ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2012-ENST-0012

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Biométrie Faciale 3D »

présentée et soutenue publiquement par

Nesli ERDOGMUS

le jour mois année

Utilization of 3D Data in Face Recognition

Directeur de thèse : **Jean-Luc Dugelay**

Jury

M. Bulent SANKUR, Professeur, Université Bosphore	Rapporteur
M. Marc ANTONINI, Directeur de Recherche, Laboratoire I3S 2000, CNRS	Rapporteur
M. Joseph COLINEAU, Expert Senior du Groupe de Recherches, THALES	Examinateur
M. Igor S. PANDZIC, Assistant Professeur, Université Zagreb	Examinateur
M. Mohamed DAOUDI, Professeur, Telecom Lille 1	Examinateur

TELECOM ParisTech
école de l'Institut Télécom - membre de ParisTech

T H È S E

UTILIZATION OF 3D DATA IN FACE RECOGNITION

Nesli Erdoğmuş

A doctoral dissertation submitted to:
INSTITUT Mines-Télécom
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
Major subject: Image Processing, Biometrics

Approved by the following examining committee:

Supervisor:	Prof. Jean-Luc Dugelay
President of the jury:	Prof. Mohamed Daoudi
Examiner:	Prof. Bülent Sankur
Examiner:	Prof. Marc Antonini
Member:	Asst. Prof. Igor Pandzic
Member:	Joseph Colineau

20th March 2012
REF: 2012-ENST-012

Abstract

As the need and investment for security applications grow vastly, one particular topic has certainly been attracting higher attention than many other pattern recognition fields: biometrics – recognition of humans based on their physical and/or behavioral traits. Among those traits, face stands out with its favorable reconcilement between accessibility and reliability. In this doctoral dissertation, we principally investigate 3D shape which is a relatively more recent modality in face recognition. Being advantageous to its 2D counterpart for being intrinsically invariant to illumination and pose changes, 3D face recognition still encounters major challenges such as acquisition inadequacy or facial surface deformations due to expressions or occlusion. Bearing those in mind, a thorough study is carried out on this topic, including pre-processing of 3D facial data, automatic landmarking, facial expression and alteration simulations in 3D, 3D feature extraction and 3D regional surface analysis.

We firstly study pre-processing techniques for 3D face. The pre-processing includes extraction of the facial region, elimination of spikes and holes in the facial mesh and finally smoothing. Additionally, we present an automatic landmarking method that utilizes both modalities (i.e. shape and texture) according to inherent characteristics of different regions on face.

Then, we investigate the advantages of 3D face information in an asymmetrical scenario in which a combined enrollment is realized by both 2D and 3D information whereas the target images are still 2D images. Despite the decades of research, 2D face recognition systems still encounter difficulties with intra-subject variations, especially in case of few sample images which cannot capture all those possible deviations. For this problem, we propose a face recognition system that synthesizes various facial expressions for each subject to augment a given gallery set. By adopting a *Thin Plate Spline* (TPS) – based method, we warp a generic model to generate client-specific MPEG-4 compliant animatable face models for facial expression simulations and we demonstrate the beneficial effects of the proposed gallery augmentation in recognition performances.

Next, as a useful by-product of the previous framework, we propose to employ the TPS parameters that are obtained by warping the generic model as biometric signatures and analyze their discriminative properties. Those parameters which have mostly been overlooked as superfluous sediment in registration applications represent the deviations from the conventional and they are proven to possess concentrated information about the

facial shape. However, the results also revealed a frailty in the presence of facial expressions.

Subsequently, we present a regional confidence level estimation algorithm for facial surfaces based on local primitive shape distributions. Face is divided into 7 non-overlapping regions that are to be evaluated according to the presence of mesh distortions. For each segment, an *Artificial Neural Network* (ANN) is utilized to estimate a confidence level. The ANNs are trained using regional quality scores automatically measured by an *Iterative Closest Point* (ICP)-based method. Here, "quality" implies the absence of occlusions or expressions, rather than mesh resolution, noisiness, etc. In the end, we exploit the reliability score estimations to combine the regional match scores for verification and identification purposes via three different fusion methods at score level: Sum rule, Borda Count and a probabilistic approach which converts match scores into probabilities.

Finally, an emerging field of research in 3D face recognition is explored: susceptibility to evasion in a recognition setting by partially modifying the facial surface. Those modifications can be achieved via plastic surgery, prosthetic make-up, latex appliances, etc. and they can be in countless different ways and amounts. However studies on their possible impacts are still very limited. In this dissertation, we analyze how such changes on nose region can affect the face recognition performances based on experiments conducted by using several benchmark algorithms. For this purpose, a simulated face database is prepared in which nose of each probe sample is replaced with another randomly chosen one and the impact of the nose alterations on the recognition performances is assessed.

