested on the FRGC v2.0 dataset, the best verification rates (ROC III) at 0.1% FAR using 2D alone, 3D alone and multi-modal are 39.37%, 79.66%, and 82.72% respectively, all of which are based on the Mahalanobis Cosine Average (MahCosAvg) distance.

Based on the PCA algorithm, Yuan et al. [Yuan *et al.* 2005] also project all the 2D texture images and 3D shape images into a low-dimensional subspace. In the following step, they apply fuzzy clustering and parallel neural networks to recognize faces. Experimental results are computed on a dataset composed of 70 persons with different pose changes, they demonstrate that using 3D data, pose can be efficiently corrected, however, they do not mention any performance to show the effectiveness of the combination of both modalities as well as the accuracy difference between commonly used nearest neighbor based classifier and the proposed neural networks.

Abate et al. [Abate *et al.* 2006a] address multi-modal face recognition through 3D geometry and 2D texture features by means of two different metrics: Augmented Normal Map and PCA. Augmented Normal Map includes shape (surface normals represented as 24 bit color pixels) and texture info (additional 8 bit for skin color) into a 32 bit image. The method operates in a hierarchical way. It firstly performs a fast one-to-many comparison of facial geometry exploiting normal map metric. Then, to further improve recognition precision and reliability, best rank faces are compared to probe by PCA resulting in a final decision. They present preliminary experimental results on a dataset of 101 textured 3D faces.

Mian et al. [Mian *et al.* 2006a] combine four categories of similarity measurements to recognize textured 3D faces, i.e. eye and forehead region (3D), nose region (3D), entire range face (3D), and entire texture face (2D). The former two are computed using ICP based registration, whilst the latter two are calculated by PCA holistic matching. Product rule is adopted for final fusion, which achieves a 100% verification rate at 0.06% FAR on the FRGC v1.0 dataset. Mian et al. [Mian *et al.* 2007] extend this work to a more complex and robust system as shown in Fig. 4.1. To calculate similarity scores between 3D facial surfaces, Spherical Face Representation (SFR) is proposed, which can be imagined as the quantization of 3D facial point-cloud into spherical bins centered at the nose tip. SIFT is applied

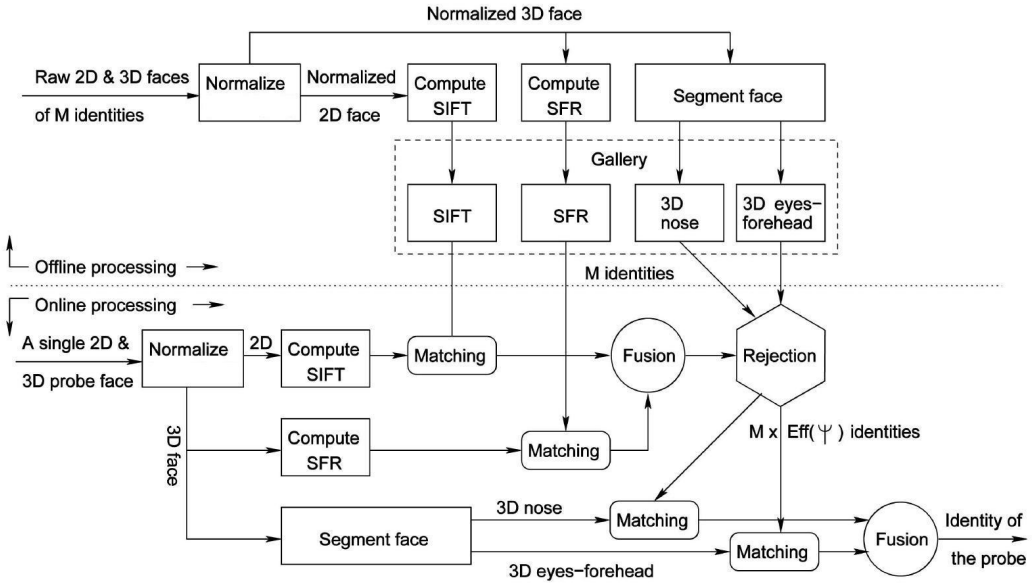


Figure 4.1: The block diagram of the recognition algorithm proposed by Mian et al [Mian *et al.* 2007].

directly to compute the accuracy of 2D part. A weighted sum rule is then used to fuse both measurements for a coarse classification. Possible candidates are finally verified using the previously proposed ICP based facial rigid region (eye, forehead and nose region) matching. Evaluated on the FRGC v2.0 dataset with the experimental protocol of Neutral vs. All, they report a 97.4% rank one recognition rate and a 99.3% verification rate at 0.1% FAR.

Mian et al. [Mian *et al.* 2008] implement another feature-based multi-modal face recognition system aiming to overcome the variations of facial expressions, illumination, pose, occlusions and make-up. For facial shape matching, they first propose a technique which can repeatably detect keypoints at locations where shape variation is high in 3D faces, and a unique 3D coordinate basis then can be defined locally at each keypoint facilitating the extraction of highly descriptive pose invariant features. A 3D feature is extracted by fitting a surface to the neighborhood of a keypoint and sampling it on a uniform grid, and features from a probe and gallery face are finally projected to the PCA subspace and matched. The set of matched keypoints are used to build two graphs from which the similarity between two faces is measured. In the 2D domain, they employ the SIFT features and performed fusion at the feature and score-level. The proposed algorithm achieved 96.1% identification rate and 98.6% verification rate at 0.1% FAR on the complete FRGC v2.0 dataset.

Different from most existing methods of 2D + 3D multi-modal face recognition that do fusion at the late-stage feature, score, or decision level, Kusuma and Chua [Kusuma & Chua 2008] propose an image-level fusion method that explores to reduce the dependency between modalities for face recognition. Facial cues from 2D

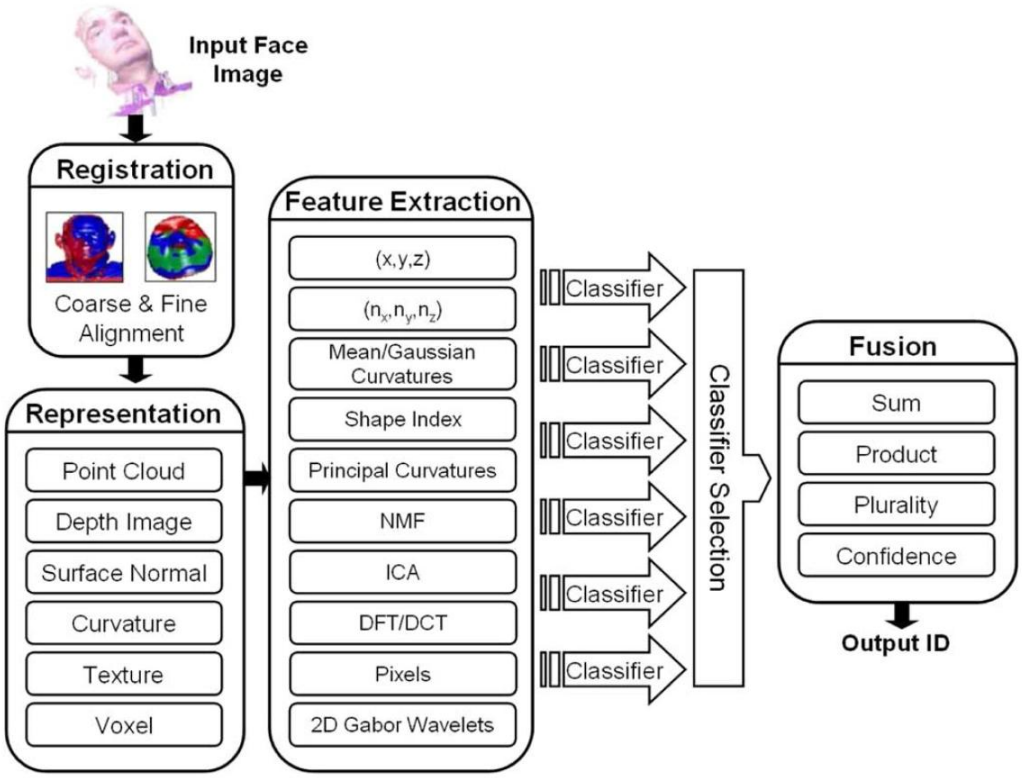


Figure 4.2: Overall structure of the face recognition system proposed by Gokberk et al [Gokberk *et al.* 2008].

and 3D images are combined into more independent and discriminating data by finding fusion axes that pass through the most uncorrelated information in the images. Experimental results based on a database of 1280 textured 3D facial samples from 80 persons show that the proposed image-level fusion outperforms the feature- and score-level fusion.

Gokberk et al. [Gokberk *et al.* 2008] design a comprehensive system to present an extensive study of 3D face recognition algorithms and examine the benefits of various score-, rank-, and decision-level fusion rules (Fig. 4.2). They discuss various feature extraction techniques applied to both texture and shape modalities, and compare different classifier combination methods. They also present a dynamic confidence estimation algorithm to boost fusion performance. In experiments performed on the FRGC v1.0 and v2.0 databases, they give answers to: 1) the relative importance of the face data format to the types of features extracted; 2) the impact of the gallery size; 3) the conditions, under which subspace methods are preferable, and the compression factor; 4) the most advantageous fusion level and fusion methods; 5) the role of confidence votes in improving fusion and the style of selecting experts in the fusion; and 6) the consistency of the conclusions across different databases.

Xu et al. [Xu *et al.* 2009] research into the contributions that depth and intensity

information makes to face recognition when facial expression and pose changes are taken into account. In their system, local facial features described by Gabor wavelets are extracted from corresponding depth and intensity images generated from 3D face models after fine alignment. Then a hierarchical selecting scheme embedded in LDA and AdaBoost is designed to select the most effective and most robust features to further construct a strong classifier. Experiments are performed on the FRGC v2.0 database, and using the protocol of Neutral vs. All, a verification rate of 97.5% at 0.1% FAR is reported.

4.3 Biological Vision-based Facial Description

In order to improve the distinctiveness of human faces and offering certain tolerance to lighting and pose changes, in this section, we introduce a novel biological vision-based facial description which can be applied to both facial range and texture images.

The proposed facial representation is inspired by the study of Edelman et al. [Edelman *et al.* 1997], who proposed a representation concept of complex neurons in primary visual cortex. These complex neurons respond to a gradient at a particular orientation and spatial frequency, but the location of gradient is allowed to shift over a small receptive field rather than being precisely localized. Our facial representation implements this idea into practice.

4.3.1 Description of The Complex Neuron Response

The proposed description aims at simulating the response of complex neurons, based on a convolution of gradients in specific orientations in a pre-defined circular neighborhood. The radius value can be varied experimentally for different applications.

Specifically, given an input image I, a certain number of gradient maps L_1, L_2, ..., L_o, one for each quantized direction o, are first computed. They are defined as:

$$L_o = \left(\frac{\partial I}{\partial o}\right)^+ \tag{4.1}$$

The "+" means that only positive values are kept to preserve the polarity of the intensity changes, while the negative ones are set to zero.

Each gradient map describes gradient norms of the input original image in an orientation o at every pixel. We then simulate the response of complex neurons by convolving its gradient maps with a Gaussian kernel G, and the standard deviation of G is proportional to the value of radius of the given neighborhood area, R, as in eq. 4.2.

$$\rho_o^R = G_R * L_o \tag{4.2}$$

The purpose of the convolution with Gaussian kernels is to allow the gradients to shift in a neighborhood without abrupt changes.

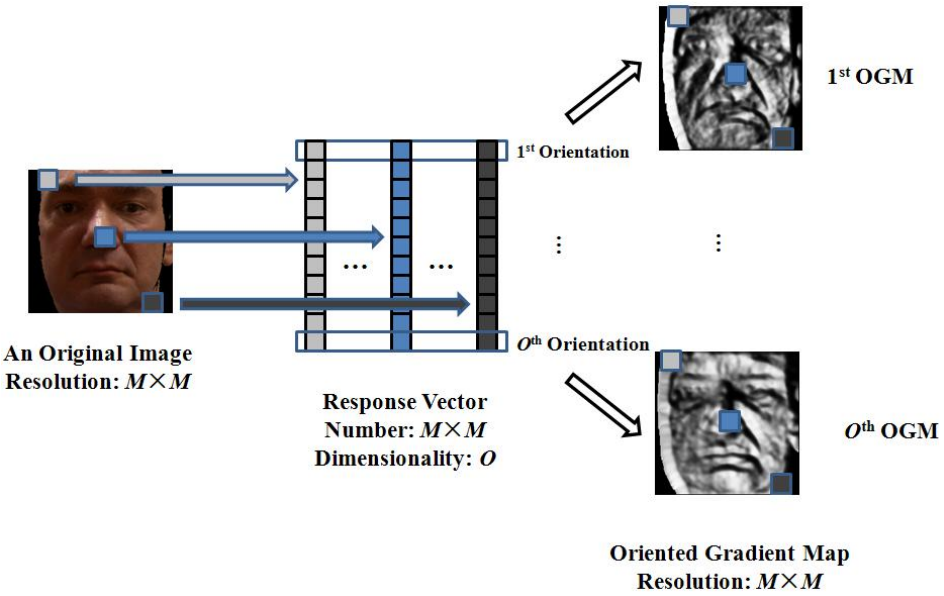


Figure 4.3: An illustration of the oriented gradient map for each of the quantized orientations o.

At a given pixel location (x, y), we collect all the values of the convolved gradient maps at that location and form the vector $\rho^R(x, y)$, and it thus has a response value of complex neurons for each orientation o.

$$\rho^R(x, y) = \left[\rho_1^R(x, y), \cdots, \rho_O^R(x, y)\right]^t \tag{4.3}$$

This vector, $\rho^R(x, y)$, is further normalized to unit norm vector, which is called response vector and denoted by $\underline{\rho}^R$ in the following parts of this chapter.

4.3.2 Facial Description by Response Vectors

Now facial range and texture images can be represented by their perceived values of complex neurons according to their response vectors. Specifically, given a facial range or texture image I, we can generate a new Oriented Gradient Map (OGM) J_o using complex neurons for each orientation o defined as:

$$J_o(x, y) = \underline{\rho}_o^R(x, y) \tag{4.4}$$

Fig. 4.3 shows such a process. In our work, we generate 8 OGMs for 8 quantized directions respectively. Instead of the original facial range and texture images, the OGMs are further fed to SIFT-based local feature matching for face recognition.

4.3.3 The Properties of Distinctiveness and Invariance

The generated OGMs potentially offer high distinctiveness because they highlight the details of local shape and texture variations. Meanwhile, they also possess the

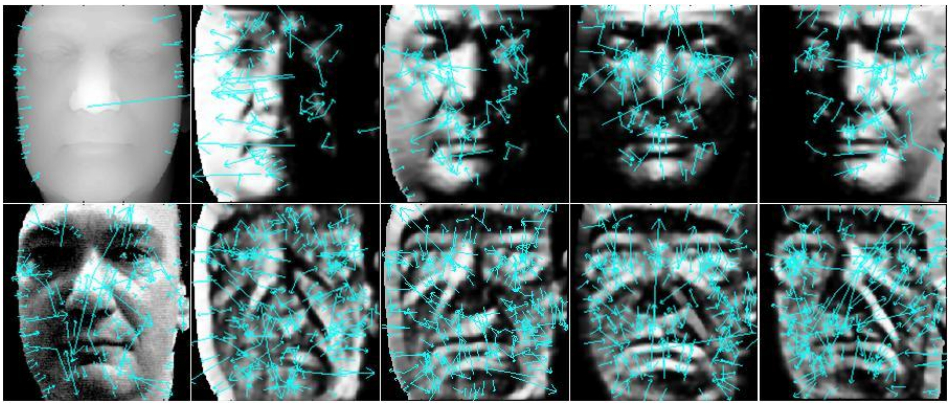


Figure 4.4: SIFT-based keypoint detection examples. The upper row lists an original facial range image and its oriented gradient range maps for the first four orientations; while the bottom one displays an original facial texture image as well as its first four oriented gradient texture maps.

property of being robust to affine lighting and geometric transformations.

As applied to facial texture images, the OGMs offer the property of being robust to affine illumination transformations. Indeed, an OGM, J_o, is simply the normalized convolved gradient map of the facial image at a given orientation o according to eq. 4.4, while affine lighting variations usually add a constant intensity value on images, as a result, it does not affect the computation of gradients. On the other hand, a change of image contrast in which the intensities of all the pixels are multiplied by a constant will result in the multiplication of gradient calculation; however, this contrast change will be cancelled by the normalization of the response vector.

Similarly, the OGMs of facial range images which contain 3D shape information are also invariant to affine geometric transformation leading to certain tolerance to pose changes.

The proposed OGMs can be made even rotation invariant if we choose to quantize directions starting from the principal direction of all gradients in the neighborhood. Nevertheless, we do not perform such a rotation normalization step to save computational cost, since 3D face models are generally in an upright frontal position in user cooperative applications.

4.4 SIFT based Local Feature Matching

When the OGMs of all quantized orientations are achieved from both range and texture facial images, a local feature matching step is carried out on these widely-used SIFT based features [Lowe 2004] extracted from OGMs for similarity score calculation, since it is well known that local feature based matching scheme is generally more robust to occlusion and pose changes.

SIFT operates on each OGM separately. As OGMs highlight local texture and shape changes of generally smooth facial images, many more keypoints can be detected than the ones when we directly apply SIFT on the original facial range and

texture images. On the FRGC v2.0 database, it shows that the average number of keypoint detected on original facial range and texture images are only 41 and 67 respectively; while using OGMs for keypoint detection, the average numbers rise up to 116 and 304. Figure 4.4 shows an example of SIFT-based keypoint detection.

The similarity measurement based on the number of matched points between two facial range or texture OGMs, $N_{R(R,o)}$ or $N_{T(R,o)}$, is with a positive polarity (a bigger value means a better matching).

4.5 Optimized Weighted Score Sum Fusion

Given a textured 3D face model as a probe, the previous SIFT-based local matching produces a similarity measure in each orientation of both facial texture and range images in the gallery set. These similarity measurements need to be further fused to deliver a final similarity score. In this chapter, we develop a fusion scheme by using a weighted sum rule, as score level fusion has been extensively used in the literature of 2D/3D face recognition and has proved its efficiency in a number of non-trivial pattern recognition problems [Ruta & Gabrys 2005].

Formally, a weighted sum rule is defined as:

$$S = \sum_{i=1}^{N} w_i * S_i \tag{4.5}$$

where S_i is a similarity score; w_i is its corresponding weight; and N is the number of modalities used for generating similarity scores. A bigger weight value indicates a higher importance; and a smaller one indicates a lower importance.

The proposed weighting strategy is learning-based, by using a genetic algorithm [Said 2005] seeking an optimal set of weights through applications of selection, mutation, and recombination of a population.

Figure 4.5 shows the process of learning optimal weighting. First, a population is created by randomly generating individual "chromosomes". The chromosome length is the same with the number of variables (weights) corresponding to the number of similarity score matrices. Given N similarity measurement matrices generated by the SIFT-based matching on different OGMs, each of chromosomes thus possesses N gene positions representing N different weights. At each iteration, a normalization process is first carried out to keep the sum of all the weights as one.

The chromosomes are used to encode trial solutions in a genetic algorithm. Iterative selection, crossover, and mutation are then exploited to make evolution of the population. At each generation, a new set of chromosomes is generated based on the fittest genes of previous generation to achieve a better solution. This fitness is calculated according to the produced similarity measure matrix in terms of recognition accuracy. Stochastic Universal Sampling [Baker 1987] is applied to select chromosomes and to generate offspring. The operation of crossover leads to generate better offspring by exchanging characteristics of their parents. It enables the most efficient characteristics to be concentrated in an individual. The mutation randomly varies the genetic representation of an individual by adding a random value and tends to

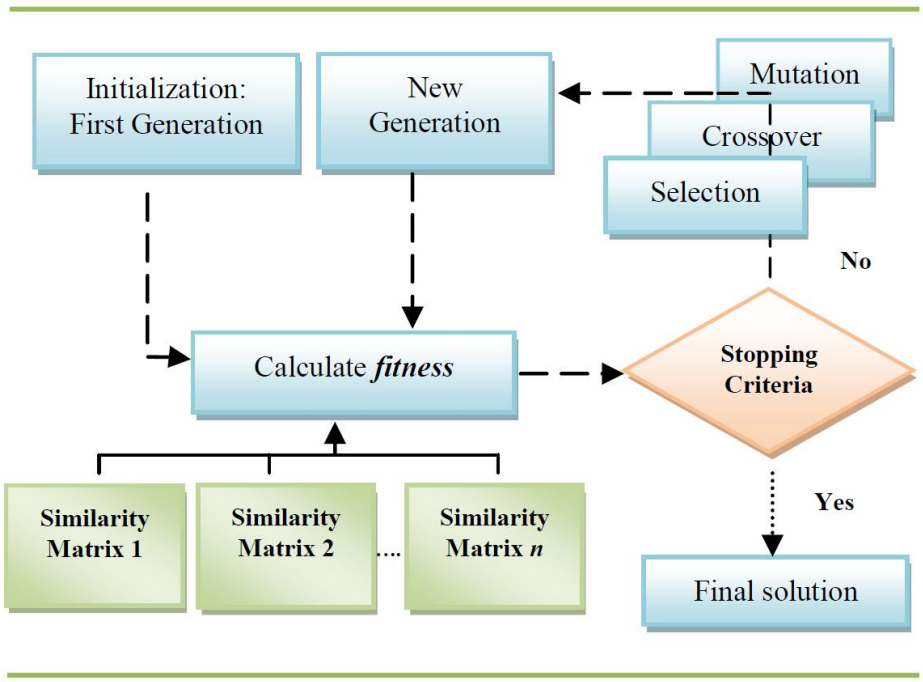


Figure 4.5: The process of the proposed learning strategy for optimized weighting.

inhibit the possibility of converging to a local optimum, rather than the global one. The evolution operates until a desired solution is achieved, or a pre-specified number of iterations are finished. The final solution with a higher fitness represents the best vector of weights.

4.6 Experimental Results

To evaluate the effectiveness of the proposed textured 3D face recognition system, we carried out the experiments on the FRGC v2.0 [Phillips *et al.* 2005] and 3DTEC datasets, and their general introductions can be found in the first chapter.

4.6.1 Experiment Design

The experiments were conducted mainly on FRGC v2.0, one of the most comprehensive databases. It contains 4007 textured 3D face models of 466 different subjects. One facial range image as well as its 2D texture counterpart were extracted from each 3D face model. For 3D face data, a preprocessing step was introduced to remove spikes with a median filter and fill holes by using cubic interpolation. For 2D facial texture images, histogram equalization was used to reduce influence caused by non-affine lighting variations. Due to the properties of the OGMs and the use of local feature-based matching, we did not perform any registration on 3D face models since they are all in a nearly frontal pose. It is different from the work in [Mian *et al.* 2007, Xu *et al.* 2009] etc. which also made use of both facial range and

texture data. The faces were cropped by using a basic bounding box based on the mask provided by a 3D scanner indicating if the point is valid or not in that position. Cropped faces thus have pose, lighting and expression changes as well as occlusions caused by hair. The FRGC v1.0 database consisting of 943 neutral expression face models was adopted by our adaptive fusion approach for training optimized weights which were then used for test stage.

In the experiments, we adopted the same protocol of the state of the art for the purpose of fair comparison. The first 3D face model with a neutral facial expression of each subject made up a gallery set of 466 subjects. The remaining face models (4007-466=3541) were treated as probes.

We preliminarily explore the distinctiveness of some basic features extracted from 2D facial texture images or 3D face models as well as their possible complementary contribution in a fusion process for face recognition.

We further designed three experiments for the proposed OGM based method: the first is to discuss the impact of the neighborhood area radius R on the final performance; the second is to test the proposed approach both in face recognition and verification compared with the state of the art; the last one is to evaluate the robustness of the proposed approach to facial expression variations.

In addition, we carried out the experiments on the 3DTEC dataset to investigate the power of the proposed method in distinguishing from twins. The 3DTEC dataset consists of 107 pairs of identical twins, each of which owns two 3D face models with the same format as the ones in FRGC v2.0. The same preprocessing operations are thus adopted. We computed the performance in the scenarios of face recognition and verification as well.

4.6.2 Basic Feature Comparison

In this experiment, we discuss the discriminative ability of several basic individual facial features. From 2D facial texture images, original intensity-, Gabor filter- and LBP- based features are extracted. While in 3D part, we consider geometric features containing normal, binormal, tangent vector and curvature related features which have the potential for a higher accuracy to describe surface based events and are well suited to represent the properties of facial regions, such as two cheeks, forehead, and chin. Four categories of curvature-based features are utilized. The first two types rely on the principal directions corresponding to the maximum and minimum curvatures [Tanaka *et al.* 1998]. The latter two are their derivatives, i.e., mean (H) and Gaussian (K) curvatures [Maurer *et al.* 2005]. Furthermore, we study another type of 3D features based on the anthropometric approach which advocates extracting a signature from some anthropometric points considered the most relevant; these points should be stable and discriminative. Chang et al. [Chang *et al.* 2006] have illustrated that the face region surrounding nose was very stable. In [Arca *et al.* 2007], Arca et al. used segments of 2D face around the eyes and nose to extract 3D profiles. Inspired by these tasks and the one by Perrot [Perrot 1996] on facial anthropometric measurements, a feature vector is generated including distance representing the edge length of every two anthropometric points; the angle formed by two edges with one same endpoint; and the ratio between horizontal nose width and vertical nose

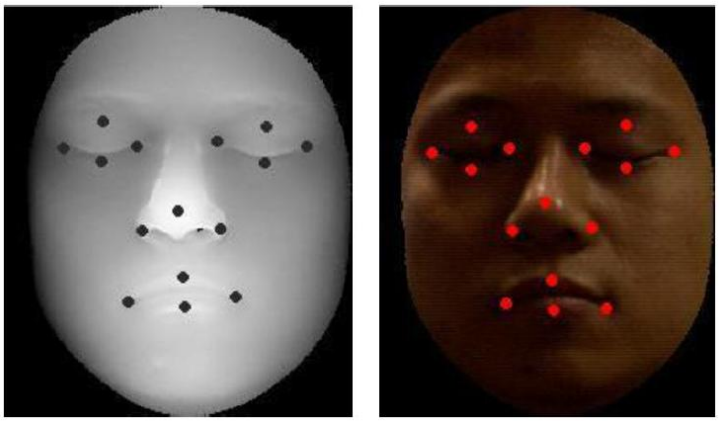


Figure 4.6: The preprocessed face sample with manual landmarks.

height. It is important to note that the measurements are on an absolute scale (mm) instead of a relative scale (pixels).

Since we focus on study of feature distinctiveness and the benefit of their fusion for face recognition, we manually landmarked 15 fiducial points onto 3D face models as accurately as possible (as shown in Fig. 4.6) to reduce the impact by erroneous registration. Based on these landmarks, a coarse to fine alignment process is used to correct 3D pose. The coarse step utilizes eleven facial landmarks located on the upper part of the face model and applies SVD to recover 3D rotation and translation in a rigid transformation. At the fine step, ICP is applied for surface matching and to improve the estimates of translation and rotation parameters.

All extracted features are then fed into Sparse Representation Classifier (SRC) to calculate similarity scores. Table 4.1 lists rank-one recognition rates of individual features using SRC on the FRGC v2.0 dataset. It is worth noting that in the Gabor filters, 5 scales and 8 orientations are exploited; while in the LBP operator, we set the number of sampling points at 16 and the radius of neighborhood at 2.

From Table 4.1, we can see that all the individual 3D and 2D features do not provide enough distinctiveness for representing faces, only displaying 79.72% with the best 3D feature by tangent vectors and 77.89% with the best 2D feature by Gabor filters. Table 4.1 also displays the performance of each feature in the case of Neutral vs. Neutral. As we can see, all the features, either 3D or 2D, achieve better performance than the case of Neutral vs. All. Still, none of these 3D or 2D features reports enough distinctiveness for face recognition, and the best rank-one recognition rate is 89.64% by tangent vectors. With an average gain of about 10 percent, the most significant improvement is achieved by 3D features, confirming the intuition that 3D facial shape descriptors are much more sensitive to facial expression variations than 2D features. This intuition is further confirmed by figures which evidences performance degradation is much greater for 3D features than the 2D ones in the case of Neutral vs. Non-Neutral.

Before fusion, it is important to normalize scores achieved by different types of features aiming at mapping scores into a common scale and range.

Table 4.1: Rank-one recognition rates of individual features using SRC on the FRGC v2.0 database according to three experimental protocols: (I) Neutral vs. All; (II) Neutral vs. Neutral; and (III) Neutral vs. Non-neutral.

Different Features	I	II	III
2D: Gabor Filters	77.89%	84.59%	69.40%
2D: LBP	71.82%	78.83%	63.06%
2D: Intensity Image	49.82%	56.85%	40.91%
3D: Anthropometric Measurements	46.48%	55.99%	34.44%
3D: Gaussian Curvature	59.02%	73.67%	40.46%
3D: Mean Curvature	71.62%	85.25%	54.35%
3D: Maximum Curvature	67.81%	82.57%	49.10%
3D: Minimum Curvature	66.73%	81.05%	48.59%
3D: Binormal Vectors	70.63%	84.49%	53.07%
3D: Normal Vectors	70.01%	83.78%	52.56%
3D: Tangent Vectors	79.72%	89.64%	67.16%

Specifically, we denote the raw matching similarity score as s, from a set S of scores produced by different features, and its corresponding normalized score as n. Three normalization approaches are explored and compared.

Min-Max (MM) [Godil *et al.* 2005] [Mian *et al.* 2006b] [Snelick *et al.* 2005] maps matching scores to the range of [0, 1]. The function $max(s)$ and $min(s)$ specify the maximum and minimum of the score range respectively:

$$n = \frac{s - \min(s)}{\max(s) - \min(s)} \tag{4.6}$$

Z-Score (ZS) [Godil *et al.* 2005] [Snelick *et al.* 2005] transforms all the similarity scores to a distribution with the mean of 0 and the standard deviation of 1. The operators $mean(s)$ and $std(s)$ denote the arithmetic mean and standard deviation respectively:

$$n = \frac{s - mean(s)}{std(s)} \tag{4.7}$$

Tanh (TH) [Godil *et al.* 2005] [Snelick *et al.* 2005] provided by so-called robust statistical techniques, converts matching scores to the [0, 1] range:

$$n = \frac{1}{2}[\tanh(\frac{s - mean(s)}{std(s)} \times 0.01) + 1] \tag{4.8}$$

Their combination is achieved at the score level (Sum Rule) as well as decision level, and their results are shown in Table 4.2. This result further confirms that 3D and 2D features are complementary modalities whose fusion can improve the final decision.

Table 4.2: Rank-one recognition rates of different score normalization methods with different fusion schemes.

Fusion Scheme	Max-Min	Z-Score	Tanh
Decision Level Fusion	88.87%		
Score Level Fusion	87.49%	85.12%	86.44%

4.6.3 Identification and Verification

Unlike the previous experiment, to evaluate our proposed approach, we did not use any manual landmarks and perform registration. In this experiment, we set the radius value of OGMs at 1.0 and 1.5 for facial range and texture images respectively. In the literature, many tasks addressed the problem of 2D or 3D face recognition and used the FRGC v2.0 database for evaluation. Table 4.3 lists performance comparisons between the proposed approach and several existing features only utilizing 2D texture images for face recognition; while Table 4.4 lists a comparison between our method and the state of the art results only using facial range images or 3D face models for the same task.

It should be noted that as shown in Table 4.3, the results based on eigenface, LBP histogram, and Gabor filters are directly cited from the last experiment, which were achieved based on optimized parameters and the Sparse Representation Classifier (SRC). When SIFT-based matching was directly applied on original texture images, the accuracy is 79.3%, while if we exploited the proposed OGM instead, the accuracy was improved to 95.9%, which highlights its effectiveness to describe local texture changes. We also investigated the LBP face (8 sampling points and the same radius value as in OGMs) in the same framework, but it only reported a recognition rate of 44.8%, and the reason probably lies in the fact that illumination changes add much noise which degrades the performance of LBP.

On the other hand, Table 4.4 compares our approach with several existing systems for 3D face recognition. Similarly, we also operated SIFT-based matching on raw facial range images, but it did not achieve a reasonable performance, because the detected keypoints on original range images are too limited, and usually located on the face border as shown in Figure 4.4. While, in this case, LBP face enhances the distinctiveness of facial range image, and improves the accuracy to 80.1%. As we can see in Table 4.4, the performance of the proposed approach is comparable to those of the state of the art in 3D face recognition.

Table 4.3 and 4.4 both illustrate that the proposed OGMs enhanced local texture and shape variations leading to satisfying recognition results.

Table 4.5 compares the proposed approach with the state of the art on both face recognition and verification tasks by using textured 3D face models. In recognition, the rank-one recognition rate of the proposed approach outperforms all the others, while in the scenario of verification, the achieved verification rate at FAR = 0.1% is only slightly weaker than that in [Mian *et al.* 2007]. However, in this work, we did not perform costly 3D face registration in the preprocessing step while an ICP based fine registration step was always required as in [Mian *et al.* 2007] etc.

Table 4.3: Comparisons with the state-of-the-art using only 2D face data.

2D Approaches	Rank-One Recognition Rate
Eigenface	49.8%
LBP Histogram	71.8%
Gabor	77.9%
Original Texture + SIFT	79.3%
Texture LBP Face + SIFT	44.8%
Texture OGMs + SIFT	95.9%

Table 4.4: Comparisons with the state-of-the-art using only 3D face data.

3D Approaches	Rank-One Recognition Rate
Chang et al. [Chang *et al.* 2005a]	91.9%
Cook et al. [Cook *et al.* 2006a]	94.6%
Kakadiaris et al. [Kakadiaris *et al.* 2007]	97.0%
Mian et al. [Mian *et al.* 2007]	96.2%
Mian et al. [Mian *et al.* 2008]	93.5%
Huang et al. [Huang *et al.* 2010b]	79.3%
Alyuz et al. [Alyuz *et al.* 2010]	97.5%
Huang et al. [Huang *et al.* 2011a]	97.2%
Original Range + SIFT	NA
Range LBP Face + SIFT	80.1%
Range OGMs + SIFT	95.9%

4.6.4 Radius Analysis of Neighborhood Area

Recall that complex neurons respond to gradient information within a given neighborhood which is defined as a circular region in our implementation. In our experiments, we tested different values of radius R and studied their impact on final performance of both facial texture as well as range images. Table 4.6 and 4.7 summarizes the results with different radius values applied to texture OGMs and range OGMs respectively in eight orientations. Meanwhile, in order to illustrate the effectiveness of the proposed fusion scheme, we compared its results with those calculated using the technique as in [Mian *et al.* 2008]. In Table 4.6 and 4.7, "Fusion I" denotes the results with the weighted sum rule in [Mian *et al.* 2008]; while "Fusion II" lists performance using the proposed optimized weighted sum fusion strategy.

As we can see in Table 4.6 and 4.7, using the neighborhood area with a smaller radius generally achieves a better result. For facial range images, when the radius value was set at 1.0, the best performance was reached. On the other hand, for facial texture images, the highest accuracy was obtained when $R = 1.5$. Compared with the fusion technique marked as "Fusion I", the proposed optimized fusion (Fusion II)

Table 4.5: Comparison with the state-of-the-arts using textured 3D face data.

Textured 3D Systems	Rank-one RR	VR@FAR=0.1%
Maurer et al. [Maurer *et al.* 2005]	NA	95.8%
Husken et al. [Husken *et al.* 2005]	NA	97.3%
Mian et al. [Mian *et al.* 2007]	96.1%	98.6%
Mian et al. [Mian *et al.* 2008]	97.4%	99.3%
Gokberk et al. [Gokberk *et al.* 2008]	95.5%	NA
Xu et al. [Xu *et al.* 2009]	NA	97.5%
Ben Soltana et al. [Ben Soltana *et al.* 2010]	95.5%	97.0%
Texture OGM + SIFT	95.9%	97.3%
Range OGM + SIFT	95.5%	97.1%
Multi-Modal OGM +SIFT	98.0%	98.9%

strategy always slightly performs better. Considering that the FRGC v1.0 dataset does not include the variations of facial expressions, which is different from the face models in the test set, i.e. the FRGC v2.0 dataset, the weights for combination can be further improved by using a dataset with more expressive face models.

Table 4.6: Results when using different neighborhood area radius R on texture faces.

Texture	$R = 1.0$	$R = 1.5$	$R = 2.0$	$R = 2.5$	$R = 3.0$	$R = 3.5$
OGM_1	78.00%	81.42%	82.60%	83.23%	83.56%	83.20%
OGM_2	83.08%	84.69%	86.19%	85.31%	84.61%	83.51%
OGM_3	85.17%	87.18%	87.49%	87.97%	87.63%	87.09%
OGM_4	86.22%	87.49%	88.62%	87.77%	87.49%	86.33%
OGM_5	80.06%	81.90%	83.65%	82.43%	82.15%	80.66%
OGM_6	80.74%	82.97%	84.89%	86.02%	85.77%	85.03%
OGM_7	84.44%	86.02%	85.26%	85.63%	84.67%	82.01%
OGM_8	84.19%	85.63%	86.39%	87.46%	86.64%	85.99%
Fusion I	95.31%	95.74%	95.51%	95.51%	95.54%	94.38%
Fusion II	95.45%	95.85%	95.54%	95.71%	95.76%	94.78%

4.6.5 Evaluation on Facial Expression Variations

In this experiment, the probe face scans were divided into two subsets according to their expression labels to evaluate its insensitiveness to facial expression variations as we did in the last chapter. The first subset contains the facial scans with the neutral expression; while the other with face scans possessing non-neutral expressions. Therefore, besides the experiment of Neutral vs. All, two additional experiments of Neutral vs. Neutral and Neutral vs. Non-Neutral were carried out as well. In the

Table 4.7: Results when using different neighborhood area radius R on range faces.

Range	$R = 1.0$	$R = 1.5$	$R = 2.0$	$R = 2.5$	$R = 3.0$	$R = 3.5$
OGM_1	79.95%	80.54%	80.68%	78.96%	76.76%	74.10%
OGM_2	84.33%	83.95%	82.91%	81.30%	78.68%	76.67%
OGM_3	89.04%	88.84%	87.97%	87.01%	85.37%	84.16%
OGM_4	85.09%	85.43%	83.71%	82.10%	79.70%	78.17%
OGM_5	82.91%	83.39%	82.58%	81.05%	78.62%	76.28%
OGM_6	89.24%	88.93%	87.18%	85.43%	83.96%	82.15%
OGM_7	84.75%	84.24%	82.01%	79.53%	76.73%	72.83%
OGM_8	88.20%	88.68%	86.90%	85.65%	84.07%	81.73%
Fusion I	95.14%	94.55%	93.67%	92.26%	91.10%	90.06%
Fusion II	95.48%	94.94%	94.07%	92.54%	91.67%	90.43%

Table 4.8: Rank-one face recognition rates using the facial expression protocol on the FRGC v2.0 dataset.

Different Systems	**Subset I**	**Subset II**	**Degradation**
Mian et al. [Mian *et al.* 2008]	99.4%	92.1%	7.3%
Ben Soltana et al. [Ben Soltana *et al.* 2010]	98.6%	90.7%	7.9%
Texture OGM + SIFT	98.8%	92.1%	6.7%
Range OGM + SIFT	98.5%	91.7%	6.8%
Multi-Modal OGM +SIFT	99.6%	96.0%	3.6%

Subset I: Neutral vs. Neutral
Subset II: Neutral vs. Non-Neutral

Neutral vs. Neutral and Neutral vs. Non-Neutral experiment, only the neutral and non-neutral probe subsets were used, respectively.

Using the same experimental protocol, we also compared the performance of the proposed method in face recognition with the one by Mian et al. [Mian *et al.* 2008] and Ben Soltana et al. [Ben Soltana *et al.* 2010] for robustness analysis on expression variations (see Table 4.8). The results of our approach are 99.6% and 96.0% for Neutral vs. Neutral and Neutral vs. Non-Neutral experiment, respectively. The recognition rate on the first subset is comparable to the state-of-the-art, while we made great progress on the second subset, displaying a rank-one recognition rate of 96.0%. The degradation when non-neutral facial expression models were included drops by 3.6%, which is much lower than 7.3% in [Mian *et al.* 2008] and 7.9% in [Ben Soltana *et al.* 2010] as both shape and texture clues were combined. These results suggest that our method tends to be insensitive to facial expression changes. Table 4.9 lists the robustness comparison in face verification task, similar conclusions can be drawn.

Table 4.9: Comparison of face verification rates at 0.001 FAR using the facial expression protocol on the FRGC v2.0 dataset.

Different Systems	VR I	VR II	VR III
Mian et al. [Mian *et al.* 2008]	97.4%	99.9%	92.7%
Texture OGM + SIFT	97.3%	99.7%	93.7%
Range OGM + SIFT	97.1%	99.4%	93.5%
Multi-Modal OGM +SIFT	98.9%	99.9%	97.1%

VR I: Neutral vs. All
VR II: Neutral vs. Neutral
VR III: Neutral vs. Non-Neutral

Meanwhile, the accuracies of 2D modality are always slightly better than those of 3D based one, suggesting that 2D modality tends to be more insensitive to facial expression variations than its 3D counterpart.

Fig. 4.7 indicates verification rates by the ROC curves in these three experiments in Table 4.9.

4.6.6 Evaluation on Twins

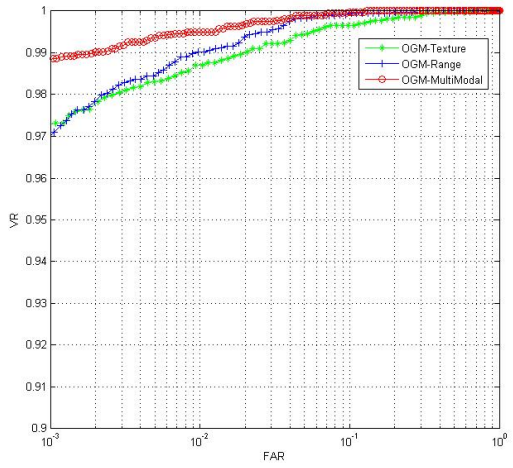
In the 3DTEC database, we arbitrarily labelled one person in each pair of twins as Twin A and another one as Twin B, and then performed four experiments with the different settings of gallery and probe sets as shown in Table 4.10.

Table 4.10: List of experiments performed on the 3DTEC dataset.

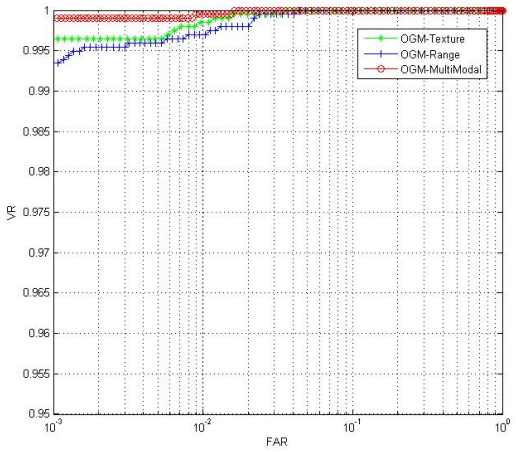
No.	Gallery	Probe
Exp.I	A Smile, B Smile	A Neutral, B Neutral
Exp.II	A Neutral, B Neutral	A Smile, B Smile
Exp.III	A Smile, B Neutral	A Neutral, B Smile
Exp.IV	A Neutral, B Smile	A Smile, B Neutral

Experiment I has all of the facial images with a smiling expression in the gallery and the ones with a neutral expression as the probe. Experiment II reverses these roles. We thus assessed the ability of the proposed approach to distinguish between twins in some gallery, when the probe and the gallery facial images have different expressions.