Résumé

Avec l'augmentation des applications de sécurité, un sujet particulier a attiré plus d'attention que de nombreux autres domaines en reconnaissance de formes : la biométrie – la reconnaissance des personnes en fonction de leur physique et/ou de leurs traits comportementaux. Parmi ces traits, le visage se distingue par un compromis favorable entre accessibilité et fiabilité. Dans cette thèse, nous avons principalement travaillé sur une modalité relativement plus récente en reconnaissance du visage : la forme 3D. Plus avantageuse que son homologue en 2D, car intrinsèquement invariant aux changements d'illumination de pose, l'analyse 3D du visage se heurte encore à des défis majeurs tels que les déformations de surface du visage dues aux expressions ou occultations. En conséquence, une étude approfondie est effectuée sur ce sujet incluant le prétraitement des données 3D du visage, l'annotation automatique, la simulation d'altération et d'expressions en 3D, l'extraction de caractéristiques 3D et l'analyse locale de surface.

Tout d'abord, nous étudions les représentations 3D du visage et des techniques de prétraitement associées. Le prétraitement comprend l'extraction de la région faciale, l'élimination des pics et des trous dans le maillage du visage et enfin le lissage. De plus, nous présentons une méthode d'annotation automatique qui utilise les deux modalités (forme et texture) en fonction des caractéristiques inhérentes des différentes régions sur le visage.

Ensuite, nous étudions les avantages du 3D dans un scénario asymétrique dans lequel un enrôlement combiné est réalisé en 2D et en 3D alors que les images cibles sont toujours des images 2D. Malgré des décennies de recherche, les systèmes de reconnaissance faciale 2D rencontrent encore des difficultés avec les variations intra-individus, en particulier lorsque les échantillons d'image sont peu nombreux et ne peuvent pas couvrir toutes les variations possibles. Pour ce problème, nous proposons un système de reconnaissance faciale qui synthétise diverses expressions faciales pour chaque individu afin d'augmenter un ensemble « galerie » donné. A l'aide d'une approche TPS nous adoptons un modèle générique pour produire des clients compatibles avec les modèles animables MPEG-4 pour les simulations des expressions faciales et nous démontrons les effets positifs liés à l'augmentation proposée de la galerie.

Ensuite, en tant qu'application dérivée du travail précédant, nous proposons d'utiliser les paramètres TPS qui sont obtenus par déformation du modèle générique comme des signatures biométriques et d'analyser leurs propriétés discriminantes. Ces paramètres, qui

ont surtout été négligés en tant que éléments superflus en recalage représentent les écarts par rapport à une valeur moyen et par conséquent, on peut démontrer qu'ils possèdent des informations significatives sur la forme du visage. Cependant, les résultats ont également révélé une fragilité en présence d'expressions faciales.

Par la suite, nous présentons un algorithme d'estimation de la confiance locale au niveau des surfaces faciales en se basant sur la distribution locale des formes. Le visage est divisé en sept régions non recouvrantes qui doivent être évaluées en fonction de la présence de distorsions du maillage. Pour chaque segment, un réseau de neurones artificiel (RNA) est utilisé pour estimer un niveau de confiance. Les RNAs sont formés en utilisant des scores régionaux de qualité mesurés automatiquement par ICP. Ici, la « qualité » implique plutôt l'absence d'occultations ou d'expressions que la résolution du maillage, le bruit, etc. A la fin, nous exploitons les scores de fiabilité pour combiner les scores d'appariements locaux à des fins de vérification et d'identification au travers de trois méthodes différentes de fusion au niveau des scores : Somme, Borda et une approche probabiliste qui convertit les scores en probabilités.

Enfin, un domaine de recherche émergent en reconnaissance faciale 3D est exploré : la sensibilité à l'évasion dans un cadre de reconnaissance en face des modifications de la surface du visage. Ces modifications peuvent être obtenues par chirurgie plastique, prothèses, maquillage, etc. Cependant des études sur leurs effets possibles sont encore très limitées. Dans cette thèse, nous analysons comment ces changements au niveau de la région du nez affectent les performances en reconnaissance de visage de plusieurs algorithmes clés. Pour cela, une base de données simulée de visages est préparée dans laquelle le nez de chaque personne à vérifier est remplacé par un autre choisi au hasard et l'impact de la modification du nez sur les performances de reconnaissance est évaluée.

Özetçe

Güvenlik uygulamalarına olan ihtiyaç ve yatırımlar büyük ölçüde artarken, özellikle bir konu diğer bir çok örüntü tanıma alanlarından çok daha fazla dikkat çekmektedir: Biyometri - insanları fiziksel ve/veya davranışsal özelliklerine dayanarak tanıma. Bu fiziksel özelliklerden biri olan yüz, erişilirlik ve güvenilirlik arasında sağladığı uzlaşma ile ön plana çıkmaktadır. Bu doktora tezinde temel olarak, yüz tanımada diğerlerine görece daha yeni bir modalite olan 3B şekil bilgisi incelenmiştir. Doğası gereği aydınlatma ve duruş değişikliklerinden etkilenmemesi dolayısıyla 2B karşılığından daha avantajlı olsa da, 3B yüz bilgisi halen veri elde etme aksaklıkları veya yüz ifadeleri ve perdelemeden kaynaklanan yüzey bozulmaları gibi önemli sorunlar ile karşı karşıyadır. Bu noktalar göz önünde bulundurularak, 3B yüz verisinin ön-işlemesi, nirengi noktalarının otomatik olarak bulunması, yüz ifadelerinin ve tahriflerinin 3B benzetimleri, 3B öznitelik çıkarma ve 3B bölgesel yüzey incelemesi konularının dahil olduğu kapsamlı bir çalışma sürdürülmüştür.

İlk olarak yüzün 3B gösterimleri ve buna bağlı ön-işleme teknikleri gözden geçirilmiştir. Ön-işleme genel olarak yüz bölgesinin çıkarılması, ani çıkıntıların ve boşlukların yok edilmesi ve yüzeyin yumuşatılması süreçlerinden oluşmaktadır. Bunlara ek olarak, farklı yüz bölgelerinin yapısal özelliklerine göre her iki modaliteyi (şekil ve doku) de kullanan bir otomatik nirengi noktası bulma yöntemi sunulmuştur.

Ardından, 3B yüz bilgisinin avantajları sistem kaydının hem 2B hem de 3B veriyle yapıldığı ancak tanıma için kullanılan test imgelerinin 2B imgeler olduğu asimetrik bir senaryo dahilinde araştırılmıştır. Onyıllardır süregelen araştırma çalışmalarına rağmen, 2B yüz tanıma sistemleri halen, özellikle mümkün olan tüm değişimleri temsil edemeyen az sayıda örneğin olduğu durumlarda, sınıf-içi değişimlerde zorluklarla karşılaşmaktadır. Bu probleme yönelik, galeri kümesini genişletme amacıyla her birey için çeşitli yüz ifadelerini sentetize eden bir yüz tanıma sistemi önerilmiştir. İnce metal plaka eğrileri (TPS) tabanlı bir yöntem benimsenerek, yüz ifadeleri benzetimleri yapmak için soysal bir model, istemciye özgü MPEG-4 uyumlu canlandırılabilir yüz modellerine bükülmüştür. Önerilen galeri büyütmenin yüz tanıma başarımına olan yararlı etkileri deneylerle gösterilmiştir.

Daha sonra, yukarıda bahsedilen uygulama çerçevesinin bir yan ürünü olarak, soysal modeli bükme esnasında elde edilen TPS parametrelerinin biyometrik imzalar olarak kullanılması önerilmiş ve ayırdedici özellikleri analiz edilmiştir. Çakıştırma uygulamalarında çoğu zaman gereksiz bir artık ürün olarak görülen bu parametreler aslında alışılagelmiş olandan (soysal model) sapmaları temsil etmektedirler. Bu noktadan hareketle yüz şekli

hakkında yoğun bilgi içerdikleri ispatlanmıştır, ancak sonuçlar aynı zamanda yüz ifadelerine olan dayanıksızlıklarını da ortaya koymuşur.

Bu sonuçları müteakip, yüz modelleri için yerel asli şekil dağılımlarına dayanan, bölgesel güvenilirlik seviyesi kestirimi algoritması geliştirilmiştir. Yüz çakışmayacak şekilde 7 bölgeye ayrılmış ve her bölge yüzey bozulumlarının miktarına göre değerlendirilmiştir. Her bir parça için, yapay bir sinir ağı (ANN) kullanılarak güvenilirlik kestirilmiştir. ANNler yinelemeli en yakın nokta (ICP) tabanlı bir yöntem kullanılarak otomatik ölçülen bölgesel kalite skorları esas alınarak eğitilmişlerdir. Burada "kalite" çözünürlük, gürültülülük, vs. gibi özellikleri değil, yüz ifadesi ve perdelemeden kaynaklı yüzey bozulmalarının miktarını esas almaktadır. Sonrasında, güvenilirlik dereceleri kestirimleri bölgesel eşleştirme skorlarının doğrulama ve kimlik belirleme amaçlı birleştirilmesinde kullanılmıştır. Bu amaç için üç farklı tümleştirme yöntemi denenmiştir: Toplam kuralı, Borda sayımı ve bölgesel eşleştirme skorlarını olasılıklara dönüştüren olasılıksal bir yöntem.

Son olarak, 3B yüz tanımada yeni yeni ağırlık kazanan bir araştırma alanı ele alınmıştır: yüzde kısmi tahrifat yaparak bir tanıma sisteminden kaçınmaya olan duygunluk. Plastik cerrahi, prostetik makyaj, takma burun gibi bir çok araç kullanılarak gerçekleştirilebilen bu tahrifatlar sayısız şekilde ve büyüklükte olabilir. Ne var ki, bu değişikliklerin olası etkileri üzerine yapılan araştırmalar oldukça sınırlıdır. Bu doktora çalışmasında, burun bölgesine uygulanacak benzer değişikliklerin çeşitli önemli yüz tanıma tekniği başarımlarına olan etkileri incelenmiştir. Bu amaç için, her örnek yüz modelindeki burun bölgesinin rasgele seçilmiş bir başka burun bölgesi ile değiştirildiği sentetik bir veri tabanı hazırlanmış, burun değişikliklerinin tanıma başarımlarındaki etkileri tayin edilmiştir.