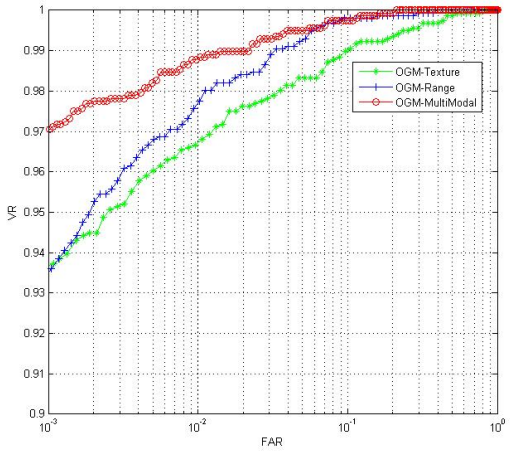
Experiment III has Twin A smiling and Twin B neutral in the gallery set but with Twin A neutral and Twin B smiling as the probe. Experiment IV reverses these roles. This models a scenario where the system does not control for the expressions of the subject in the gallery set and the probe has the same expression as his twin in the gallery but not the same expression as his image in the gallery.



(A)



(B)



(C)

Figure 4.7: ROC curves using texture OGMs, range OGMs, and Multi-Modal OGMs respectively in the experiments with neutral faces enrolled: (a) Neutral vs. All. (b) Neutral vs. Neutral. (c) Neutral vs. Non-neutral.

Table 4.11: Rank-one face recognition rates of four experiments on the 3DTEC dataset.

Rank-one Recognition Rate	RR I	RR II	RR III	RR IV
Texture OGM + SIFT	95.8%	96.3%	92.1%	92.5%
Range OGM + SIFT	91.6%	93.9%	69.2%	71.0%
Multi-Modal OGM +SIFT	96.3%	96.3%	88.8%	88.8%

Table 4.12: Face verification rates of four experiments at 0.1% FAR on the 3DTEC dataset.

Verification Rate	VR I	VR II	VR III	VR IV
Texture OGM + SIFT	96.7%	96.7%	93.0%	93.5%
Range OGM + SIFT	94.9%	94.4%	68.7%	69.2%
Multi-Modal OGM +SIFT	96.7%	96.7%	88.3%	89.7%

Table 4.11 and Table 4.12 demonstrate the accuracies of texture OGMs, range OGMs, as well as multi-modal OGMs in two scenarios respectively, i.e. face recognition and verification. In the first two experiments in both scenarios, all subjects are enrolled with a 3D facial scan that has one expression, and all recognition attempts are made with another expression. Thus, the difference in expression between gallery and probe is the same for all subjects. As we can see, the results of recognition and verification for twins exceed 90% for both texture and range OGMs, and texture OGMs perform better than range OGMs. While in the more challenging two of the four experiments, the facial expression differs between the gallery images and also between their probe images of the twins. In these experiments, texture OGMs still provide satisfying accuracies which are higher than 90%, but the performance of range OGMs is dramatically degraded, proving that, 3D modality is more sensitive to facial expression variations than 2D modality. When combining the two modalities, we find that in the first two experiments of both scenarios, the multi-modal based results are slightly improved, however, in the other two, the final performance is not as good as that of 2D modality, since the accuracy differences between two modality are too large, which can be regarded as a special case that combining 2D and 3D information is not always an effective way to improve system performance.

4.6.7 Matching Examples

Figure 4.8 depicts some examples of SIFT based local face matching using texture information. As we can see, 28 pairs of keypoints are matched between the original facial texture images of the same person, while 1 pair of keypoint is matched between those from different persons. When using the OGMs instead of the original texture information, in each pair of corresponding OGMs from the same person, we can find a certain number of matched keypoints, leading to a total number of 139 between the

same person by summing the number achieved in each OGM pair; on the other hand, the total number of texture OGM pairs from different persons is only 7, illustrating that the inter-class dissimilarity is enlarged. Meanwhile, we can also see that the matched points of each OGM pair do not have the same distribution, and they thus contain complementary information to each other.

The matching examples using depth information are shown in Fig. 4.9. A similar phenomenon can be found as in Fig. 4.8. Using the OGMs to replace the original facial range images as the system input, the wrong matching relationship between original facial range images can be corrected.

4.6.8 Computation Cost

Because more keypoints are detected from texture OGMs than from range OGMs, the computation cost of the matching process in similarity calculation between facial texture images is higher than that between facial range images. An unoptimized implementation of the proposed approach with MATLAB (R2010a) can perform about 280 and 410 matches between texture and range face pairs per second respectively, using a machine using Intel(R)Core(TM) i7 CPU (3.07 GHz) and 8 GB RAM.

4.7 Summary

This chapter proposed a novel biological vision-inspired facial representation, namely Oriented Gradient Maps (OGMs), and we then applied it to both facial texture and range images for the issue of textured 3D face recognition. As compared with other intermediate facial description, for instance, Eigenface or Fischer face, the proposed OGMs simulate the response of complex neurons to gradient information in different orientations within a given neighborhood, thereby highlighting local details of facial range and texture images and increasing their distinctiveness. When compared with LBP faces that also encode the difference between a central pixel and its neighbors, the OGMs are more informative as they take into account gradient information in several directions. The OGMs are also likely less sensitive to noise than LBP faces because gradient information is summarized within a neighborhood convolved by a Gaussian kernel. Finally, OGMs also possess the properties of being robust to affine illumination and geometric transformations. Additionally, the designed score level fusion strategy further improved the final result when combining the results of OGMs of different orientations as well as fusing the accuracies of these two modalities. The experiments carried out on the FRGC v2.0 database showed the efficiency of the proposed method. On the other hand, the experimental results achieved on the 3DTEC dataset illustrate that 3D modality is more sensitive than 2D based one.

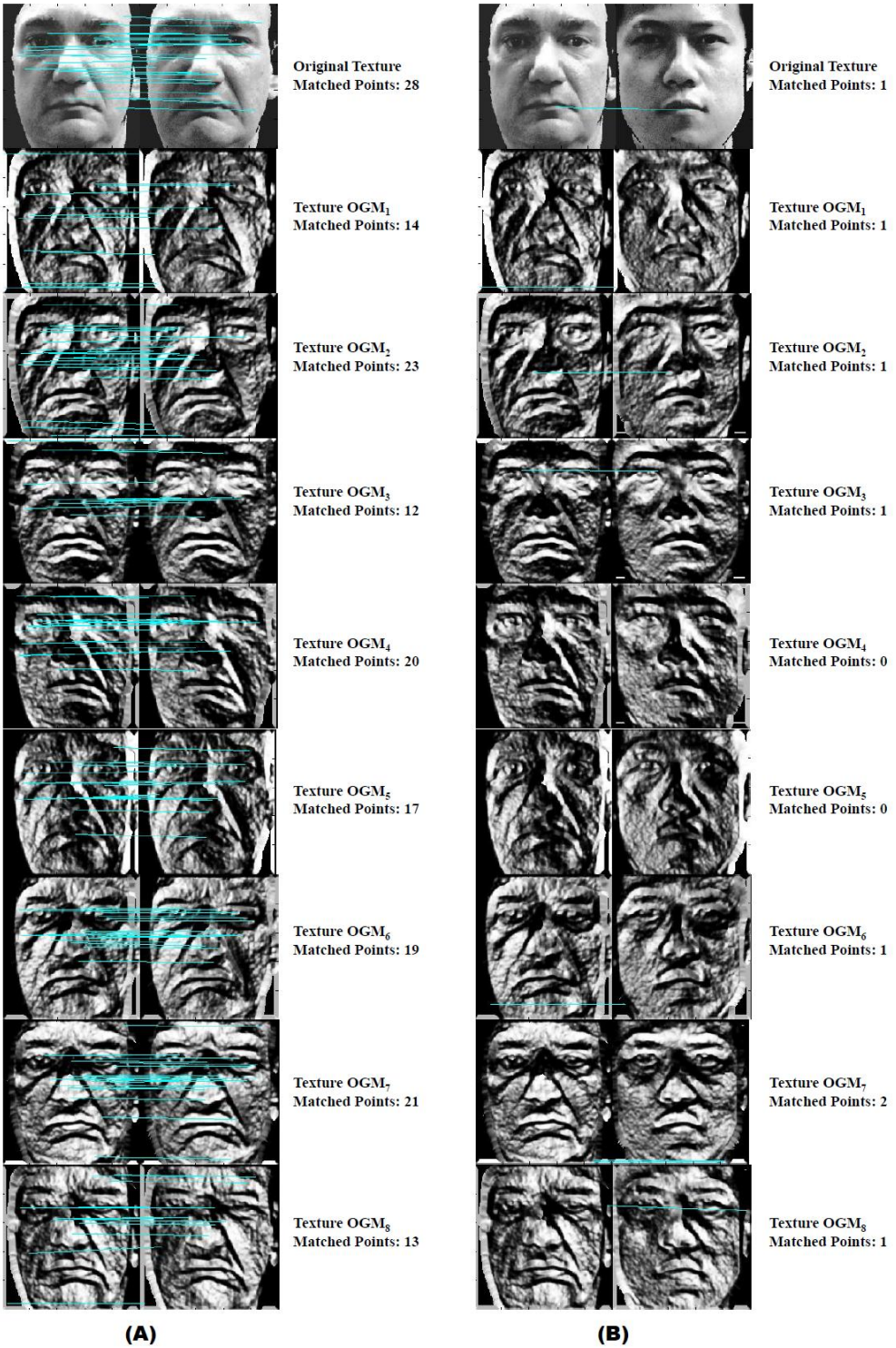


Figure 4.8: Matching examples using original facial texture images and their OGMs: (A) matching faces of the same subject; (B) matching faces of different subjects.

Figure 4.9: Matching examples using original facial range images and their OGMs: (A) matching faces of the same subject; (B) matching faces of different subjects.

Chapter 5

Asymmetric Face Recognition

5.1 Introduction

3D face recognition has emerged as a major alternative to handle the unsolved issues for reliable face recognition, i.e., illumination and pose changes, and in chapter 3 and 4, we have introduced two methods for only 3D shape and textured 3D based face recognition respectively. However, 3D methods are currently limited by their data acquisition and computation cost.

More recently, asymmetric 3D-2D face recognition has attracted increasing interests [Riccio & Dugelay 2005, Rama *et al.* 2006, Yang *et al.* 2008, Huang *et al.* 2009, Huang *et al.* 2010a], because it is expected to limit the use of 3D face data where it really helps to improve face recognition performance. In contrast to these traditional 2D, 3D and multi-modal face recognition techniques that require gallery and probe data to possess similar properties: e.g. 2D/3D, color/gray, or even to be captured by the same type of camera sensors, the asymmetric 3D-2D face recognition approaches assume that the gallery set contains 3D face data, but the probe set only consists of 2D facial images, which makes them more likely to be used in real-time environment than 3D based ones. Meanwhile, several novel applications, namely asymmetric [Riccio & Dugelay 2005, Huang *et al.* 2009, Huang *et al.* 2010a] or heterogeneous [Wang *et al.* 2009, Liao *et al.* 2009] facial image analysis, match faces between different types of data, which can be related by applying techniques [Reiter *et al.* 2006, Lei & Li 2009]. Therefore, it is possible to obtain the relationship between 2D and 3D face data.

Up to now, few tasks in the literature have addressed such a problem on asymmetric 3D-2D face recognition. Rama et al. [Rama *et al.* 2006] proposed the Partial Principle Component Analysis (P^2CA) for feature extraction and dimensionality reduction on both the cylindrical texture representation (3D) in the gallery set and 2D images in the probe set. However, the 3D face data utilized still conveyed texture information rather than shape clues. In [Riccio & Dugelay 2005], Riccio et al. used pre-defined control points to compute several geometrical invariants for 2D/3D face recognition. But it is really difficult to accurately locate these keypoints on both 2D and 3D facial data. More recently, Yang et al. [Yang *et al.* 2008] proposed a patch based Kernel CCA to learn mappings between facial range and texture images. Yet, their results still can be improved if we can represent both texture and range faces more comprehensively. Furthermore, all the tasks above partially relied on 2D faces; nevertheless, none of them provided reliable performance when lighting condition or pose status changes.

This chapter introduces an asymmetric 3D-2D face recognition approach, aiming to achieve better performance than 2D based methods and comparable accuracies as

3D based ones do while keeping computational cost under control. The experimental results on the entire FRGC v2.0 database clear illustrate that the effectiveness of the proposed method.

The contributions of this study are fivefold:

(1) A novel framework for asymmetric 3D-2D face recognition is proposed. It uses textured 3D face models for enrollment, whilst only 2D facial images for identification. Compared with the state of the art which only exploited single type of 3D data in the gallery set such as cylindrical texture images and facial range images, this framework considers more clues to improve system performance;

(2) We investigate to encode local texture and shape variations into Multi-Scale Local Binary Patterns (MS-LBP) space, where the relationship between 2D and 3D information can be better established;

(3) A new biological vision-based facial representation, namely Oriented Gradient Maps (OGMs), is further employed to represent local texture changes of 2D faces and local geometric variations of 3D faces simultaneously and comprehensively;

(4) An effective preprocessing pipeline is introduced to deal with illumination and pose changes based on Logarithmic Total variation (LTV) [Chen *et al.* 2006] and Active Appearance Model (AAM) [Cootes *et al.* 2001] respectively, which greatly improves the robustness of the entire asymmetric face recognition system;

(5) To the best of our knowledge, almost all the methods for asymmetric 3D-2D face recognition were evaluated either on a small private database or a subset of a public one; therefore, it is very difficult to compare their performance and test their robustness when a comprehensive database is utilized. All the experiments in this chapter are carried out on the complete FRGC v2.0 dataset [Phillips *et al.* 2005] for later possible comparisons.

The remainder of this chapter is organized as follows. The overall framework of both the training and test stages in the proposed method is presented in section 5.2. Section 5.3 introduces the preprocessing pipeline, and section 5.4 describes the MS-LBP and the biological vision-based facial representations. Section 5.5 depicts the asymmetric face recognition approach in detail. Experimental results are discussed and analyzed in section 5.6. Section 5.7 concludes this chapter.

5.2 Method Overview

The training and test stage frameworks are illustrated in Figure 5.1 and Figure 5.2 respectively.

At the training stage (Fig. 5.1), textured 3D face models, each of which contains a densely registered facial texture image and 3D facial point-cloud, are required. For each face model, there are totally 64 manually landmarked fiducial points. In 2D phase, AAM is constructed using illumination normalized 2D facial texture images; while in 3D phase, all point-clouds are first registered to a selected reference 3D face model, and then facial range images are extracted. According to these corresponding manual landmarks, all the texture and range faces are transformed to the mean face shape generated by AAM. The local texture changes of resulted 2D faces and local shape variations of 3D ones are accurately represented by Oriented Gradient Maps

(OGMs). Four PCA based subspaces are further produced from the sample sets of facial texture images, facial range images, as well as their different OGMs. In the end, Canonical Correlation Analysis (CCA) [Hardoon *et al.* 2004] is introduced to learn the mapping between corresponding PCA subspaces of the OGMs belonging to 2D and 3D data respectively.

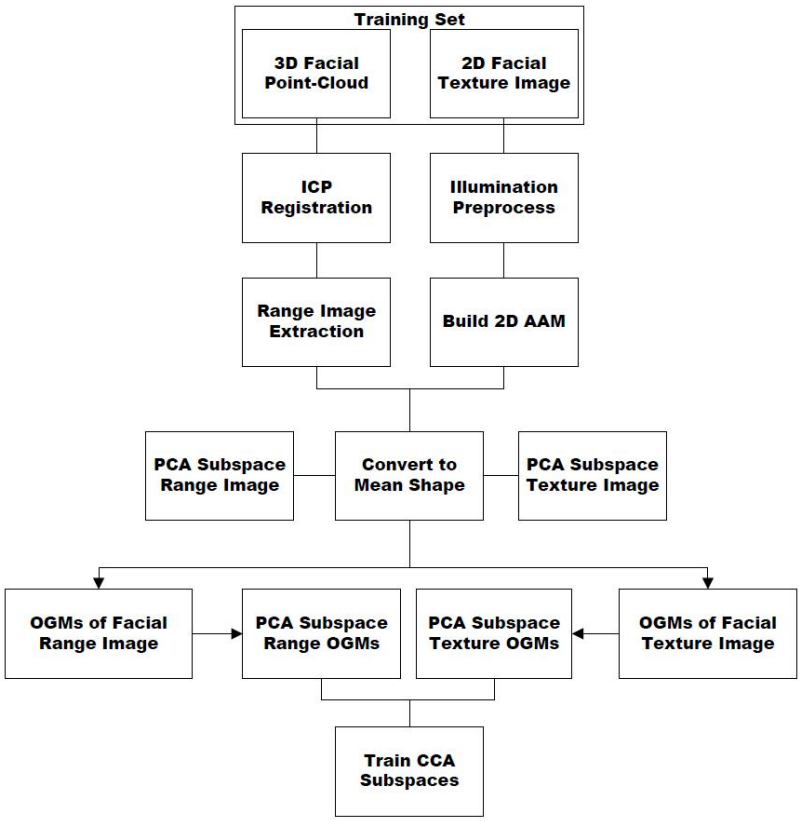


Figure 5.1: Training stage framework.

At the test stage, the preprocessing and facial representation steps are the same with those in training. Since one facial range image and its corresponding 2D counterpart are extracted from each 3D face model, the classification step hence contains two separate matching processes: 1) 3D-2D matching and 2) 2D-2D matching. As shown in Fig.5.2, for 3D-2D matching, CCA is adopted to compute similarity measurements between the OGM-based features of facial range images in the gallery set and those of facial texture images in the probe set. For 2D-2D matching, the LBP histogram based features are extracted from OGMs of facial texture images in the gallery and probe set respectively, and then Sparse Representation Classifier (SRC) [Wright *et al.* 2009] is applied to compute their similarity measures. Finally, both scores are fused for final decision.

For comparison, we also investigate the discriminative power of Multi-Scale Local Binary Pattern (MS-LBP) to describe both the local texture and shape information. Similar to OGMs, in 3D-2D matching, we firstly generate the MS-LBP Face Maps

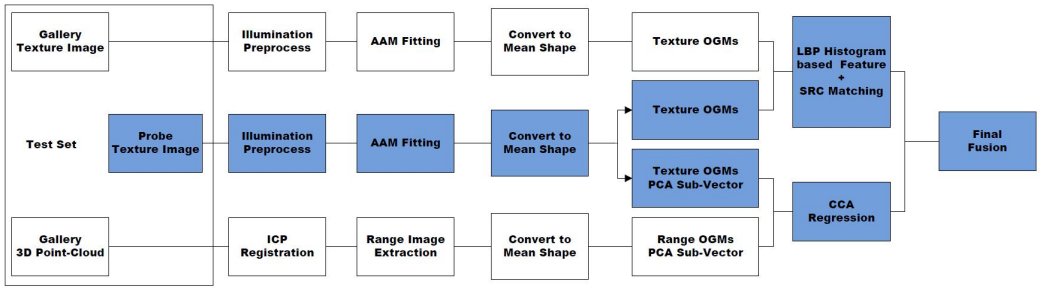


Figure 5.2: Test stage framework.

(MS-LBP-FMs) of original facial texture and range images, and the similarity scores are then calculated in CCA subspaces. While in 2D-2D matching, LBP histograms are directly extracted from original 2D facial images to measure their similarities. The similarity scores of both steps are combined to provide the final result.

Furthermore, all operations on gallery face data are off-line, including illumination normalization and AAM fitting on facial texture images; ICP-based registration for 3D facial point-clouds; transforming both facial range and texture images to the mean shape as pose correction; the OGMs generation of facial texture and range images; as well as LBP and PCA based feature extraction. Meanwhile, the operations on probe faces are on-line, containing the preprocessing pipeline, OGMs production, LBP and PCA based feature extraction, 2D-2D and 3D-2D matching score calculation as well as score combination. See Figure 5.2 for more details (on-line processes are marked in blue).

5.3 Data Preprocessing

In the process of asymmetric 3D-2D face recognition, both the 2D and 3D face data are utilized. In this section, we propose an effective pipeline for data preprocessing, containing illumination normalization and pose correction. The former part is only designed for 2D facial images, while the latter is for both 2D and 3D face data.

5.3.1 2D Preprocessing

LTV [Chen *et al.* 2006] is applied to normalize illumination variations since it works on any single image without any prior information about 3D face geometry or lighting sources. LTV not only inherits the ability from the TV-L1 model to decompose a facial image f into a large-scale output u and a small-scale output v, but also has its properties of edge-preserving and multi-scale additive signal decomposition.

The LTV model is based on a general multiplicative lighting reflectance theory:

$$I(x, y) = \rho(x, y) \cdot S(x, y) \tag{5.1}$$

Figure 5.3: The upper row presents original 2D facial images; the middle row shows illumination normalized ones; and the bottom row lists the samples after pose correction.

where I is the reflected light intensity, ρ is the surface albedo, and S is the lighting amount. Using logarithm transform on eq. 5.1, the LTV model can be described as follows:

$$f(x,y) = \log(I(x,y)) = \log(\rho(x,y)) + \log(S(x,y)) \tag{5.2}$$

$$u^* = \arg\min\{\int |\nabla u| + \lambda \left\| f - u \right\|_L\} \tag{5.3}$$

$$v^* = f - u^* \tag{5.4}$$

where $\int |\nabla u|$ is the total variation of u, and λ is a scalar threshold on scale. In eq. 5.3, minimizing $\int |\nabla u|$ would make the level sets of u have simple boundaries and minimizing $\left\| f - u \right\|_L$ would ensure the approximation of u to f. S and ρ can thus be approximately estimated by solving eq. 5.5:

$$S \approx \exp(u^*), \quad \rho \approx \exp(v^*) \tag{5.5}$$

The middle row of Figure 5.3 shows several LTV based illumination normalized samples of 2D facial images.

After the step of the LTV based illumination normalization, AAM is applied to correct pose variations of 2D facial images. A training set is required to produce an AAM. The AAM fitting approach referred to [Matthews & Baker 2004]; and its implementation is achieved thanks to the source code provided by DTU [DTU-AAM]. The bottom row of Fig. 5.3 lists several face samples after pose correction.