Acknowledgments

First, I would like to express my gratitude to my supervisor Prof. Jean-Luc Dugelay for his support and all of my fellow colleagues in the Image Group; particularly to Antitza Dantcheva, Carmelo Velardo, Usman Saeed and Neslihan Kose for their friendship and collaboration. I would like to thank the members of our small Turkish community within Eurecom for their enormous encouragement and invaluable company: Melek Onen, Erhan Yilmaz, Leyla Bilge, Turgut Mustafa Oktem, Baris Demiray and Ayse Unsal.

I am also extremely grateful to the members of the examining committee of this dissertation for their precious time dedicated to evaluating this work and for their constructive comments.

Many thanks to my family in Turkey for their sacrifice and support through all those years of studying. Kilometers apart, I could always feel their warmth and prayers in my heart.

Above all, I would like to pay my deepest gratitude to my dear husband Ufuk for always encouraging me and brightening my every day. For reading my articles and pretending that they are revolutionary, for being with me at the office, day and night, weekdays and weekend, and especially for bearing my daily grumble and chasing my disappointments away. Without him, I would be lost.

List of Contents

List of Figures

List of Tables

Chapter I. Introduction

1. Motivation

Recognition of humans has become a substantial topic today as the need and investments for security applications grow continuously. Biometric authentication enables reliable and efficient identity management systems by exploiting physical and behavioral characteristics of the subjects that are permanent, universal and easy to access. The motivation to improve the security systems based on single or multiple biometric traits rather than passwords and tokens emanates from the fact that controlling a person's identity is less precarious than controlling what he possesses or what he knows. Additionally, biometry-based procedures obviate the need to remember a PIN number or to carry a badge.

Each carrying their own limitations, numerous biometric systems exist that utilize various human characteristics such as iris, voice, face, fingerprint, gait or DNA. "Superiority" among those traits is not a realistic concept when it is parted from the application scenario. The system constraints and requirements should be taken into account as well as the purposes of use-context that include technical, social and ethical factors (Introna, et al., 2009). For instance, while fingerprint is the most wide-spread biometric from a commercial point of view (Abate, et al., 2007) - mainly due to a long history in forensics, it requires strong user collaboration. Similarly, iris recognition, although being very accurate, highly depends on the image quality and also requires active participation of the subjects.

Face recognition stands out with its favorable reconcilement between accessibility and reliability. It allows identification at relatively high distances for unaware subjects that do not have to cooperate. Given a 2D or 3D still image or a video sequence, the face recognition problem can be briefly interpreted as identification or verification of one or more persons by matching the extracted patterns with the templates stored in a database, as is the case with all biometric traits.

Despite the fact that face recognition has been drawing a never-ending interest for decades and major advances were achieved, it still suffers from intra-class variations due to various factors in real-world scenarios such as illumination, pose, expression, occlusion and age. In the Face Recognition Vendor Test (FRVT) 2002, it was demonstrated that using 2D intensity or color images, a recognition rate higher than 90% could be achieved under controlled conditions (Phillips, et al., 2003). However, with the introduction of aforementioned variations, the performances deteriorated. The obtained results motivated an acceleration of studies on alternative modalities, especially the three-dimensional (3D)

face, since it -by its nature- seems like a logical way to evade pose and illumination problems.

As 3D sensing technologies advance and the acquisition devices become more accurate and less expensive, the utilization of range data instead of / together with color data broadens. Consequently, in FRVT 2006 (Phillips, et al., 2009), an order-of-magnitude improvement in recognition performance was achieved over the FRVT 2002 from 3D face images, with FRR of 0.01 at a FAR of 0.001. Similar results were also achieved with high resolution still images under controlled illumination. However, for the uncontrolled scenario, performances again drop.

In this dissertation, we investigate possible contributions of 3D information to face recognition performances. Thus, the first motivation of our work is to explore methods to incorporate advantageous features of 3D modality into identification/verification process whilst evading its drawbacks. One of the main challenges related to 3D face is its acquisition. Even with the significant progress achieved in the last few decades, it is still not as straightforward as its 2D counterpart to obtain an accurate 3D image. Most acquisition systems with high accuracy (e.g. laser scanners) require user collaboration and may take several seconds. On the other hand, systems that extract shape information from 2D images (e.g. passive stereo approach) can take instantaneous measurements; but they highly rely on the knowledge of extrinsic parameters of the scene and intrinsic parameters of the camera to obtain a certain degree of accuracy. Taking these facts into consideration, we propose an asymmetrical system which requires a combined enrollment of both 2D and 3D information whereas the probe images are still 2D images. Via this approach, we address the problem of facial expression variations in 2D face recognition by synthetically augmenting the available gallery. To this end, realistic simulations of various expressions for the enrolled subjects are obtained by deriving MPEG-4 compliant animatable models from their 3D facial scans in the gallery. Animatable model generation is achieved by warping a generic annotated model to transform into the subject to be animated by means of TPS.

In the next step, we explore discriminative properties of the TPS parameters obtained during the warping process of the generic model. TPS is mostly utilized as a registration algorithm which establishes dense correspondences between face models and its parameters have often been considered redundant. However, if the generic model is fully warped to adapt to the target face rather than deforming it minimally as applied for registration purposes, the warping parameters take a new meaning. In this case, as they represent deviations from a common model, they may be claimed to possess dense information on facial shape.