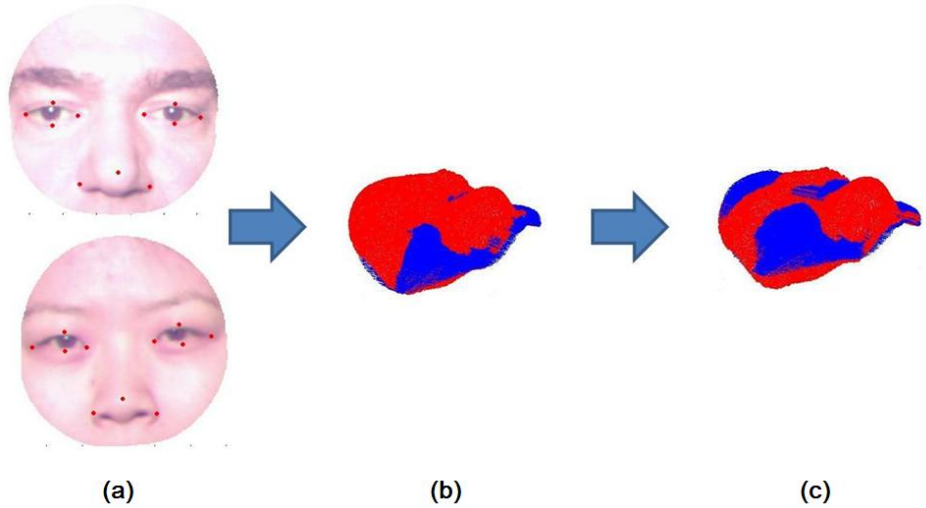


Figure 5.4: The R-ICP based Registration (best scene in color): (a) rigid region of textured 3D data; (b) coarse step; (c) fine step.

5.3.2 3D Preprocessing

3D facial surface registration is an important step in 3D face recognition for 3D pose correction. Based on the AAM fitting result on its corresponding facial texture image, the keypoints of each 3D facial point-cloud can be known. Then, Region based Iterative Closet Point (R-ICP) [Ben Amor *et al.* 2006b] is exploited after removing spikes and filling holes. R-ICP only works on the rigid regions around the nose and forehead that are considered insensitive to facial expression variations. The registration adopts a coarse-to-fine scheme. The coarse step uses 11 landmarks of all the 64 keypoints and applies SVD to recover 3D rotation and translation. At the fine step, ICP is introduced to match rigid surfaces and improve the estimate of translation and rotation parameters (see Fig. 5.4).

5.4 Facial Description

When all face data are preprocessed, to use both the 2D and 3D data sufficiently, we encode both of them into certain facial feature space to highlight their distinctiveness. In this section, we present two approaches to describe local texton variations of 2D facial images as well as local geometry variations of 3D based ones, namely Local Binary Patterns(LBP) and Oriented Gradient Maps (OGMs).

5.4.1 LBP Face based

The LBP methodology has been detailedly introduced in the second chapter. Recall that there are two ways to adopt LBP for facial representation, i.e. LBP histogram based and LBP image based. The general concept for the former one is that a facial image can be considered as a composition of micro-patterns provided by LBP. But

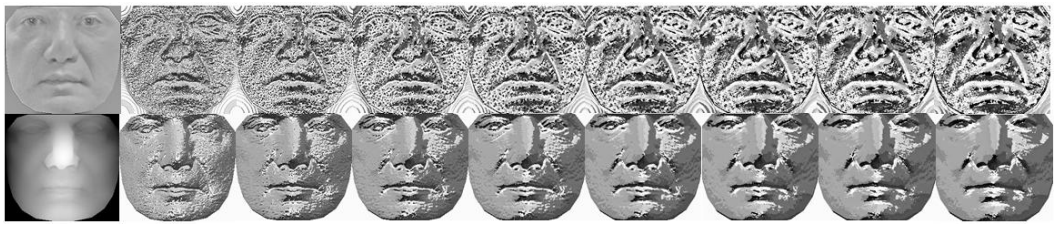


Figure 5.5: LBP faces of facial texture and range images output by the preprocessing pipeline using different neighborhoods: the images from left to right are LBP images generated based on the scale value from 1 to 8 pixels and 8 neighboring points.

LBP histograms calculated over the whole facial region encode only occurrences of micro-patterns without any indication on their locations. To take into account the facial shape information as well, the images are proposed to be divided into a certain number of local sub-regions, from which LBP histograms are then extracted. These LBP histograms are then concatenated into a single, spatially enhanced histogram, containing both local texture and global shape information of the face images. The similarity between two faces can be calculated by comparing the final LBP histogram based features. An alternative manner is to regard the decimal values of LBP codes as pixel values to generate another image, namely LBP image, to represent faces.

In 2D-2D face matching, we follow the way for traditional 2D face recognition as did in [Ahonen *et al.* 2004], and thus LBP histogram based facial representation is employed; while in 3D-2D face matching, LBP image based one is utilized. Moreover, to further improve the performance of 3D-2D matching, LBP is used to extract local texture and shape information within the neighborhood with varying sizes, and the information of different local regions is then combined to achieve a more comprehensive Multi-Scale LBP (MS-LBP) description. See Figure 5.5 for an illustration of LBP faces of facial texture and range images output by the preprocessing pipeline using different neighborhoods.

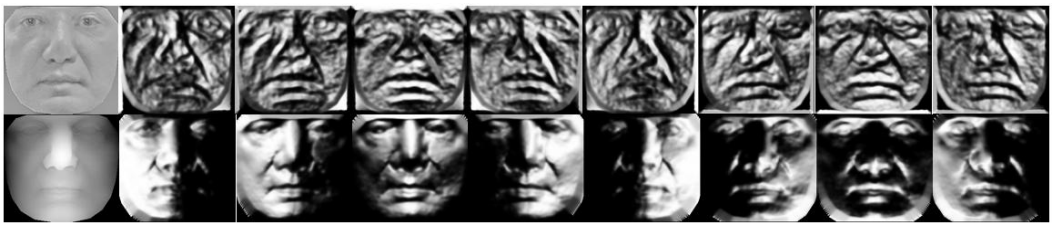


Figure 5.6: OGMs of facial texture and range images output by the preprocessing pipeline.

5.4.2 OGM Face based

For this application, we also investigate the Oriented Gradient Maps (OGMs) based facial representation, which has been proposed in the previous chapter for textured

3D face recognition.

OGMs simulate the responses of complex neurons to gradient information within a pre-defined neighborhood, and hence can represent both the local texture changes of 2D faces and local geometric variations of 3D faces simultaneously and comprehensively. More technical instructions have been described in detail in section 4.3. Fig. 5.6 shows the examples of the OGMs of facial texture and range images output by the preprocessing pipeline.

5.5 Asymmetric Face Recognition

The proposed asymmetric face recognition makes use of two types of face data, i.e. facial range images and their corresponding 2D facial images in the gallery set; while only 2D facial images are taken as probes. Therefore, two independent face matching steps are contained: 3D-2D and 2D-2D. After facial representation, each range and texture faces are described in their LBP feature space, denoted as $\{L_{R_1}, J_{R_2}, ..., J_{R_n}\}$ and $\{J_{T_1}, J_{T_2}, ..., J_{T_n}\}$; where n is the number of different neighborhoods, or in their OGM feature space, denoted as $\{J_{R_1}, J_{R_2}, ..., J_{R_o}\}$ and $\{J_{T_1}, J_{T_2}, ..., J_{T_o}\}$; where o is the number of quantized orientations.

5.5.1 3D-2D Face Matching

In this step, we compute similarity measurements between probe facial texture images and gallery facial range images in the LBP image or OGM based feature space.

Specifically, if using LBP images, to one 2D probe facial image P_T, its similarity with a gallery range face G_T is calculated by firstly computing the matching score $S^{Asy}_{L_i}$ between each LBP image pair within different neighborhoods: $L^P_{T_i}$ and $L^G_{R_i}$, i = {1, 2,..., n}, independently, and then combining them with a fusion scheme.

Similarly, the matching scores $S^{Asy}_{J_i}$ can also be calculated between each OGM pair of certain quantized direction: $J^P_{T_i}$ and $J^G_{R_i}$, i = {1, 2,..., o}, respectively, which are finally fused to give an overall result of 3D-2D face matching.

CCA [Hardoon *et al.* 2004] is a powerful analysis algorithm especially useful for relating two sets of variables, by maximizing correlation in the CCA subspace. In this chapter, it is introduced to learn the mapping between each range and texture LBP image pair (J_{T_i} and J_{R_i}) or OGM pair (J_{T_i} and J_{R_i}).

Given N pairs of sample data (x_i, y_i) of (X, Y), $i = 1, 2, ..., N$, where $X \in L_T$ or J_T, $Y \in L_R$ or J_R, with the mean value of zero. The goal of CCA is to learn a pair of directions w_x and w_y to maximize correlation between $x = w_x^T X$ and $y = w_y^T Y$. In the context of CCA, two projections: x and y are also referred to as canonical variants. Formally, the directions can be calculated as the maxima of the function:

$$\rho = \frac{E[w_x^T X Y^T w_y]}{\sqrt{E[w_x^T X X^T w_x] E[w_y^T Y Y^T w_y]}} \tag{5.6}$$

where $E[f(x, y)]$ denotes empirical expectation of $f(x, y)$.

The covariance matrix of (X, Y) is

$$C(X,Y) = E\left[\begin{pmatrix} X \\ Y \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}^T\right] = E\left[\begin{pmatrix} C_{XX} \\ C_{XY} \end{pmatrix}\begin{pmatrix} C_{YX} \\ C_{YY} \end{pmatrix}^T\right] \tag{5.7}$$

where C_{XX} and C_{YY} are within-set covariance matrices; C_{XY} and C_{YX} are between-sets covariance matrices.

Hence, ρ can be rewritten as

$$\rho = \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}} \tag{5.8}$$

Let

$$A = \begin{pmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{pmatrix}, \quad B = \begin{pmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{pmatrix} \tag{5.9}$$

It can be shown that the solution $W = (w_x^T, w_y^T)^T$ amounts to extremum points of the Rayleigh quotient:

$$r = \frac{W^T A W}{W^T B W} \tag{5.10}$$

and the solution w_x and w_y can be obtained as solutions of the generalized eigenproblem:

$$AW = BW\lambda \tag{5.11}$$

To test new pairs of variables, we project them into CCA subspace: $x' = w_x^T X'$ and $y' = w_y^T Y'$; and their similarity is calculated by eq. 5.12, and a bigger value of the measure indicates a higher similarity.

$$S(x', y') = \frac{x' \cdot y'}{\|x'\| \, \|y'\|} \tag{5.12}$$

5.5.2 2D-2D Face Matching

In this step, facial texture images in the probe set are matched with the ones in the gallery set.

Considering that the similarity measurement of 3D-2D matching and the one of 2D-2D matching will be further combined for final decision, the scores of both parts are expected to be complementary to each other. Since the similarity score achieved in 3D-2D face matching is computed in a holistic way, in this subsection, we exploit a feature-based approach, LBP histogram based technique, for face matching.

For LBP based asymmetric face recognition, in this step, LBP histogram based features are extracted from facial texture images in the gallery and probe set respectively, which are further compared the similarity score of 2D-2D face matching. For OGM based asymmetric face recognition, LBP histogram based features are extracted from the OGMs of original texture facial images for similarity computation. The similarity score of each OGM, $S_{J_i}^{Sym}$, is calculated based on LBP features of $J_{T_i}^{P}$ and $J_{T_i}^{G}$, i = {1, 2,..., o} independently, and then all of them are combined with a fusion scheme. Moreover, We employ Sparse Representation Classifier (SRC) [Wright *et al.* 2009] instead of Chi-square distance to compare LBP histogram based features of two faces.

Sparse representation for signal classification (SRSC) [Huang & Aviyente 2006] was firstly proposed to incorporate reconstruction properties, discriminative power as well as sparsity for robust classification. A general classification SRC for 2D face recognition was presented using sparse representation calculated by L1-minimization [Wright *et al.* 2009]. It usually achieves high accuracy when illumination variations and occlusions occur.

k classes and n_i feature vectors, $v_{i,j} \in R_m$ are used for training from the i_{th} class, $i = 1, 2, ..., k$ and j is the index of the training sample, $j = 1, 2, ..., n_i$. All training data from the i_{th} class are placed in a matrix $A_i = [v_{i,1}, v_{i,2}, ..., v_{i,n_i}] \in R_{m \times n_i}$. A dictionary matrix A for k classes is developed by concatenating A_i, and a new test pattern y can be represented as a linear combination of all the n training samples ($n = n_i \times k$):

$$y = Ax \tag{5.13}$$

where x is an unknown coefficient vector; from eq. 5.13, it is straightforward that only entries of x that are non-zero correspond to the class of y. Equation 5.13 can be solved according to compressed sensing as long as its solution is known sufficiently sparse. An equivalent L1-norm minimization:

$$(L1) : x_1 = \arg\min \|x\|_1 ; \quad Ax = y \tag{5.14}$$

can be solved as a good approximation to eq. 5.13. With the solution x_1 of eq. 5.14, we can compute the residual between a given probe face and each individual gallery face as:

$$R_i = \left\| y - \sum_{j=1}^{k} x_{1_{i,j}} v_{i,j} \right\|_2 \tag{5.15}$$

The identity of a given probe face is then determined as the one with the smallest residual R.

5.5.3 Score Fusion

Because the OGM based facial representation has o quantized gradient orientations, either for 2D-2D or 3D-2D matching, the similarity measurement FS_J^{Sym} or FS_J^{Asy} combines the ones of all OGMs ($S_{J_i}^{Sym}$ or $S_{J_i}^{Asy}$).

Meanwhile, in 3D-2D face matching, when applying LBP images, we calculate similarity scores for different neighborhoods, which are further combined to generate the MS-LBP based score FS_L^{Asy}.

The final similarity measurement of asymmetric 3D-2D face recognition, F_L or F_J, is achieved by further fusing FS_L^{Sym} and FS_L^{Asy} or by fusing FS_J^{Sym} and FS_J^{Asy}.

All the fusion steps in this chapter are computed according to a simple sum rule. Before that, all the scores are normalized to the interval of [0, 1] by using min-max normalization and different polarities should be reversed.

5.6 Experimental Results

The experiments were evaluated on the FRGC v2.0 dataset [Phillips *et al.* 2005]. It is one of the most comprehensive and popular databases, made up of 4007 textured 3D face models of 466 subjects. After the preprocessing pipeline, all the facial range and texture images were converted to a pre-defined mean shape, which were further cropped to 175×190 pixels as the input of the subsequent experiments.

5.6.1 Experiment Design

The first 3D face model containing one facial texture and range image with a neutral expression from each subject formed a gallery set of 466 samples. The remaining texture faces (4007-466=3541) were treated as probes.

We designed four special experiments: the first one is to discuss the effectiveness of the preprocessing pipeline; the second is to test the proposed facial representation, i.e. LBP and OGM, both in 2D-2D and 3D-2D face matching; the third one is to analyze the impact of the neighborhood area radius R of OGMs on final performance; and the last is to show the results of the asymmetric 3D-2D face recognition when combining the two separate matching steps.

5.6.2 The Effectiveness of Preprocessing Pipeline

To illustrate the effectiveness of our preprocessing pipeline, we launched experiments on cropped facial images without the preprocessing pipeline for both 2D-2D matching and 3D-2D matching. Specifically, for 2D-2D matching, a holistic method, PCA, and a feature-based one, LBP, were used for feature extraction, followed by SRC for classification. While for 3D-2D matching, CCA was employed to classify PCA and LBP based features. In the LBP operator, we set the number of sampling points at 8 and the distance between the central pixel and its neighbors at 2; meanwhile the strategy of uniform pattern was adopted.

From Table 5.1, we can see that for both types of facial features, the proposed preprocessing pipeline is effective to improve the performance in 2D-2D face matching. The improvement on LBP-based histograms is not as great as that based on

Table 5.1: Rank-one recognition rates of 2D-2D face matching on original cropped facial images and the ones after preprocessing pipeline using PCA and LBP based feature classified by SRC.

2D-2D Face Matching	PCA	LBP Histogram
Original Faces	46.68%	75.74%
Preprocessed Faces	78.54%	79.95%

Table 5.2: Rank-one recognition rates of 3D-2D face matching on original cropped facial images and the ones after preprocessing pipeline using PCA and LBP based feature classified by CCA.

3D-2D Face Matching	PCA	LBP Image
Original Faces	36.18%	42.42%
Preprocessed Faces	81.70%	76.36%

PCA facial features, because local features generally require less normalization than the holistic ones do.

As regarding the effectiveness of the preprocessing pipeline in 3D-2D matching, the conclusion accords with that in the step of 2D-2D face matching and Table 5.2 clearly demonstrates that it largely improves the results of both types of features.

To sum up, the proposed preprocessing method is effective to improve the results of the PCA and LBP based facial features in both 2D-2D and 3D-2D face matching.

5.6.3 The Performance of Individual Matching Steps

In this experiment, we executed both 2D-2D and 3D-2D face matching on the LBP and OGM features of facial range and texture images to observe and compare their performance.

Using LBP features, in the step of 2D-2D face matching, we set the number of sampling points in the LBP operator at 8, and the radius value varied from 1 to 8 pixels. These LBP operators were directly applied to facial texture images. While in 3D-2D matching, using the same LBP operators as in 2D-2D matching, Multi-Scale LBP images were generated from the original texture and range images respectively.

Employing OGM features, each facial range or texture image was represented by a set of OGMs; therefore, as described in Section 5.4, both 2D-2D and 3D-2D face matching were applied to OGMs of facial range and texture images instead of the original ones. To calculate OGMs, we set the radius value of OGMs at 1 and the number of orientations at 8. In its 2D-2D matching step, only the operator $LBP^{U_2}_{(8,2)}$ was utilized, and we did not perform MS-LBP in order to save computational cost.

Table 5.4 shows that in 2D-2D face matching, OGM based result (Sym26) which combines the information of eight orientations is much better than the ones based on

Table 5.3: The rank-one recognition rate of 3D-2D face matching based on different facial features using CCA.

3D-2D Matching	**Accuracy**
(Asy01) OGM1 + PCA	0.8772
(Asy02) OGM2 + PCA	0.8726
(Asy03) OGM3 + PCA	0.8582
(Asy04) OGM4 + PCA	0.8644
(Asy05) OGM5 + PCA	0.8630
(Asy06) OGM6 + PCA	0.8690
(Asy07) OGM7 + PCA	0.8421
(Asy08) OGM8 + PCA	0.8582
(Asy09) $LBP^{U2}_{(8,1)}$ Image + PCA	0.6461
(Asy10) $LBP^{U2}_{(8,2)}$ Image + PCA	0.7636
(Asy11) $LBP^{U2}_{(8,3)}$ Image + PCA	0.8128
(Asy12) $LBP^{U2}_{(8,4)}$ Image + PCA	0.8328
(Asy13) $LBP^{U2}_{(8,5)}$ Image + PCA	0.8492
(Asy14) $LBP^{U2}_{(8,6)}$ Image + PCA	0.8585
(Asy15) $LBP^{U2}_{(8,7)}$ Image + PCA	0.8540
(Asy16) $LBP^{U2}_{(8,8)}$ Image + PCA	0.8571
(Asy17) OGMs (Asy01-Asy08)	**0.9404**
(Asy18) MS-LBP (Asy09-Asy16)	0.8700
(Asy19) Original Face + PCA	0.8170

single LBP (Sym17-Sym24) and PCA (Sym28) features, and is even superior to that of MS-LBP by fusing all single LBP results (Sym27). Except the neighborhood of 1 pixel (Sym17) and 8 pixels (Sym24), all the other ones based single LBP histogram features (Sym18-Sym23) are still better than PCA features (Sym28) when working on well normalized facial images, and certainly, the performance of MS-LBP largely surpasses that of PCA.

As applied to each OGM, the PCA based method (Sym09-Sym16) achieves better results than LBP does (Sym01-Sym08). However, when combining the similarity scores of all the eight orientations of the OGMs to calculate the final accuracy, LBP (Sym25) slightly outperforms PCA (Sym26). To both features, i.e. PCA and LBP based ones, operating on each OGM and then fusing the similarity measurements generates much better performance than directly applying on the preprocessed faces.

In 3D-2D matching, by encoding both the local texture and geometry variations of textured 3D face models, both MS-LBP (Asy18) and OGM (Asy17) provide higher accuracies than PCA (Asy19) does. The performance combining the similarities of eight OGMs is better than the MS-LBP image-based one which fuses the scores of eight different scales as illustrated in Table 5.3. Once again, it proves that combining information provided by the LBP operators at various scales to achieve the MS-LBP

facial representation is a promising way to improve the performance of the single scale LBP.

Table 5.4: The rank-one recognition rate of 2D-2D face matching based on different facial features using SRC.

2D-2D Matching	**Accuracy**
(Sym01) OGM1 + $LBP^{U_2}_{(8,2)}$ Histograms	0.8114
(Sym02) OGM2 + $LBP^{U_2}_{(8,2)}$ Histograms	0.8082
(Sym03) OGM3 + $LBP^{U_2}_{(8,2)}$ Histograms	0.8018
(Sym04) OGM4 + $LBP^{U_2}_{(8,2)}$ Histograms	0.8116
(Sym05) OGM5 + $LBP^{U_2}_{(8,2)}$ Histograms	0.8068
(Sym06) OGM6 + $LBP^{U_2}_{(8,2)}$ Histograms	0.8150
(Sym07) OGM7 + $LBP^{U_2}_{(8,2)}$ Histograms	0.8130
(Sym08) OGM8 + $LBP^{U_2}_{(8,2)}$ Histograms	0.8170
(Sym09) OGM1 + PCA	0.8879
(Sym10) OGM2 + PCA	0.8924
(Sym11) OGM3 + PCA	0.8834
(Sym12) OGM4 + PCA	0.8743
(Sym13) OGM5 + PCA	0.8772
(Sym14) OGM6 + PCA	0.8848
(Sym15) OGM7 + PCA	0.8577
(Sym16) OGM8 + PCA	0.8786
(Sym17) Original Face + $LBP^{U_2}_{(8,1)}$ Histograms	0.6854
(Sym18) Original Face + $LBP^{U_2}_{(8,2)}$ Histograms	0.7995
(Sym19) Original Face + $LBP^{U_2}_{(8,3)}$ Histograms	0.8210
(Sym20) Original Face + $LBP^{U_2}_{(8,4)}$ Histograms	0.8503
(Sym21) Original Face + $LBP^{U_2}_{(8,5)}$ Histograms	0.8472
(Sym22) Original Face + $LBP^{U_2}_{(8,6)}$ Histograms	0.8342
(Sym23) Original Face + $LBP^{U_2}_{(8,7)}$ Histograms	0.7933
(Sym24) Original Face + $LBP^{U_2}_{(8,8)}$ Histograms	0.7388
(Sym25) OGMs (Sym01-08) + $LBP^{U_2}_{(8,2)}$ Histograms	**0.9390**
(Sym26) OGMs (Sym09-16) + PCA	0.9365
(Sym27) Original Face + MS-LBP (Sym17-24)	0.8918
(Sym28) Original Face + PCA	0.7854

These experiments of both matching steps clearly show that the proposed MS-LBP and OGM based descriptions improve the distinctiveness of original facial images, and thereby lead to the great improvement on the final recognition accuracy. When comparing OGM with MS-LBP, OGM works even better.

5.6.4 Radius Analysis of OGM Neighborhood

In our experiments, we evaluated different values of radius R and studied its impacts on the performance of both 2D-2D and 3D-2D face matching. In 2D-2D matching, the PCA approach was applied to OGMs for feature extraction, and SRC was further utilized for classification. In 3D-2D face matching, CCA was introduced to calculate the similarity scores.

From Fig. 5.7, it is obvious that in both steps of 2D-2D and 3D-2D face matching, when we increase the radius value of the OGM neighborhood, the performance of each OGM and their fusion degrades. The best results are achieved as the radius value is set at 1 in both the matching steps.

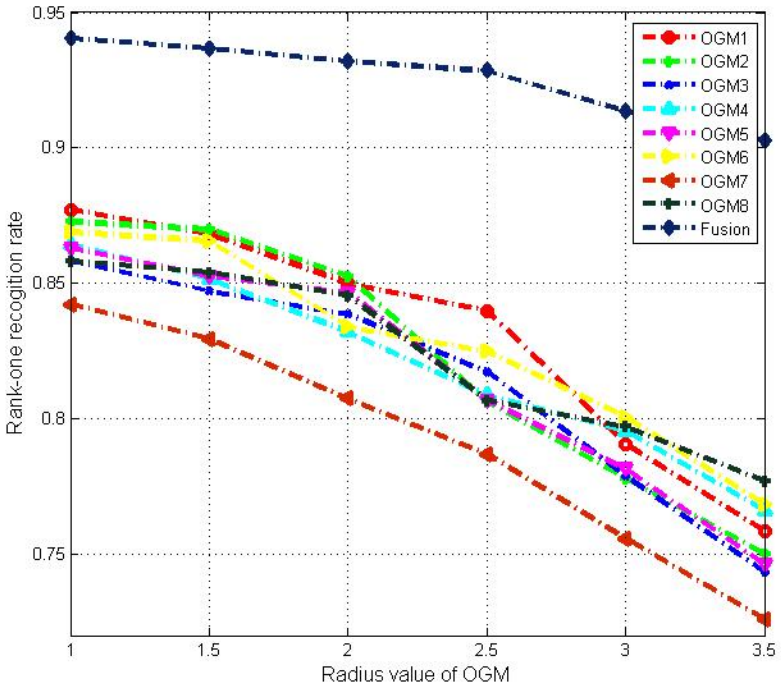
Table 5.5 and 5.6 list the exact accuracies of each OGM using different values of radius in 2D-2D matching and 3D-2D matching respectively.

Table 5.5: Results when using different sizes of OGM neighborhood in 2D-2D face matching.

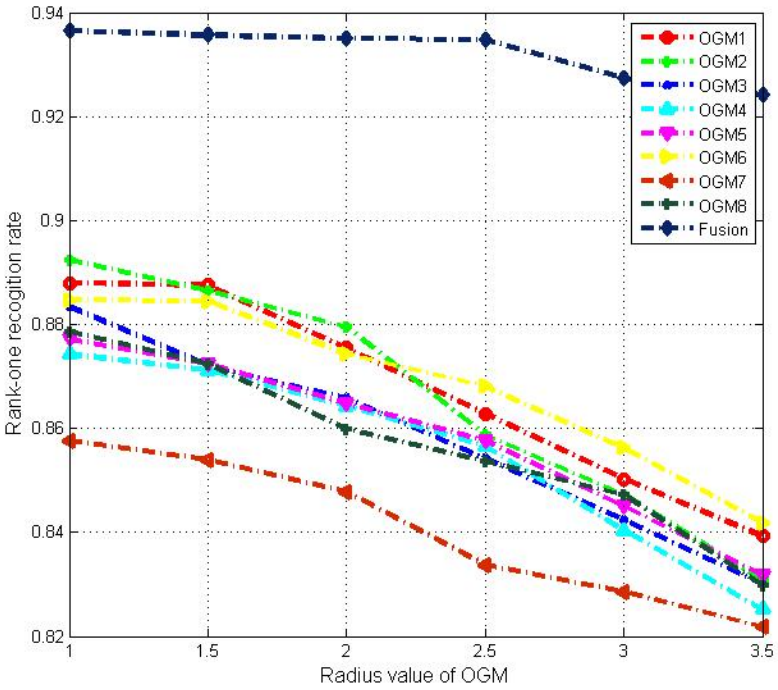
2D-2D	$R = 1.0$	$R = 1.5$	$R = 2.0$	$R = 2.5$	$R = 3.0$	$R = 3.5$
OGM_1	88.79%	88.76%	87.55%	86.28%	85.03%	83.93%
OGM_2	89.24%	88.65%	87.94%	85.88%	84.72%	83.06%
OGM_3	88.34%	87.21%	86.59%	85.43%	84.24%	83.00%
OGM_4	87.43%	87.12%	86.42%	85.65%	84.04%	82.52%
OGM_5	87.72%	87.24%	86.47%	85.77%	84.50%	83.17%
OGM_6	88.48%	88.45%	87.43%	86.81%	85.63%	84.19%
OGM_7	85.77%	85.40%	84.78%	83.37%	82.86%	82.18%
OGM_8	87.86%	87.24%	85.99%	85.37%	84.72%	82.97%
Fusion	93.65%	93.56%	93.50%	93.48%	92.74%	92.40%

Table 5.6: Results when using different sizes of OGM neighborhood in 3D-2D face matching.

3D-2D	$R = 1.0$	$R = 1.5$	$R = 2.0$	$R = 2.5$	$R = 3.0$	$R = 3.5$
OGM_1	87.72%	86.84%	85.00%	83.99%	79.07%	75.85%
OGM_2	87.26%	86.98%	85.26%	80.63%	77.80%	75.01%
OGM_3	85.82%	84.72%	83.87%	81.73%	77.89%	74.36%
OGM_4	86.44%	85.15%	83.20%	80.88%	79.55%	76.62%
OGM_5	86.30%	85.26%	84.64%	80.68%	78.14%	74.61%
OGM_6	86.90%	86.56%	83.39%	82.49%	80.06%	76.84%
OGM_7	84.21%	82.94%	80.74%	78.68%	75.57%	72.61%
OGM_8	85.82%	85.40%	84.55%	80.68%	79.70%	77.69%
Fusion	94.04%	93.67%	93.19%	92.83%	91.36%	90.26%



(a)



(b)

Figure 5.7: The curves of performance based on different radii of each OGM as well as their fusion accuracy: (a) 3D-2D matching; (b) 2D-2D matching.

Table 5.7: Final performance of asymmetric 3D-2D face recognition when combining 2D-2D and 3D-2D matching steps.

Different Fusion	Recognition Rate
(F01): (Sym28) + (Asy19)	0.8901
(F02): (Sym27) + (Asy18)	0.9127
(F03): (Sym25) + (Asy17)	**0.9537**

5.6.5 Identification of Asymmetric Face Recognition

In order to achieve the overall performance of the proposed asymmetric 3D-2D face recognition, the similarity measurements of the two separate matching steps, i.e., 2D-2D and 3D-2D matching, are finally fused. The accuracies of different combinations are listed in Table 5.7: in both the steps of 2D-2D and 3D-2D face matching, (F01) denotes that PCA based features are used; (F02) denotes MS-LBP based features are used; (F03) denotes OGM features are used.

As we can see from Table 5.7 the best performance is achieved by fusing (Sym25) and (Asy17). Specifically, the result of (Sym25) is based on the LBP histogram based facial features extracted from OGMs for 2D-2D matching, and (Asy17) is computed by CCA using the OGMs of facial range and texture images. The MS-LBP based accuracy is in the second place. These facts clearly prove that both LBP and OGM based facial description is effective to improve the distinctiveness of original texture or depth faces. Furthermore, to OGMs, two different types of facial features in two face matching steps, i.e. (1) local ones, LBP histograms in 2D-2D matching and (2) holistic ones, PCA in 3D -2D matching are complementary to each other, and their fusion result is better than either of them.

According to the experimental results, the proposed OGM-based method is competent in representing 2D and 3D facial images, leading to satisfying performance in both 2D-2D and 3D-2D matching. Meanwhile, the rank-one recognition rate of the entire asymmetric face recognition system is up to 95.4% obtained on the entire FRGC v2.0 database, which is better than those of most 2D-2D face recognition techniques, and not far behind 3D-3D based ones (about 97%).

5.7 Summary

This chapter presented a novel scenario for face recognition, asymmetric 3D-2D face recognition, making use of textured 3D face models for enrollment while only 2D facial images as probes. It consists of two separate face matching steps, i.e. 2D-2D and 3D-2D face matching, and their fusion. To describe local texton as well as shape variations of textured 3D face models, we firstly investigated Multi-Scale LBP (MS-LBP) based facial representation. Specifically, in 2D-2D face matching, MS-LBP histogram based features were extracted and then fed into SRC for classification, while in 3D-2D face matching, CCA was applied to learn the mapping between MS-

LBP images of range and texture faces. Then, we improved its performance by a biological vision-based facial description, namely OGMs. It simulates the response of complex neurons to gradient information within a pre-defined neighborhood, and thus can describe local texture changes of 2D faces and local shape variations of 3D faces. Due to its property of being highly distinctive, OGMs improve the results of both matching steps of asymmetric face recognition, i.e. (1) 3D-2D matching using CCA; (2) 2D-2D matching using LBP histogram based features and SRC. Some comparative experiments were carried out on the complete FRGC v2.0 database, and the results clearly demonstrate that the asymmetric face recognition outperforms the traditional 2D intensity image based techniques, while achieves comparable results to 3D face model based ones but with computational cost under control. Moreover, the performance also highlights the effectiveness of MS-LBP and OGM for facial representation and their successful application to asymmetric face recognition.

Chapter 6

Conclusions and Future Work

6.1 Contributions

Nowadays, 3D face models have emerged as a major alternative to solve the unsolved issues in 2D face recognition, i.e. illumination and pose changes, because they convey much more useful information than 2D facial images do. This Ph.D. thesis mainly concentrates on the problem of face recognition by using three dimensional data. In the first half, we extensively review the milestone techniques in 2D intensity image based face recognition as well as the recent development in 3D domain. We further dedicate to three scenarios in face recognition, including only 3D shape based face recognition, textured 3D face recognition, and asymmetric 3D-2D face recognition.

The contributions in this research work are as follows.

6.1.1 Only Shape based 3D Face Recognition

In only 3D shape-based face recognition, we present an effective and efficient method by using a novel geometric facial representation along with a local feature matching scheme. The proposed facial description is based on a set of facial depth maps extracted by multi-scale extended Local Binary Patterns (eLBP) and enables accurate and fast representation of local shape changes; it thus enhances the distinctiveness of generally smooth and similar facial range images. The following matching step is SIFT-based and performs in a local manner which is further improved by facial component and configuration constraints. It hence can robustly associate keypoints between two facial representations of the same individual. This approach displays a rank-one recognition rate up to 97.6% and a verification rate of 98.4% at a 0.001 FAR respectively on the FRGC v2.0 database without any registration. Furthermore, experimental results on the Gavab dataset demonstrate that the entire system is also robust to partially occluded faces when only aided by a coarse alignment process. Finally, additional experiments show that this approach is insensitive to facial expression variations and has a good tolerance to data degradations, such as Gaussian noise, decimation, and random holes.

6.1.2 Textured 3D Face Recognition

In textured 3D face recognition, we propose a novel biological vision-based facial description, namely Oriented Gradient Maps (OGMs), aiming to highlight intra-class and inter-class variations of both facial range and texture images. These generated OGMs simulate the response of complex neurons to gradient information within a pre-defined neighborhood and possess the properties of being highly distinctive and robust to affine illumination and geometric transformation as well. Based on such

an intermediate facial representation, SIFT-based matching is further carried out to calculate similarity scores between a given probe face and the gallery ones. Because the facial representation generates an OGM for each quantized gradient orientation of facial range and texture images, we then introduce a score level fusion strategy which optimizes the weights using a genetic algorithm in a learning step. Tested on the entire FRGC v2.0 database, the rank-one recognition rate using only 3D or 2D modality is 95.5% and 95.9%, respectively; while fusing both modalities, i.e. range and texture-based OGMs, the final accuracy is 98.0%, demonstrating the effectiveness of the biological vision-based facial description and the optimized weighted sum fusion. Moreover, the experiment on the 3DTEC database proves that 3D modality is more sensitive to facial expression variations than 2D modality.

6.1.3 Asymmetric 3D-2D Face Recognition

In order to limit the use of 3D data where they really help to improve the accuracy of face recognition and thereby avoid the high cost in data acquisition and computation, we design a novel scenario for face recognition, namely asymmetric 3D-2D face recognition, making use of textured 3D face models in the gallery set while only 2D facial images as probes. According to such a protocol, this scenario consists of two separate face matching steps, i.e. 2D-2D and 3D-2D face matching, and their combination. To describe local texton and geometry variations of textured 3D face models, we first investigate Multi-Scale LBP (MS-LBP) based facial representation. Specifically, in 2D-2D matching, MS-LBP histogram based features are extracted from original facial texture images and then fed into SRC for classification, while in 3D-2D matching, CCA is applied to learn the mapping between corresponding MS-LBP images of range and texture faces. We then improve its performance by OGMs. Since they simulate the response of complex neurons to gradient information within a given neighborhood, they thus can well describe local texton changes of texture faces and local shape variations of range faces. Due to its property of being highly distinctive, OGMs improve the results in both matching steps of asymmetric face recognition, i.e. (1) 3D-2D matching exploiting CCA; (2) 2D-2D matching using LBP histogram based features and SRC. Some comparative experiments are carried out on the complete FRGC v2.0 database, and the performance clearly demonstrates that the proposed asymmetric face recognition outperforms the traditional 2D intensity image based techniques, while achieves comparable results to 3D face model based ones but with computational cost under control. Moreover, these experiments also highlight the effectiveness of MS-LBP and OGM for facial representation and their successful application to asymmetric face recognition.

6.2 Discussions on Different Scenarios

As presented in this thesis, we proposed three approaches in three scenarios of face recognition respectively, i.e. only 3D shape based face recognition, textured 3D face recognition, and asymmetric face recognition. They make use of different types of 3D data either in the offline enrollment step (gallery set) or the online identification step

Table 6.1: The comparison among the three proposed approaches in three different face recognition scenarios respectively.

Different Scenarios	Gallery	Probe	Recognition Rate
Only 3D shape based	Shape	Shape	97.6% (Rank-one)
Textured 3D	Shape and Texture	Shape and Texture	98.0% (Rank-one)
Asymmetric 3D-2D	Shape and Texture	Texture	95.4% (Rank-one)

(probe set). We compare their performance in the application of face recognition on the FRGC v2.0 database.

From Table 6.1, we can see that textured 3D face recognition achieves the best result because it exploits both the texture and shape clues, and according to the experiments conducted on the FRGC v2.0 database, we can draw the conclusion that textured 3D face recognition provides more accurate decisions. While by calculating similarity scores using geometric information, 3D shape based face recognition takes the second place, and the performance is only slightly inferior to that of the proposed multi-modal face recognition system. The asymmetric 3D-2D face recognition does not work as well as the other two, but its accuracy is not very far behind them.

When considering the efficiency of these three scenarios, although the computational cost of only shape based face recognition and textured 3D face recognition can be largely reduced based on the parallel system design, e.g. GPU computing, they are also limited by the high expenditure of data acquisition and even preprocessing operations. With out a breakthrough of hardware, they cannot be implemented in real time (response time less than 1 second) in the near further. In contrast to them, asymmetric 3D-2D face recognition depending only on 2D facial images for online identification probably gives an alternative solution which utilizes 3D information but keep the cost under control.

6.3 Perspectives for Future Work

In this section, possible extensions of this thesis that we envisage are presented.

6.3.1 Advances in 3D Landmarking

Facial landmarks are points of correspondence on faces that matches between and within populations, and they are expected to possess consistent reproducibility even in adverse situations such as facial expression and occlusions. These facial landmarks generally include nose tip, inner eye corners, outer eye corners, mouth corners, etc. Locating landmarks on 3D faces is a fundamental step in 3D face analysis, especially for 3D face recognition and 3D facial expression recognition. Most techniques of 3D face recognition demand a certain number of landmark points to provide a good initial condition for an ICP-like surface registration.

Even though, the methods of (textured) 3D face recognition proposed in this the-

sis do not require any registration for nearly frontal faces, they need a coarse alignment when facial scans have severe pose changes, and three landmarks (nose tip and inner eye corners) should be located before that. The method [Szeptycki *et al.* 2009] is applied for this aim, and it is effective when all these landmarks are available at the same time. To some facial scans with extreme pose variations, i.e. left and right profiles, we still have to manually landmarked several points to roughly correct their pose.

Current landmarking solutions are mainly based on curvature or symmetry analysis of facial surfaces, which enable at least finding out the positions of three most stable points: nose tip and inner eye corners with an excellent precision (above 98% accuracy within error of 10mm). However, operating on the profiles, they always fail, because half of the facial surface is missing, its symmetry disappears, and the ear region also brings in some points whose curvature values are similar to the target ones. Both the reasons impede the accurate landmarking result. Therefore, a more powerful technique should be explored to precisely landmark on facial profiles.

6.3.2 Importance Analysis of Facial Regions

It has been investigated in several tasks that different facial regions are of different importance to the face recognition accuracy, because some of them are relative rigid to facial expression variations while some of them are not. Base on this conclusion, we can thereby improve the performance of the first two scenarios in 3D face recognition explored in this thesis, i.e. only 3D shape based face recognition and textured 3D face recognition.

These impacts of different facial regions on face recognition results can be measured by corresponding weights learned from a special database which contains rich facial expression variations, such as the BU-3DFE [Yin *et al.* 2006b] and Bosphorus [Savran *et al.* 2008] datasets. For the SIFT-based local matching used in these two applications, we can assign different weights to the matches between keypoints from different facial regions to see whether it will lead to an improvement in performance.

6.3.3 Further Investigation in Asymmetric Face Recognition

In the framework of asymmetric face recognition, we apply AAM based pose correction to facial texture images in the probe set, which is especially critical to the step of 3D-2D face matching. In our case, considering that the AAM model is built using nearly frontal faces, the following AAM fitting process is only evaluated on the same type of facial texture images with good resolutions, which is an ideal condition for carrying out experiments.

When implement it in the real world, e.g. to identify faces captured by cameras designed for video surveillance or internet surfing, we will definitely encounter two main difficulties: i.e. pose (yaw and pitch) changes as well as low resolutions. For the former one, a potential solution is to construct several individual AAMs in different poses pre-defined to cover a comprehensive distribution of pose variations similar to the approach in [Ramnath *et al.* 2008]; while to address the latter one, the quality of 2D facial images output by the previous detection step should be assessed to make

the system more robust and flexible.

Chapter 7

Publications

During my Ph.D. studying, I have published twelve publications in international conferences and one in an international journal. Two journal papers and one French patent are currently under review. In addition, one journal paper is being finalized.

International Conference:

1. **D. Huang**, M. Ardabilian, Y. Wang, L. Chen: Asymmetric 3D/2D face recognition based on LBP facial representation and canonical correlation analysis, IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 2009;

2. W. Ben Soltana, **D. Huang**, M. Ardabilian, L. Chen, C. Ben Amar: Comparison of 2D/3D features and their adaptive score level Fusion for 3D Face Recognition, International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT), Paris, France, 2010;

3. **D. Huang**, M. Ardabilian, Y. Wang, L. Chen: Automatic asymmetric 3D-2D face recognition, International Conference on Pattern Recognition (ICPR), Istanbul, Turkey, 2010;

4. X. Zhao, **D. Huang**, E. Dellandréa, L. Chen: Automatic 3D facial expression recognition based on a bayesian belief net and a statistical facial feature model, International Conference on Pattern Recognition (ICPR), Istanbul, Turkey, 2010;

5. **D. Huang**, G. Zhang, M. Ardabilian, Y. Wang, L. Chen: 3D face recognition using distinctiveness enhanced facial representation and local feature hybrid matching, IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), Washington DC, USA, 2010;

6. **D. Huang**, K. Ouji, M. Ardabilian, Y. Wang, L. Chen: 3D face recognition based on local shape patterns and sparse repre-sentation classifier, International Conference on MultiMedia Modeling (MMM), Taipei, Taiwan, 2011;

7. **D. Huang**, M. Ardabilian, Y. Wang, L. Chen: A novel geometric facial representation based on multi-scale extended local binary patterns, IEEE International Conference on Automatic Face and Gesture Recognition (FG), Santa Barbara, USA, 2011;

8. **D. Huang**, W. Ben Soltana, M. Ardabilian, Y. Wang, L. Chen: Textured 3D face recognition using biological vision-based facial representation and optimized weighted sum fusion, IEEE Conference on Computer Vision and Pattern Recognition Workshop on Biometrics (CVPRW), Colorado Spring, USA, 2011;

9. W. Ben Soltana, **D. Huang**, M. Ardabilian, L. Chen, C. Ben Amar: A mixture of gated experts optimized using simulated annealing for 3D face recognition, IEEE International Conference on Image Processing (ICIP), Brussels, Belgium, 2011;

10. H. Li, **D. Huang**, P. Lemaire, J.-M. Morvan, L. Chen: Expression robust 3D face recognition via mesh-based histograms of multiple order surface differential quantities, IEEE International Conference on Image Processing (ICIP), Brussels, Belgium, 2011.