Our analysis on warping parameters proves them to be strong biometric signatures when imaging conditions are the same for gallery and probe sets, but not able to handle expression variations. This approach assumes the face to be rigid and hence fails to deal with confounding effects of these changes. Recently, 3D research is largely committed to the problem of expression variations; however it is still recognized as one of the most

difficult challenges of 3D face recognition (Smeets, et al., 2010). Consequently, we expand our research onto that matter.

Due to being local distortions on the facial mesh, expression and occlusion variations are often proposed to be handled via region-based methods which suggest dividing the face into multiple parts for more robust comparison. Computed similarity scores from all regions are often fused using a weighted similarity metric which attaches more importance to more static regions (e.g. (Ben Amor, et al., 2008), (Smeets, et al., 2010)). In our work, instead of making assumptions about reliability (may also be referred as confidence throughout the text) of different facial regions, we propose to estimate it based on surface characteristics. Using a training set of 3D face models with different expressions and occlusions, an Artificial Neural Network is trained for each region to assess the presence of such distortions from the distribution of the vertices' primitive shape categories (Due Trier, et al., 1995). Initial calculation of the reliability scores for training is achieved by registering each region to its neutral and clean equivalent which belongs to the same person, by Iterative Closest Point (ICP) algorithm and calculating the residual distance between corresponding points. Once the regional confidence levels are obtained, they are incorporated in the fusion of regional match scores. Three different fusion approaches are evaluated and compared for this scheme: Sum rule, Borda count and probabilistic approach. Experimental results demonstrate improved robustness against expression and/or occlusion induced mesh distortions.

Unfortunately, variations that modify the texture and the shape of the face are not limited to expressions and occlusions. There are other causes of variation such as ageing, body-mass index (BMI) differences or plastic surgeries. In particular, becoming more advanced and affordable, plastic surgery procedures introduce new challenges in designing future face recognition systems (Singh, et al., 2010), yet very few studies exist which address those. In this dissertation, we undertake this problem with a wider perspective as facial alterations, since plastic surgery is just one of many ways to change the appearance of face. We choose to focus on nose region as it is frequently assumed to be robust to variations and given highest significance in region-based approaches mentioned previously. In order to analyze the impact of nose alterations, a synthetic face database is prepared in which nose in each sample is replaced with another randomly chosen one and performance comparisons of several key face recognition techniques are accomplished.

2. Original Contributions

The original contributions of this thesis can be summarized in four clauses: realistic 3D animation of facial expressions, utilization of TPS parameters as biometric signatures, regional reliability estimation for 3D face and simulated nose alteration database.

Learning or modeling the non-rigid behavior of expressions have been proposed in many studies which mostly aim the removal of expressions from the probe images (Wang, et al., 2007) or addition of expressions to the gallery images (Lu, et al., 2006) for robust

3D matching. There are two disadvantages of this approach: Need for additional data to train the system for a big range of possible expressions and the validity of transferring expressions. Obtaining client-specific animatable models for the proposed system is simple and executed only once during enrollment. Additionally, simulating the expressions is very straightforward with animation engines thanks to the MPEG-4 compliance.

Another contribution is the analysis of TPS parameters for their discriminative properties. Those features, mostly neglected as a residue of registration step, are proven to be highly informative despite the low complexity of their extraction and small size.

Additionally, a scheme is presented to evaluate different regions of face according to the presence of distortions for fusion of regional match scores instead of suppositions of stability. Residual error of ICP registration between two samples (one being the reference sample) of the same person is proposed to be utilized as indicators of a region's reliability.

Finally, an almost untouched field in 3D face recognition is explored where facial alterations are simulated for nose region and their impact on recognition performances is determined.

3. Outline

The outline of the dissertation is follows:

- In Chapter II, we provide an introduction to 3D face recognition, reviewing full spectrum of 3D face processing technology, the state of the art for the relevant work, current challenges and possible future directions.

- In Chapter III, we present our pre-processing and automatic landmarking approach for 3D face, which is utilized repeatedly for the rest of the study.

- In Chapter IV, we introduce our asymmetrical scenario in which a combined enrollment is realized by both 2D and 3D information whereas the target images are still 2D images. Our method to generate client-specific MPEG-4 compliant animatable models is detailed and the positive effect of gallery augmentation on recognition performances is reported.

- In Chapter V, discriminative properties of warping (TPS) parameters are analyzed.

- In Chapter VI, we propose a method for regional confidence score estimation for 3D face models. Different methods for fusion of match scores which are based on estimations rather than assumptions are evaluated.

- In Chapter VII, the impact of nose alterations on the recognition performances is investigated. Our method to simulate nose alterations is explained in detail and the experimental results are given to show the effect of the applied modifications.

- In Chapter VIII, the conclusions are provided.

Chapter II. 3D Face Recognition

1. Introduction

While the requirements of human identification experiences drastic changes with the incredibly fast growth and mobilization of populations, biometrics technology is regarded as the most promising solution. With the addition of improvements in sensor and computer technologies, it is the prime candidate to be employed in a wide range of applications from forensics to border control.