11. H. Li, **D. Huang**, J.-M. Morvan, L. Chen: Learning Weighted Sparse Representation of Encoded Facial Normal Information for Expression-Robust 3D Face Recognition, IEEE International Joint Conference on Biometrics (IJCB), Washington DC, USA, 2011;

12. V. Vijayan, K. Bowyer, P. Flynn, **D. Huang**, L. Chen, M. Hansen, S. Shah, Omar Ocegueda, and I. Kakadiaris: Twins 3D Face Recognition Challenge, IEEE International Joint Conference on Biometrics (IJCB), Washington DC, USA, 2011.

International Journal:

1. **D. Huang**, C. Shan, M. Ardabilian, Y. Wang, L. Chen: Local binary patterns and its application to facial image analysis: a survey, IEEE Transactions on Systems, Man, and Cybernetics, Part C (TSMC-C): Applications and Reviews, 2011.

International Journal under Review:

1. **D. Huang**, M. Ardabilian, Y. Wang, L. Chen: 3D Face Recognition using eLBP-based Facial Description and Local Feature Hybrid Matching, Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI);

2. **D. Huang**, M. Ardabilian, Y. Wang, L. Chen: Automatic Asymmetric 3D-2D Face Recognition, IEEE Transactions on Information Forensics and Security (TIFS).

International Journal to be Finalized:

1. **D. Huang**, M. Ardabilian, Y. Wang, L. Chen: Textured 3D Face Recognition based on Oriented Gradient Maps and optimized weighted sum fusion.

French Patent:

1. L. Chen and **D. Huang**: Oriented Gradient Maps (OGM) and SIFT-based Matching for Biometric Applications, B20230.

Bibliography

[Abate *et al.* 2005] A. F. Abate, M. Nappi, S. Ricciardi and G. Sabatino. *Fast 3D face recognition based on normal map.* In IEEE International Conference on Image Processing, volume 2, pages 946–949, Sep. 2005. 38, 54

[Abate *et al.* 2006a] A. F. Abate, M. Nappi, S. Ricciardi and G. Sabatino. *Multimodal face recognition by means of augmented normal map and PCA.* In IEEE International Conference on Image Processing, pages 649–652, Oct. 2006. 85

[Abate *et al.* 2006b] A. F. Abate, M. Nappi, D. Riccio and G. Sabatino. *3D face recognition using normal sphere and general Fourier descriptor.* volume 3, pages 1183–1186, 2006. 38, 54

[Achermann & Bunke 2000] B. Achermann and H. Bunke. *Classifying range images of human faces with Hausdorff distance.* In International Conference on Pattern Recognition, volume 2, pages 809–813, 2000. 36, 37, 38, 42

[Achermann *et al.* 1997] B. Achermann, X. Jiang and H. Bunke. *Face recognition using range images.* In International Conference on Virtual Systems and MultiMedia, pages 129–136, 1997. 36, 38, 40, 59

[Ahonen *et al.* 2004] T. Ahonen, A. Hadid and M. Pietikainen. *Face Recognition with Local Binary Patterns.* In European Conference on Computer Vision, pages 469–481, 2004. viii, 26, 27, 61, 81, 113

[Ahonen *et al.* 2006] T. Ahonen, A. Hadid and M. Pietikainen. *Face Description with Local Binary Patterns: Application to Face Recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pages 2037–2041, Dec. 2006. viii, 2, 15, 26

[Al-Osaimi *et al.* 2009] F. R. Al-Osaimi, M. Bennamoun and A. S. Mian. *An expression deformation approach to non-rigid 3D face recognition.* International Journal of Computer Vision, vol. 81, no. 3, pages 302–316, 2009. 39, 45, 56

[Alyuz *et al.* 2008] N. Alyuz, B. Gokberk, H. Dibeklioglu and L. Akarun. *Component-based registration with curvature descriptors for expression insensitive 3D face recognition.* In IEEE International Conference on Automatic Face Gesture Recognition, pages 1–6, Sep. 2008. 39

[Alyuz *et al.* 2010] N. Alyuz, B. Gookberk and L. Akarun. *Regional registration for expression resistant 3-D face recognition.* IEEE Transactions on Information Forensics and Security, vol. 5, no. 3, pages 425–440, Sep. 2010. viii, 39, 51, 56, 73, 74, 97

[Amberg *et al.* 2007] B. Amberg, S. Romdhani and T. Vetter. *Optimal Step Nonrigid ICP Algorithms for Surface Registration.* In IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, Jun. 2007. 45

[Amberg *et al.* 2008] B. Amberg, R. Knothe and T. Vetter. *Expression invariant 3D face recognition with a Morphable Model.* In IEEE International Conference on Automatic Face Gesture Recognition, pages 1–6, Sep. 2008. 39, 44

[Antini *et al.* 2006] G. Antini, S. Berretti, A. Del Bimbo and P. Pala. *3D Face Identification Based on Arrangement of Salient Wrinkles.* In IEEE International Conference on Multimedia and Expo, pages 85–88, 2006. 38, 47

[Arca *et al.* 2007] S. Arca, R. Lanzarotti and G. Lipori. *Face recognition based on 2D and 3D features.* In International Conference on Knowledge-based Intelligent Information and Engineering Systems, pages 455–462, Berlin, Heidelberg, 2007. Springer-Verlag. 93

[Baker 1987] J. E. Baker. *Reducing bias and inefficiency in the selection algorithm.* In International Conference on Genetic Algorithms on Genetic Algorithms and Their Application, pages 14–21, 1987. 91

[Bartlett *et al.* 2002] M. S. Bartlett, J. R. Movellan and T. J. Sejnowski. *Face recognition by independent component analysis.* IEEE Transactions on Neural Networks, vol. 13, pages 1450–1464, 2002. vii, 15, 17, 19

[Belhumeur *et al.* 1997] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman. *Eigenfaces vs. Fisherfaces: recognition using class specific linear projection.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pages 711–720, Jul. 1997. 2, 15, 19, 81

[Ben Amor *et al.* 2006a] B. Ben Amor, M. Ardabilian and L. Chen. *Enhancing 3D face recognition by mimics segmentation.* In International Conference on Intelligent Systems Design and Applications, pages 150–155, 2006. 38, 50

[Ben Amor *et al.* 2006b] B. Ben Amor, M. Ardabilian and L. Chen. *New experiments on ICP-based 3D face recognition and authentication.* In International Conference on Pattern Recognition, pages 1195–1199, 2006. 38, 112

[Ben Soltana *et al.* 2010] W. Ben Soltana, D. Huang, M. Ardabilian, L. Chen and C. Ben Amar. *Comparison of 2D/3D features and their adaptive score level fusion for 3D face recognition.* In International Symposium on 3D Data Processing, Visualization and Transmission, 2010. 8, 98, 99

[Bentley 1975] J. Bentley. *Multidimensional binary search trees used for associative searching.* The Communications of ACM, vol. 18, pages 509–517, Sep. 1975. 40

[Berretti *et al.* 2006] S. Berretti, A. Del Bimbo and P. Pala. *3D face recognition by modeling the arrangement of concave and convex regions.* In International Workshop on Adaptive Multimedia Retrieval: User, Context, and Feedback, pages 108–118, 2006. 76

[Berretti *et al.* 2008] S. Berretti, A. Del Bimbo and P. Pala. *3D face recognition by spatial arrangement of iso-geodesic surfaces.* In 3DTV Conference: The True

Vision - Capture, Transmission and Display of 3D Video, pages 365–368, May 2008. 39, 48

[Berretti *et al.* 2010] S. Berretti, A. Del Bimbo and P. Pala. *3D face recognition using isogeodesic stripes.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 12, pages 2162–2177, Dec. 2010. 37, 39, 48

[Besl & McKay 1992] P. J. Besl and H. D. McKay. *A method for registration of 3-D shapes.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pages 239–256, Feb. 1992. 40, 59

[Beumier & Acheroy 2000] C. Beumier and M. Acheroy. *Automatic 3D face authentication.* Image and Vision Computing, vol. 18, no. 4, pages 315–321, 2000. 38, 59

[Beumier & Acheroy 2001] C. Beumier and M. Acheroy. *Face verification from 3D and grey level clues.* Pattern Recognition Letters, vol. 22, no. 12, pages 1321–1329, 2001. 10, 82

[Beymer 1994] D. J. Beymer. *Face recognition under varying pose.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 756–761, Jun. 1994. vii, 3

[Bicego *et al.* 2006] M. Bicego, A. Lagorio, E. Grosso and M. Tistarelli. *On the use of SIFT features for face authentication.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop, volume 0, pages 35–40, 2006. 15, 31

[Blanz & Vetter 1999] V. Blanz and T. Vetter. *A morphable model for the synthesis of 3D faces.* In Annual Conference on Computer Graphics, pages 187–194, 1999. 10, 44

[Bowyer *et al.* 2006] K. W. Bowyer, K. Chang and P. Flynn. *A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition.* Computer Vision and Image Understanding, vol. 101, pages 1–15, 2006. 1, 3, 6, 7, 37, 59

[Bronstein *et al.* 2003] A. M. Bronstein, M. M. Bronstein and R. Kimmel. *Expression-invariant 3D face recognition.* In International conference on Audio- and Video-based Biometric Person Authentication, pages 62–70, 2003. 59, 83

[Bronstein *et al.* 2004a] A. M. Bronstein, M. M. Bronstein, E. Gordon and R. Kimmel. *Fusion of 2D and 3D data in three-dimensional face recognition.* In IEEE International Conference on Image Processing, volume 1, pages 87–90, Oct. 2004. 83

[Bronstein *et al.* 2004b] A. M. Bronstein, M. M. Bronstein, A. Spira and R. Kimmel. *Face recognition from facial surface metric.* In European Conference on Computer Vision, pages 225–237, 2004. 37, 38

[Bronstein *et al.* 2005] A. M. Bronstein, M. M. Bronstein and R. Kimmel. *Three-dimensional face recognition.* International Journal of Computer Vision, vol. 64, no. 1, pages 5–30, 2005. 37, 38, 54, 60

[Bronstein *et al.* 2007] A. M. Bronstein, M. M. Bronstein and R. Kimmel. *Expression-invariant representations of faces.* IEEE Transactions on Image Processing, vol. 16, no. 1, pages 188–197, Jan. 2007. 37, 39

[Burr *et al.* 1989] D. C. Burr, M. C. Morrone and D. Spinelli. *Evidence for edge and bar detectors in human vision.* Vision Research, vol. 29, no. 4, pages 419–431, 1989. 23

[Cardinaux *et al.* 2003] F. Cardinaux, C. Sanderson and S. Marcel. *Comparison of MLP and GMM classifiers for face verification on XM2VTS.* In International Conference on Audio- and Video-based Biometric Person Authentication, numéro 10, 2003. 32

[Cartoux *et al.* 1989] J. Y. Cartoux, J. T. LaPreste and M. Richetin. *Face authentication or recognition by profile extraction from range images.* In Workshop on Interpretation of 3D Scenes, pages 194–199, 1989. 37, 47

[Castellani *et al.* 2008] U. Castellani, M. Cristani, X. Lu, V. Murino and A. K. Jain. *HMM-based geometric signatures for compact 3D face representation and matching.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Biometrics, pages 1–6, Jun. 2008. 39, 46

[Chan *et al.* 2007] C. Chan, J. Kittler and K. Messer. *Multi-scale local binary pattern histograms for face recognition.* In International Conference on Biometrics, pages 809–818, 2007. 61, 65

[Chang *et al.* 2003a] K. I. Chang, K. W. Bowyer and P. J. Flynn. *Face recognition using 2D and 3D facial data.* In ACM Workshop on Multimodal User Authentication, pages 25–32, 2003. 7, 82

[Chang *et al.* 2003b] K. I. Chang, K. W. Bowyer and P. J. Flynn. *Multi-Modal 2D and 3D biometrics for face recognition.* In IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pages 187–194, 2003. 82

[Chang *et al.* 2005a] K. I. Chang, K. W. Bowyer and P. J. Flynn. *Adaptive rigid multi-region selection for handling expression variation in 3D face recognition.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, pages 157–157, 2005. 37, 38, 50, 73, 97

[Chang *et al.* 2005b] K. I. Chang, K. W. Bowyer and P. J. Flynn. *Effects on facial expression in 3D face recognition.* In SPIE Conference on Biometric Technology for Human Identification, volume 5779, pages 132–143, 2005. 38, 50, 59

[Chang *et al.* 2005c] K. I. Chang, K. W. Bowyer and P. J. Flynn. *An evaluation of multimodal 2D + 3D face biometrics.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 4, pages 619–624, Apr. 2005. 82

[Chang *et al.* 2006] K. I. Chang, K. W. Bowyer and P. J. Flynn. *Multiple nose region matching for 3D face recognition under varying facial expression.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 10, pages 1695–1700, Oct. 2006. 37, 38, 50, 93

[Chen *et al.* 2000] L.-F. Chen, H.-Y. Mark-Liao, M.-T. Ko, J.-C. Lin and G.-J. Yu. *A new LDA-based face recognition system which can solve the small sample size problem.* Pattern Recognition, vol. 33, no. 10, pages 1713–1726, 2000. 20

[Chen *et al.* 2006] T. Chen, W. Yin, X. Zhou, D. Comaniciu and T. S. Huang. *Total variation models for variable lighting face recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 9, pages 1519–1524, Sep. 2006. 108, 110

[Chua *et al.* 2000] C. Chua, F. Han and Y. Ho. *3D human face recognition using point signature.* In IEEE International Conference on Automatic Face and Gesture Recognition, pages 233–238, 2000. 37, 38, 50, 59, 60

[Colbry *et al.* 2005] D. Colbry, G. Stockman and A. Jain. *Detection of anchor points for 3D face verication.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, pages 118–118, Jun. 2005. 56

[Colombo *et al.* 2005] A. Colombo, C. Cusano and R. Schettini. *A 3D face recognition system using curvature-based detection and holistic multimodal classification.* In International Symposium on Image and Signal Processing and Analysis, pages 179–184, Sep. 2005. 84

[Conde *et al.* 2007] C. Conde, A. Serrano, L. J. Rodriguez-Aragon and E. Cabello. *An automatic 2D, 2.5D & 3D score-based fusion face verification system.* In IEEE International Conference on Application-Specific Systems, Architectures and Processors, pages 214–219, 2007. 10

[Cook *et al.* 2005] J. Cook, V. Chandran, S. Sridharan and C. Fookes. *Gabor filter bank representation for 3D face recognition.* In Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications, page 4, 2005. 85

[Cook *et al.* 2006a] J. Cook, V. Chandran and C. Fookes. *3D face recognition using Log-Gabor templates.* In British Machine Vision Conference, 2006. 39, 73, 74, 97

[Cook *et al.* 2006b] J. Cook, C. Mccool, V. Chandran and S. Sridharan. *Combined 2D/3D face recognition using Log-Gabor templates.* In IEEE International Conference on Video and Signal Based Surveillance, page 83, Nov. 2006. 85

[Cook *et al.* 2007] J. Cook, M. Cox, V. Chandran and S. Sridharan. *Robust 3D face recognition from expression categorisation.* In International Conference on Biometrics, pages 271–280, 2007. 39, 50

[Cootes *et al.* 1998] T. F. Cootes, G. J. Edwards and C.J. Taylor. *Active Appearance Models.* In European Conference on Computer Vision, pages 484–498, 1998. 34

[Cootes *et al.* 2001] T. F. Cootes, G. J. Edwards and C. J. Taylor. *Active appearance models.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pages 681–685, Jun. 2001. viii, 33, 108

[Daniyal *et al.* 2009] F. Daniyal, P. Nair and A. Cavallaro. *Compact signatures for 3D face recognition under varying expressions.* In IEEE International Conference on Advanced Video and Signal Based Surveillance, pages 302–307, Sep. 2009. viii, 39, 45, 46

[Daugman 1980] J. G. Daugman. *Two-dimensional spectral analysis of cortical receptive field profiles.* Vision Research, vol. 20, no. 10, pages 847–856, 1980. 22

[Daugman 1985] J. G. Daugman. *Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters.* Journal of The Optical Society of America A: Optics Image Science and Vision, vol. 2, no. 7, pages 1160–1169, 1985. 22

[Daugman 1988] J. G. Daugman. *Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression.* IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 36, no. 7, pages 1169–1179, 1988. 22

[del Solar *et al.* 2009] J. Ruiz del Solar, R. Verschae and M. Correa. *Recognition of faces in unconstrained environments: a comparative study.* Eurasip Journal on Advances in Signal Processing, vol. 2009, no. 1, pages 1–20, 2009. 36

[Drira *et al.* 2010] H. Drira, B. Ben Amor, M. Daoudi and A. Srivastava. *Pose and expression-invariant 3D face recognition using elastic radial curves.* In British Machine Vision Conference, pages 90.1–90.11, 2010. vii, viii, 12, 39, 49, 75, 76, 77

[DTU-AAM] DTU-AAM. http://www2.imm.dtu.dk/ aam/. 111

[Edelman *et al.* 1997] S. Edelman, N. Intrator and T. Poggio. Complex cells and object recognition. http://kybele.psych.cornell.edu/ edelman/archive.html, 1997. 88

[Edwards *et al.* 1998] G. J. Edwards, T. F. Cootes and C. J. Taylor. *Face recognition using active appearance models.* In European Conference on Computer Vision, pages 581–595, 1998. 15, 32

[Faltemier *et al.* 2007] T. C. Faltemier, K. W. Bowyer and P. J. Flynn. *Using a Multi-Instance Enrollment Representation to Improve 3D Face Recognition.* In IEEE International Conference on Biometrics: Theory, Applications, and Systems, pages 1–6, Sept. 2007. 39

[Faltemier *et al.* 2008a] T. C. Faltemier, K. W. Bowyer and P. J. Flynn. *A region ensemble for 3-D face recognition.* IEEE Transactions on Information Forensics and Security, vol. 3, no. 1, pages 62–73, 2008. 37, 39, 50, 56, 69, 73, 74

[Faltemier *et al.* 2008b] T. C. Faltemier, K. W. Bowyer and P. J. Flynn. *Using multi-instance enrollment to improve performance of 3D face recognition.* Computer Vision and Image Understanding, vol. 112, no. 2, pages 114–125, 2008. 37, 39

[Faltemier 2007] T. C. Faltemier. *Flexible and robust 3D face recognition.* PhD thesis, The University of Notre Dame, 2007. 6

[Feng *et al.* 2006] S. Feng, H. Krim, I. Gu and M. Viberg. *3D face recognition using affine integral invariants.* In IEEE International Conference on Acoustics, Speech and Signal Processing, volume 2, pages 189–192, May 2006. 38, 48

[Field 1987] D. J. Field. *Relations between the statistics of natural images and the response properties of cortical cells.* Journal of The Optical Society of America A: Optics Image Science and Vision, vol. 4, no. 12, pages 2379–2394, 1987. 23

[Fu *et al.* 2010] Y. Fu, G. Guo and T. S. Huang. *Age Synthesis and Estimation via Faces: A Survey.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 11, pages 1955–1976, Nov. 2010. vii, 5

[G. G. Gordon 1992] Gaile Gibson G. G. Gordon. *Face recognition from depth and curvature.* PhD thesis, Harvard University, 1992. 37

[Gao *et al.* 2008] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang and D. Zhao. *The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations.* IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 38, no. 1, pages 149–161, Jan. 2008. vii, 4

[Godil *et al.* 2005] A. Godil, Y. Ressler and P. Grother. *Face recognition using 3d facial shape and color map information: comparison and combination.* In SPIE Conference on Biometric Technology for Human Identification, pages 351–361, 2005. 84, 95

[Gokberk *et al.* 2005] B. Gokberk, A. A. Salah and L. Akarun. *Rank-based decision fusion for 3D shape-based face recognition.* In International Conference on Audio- and Video-based Biometric Person Authentication, pages 1019–1028, 2005. 37, 38, 52

[Gokberk *et al.* 2008] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun and B. Sankur. *Representation plurality and fusion for 3-D face recognition.* IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 38, no. 1, pages 155–173, Feb. 2008. ix, 87, 98

[Gordon 1991] G. G. Gordon. *Face recognition based on depth maps and surface curvature.* In SPIE Conference on Geometric Method in Computer Vision, volume 1570, pages 234–247, 1991. 37

[Gordon 1992] G. G. Gordon. *Face recognition based on depth and curvature features.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 808–810, Jun. 1992. 6, 37, 52, 59

[Gross *et al.* 2006] R. Gross, I. Matthews and S. Baker. *Active appearance models with occlusion.* Image and Vision Computing, vol. 24, pages 593–604, 2006. 34

[Guan & Zhang 2008] P. Guan and L. Zhang. *3D face recognition based on facial structural angle and local region map.* In IEEE International Conference on Multimedia and Expo, pages 41–44, Apr. 2008. 39, 53

[Gupta *et al.* 2007a] S. Gupta, J. K. Aggarwal, M. K. Markey and A. C. Bovik. *3D face recognition founded on the structural diversity of human faces.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 1–7, Jun. 2007. 39, 46, 59

[Gupta *et al.* 2007b] S. Gupta, M. P. Sampat and Z. Wang. *Facial range image matching using the complex wavelet structural similarity metric.* In IEEE International Workshop on Applications of Computer Vision, page 4, Feb. 2007. 39, 51

[Hadid *et al.* 2004] A. Hadid, M. Pietikainen and T. Ahonen. *A discriminative feature space for detecting and recognizing faces.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 797–804, 2004. 59

[Hardoon *et al.* 2004] D. R. Hardoon, S. Szedmak and J. Shawe-Taylor. *Canonical correlation analysis: an overview with application to learning methods.* Neural Computation, vol. 16, no. 12, pages 2639–2664, 2004. 109, 114

[He *et al.* 2005] X. He, S. Yan, Y. Hu, P. Niyogi and H. Zhang. *Face recognition using Laplacianfaces.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pages 328–340, Mar. 2005. 21

[Heisele *et al.* 2003] B. Heisele, P. Ho, J. Wu and T. Poggio. *Face recognition: component-based versus global approaches.* Computer Vision and Image Understanding, vol. 91, no. 1-2, pages 6–21, 2003. 35

[Heseltine *et al.* 2004a] T. Heseltine, N. Pears and J. Austin. *Three-dimensional face recognition: a fishersurface approach.* In International Conference on Image Analysis and Recognition, pages 684–691, 2004. 38, 40

[Heseltine *et al.* 2004b] T. Heseltine, N. Pears and J. Austin. *Three-dimensional face recognition: an eigensurface approach.* In IEEE International Conference on Image Processing, volume 2, pages 1421–1424, Oct. 2004. 38, 40

[Heseltine 2005] T. Heseltine. *Face recognition: two-dimensional and three-dimensional techniques.* PhD thesis, The University of York, 2005. 10

[Hesher *et al.* 2003] C. Hesher, A. Srivastava and G. Erlebacher. *A novel technique for face recognition using range imaging.* In International Symposium on Signal Processing and Its Applications, volume 2, pages 201–204, Jul. 2003. 6, 38, 40, 59

[Hietmeyer 2000] R. Hietmeyer. *Biometric Identification promises fast and secure processing of airline passengers.* Internation Civil Aviation Organization Journal, vol. 55, no. 9, pages 10–11, 2000. vii, 1, 2

[Huang & Aviyente 2006] K. Huang and S. Aviyente. *Sparse representation for signal classification.* In Annual Conference on Neural Information Processing Systems, pages 609–616, 2006. 116

[Huang *et al.* 2006] Y. Huang, Y. Wang and T. Tan. *Combining statistics of geometrical and correlative features for 3D face recognition.* In British Machine Vision Conference, 2006. 39, 60, 61, 63

[Huang *et al.* 2007] Y. Huang, Y. Wang and T. Tan. *Discriminating 3D faces by statistics of depth differences.* In Asian Conference on Computer Vision, volume 2, pages 690–699, 2007. 37, 39, 52

[Huang *et al.* 2009] D. Huang, M. Ardabilian, Y. Wang and L. Chen. *Asymmetric 3D/2D face recognition based on LBP facial representation and canonical correlation analysis.* In IEEE International Conference on Image Processing, pages 3325–3328, 2009. 107

[Huang *et al.* 2010a] D. Huang, M. Ardabilian, Y. Wang and L. Chen. *Automatic asymmetric 3D-2D face recognition.* In International Conference on Pattern Recognition, pages 1225–1228, 2010. 107

[Huang *et al.* 2010b] D. Huang, G. Zhang, M. Ardabilian, Y. Wang and L. Chen. *3D face recognition using distinctiveness enhanced facial representations and local feature hybrid matching.* In IEEE International Conference on Biometrics: Theory Applications and Systems, pages 1–7, Sep. 2010. 73, 74, 97

[Huang *et al.* 2011a] D. Huang, M. Ardabilian, Y. Wang and L. Chen. *A novel geometric facial representation based on multi-scale extended local binary patterns.* In International Conference on Automatic Face and Gesture Recognition, pages 1–7, 2011. 97

[Huang *et al.* 2011b] D. Huang, W. Ben Soltana, M. Ardabilian, Y. Wang and L. Chen. *Textured 3D face recognition using biological vision-based facial*

representation and optimized weighted sum fusion. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Biometrics, 2011. 8

[Huang *et al.* 2011c] D. Huang, C. Shan, M. Ardabilian, Y. Wang and L. Chen. *Local binary patterns and its application to facial image analysis: a survey.* IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 41, no. 4, pages 1–17, 2011. 28

[Huang *et al.* 2011d] D. Huang, C. Shan, M. Ardabilian, Y. Wang and L. Chen. *Local binary patterns and its application to facial image analysis: a Survey.* IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, pages 1–17, 2011. 60

[Husken *et al.* 2005] M. Husken, M. Brauckmann, S. Gehlen and C.v. d. Malsburg. *Strategies and benefits of fusion of 2D and 3D face recognition.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, page 174, Jun. 2005. 73, 74, 84, 98

[Huttenlocher *et al.* 1993] D. P. Huttenlocher, G. A. Klanderman and W. J. Rucklidge. *Comparing images using the Hausdorff distance.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 9, pages 850–863, Sep. 1993. 41, 42

[Hwang *et al.* 2007] J. Hwang, W. L. Woo and S. S. Dlay. *A novel FNP-pose estimation for three-dimensional face recognition using DPCA under facial expression.* In International Conference on Digital Signal Processing, pages 235–239, Jul. 2007. 39

[Hyvarinen & Oja 2000] A. Hyvarinen and E. Oja. *Independent component analysis: algorithms and applications.* Neural Networks, vol. 13, pages 411–430, May 2000. 17

[Hyvarinen 1999] A. Hyvarinen. *Fast and robust fixed-point algorithms for independent component analysis.* IEEE Transactions on Neural Networks, vol. 10, pages 626–634, 1999. 18

[Irfanoglu *et al.* 2004] M. O. Irfanoglu, B. Gokberk and L. Akarun. *3D Shape-based face recognition using automatically registered facial surfaces.* In International Conference on Pattern Recognition, pages 183–186, 2004. 38

[Jahanbin *et al.* 2008] S. Jahanbin, H. Choi, Y. Liu and A. C. Bovik. *Three dimensional face recognition using iso-geodesic and iso-depth curves.* In IEEE International Conference on Biometrics: Theory, Applications and Systems, pages 1–6, Oct. 2008. 39, 48

[Jain *et al.* 2004] A. K. Jain, A. Ross and S. Prabhakar. *An introduction to biometric recognition.* IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pages 4–20, Jan. 2004. 3

[Jones & Palmer 1987] J. Jones and L. Palmer. *An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex.* Journal of Neurophysiology, pages 1233–1258, 1987. 23

[Kakadiaris *et al.* 2007] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis and T. Theoharis. *Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, pages 640–649, 2007. 37, 39, 41, 54, 56, 69, 73, 74, 97

[Kang *et al.* 2002] H. Kang, T. F. Cootes and C. J. Taylor. *A Comparison of face verification algorithms using appearance models.* In British Machine Vision Conference, pages 477–486, 2002. 34, 35

[Kirby & Sirovich 1990] M. Kirby and L. Sirovich. *Application of the Karhunen-Loeve procedure for the characterization of human faces.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 1, pages 103–108, 1990. 16

[Koenderink & Doorn 1992] J. J. Koenderink and A. J. Doorn. *Surface shape and curvature scales.* Image and Vision Computing, vol. 10, no. 8, pages 557–564, 1992. 63

[Kong *et al.* 2005] H. Kong, L. Wang, E. Teoh, J. Wang and R. Venkateswarlu. *A framework of 2D fisher discriminant analysis: application to face recognition with small number of training samples.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 1083–1088, 2005. 20

[Koudelka *et al.* 2005] M. L. Koudelka, M. W. Koch and T. D. Russ. *A prescreener for 3D face recognition using radial symmetry and the Hausdorff fraction.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, pages 168–168, June 2005. 38, 53

[Krizaj *et al.* 2010] J. Krizaj, V. Struc and N. Pavesic. *Adaptation of SIFT features for robust face recognition.* In International Conference on Image Analysis and Recognition, pages 394–404, 2010. 31

[Kusuma & Chua 2008] G. P. Kusuma and C. Chua. *Image level fusion method for multimodal 2D + 3D face recognition.* In International Conference on Image Analysis and Recognition, pages 984–992, 2008. 86

[Lades *et al.* 1993] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. v. d. Malsburg, R. P. Wurtz and W. Konen. *Distortion invariant object recognition in the dynamic link architecture.* IEEE Transactions on Computers, vol. 42, no. 3, pages 300–311, Mar. 1993. 22

[Lee & Milios 1990] J. C. Lee and E. Milios. *Matching range images of human faces.* In International Conference on Computer Vision, pages 722–726, Dec. 1990. 37, 49

[Lee & Shim 2004] Y. Lee and J. Shim. *Curvature based human face recognition using depth weighted Hausdorff distance.* In International Conference on Image Processing, pages 1429–1432, 2004. 38, 42, 59

[Lee *et al.* 2003] Y. Lee, K. Park, J. Shim and T. Yi. *3D face recognition using statistical multiple features for the local depth information.* In International Conference on Vision Interface, 2003. 38, 50

[Lee *et al.* 2005] Y. Lee, H. Song, U. Yang, H. Shin and K. Sohn. *Local feature based 3D face recognition.* In International Conference Audio- and Video-Based Biometric Person Authentication, pages 909–918, 2005. 38, 53

[Lei & Li 2009] Z. Lei and S. Z. Li. *Coupled spectral regression for matching heterogeneous faces.* In IEEE Conference on Computer Vision and Pattern Recognition, pages 1123–1128, June 2009. 107

[Levine & Rajwade 2006] M. D. Levine and A. Rajwade. *Three-dimensional view-invariant face recognition using a hierarchical pose-normalization strategy.* Machine Vision and Applications, vol. 17, no. 5, pages 309–325, 2006. 38

[Li & Jain 2005] S. Z. Li and A. K. Jain. Handbook of face recognition. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005. vii, viii, 2, 20, 34, 35

[Li & Zhang 2007] X. Li and H. Zhang. *Adapting geometric attributes for expression-invariant 3D face recognition.* In IEEE International Conference on Shape Modeling and Applications, pages 21–32, Jun. 2007. 39, 53

[Li *et al.* 2005a] C. Li, A. Barreto, J. Zhai and C. Chin. *Exploring face recognition by combining 3D profiles and contours.* In IEEE Southeast Conference, pages 576–579, Apr. 2005. 38, 47

[Li *et al.* 2005b] S. Li, C. Zhao, M. Ao and Z. Lei. *Learning to fuse 3D + 2D based face recognition at both feature and decision levels.* In International Workshop on Analysis and Modelling of Faces and Gestures, pages 44–54, 2005. 37, 60, 61, 83

[Li *et al.* 2007] S. Z. Li, R. Chu, S. Liao and L. Zhang. *Illumination invariant face recognition using near-infrared images.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, pages 627–639, Apr. 2007. 4

[Li *et al.* 2009] X. Li, T. Jia and H. Zhang. *Expression-insensitive 3D face recognition using sparse representation.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 2575–2582, Jun. 2009. 39, 41, 53, 76

[Liao *et al.* 2007] S. Liao, X. Zhu, Z. Lei, L. Zhang and S. Li. *Learning multi-scale block local binary patterns for face recognition.* In International Conference on Biometrics, pages 828–837, 2007. 28

[Liao *et al.* 2009] S. Liao, D. Yi, Z. Lei, R. Qin and S. Z. Li. *Heterogeneous face recognition from local structures of normalized appearance.* In International Conference on Biometrics, pages 209–218, 2009. 107

[Lin *et al.* 2006] W. Lin, K. Wong, N. Boston and Y. Hu. *3D human face recognition using summation invariants.* In IEEE International Conference on Acoustics, Speech and Signal Processing, volume 2, pages 341–344, May 2006. 38, 48

[Lin *et al.* 2007] W. Lin, K. Wong, N. Boston and Y. Hu. *3D face recognition under expression variations using similarity metrics fusion.* In IEEE International Conference on Multimedia and Expo, pages 727–730, Jul. 2007. 39

[Liu & Wechsler 2000] C. Liu and H. Wechsler. *Evolutionary pursuit and its application to face recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 6, pages 570–582, 2000. 24

[Liu & Wechsler 2002] C. Liu and H. Wechsler. *Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition.* IEEE Transactions on Image Processing, vol. 11, no. 4, pages 467–476, Apr. 2002. viii, 23, 25

[Liu *et al.* 2002] Q. Liu, R. Huang, H. Lu and S. Ma. *Face recognition using kernel-based fisher discriminant analysis.* In IEEE International Conference on Automatic Face and Gesture Recognition, pages 197–201, May 2002. 21

[Llonch *et al.* 2008] R. S. Llonch, E. Kokiopoulou, I. Tosic and P. Frossard. *3D face recognition using sparse spherical representations.* In International Conference on Pattern Recognition, pages 1–4, 2008. 39, 55

[Llonch *et al.* 2010] R. S. Llonch, E. Kokiopoulou, I. Tosic and P. Frossard. *3D face recognition with sparse spherical representations.* Pattern Recognition, vol. 43, no. 3, pages 824–834, 2010. 39

[Lowe 2004] D. G. Lowe. *Distinctive image features from scale-invariant keypoints.* International Journal of Computer Vision, vol. 60, no. 2, pages 91–110, 2004. viii, 28, 29, 30, 31, 66, 67, 90

[Lu & Jain 2003] X. Lu and A. K. Jain. *Resampling for face recognition.* In International Conference on Audio- and Video-Based Biometric Person Authentication, pages 869–877, 2003. 21

[Lu & Jain 2005a] X. Lu and A. K. Jain. *Deformation Analysis for 3D Face Matching.* In IEEE Workshop on Applications of Computer Vision, pages 99–104, 2005. 7, 37, 38, 43

[Lu & Jain 2005b] X. Lu and A. K. Jain. *Integrating range and texture information for 3D face recognition.* In IEEE International Workshop on Applications of Computer Vision, pages 156–163, 2005. 84

[Lu & Jain 2006] X. Lu and A. K. Jain. *Deformation modeling for robust 3D face matching.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 2, pages 1377–1383, 2006. 37, 38, 44, 59

[Lu & Jain 2008] X. Lu and A. K. Jain. *Deformation modeling for robust 3D face matching.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 8, pages 1346–1357, Aug. 2008. viii, 37, 39, 43, 44

[Lu *et al.* 2003] J. Lu, K. N. Plataniotis and A. N. Venetsanopoulos. *Face recognition using LDA-based algorithms.* IEEE Transactions on Neural Networks, vol. 14, pages 195–200, 2003. 20

[Lu *et al.* 2004a] X. Lu, D. Colbry and A. K. Jain. *Matching 2.5D scans for face recognition.* In International Conference on Biometric Authentication, pages 30–36, 2004. 37, 38, 84

[Lu *et al.* 2004b] X. Lu, D. Colbry and A. K. Jain. *Three-dimensional model based face recognition.* In International Conference on Pattern Recognition, pages 362–366, 2004. 37, 38, 41, 43

[Lu *et al.* 2004c] X. Lu, R.-L. Hsu, A. K. Jain, B. Kamgar-Parsi and B. Kamgar-Parsi. *Face Recognition with 3D model-based synthesis.* In International Conference on Biometric Authentication, pages 139–146, 2004. 21

[Lu *et al.* 2006] X. Lu, A. K. Jain and D. Colbry. *Matching 2.5D face scans to 3D models.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 1, pages 31–43, 2006. viii, 37, 41, 59, 60

[Lucey & Chen 2004] S. Lucey and T. Chen. *A GMM parts based face representation for improved verification through relevance adaptation.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 2, pages 855–861, 2004. 32

[Luo *et al.* 2007] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade and B. Lu. *Person-specific SIFT features for face recognition.* In IEEE International Conference on Acoustics, Speech, and Signal Processing, 2007. 31, 36, 67

[Maes *et al.* 2010] C. Maes, T. Fabry, J. Keustermans, D. Smeets, P. Suetens and D. Vandermeulen. *Feature detection on 3D face surfaces for pose normalisation and recognition.* In IEEE International Conference on Biometrics: Theory Applications and Systems, pages 1–6, Sep. 2010. 39, 46

[Mahoor & Abdel-Mottaleb 2007] M. H. Mahoor and M. Abdel-Mottaleb. *3D face recognition based on 3D ridge lines in range data.* In IEEE International Conference on Image Processing, volume 1, pages 137–140, Oct. 2007. 39, 48

[Mahoor & Abdel-Mottaleb 2009] M. H. Mahoor and M. Abdel-Mottaleb. *Face recognition based on 3D ridge images obtained from range data.* Pattern Recognition, vol. 42, pages 445–451, Mar. 2009. 39, 76

[Malassiotis & Strintzis 2004] S. Malassiotis and M. G. Strintzis. *Pose and illumination compensation for 3D face recognition.* In IEEE International Conference on Image Processing, pages 91–94, 2004. 83

[Marcelja 1980] S. Marcelja. *Mathematical description of the responses of simple cortical cells.* Journal of The Optical Society of America, vol. 70, no. 11, pages 1297–1300, Nov. 1980. 22

[Matthews & Baker 2004] I. Matthews and S. Baker. *Active appearance models revisited.* International Journal of Computer Vision, vol. 60, no. 2, pages 135–164, 2004. 34, 111

[Maurer *et al.* 2005] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West and G. Medioni. *Performance of geometrix activeID^{TM} 3D face recognition engine on the FRGC Data.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, page 154, Jun. 2005. 73, 74, 84, 93, 98

[Mayo & Zhang 2009] M. Mayo and F. Zhang. *3D face recognition using multiview keypoint matching.* In IEEE International Conference on Advanced Video and Signal Based Surveillance, pages 290–295, Sep. 2009. 39, 46

[McCool *et al.* 2008] C. McCool, V. Chandran, S. Sridharan and C. Fookes. *3D face verification using a free-parts approach.* Pattern Recognition Letter, vol. 29, pages 1190–1196, Jul. 2008. 39, 51

[McCool *et al.* 2010] C. McCool, J. Sanchez-Riera and S. Marcel. *Feature distribution modelling techniques for 3D face verification.* Pattern Recognition Letters, vol. 31, no. 11, pages 1324–1330, 2010. 39, 51

[Medioni & Waupotitsch 2003] G. Medioni and R. Waupotitsch. *Face Modeling and Recognition in 3-D.* In IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pages 232–233, 2003. 6, 7, 38, 41, 84

[Messer *et al.* 2003] K. Messer, J. Kittler, M. Sadeghi, S. Marcel, C. Marcel, S. Bengio, F. Cardinaux, C. Sanderson, J. Czyz, L. Vandendorpe, S. Srisuk, M. Petrou, W. Kurutach, A. Kadyrov, R. Paredes, E. Kadyrov, B. Kepenekci, F. B. Tek, G. B. Akar, N. Mavity and F. Deravi. *Face Verification Competition on the XM2VTS Database.* In International Conference on Audio and Video Based Biometric Person Authentication, pages 964–974, 2003. 10

[Mian *et al.* 2005] A. S. Mian, M. Bennamoun and R. A. Owens. *Matching tensors for pose invariant automatic 3d face recognition.* In IEEE Computer Society

Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, pages 120–120, 2005. 37, 38, 46

[Mian *et al.* 2006a] A. S. Mian, M. Bennamoun and R. A. Owens. *2D and 3D multimodal hybrid face recognition.* In European Conference on Computer Vision, volume 3, pages 344–355, 2006. 37, 85

[Mian *et al.* 2006b] A. S. Mian, M. Bennamoun and R. A. Owens. *Face recognition using 2D and 3D multimodal local features.* In International Symposium on Visual Computing, volume 1, pages 860–870, 2006. 95

[Mian *et al.* 2007] A. S. Mian, M. Bennamoun and R. Owens. *An efficient multimodal 2D-3D hybrid approach to automatic face recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 11, pages 1927–1943, Nov. 2007. ix, 37, 69, 73, 74, 85, 86, 92, 96, 97, 98

[Mian *et al.* 2008] A. S. Mian, M. Bennamoun and R. A. Owens. *Keypoint detection and local feature matching for textured 3D face recognition.* International Journal of Computer Vision, vol. 79, no. 1, pages 1–12, 2008. 37, 69, 73, 74, 75, 86, 97, 98, 99, 100

[Moghaddam 1999] B. Moghaddam. *Principal manifolds and Bayesian subspaces for visual recognition.* In IEEE International Conference on Computer Vision, volume 2, pages 1131–1136, 1999. 15

[Moghaddam 2002] B. Moghaddam. *Principal manifolds and probabilistic subspaces for Visual Recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 6, pages 780–788, 2002. 20

[Moreno & Sanchez 2004] A. B. Moreno and A. Sanchez. *GavabDB: A 3D Face Database.* In COST Workshop on Biometrics on the Internet: Fundamentals, Advances and Applications, pages 77–82, 2004. 10, 61, 69

[Moreno *et al.* 2003] 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, 2003. 38, 50, 59

[Moreno *et al.* 2005] A. B. Moreno, A. Sanchez, J. Velez and J. Diaz. *Face recognition using 3D local geometrical features: PCA vs. SVM.* In IEEE International Symposium on Image and Signal Processing and Analysis, pages 185–190, Sep. 2005. 38, 53, 76

[Mousavi *et al.* 2008] M. Mousavi, K. Faez and A. Asghari. *Three dimensional face recognition using SVM classifier.* In ACIS International Conference on Computer and Information Science, pages 208–213, 2008. 76

[Murphy-Chutorian & Trivedi 2009] E. Murphy-Chutorian and M. M. Trivedi. *Head Pose Estimation in Computer Vision: A Survey.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 4, pages 607–626, Apr. 2009. vii, 3

[Nagamine *et al.* 1992] T. Nagamine, T. Uemura and I. Masuda. *3D facial image analysis for human identification.* In International Conference on Pattern Recognition, pages 324–327, 1992. 37, 47, 59

[Nefian & Hayes III 1999] A. V. Nefian and M. H. Hayes III. *Face Recognition Using An Embedded HMM.* In International Conference on Audio- and Video-based Biometric Person Authentication, 1999. 32