Biometrics

Biometrics is the field of recognizing a person automatically based on single or multiple intrinsic traits associated with him/her. Those traits may be physical (e.g. face, fingerprint, iris) or behavioral (e.g. gait, lip motion, signature). Biometric authentication has mainly two advantages over traditional methods like passwords or ID cards:

- Reliability: It is relatively more difficult to forge biometric identifiers. Passwords can be disclosed or ID cards can be stolen of falsified. However, an individual's irises or fingerprints are reasonably more secure since they cannot be easily stolen, lost or shared.

- Ease of use: Biometric authentication obviates the necessity to carry a badge or remember a password.

A typical biometric system is composed of four modules (Jain, et al., 2004): *Sensor module* which captures biometric data (e.g. face image, palmprint) from an individual; *feature extraction module* which processes the output of the sensor to extract a discriminatory feature set (e.g. fingerprint minutiae, iris code); *matching module* in which the extracted features are compared against stored biometric templates and match (similarity) scores are generated; and *decision-making module* which achieves recognition based on the acquired match scores.

The stored biometric templates in the database constitute the *gallery* which contains reference samples of the people enrolled to the system. During the e*nrollment* process, subjects present one or more biometric data samples from which the biometric templates are generated. Contrary to recognition, enrollment is generally performed in a well-controlled environment.

Operational Modes

The biometric data captured for recognition is called a *probe* sample. Biometric systems may process it either in *verification* or *identification* mode. In the verification mode, the probe

is matched with the claimed template (one-to-one comparison) for validation and it is either accepted or rejected. In the identification mode, it is matched with all biometric references in the gallery (one-to-many comparison) to answer the question of whom this biometric data belong to. If the biometric sample in the probe is always someone in the gallery, it is called *closed-set identification*. Otherwise, if we do not know in advance whether the person to be identified is in the reference database, it is called *open-set identification*.

Evaluation

Each operational mode has its own set of performance metrics. In the verification mode, if the match score meets the threshold criteria, the identity claim is accepted; otherwise, it is rejected. This setting leads to four possible outcomes:

1. True accept: The person is who he claims to be (*genuine*) and his claim is verified.
2. True reject: The person is not who he claims to be (*impostor*) and his claim is disproved.
3. False accept: The person is not who he claims to be and his claim is verified.
4. False reject: The person is who he claims to be and his claim is disproved.

Threshold-based decisions always introduce a tradeoff to be considered. In the case of verification, if the operating threshold is too high, *False Reject Rate* (FRR) might increase since more legitimate claims would be rejected. If the threshold is too low, acceptance of false claims will be more likely, increasing the *False Accept Rate* (FAR). This relationship is shown with a *Receiver Operating Characteristic* (ROC) graph which represents the probability of true acceptance versus probability of false acceptance.

ROC curves are also used to measure performances of open-set identification systems. Instead of true acceptance rate, detection and identification rate is calculated and plotted against FAR. Detection and identification rate is the percentage of the probe samples represented in the gallery that are correctly accepted and identified.

With the identification task, *rank* phenomenon is introduced. In rank-1 case, a probe is identified as the first ID in the list of subjects sorted in decreasing order of computed similarity scores. Correspondingly, rank-n systems examine the top n matches. Thereby, detection and identification rate is also a function of rank. Performance of an open-set identification system can be represented in 3D surface for which the axes are detection and identification rate, FAR and rank. But mostly, 2D slices of this surface are utilized by holding either rank or FAR constant.

On the other hand, in closed-set identification which is a special case of open-set identification, the performance is computed as a function of rank only and reported on *Cumulative Match Characteristic* (CMC) curves. CMC plots the size of the rank order list against the identification rate.

Achievable recognition accuracy, often evaluated by the metrics mentioned above, is just one of many important issues that should be considered in a practical biometric

system, such as computational speed, system cost, security, privacy and usability (Bolle, et al., 2004).

Security reflects how vulnerable the biometric system is to fraudulent attacks like spoofing or tampering biometric data. Together with the accuracy, they determine the system effectiveness. Privacy and usability issues together affect the acceptability of the system, which indicates how well individuals accept to have their biometric trait captured and cooperate. Ideally, balance must be struck between user acceptance and effectiveness; however, this is not a very straightforward task. Many highly effective biometric systems have to be avoided because they are considered intrusive. For example, retinal pattern recognition is one of the most accurate and reliable of the biometric technologies, but it has very low acceptability because retina scans require contact with the sensor and additionally, they can reveal some medical conditions like pregnancy, raising both hygiene and privacy issues.

Face Recognition

Face recognition - the most commonly used biometric characteristic by humans, has moved to the forefront by offering a good compromise between effectiveness and acceptability. Besides being non-invasive and natural, it holds the promise of recognition at a distance, requiring neither the cooperation nor the knowledge of the subject.

In a typical face recognition system, given an image, firstly face is detected and segmented from the background. Next, several landmarks are localized and used in geometric normalization of the face. This is followed by facial feature extraction and finally matching.

Since the first face recognition system presented in 1973 (Kanade, 1973), it has become a very popular area in computer vision on account of rapid advancements in image capture devices, increased computational power and large variety of its applications. With the increasing popularity, face recognition systems reached recognition rates greater than 80% in constrained situations (FRVT 2002 - (Phillips, et al., 2003)) even exceeding human performance, especially in the case of very large galleries.

However, FRVT (Face Recognition Vendor Test) 2002 also made it clear that in real world scenarios where face images can include a wide range of variations, performances degrade very rapidly. As of those variations, one can principally list illumination, pose, facial expression, occlusion and age.

Pose and illumination variations are two major sources of degradation in recognition performances. As the pose of a subject or the direction of illumination deviates from the frontal, it often causes face image differences that are larger than what conventional face recognition can tolerate. Extensive efforts have been put to achieve pose/illumination invariant face recognition. As an example, two prominent methods that have been proposed can be briefly mentioned:

(Georghiades, et al., 2001), present a generative appearance-based method which synthesizes novel images with different viewpoints and lighting conditions using illumination cone model (ICM). In this approach, multiple images are required under certain restrictions. Additionally, specular reflection is discarded in facial surface approximation.

Another method is proposed in (Blanz, et al., 2003), in which the process of image formation is simulated and 3D shape and texture of faces are estimated from single images by fitting a 3D morphable model (3DMM). However, this approach is reported unstable as identity-related shape and texture coefficients are affected during cost function minimization (Zhang, et al., 2009). Additionally, 7 facial points are required for initialization and computationally, it is too slow for real-time applications (Heo, et al., 2008).

Even though many other promising methods -each carrying its own limitations- can be listed, uninterrupted efforts are still needed to fully achieve the goal of pose and illumination invariant face recognition. Those problems motivated 3D face recognition approach which fundamentally differs from 2D mainly because 3D face acquisition is –to some extent- invariant to illumination changes and pose variations can be easily handled by rotating faces in 3D.

3D Shape Based Face Recognition

Although it has been almost four decades since the earliest works on 3D face recognition (Harmon, 1973), it was not until approximately 10 years ago that it started attracting a considerable amount of attention. This is mainly because acquisition technology and computer power improved remarkably, enabling efficient 3D face recognition systems. These developments also gave way to numerous comprehensive public databases of 3D faces that strongly influenced and inspired the course of research.

In this chapter, first, we briefly introduce the past and the present of 3D imaging techniques. Next, an overview of facial surface representations is given. Finally, after presenting a survey of the literature on automatic face recognition using 3D data, current challenges and trends is discussed.

2. 3D Imaging

A number of range data acquisition (i.e. 3D scanning) techniques are available; each having its own advantages, limitations and cost. These techniques are used to create 3D point clouds, sampled from the surface of the subject being scanned and they can be classified mainly into two types: non-contact scanners which can be further divided as passive and active, and contact scanners which detect the range through physical touch. The latter is out of our scope due to inability of utilization in face recognition field. Passive (non-contact) and active (non-contact) range scanning technologies will be described in detail in the following sections.

2.1. Passive Sensing

Largely inspired by human visual system, passive sensing technologies attempt to determine the shape of the scene from single or multiple images, exploiting a wide range of visual cues such as shading, perspective, focus, stereo, motion parallax, occluding contours and so on. Assuming that the sensor simply records light that already exists in the scene, 3D scene description is derived from 2D images by analyzing the reflectance properties.

Commonly referred as shape-from-X techniques, passive sensing methods recover shape from image intensity (irradiance) values without emitting any kind of radiation themselves but relying on ambient illumination reflection. Examples are shape from shading (Horn, 1970) which determines shape from a gradual variation of shading in the image, photometric stereo (Woodham, 1978) which utilizes successive images of different direction of incident illumination and constant viewing direction, integral imaging (Frauel, et al., 2004) which captures 3D image data from multiple perspective 2D intensity images, photometric sampling (Nayar, et al., 1990)and shape from inter-reflections (Nayar, et al., 1990).

The most important advantage of these approaches is that it enables recognition of non-cooperative persons at a distance, longer than what is generally allowed by active sensors. 3D face model is generated from sequence of images and utilized in person identification at 3, 6 and 9 meters by (Medioni, et al., 2007). Later, a framework for face recognition at 15 meters based on sparse-stereo reconstruction is proposed by (Rara, et al., 2009).

On the other hand, the accuracy of these methods is generally limited by numerous factors. Firstly, the geometric relationship between the subject and the sensing devices is required to be measured precisely. This means, during the capture of the scene the environment is expected to be highly controlled. Additionally, reflectance characteristics of the surfaces and image formation models may be over-simplified, for example by Lambertian surface assumptions or omissions of mutual illuminations. These factors result in low-accuracy reconstruction. In order to overcome this issue, model-based approaches emerged (Zhang, et al., 2004) to confine 3D reconstruction of face in regions with uniform appearance.

Specific to objects that can be modeled using their common shape properties, like face, 3D model fitting is another way to acquire their shape information. (Blanz, et al., 1999) present a parametric face modeling technique based on linear combination of large number of 3D face scans. The constructed "morphable model" is able to generate 3D models of faces using a single 2D image by transforming this task into an optimization problem.