[Ojala *et al.* 1996] T. Ojala, M. Pietikainen and D. Harwood. *A comparative study of texture measures with classification based on feature distributions.* Pattern Recognition, vol. 29, no. 1, pages 51–59, Jan. 1996. 25

[Ojala *et al.* 2002] T. Ojala, M. Pietikainen and T. Maenpaa. *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pages 971–987, 2002. 26, 27, 65

[Okada *et al.* 1998] K. Okada, J. Steffens, T. Maurer, H. Hong, E. Elagin, H. Neven and C. v. d. Malsburg. *The Bochum/USC face recognition system and how it fared in the FERET phase III test.* In Face Recognition: From Theory to Applications, 1998. 84

[Ouji *et al.* 2009] K. Ouji, B. Ben Amor, M. Ardabilian, L. Chen and F. Ghorbel. *3D face recognition using R-ICP and geodesic coupled approach.* In International Conference on Multimedia Modeling, pages 390–400, 2009. 39, 60

[Pan & Wu 2005] G. Pan and Z. Wu. *3d face recognition from range data.* International Journal of Image Graphics, vol. 5, no. 3, pages 573–594, 2005. 52

[Pan *et al.* 2003a] G. Pan, Y. Wu and Z. Wu. *Investigating Profile extraction from range data for 3D recognition.* In IEEE International Conference on Systems, Man and Cybernetics, volume 2, pages 1396–1399, 2003. 37, 38

[Pan *et al.* 2003b] G. Pan, Y. Wu, Z. Wu and W. Liu. *3D Face Recognition by Profile and Surface Matching.* In IEEE/INNS International Joint Conference on Neural Networks, 2003. 37, 38, 47, 52

[Pan *et al.* 2003c] G. Pan, Z. Wu and Y. Pan. *Automatic 3D face verification from range data.* In IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 3, pages 193–196, 2003. 37, 38, 42

[Pan *et al.* 2005] G. Pan, S. Han, Z. Wu and Y. Wang. *3D face recognition using mapped depth images.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, 2005. 37, 38, 54

[Papatheodorou & Rueckert 2004] T. Papatheodorou and D. Rueckert. *Evaluation of automatic 4D face recognition using surface and texture registration.* In IEEE International Conference on Automatic Face and Gesture Recognition, pages 321–326, May 2004. 83

[Passalis *et al.* 2005] G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici and N. Murtuza. *Evaluation of 3D face recognition in the presence of facial expressions: an annotated deformable model approach.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, pages 171–171, 2005. 37, 38, 54, 73, 75

[Passalis *et al.* 2007] G. Passalis, I. A. Kakadiaris and T. Theoharis. *Intraclass retrieval of nonrigid 3D objects: application to face recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 2, pages 218–229, Feb. 2007. 37, 39

[Pentland *et al.* 1994] A. Pentland, B. Moghaddam and T. Starner. *View-based and modular eigenspaces for face recognition.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994. 20

[Perrot 1996] R. Perrot. *Use of anthropological Methods in the identification of unknown individuals : human remains and armed robbers.* In Meeting of the International Association of Forensic Sciences, Aug. 1996. 93

[Phillips *et al.* 2000] P. J. Phillips, H. Moon, S. A. Rizvi and P. J. Rauss. *The FERET evaluation methodology for face-recognition algorithms.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pages 1090–1104, Oct. 2000. 1, 24

[Phillips *et al.* 2005] P. J. Phillips, P. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek. *Overview of the face recognition grand challenge.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pages 947–954, Jun. 2005. 1, 9, 11, 61, 69, 92, 108, 117

[Phillips *et al.* 2011] P. J. Phillips, P. J. Flynn, K. W. Bowyer, R. W. V. Bruegge, P. J. Grother, G. W. Quinn and M. Pruitt. *Distinguishing identical twins by face recognition.* In IEEE International Conference on Automatic Face and Gesture Recognition, pages 185–192, 2011. 12

[Pietikainen *et al.* 2000] M. Pietikainen, T. Ojala and Z. Xu. *Rotation-invariant texture classification using feature distributions.* Pattern Recognition, vol. 33, no. 1, pages 43–52, 2000. 27

[Queirolo *et al.* 2008] C. Queirolo, L. Silva, O. Bellon and M. Segundo. *3D face recognition using the Surface Interpenetration Measure: a comparative evaluation on the FRGC database.* In International Conference on Pattern Recognition, pages 1–5, 2008. 39

[Queirolo *et al.* 2010] C. Queirolo, L. Silva, O. Bellon and M. Segundo. *3D face Recognition using simulated annealing and the surface interpenetration Measure.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 2, pages 206–219, 2010. 39, 50, 56

[Rama *et al.* 2006] A. Rama, F. Tarres, D. Onofrio and S. Tubaro. *Mixed 2D-3D information for pose estimation and face recognition.* In IEEE International Conference on Acoustics, Speech and Signal Processing, volume 2, page II, May 2006. 107

[Ramnath *et al.* 2008] K. Ramnath, S. Koterba, J. Xiao, C. Hu, I. Matthews, S. Baker, J. F. Cohn and T. Kanade. *Multi-view AAM fitting and construction.* International Journal of Computer Vision, vol. 76, no. 2, pages 183–204, 2008. 128

[Raudys & Jain 1991] S. Raudys and A. K. Jain. *Small sample size effects in statistical pattern recognition: recommendations for practitioners.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, pages 252–264, 1991. 21

[Reiter *et al.* 2006] M. Reiter, R. Dormer, G. Langs and H. Bischof. *3D and infrared face reconstruction from RGB data using canonical correlation analysis.* In International Conference on Pattern Recognition, volume 1, pages 425–428, 2006. 107

[Riccio & Dugelay 2005] D. Riccio and J.-L. Dugelay. *Asymmetric 3D/2D processing: a novel approach for face recognition.* In International Conference on Image Analysis and Processing, pages 986–993, 2005. 107

[Roweis & Saul 2000] S. T. Roweis and L. K. Saul. *Nonlinear dimensionality reduction by locally linear embedding.* Science, vol. 290, pages 2323–2326, 2000. 21

[Russ *et al.* 2004] T. D. Russ, K. W. Koch and C. Q. Little. *3D facial recognition: a quantitative analysis.* In Annual Meeting of the Institute of Nuclear Materials Management, 2004. 38, 42, 43

[Russ *et al.* 2005] T. D. Russ, M. W. Koch and C. Q. Little. *A 2D range Hausdorff approach for 3D face recognition.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, pages 169–169, Jun. 2005. 38, 59

[Ruta & Gabrys 2005] D. Ruta and B. Gabrys. *Classifier selection for majority voting.* Information Fusion, vol. 6, no. 1, pages 63–81, 2005. 91

[Said 2005] Y. H. Said. *On genetic algorithms and their applications.* In C. R. Rao, E. J. Wegman and J. L. Solka, editeurs, Data Mining and Data Visualization, volume 24 of *Handbook of Statistics*, pages 359–390. Elsevier, 2005. 91

[Samaria & Harter 1994] F. Samaria and A. Harter. *Parameterisation of a stochastic model for human face identification.* In Workshop on Applications of Computer Vision, 1994. 17, 32

[Samir *et al.* 2005] C. Samir, J.-P. Vandeborre and M. Daoudi. *Automatic 3D face recognition using topological techniques.* In IEEE International Conference on Multimedia and Expo, pages 450–453, Jul. 2005. 38

[Samir *et al.* 2006] C. Samir, A. Srivastava and M. Daoudi. *Three-dimensional face recognition using shapes of facial curves.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 11, pages 1858–1863, Nov. 2006. viii, 38, 47, 48, 59

[Savran *et al.* 2008] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur and L. Akarun. *Bosphorus database for 3D face analysis.* In COST Workshop on Biometrics and Identity Management, 2008. 10, 128

[Scheenstra *et al.* 2005] A. Scheenstra, A. Ruifrok and R. C. Veltkamp. *A survey of 3D face recognition methods.* In International Conference on Audio- and Video-Based Biometric Person Authentication, pages 891–899, 2005. 59

[Shan & Gritti 2008] C. Shan and T. Gritti. *Learning discriminative LBP-histogram bins for facial expression recognition.* In British Machine Vision Conference, 2008. 27, 61, 65

[Shan *et al.* 2009] C. Shan, S. Gong and P. W. Mcowan. *Facial expression recognition based on local binary patterns: a comprehensive study.* Image and Vision Computing, vol. 27, no. 6, pages 803–816, 2009. viii, 28

[Smeets *et al.* 2009] D. Smeets, T. Fabry, J. Hermans, D. Vandermeulen and P. Suetens. *Isometric deformation modeling using singular value decomposition for 3D expression-invariant face recognition.* In IEEE International Conference on Biometrics: Theory, Applications, and Systems, pages 1–6, Sep. 2009. 39

[Snelick *et al.* 2005] R. Snelick, U. Uludag, A. Mink, M. Indovina and A. K. Jain. *Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pages 450–455, Mar. 2005. 95

[Socolinsky *et al.* 2003] D. A. Socolinsky, A. Selinger and J. D. Neuheisel. *Face recognition with visible and thermal infrared imagery.* Computer Vision and Image Understanding, vol. 91, no. 1-2, pages 72–114, 2003. 4

[Sun & Yin 2005] Y. Sun and L Yin. *3D face recognition using two views face modeling and labeling.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Advanced 3D Imaging for Safety and Security, page 117, Jun. 2005. 38

[Szeptycki *et al.* 2009] P. Szeptycki, M. Ardabilian and L. Chen. *A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking.* In IEEE International Conference on Biometrics: Theory, Applications, and Systems, pages 1–6, Sep. 2009. 56, 70, 128

[Tan & Triggs 2007] X. Tan and B. Triggs. *Enhanced local texture feature sets for face recognition under difficult lighting conditions.* In International Workshop on Analysis and Modelling of Faces and Gestures, pages 168–182, Oct. 2007. 28

[Tan *et al.* 2006] X. Tan, S. Chen, Z. Zhou and F. Zhang. *Face recognition from a single image per person: a survey.* Pattern Recognition, vol. 39, no. 9, pages 1725–1745, 2006. 35

[Tanaka & Ikeda 1996] H. T. Tanaka and M. Ikeda. *Curvature-based face surface recognition using spherical correlation - principal directions for curved object recognition.* In International Conference on Pattern Recognition, volume 3, pages 638–642, Aug. 1996. 38

[Tanaka *et al.* 1998] H. T. Tanaka, M. Ikeda and H. Chiaki. *Curvature-based face surface recognition using spherical correlation. Principal directions for curved object recognition.* In IEEE International Conference on Automatic Face and Gesture Recognition, pages 372–377, Apr. 1998. 38, 50, 59, 93

[Tenenbaum *et al.* 2000] J. B. Tenenbaum, V. Silva and J. C. Langford. *A global geometric framework for nonlinear dimensionality reduction.* Science, vol. 290, no. 5500, pages 2319–2323, 2000. 21

[Troje & Bulthoff 1996] N. F. Troje and H. H. Bulthoff. *Face recognition under varying poses: the role of texture and shape.* Vision Research, vol. 36, no. 12, pages 1761–1771, 1996. 10

[Tsalakanidou *et al.* 2003] F. Tsalakanidou, D. Tzovaras and M. G. Strintzis. *Use of depth and colour eigenfaces for face recognition.* Pattern Recognition Letters, vol. 24, no. 9-10, pages 1427–1435, 2003. 82

[Tsalakanidou *et al.* 2004] F. Tsalakanidou, S. Malassiotis and M. G. Strintzis. *Integration of 2D and 3D images for enhanced face authentication.* In IEEE international conference on Automatic face and gesture recognition, pages 266–271, 2004. 83

[Tsalakanidou *et al.* 2005] F. Tsalakanidou, S. Malassiotis and M. G. Strintzis. *Face localization and authentication using color and depth images.* IEEE Transactions on Image Processing, vol. 14, no. 2, pages 152–168, Feb 2005. 83

[Turk & Pentland 1991] M. Turk and A. Pentland. *Eigenfaces for face recognition.* Journal of Cognitive Neuroscience, vol. 3, no. 1, pages 71–86, 1991. 1, 15, 16, 81

[Uchida *et al.* 2005] N. Uchida, T. Shibahara, T. Aoki, H. Nakajima and K. Kobayashi. *3D face recognition using passive stereo vision.* In IEEE International Conference on Image Processing, volume 2, pages 950–953, Sep. 2005. 38

[Vapnik 1995] V. Vapnik. The nature of statistical learning theory. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1995. 21

[Vasilescu & Terzopoulos 2002] M. A. O. Vasilescu and D. Terzopoulos. *Multilinear analysis of image ensembles: TensorFaces.* In European Conference on Computer Vision, pages 447–460, 2002. 20

[Vetter & Poggio 1997] T. Vetter and T. Poggio. *Linear object classes and image synthesis from a Single example image.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pages 733–742, 1997. 21

[Viola & Jones 2004] P. Viola and M. J. Jones. *Robust real-time face detection.* International Journal on Computer Vision, vol. 57, pages 137–154, May 2004. 22

[Wang *et al.* 2002] Y. Wang, C. Chua and Y. Ho. *Facial feature detection and face recognition from 2D and 3D images.* Pattern Recognition Letters, vol. 23, no. 10, pages 1191–1202, 2002. 82

[Wang *et al.* 2006] Y. Wang, G. Pan, Z. Wu and Y. Wang. *Exploring facial expression effects in 3D face recognition using partial ICP.* In Asian Conference on Computer Vision, volume 1, pages 581–590, 2006. 39, 41

[Wang *et al.* 2007a] S. Wang, Y. Wang, M. Jin, X. D. Gu and D. Samaras. *Conformal geometry and its applications on 3D shape matching, recognition, and stitching.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 7, pages 1209–1220, Jul. 2007. 39, 55

[Wang *et al.* 2007b] Y. Wang, G. Pan and Z. Wu. *3D face recognition in the presence of expression: a guidance-based constraint deformation approach.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 1–7, Jun. 2007. 37, 39, 44, 73

[Wang *et al.* 2008] Y. Wang, X. Tang, J. Liu, G. Pan and R. Xiao. *3D face recognition by local shape difference boosting.* In European Conference on Computer Vision, pages 603–616, 2008. 39, 52

[Wang *et al.* 2009] R. Wang, J. Yang, D. Yi and S. Z. Li. *An analysis-by-synthesis method for heterogeneous face biometrics.* In International Conference on Biometrics, pages 319–326, 2009. 107

[Wang *et al.* 2010] Y. Wang, J. Liu and X. Tang. *Robust 3D face recognition by local shape difference Boosting.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 10, pages 1858–1870, 2010. 39, 52, 56, 60, 73, 74

[Wei *et al.* 2007] X. Wei, P. Longo and L. Yin. *Automatic facial pose determination of 3D range data for face model and expression identification.* In International Conference on Biometrics, pages 144–153, 2007. 39, 44

[Wiskott *et al.* 1997] L. Wiskott, J.-M. Fellous, N. Kuiger and C. von der Malsburg. *Face recognition by elastic bunch graph matching.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pages 775–779, Jul. 1997. 2, 15, 24

[Wolf *et al.* 2008] L. Wolf, T. Hassner and Y. Taigman. *Descriptor based methods in the wild.* In Real-Life Images Workshop at the European Conference on Computer Vision, Oct. 2008. 28

[Wright *et al.* 2009] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma. *Robust face recognition via sparse representation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pages 210–227, 2009. 109, 116

[Wu *et al.* 2003] Y. Wu, G. Pan and Z. Wu. *Face authentication based on multiple profiles extracted from range data.* In International Conference on Audio- and video-based Biometric Person Authentication, pages 515–522, 2003. 38, 47

[Wu *et al.* 2004] Z. Wu, Y. Wang and G. Pan. *3D face recognition using local shape map.* In IEEE International Conference on Image Processing, volume 3, pages 2003–2006, Oct. 2004. 38, 45

[Xu *et al.* 2004] C. Xu, Y. Wang, T. Tan and L. Quan. *Automatic 3D face recognition combining global geometric features with local shape variation information.* In IEEE International Conference on Automatic Face and Gesture Recognition, pages 308–313, 2004. 37, 38, 52

[Xu *et al.* 2009] C. Xu, S. Li, T. Tan and L. Quan. *Automatic 3D face recognition from depth and intensity Gabor features.* Pattern Recognition, vol. 42, no. 9, pages 1895–1905, 2009. 37, 87, 92, 98

[Yamaguchi *et al.* 1998] O. Yamaguchi, K. Fukui and K. Maeda. *Face Recognition Using Temporal Image Sequence.* In IEEE International Conference on Automatic Face and Gesture Recognition, pages 318–323, 1998. 4

[Yan *et al.* 2007] S. Yan, H. Wang, X. Tang and T. S. Huang. *Exploring feature Descriptors for face recognition.* In IEEE International Conference on Acoustics, Speech and Signal Processing, volume 1, pages 629–632, Apr. 2007. 61, 65

[Yang *et al.* 2004a] J. Yang, D. Zhang, A. F. Frangi and J. Yang. *Two-dimensional PCA: a new approach to appearance-based face representation and recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, pages 131–137, 2004. 20

[Yang *et al.* 2004b] P. Yang, S. Shan, W. Gao, S. Li and D. Zhang. *Face recognition using Ada-Boosted Gabor features.* In IEEE International Conference on Automatic Face and Gesture Recognition, pages 356–361, 2004. 24

[Yang *et al.* 2008] W. Yang, D. Yi, Z. Lei, J. Sang and S. Z. Li. *2D-3D face matching using CCA.* In IEEE International Conference on Automatic Face and Gesture Recognition, pages 1–6, 2008. 107

[Yang 2002a] M.-H. Yang. *Face recognition using extended isomap.* In International Conference on Image Processing, volume 2, pages 117–120, 2002. 21

[Yang 2002b] M.-H. Yang. *Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods.* In IEEE International Conference on Automatic Face and Gesture Recognition, pages 215–220, may 2002. 21

[Yin *et al.* 2006a] L. Yin, X. Wei, P. Longo and A. Bhuvanesh. *Analyzing facial expressions using intensity-variant 3D data for human computer interaction.* In International Conference on Pattern Recognition, volume 1, pages 1248–1251, 2006. 44

[Yin *et al.* 2006b] L. Yin, X. Wei, Y. Sun, J. Wang and M. J. Rosato. *A 3D facial expression database for facial behavior research.* In International Conference on Automatic Face and Gesture Recognition, pages 211–216, 2006. 9, 128

[Yu & Yang 2001] H. Yu and J. Yang. *A direct lda algorithm for high-dimensional data with application to face recognition.* Pattern Recognition, vol. 34, pages 2067–2070, 2001. 20

[Yuan *et al.* 2005] X. Yuan, J. Lu and T. Yahagi. *A method of 3D face recognition based on principal component analysis algorithm.* In IEEE International Symposium on Circuits and Systems, pages 3211–3214, May 2005. 85

[Zeng *et al.* 2010] W. Zeng, D. Samaras and X. Gu. *Ricci flow for 3D shape analysis.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 4, pages 662–677, 2010. viii, 55

[Zhang *et al.* 2005] W. Zhang, S. Shan, W. Gao, X. Chen and H. Zhang. *Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition.* In IEEE International Conference on Computer Vision, pages 786–791, 2005. 28

[Zhang *et al.* 2006] L. Zhang, A. Razdan, G. E. Farin, J. Femiani, M. Bae and C. Lockwood. *3D face authentication and recognition based on bilateral symmetry analysis.* The Visual Computer, vol. 22, no. 1, pages 43–55, 2006. 39, 48

[Zhao & Chellappa 2000] W. Zhao and R. Chellappa. *SFS based view synthesis for robust face recognition.* In IEEE International Conference on Automatic Face and Gesture Recognition, pages 285–293, 2000. 21

[Zhao *et al.* 2003] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld. *Face recognition: A literature survey.* ACM Computing Survey, vol. 35, pages 399–458, Dec. 2003. 3, 15

[Zhao *et al.* 2009] X. Zhao, P. Szeptycki, E. Dellandrea and L. Chen. *Precise 2.5D facial landmarking via an analysis by synthesis approach.* In International Workshop on Applications of Computer Vision, pages 1–7, Dec. 2009. 56

[Zhong *et al.* 2007] C. Zhong, Z. Sun and T. Tan. *Robust 3D face recognition using learned visual codebook.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 1–6, Jun. 2007. 37, 39, 50

[Zhong *et al.* 2008] C. Zhong, Z. Sun, T. Tan and Z. He. *Robust 3D face recognition in uncontrolled environments.* In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 1–8, 2008. 37, 39

[Zou *et al.* 2007a] J. Zou, Q. Ji and G. Nagy. *A comparative study of local matching approach for face recognition.* IEEE Transactions on Image Processing, vol. 16, no. 10, pages 2617–2628, 2007. vii, 15, 22, 36

[Zou *et al.* 2007b] L. Zou, S. Cheng, Z. Xiong, M. Lu and K. R. Castleman. *3-D face recognition based on warped example faces.* IEEE Transactions on Information Forensics and Security, vol. 2, no. 3, pages 513–528, Sep. 2007. 39, 44

dernière page de la thèse

AUTORISATION DE SOUTENANCE

Vu les dispositions de l'arrêté du 7 août 2006,

Vu la demande du Directeur de Thèse

Monsieur L. CHEN

et les rapports de

Monsieur J-L. DUGELAY
Professeur - Eurecom - MultiMedia Communications dept., 2229, Route des Cretes - BP 193
06904 SOPHIA-ANTIPOLIS cedex

Et de

Monsieur A. HADID
Senior chercheur - Adjunct professor - University of Oulu - Machine Vision Group
DEPt. Of Electrical and Information Engineering - P.O. Box 4500 - FIN-90014 University of Oulu - FINLAND

Monsieur HUANG Di

est autorisé à soutenir une thèse pour l'obtention du grade de **DOCTEUR**

Ecole doctorale INFOMATHS

Fait à Ecully, le 7 septembre 2011

P/Le Directeur de l'E.C.L.
La Directrice des Etudes

ECOLE CENTRALE DE LYON
SCOLARITE
3ème Cycle

M-A. GALLAND