Despite these efforts, low accuracy and reliability in passive sensing –especially in face recognition domain; gives rise to the popularity of active sensing which is additionally boosted by their advancing technology and declining cost.

2.2. Active Sensing

Several active sensing technologies have been developed that utilize some kind of emission, such as laser, white/infrared structured/modulated light and x-ray. There are two main forms of active range sensing: time-of-flight and triangulation.

In time-of-flight systems, the scanner sends a laser beam and measures the time that it takes to return back. This can be achieved using different methods, such as emitting pulsed laser and employing digital time counters, emitting modulated light and employing phase detectors or via range gated imagers which allows single-shot capture of moving scenes by emitting a plane of light and employing a bank of sensors. The measured time multiplied by the speed of light in air gives the distance of the surface point that reflected the beam back, to the sensor. This calculation requires only a small amount of processing power and it can be achieved at a very high speed. Additionally, time-of-flight sensors can operate over very long distances. On the other hand, the accuracy highly depends on timing precision. For these reasons, this type of range sensors is often utilized to scan large structures like buildings.

Triangulation systems also use a focused beam of light to probe the environment and by tracing a line of sight through the illuminated pixel; they retrieve the depth of the incident point. Mostly, a laser stripe is swept across the scene instead of a single point, prohibiting the capture of moving objects. Their precision is higher compared to time-of-flight systems; however, they have a limited range.

In addition to laser stripe or beam, structured-light (an area pattern) can be projected on the scanned object. The distortions recorded by a camera slightly offset from the pattern projector, can then be converted into 3D coordinates using a technique similar to triangulation. In this method, the primary challenge is to uniquely identify the pattern components.

Very recently, the first consumer-grade structured-light application which uses a pattern of infrared projection is introduced to the market by Microsoft. With its availability and affordability, Kinect reached 8 million sales in two months and is expected to cause a revolution in vision applications in a wide range of domains, from biometrics to HCI.

One additional method to recover 3D shape is conoscopic holography which is based on light propagation effects in uniaxial crystals. The distance to the measured surface is determined via the frequency analysis of the diffraction pattern that is produced by the reflection of a laser beam through a conoscopic crystal. Fourier Transform Profilometry that has been developed on such a system using potentially low cost components, is reported to operate with subjects moving at $\leq$1.5 m/s at $\leq$25 meters range with <1.4 mm lateral resolution (at maximum range) and <1 mm range precision at 1 Hz capture rate (Hoft, et al., 2011).

Under uncontrolled illumination conditions like outdoors, active sensor technologies suffer from a common problem of background light which can make the projected

pattern less visible and cause loss of accuracy. Consequently, many existing active 3D sensors are restricted to indoor applications. On the other hand, especially for environments without direct sunlight, 3D data construction of the scene can be achieved within seconds with very high reliability and accuracy.

3. Surface Representations for 3D Face

In this section, we give an overview of surface representations for 3D face within biometrics context. 3D acquisition systems often capture the facial shape in combination with associated texture image. They may output the measurements using several different data types, which are inter-convertible.

Without doubt, the simplest form to represent the facial surface which is also the most common output generated by 3D scanners is the *point cloud* representation. It contains an unordered set of 3D coordinates that belong to the points which lie on the facial surface. It used to be considered as a sparse approximation of the real surface; but this opinion loses a part of its validity each day as the storage and processing capabilities increase. Finer and finer levels of detail can be kept and utilized for recognition today, without any concern for memory shortage. This convenience of use also facilitated point cloud algorithms to gather speed lately. Subsets of the whole point cloud are also proposed as sparser representations to approximate the facial shape, such as contour and profile curves.

The inclusion of edge (between two points) list and face (closed set of edges) list together with the 3D point coordinates brings us to *mesh* representation. Apart from simplifying rendering for visualization, it also enables measurements of geodesic distances between facial points using the additional connectivity information. Given a point cloud, constructing a mesh of triangles, quadrilaterals or other simple convex polygons cloud can be dealt with numerous algorithms, the most powerful one being the power crust algorithm by (Amenta, et al., 2001). Additionally, methods to build a regular facial mesh model based on the scattered point cloud have been introduced (Xu, et al., 2004).

Another popular method to represent the facial shape is to project the depth data onto a 2D image in (u,v) domain, which results in a parametric surface representation where x = u and y = v. Often referred as *depth map*, range map or 2.5D, it represents the depth of regularly sampled points along z-direction. This representation which is used by many 3D scanners as the output format has the advantage of applicability to 2D image processing techniques. On the other hand, since they contain the visible points from a single view, depth maps have limited depiction compared to 3D data.

The construction of depth maps can be extended by using different regular sampling techniques, especially in an effort to overcome view-point limitation which causes problems in cheek and ear regions. Parameterization of the face using a spherical coordinate system that measures the distance at each point from the central vertical axis to the facial surface is an example of such methods. Canonical depth maps (Colbry, et al.,

2007) and geometry maps (Kakadiaris, et al., 2007) are two other extensions for a regularly sampled 2D grid representations, which, instead of a plane, utilize a parabolic cylinder and an annotated face model with a continuous global UV parameterization, respectively. One last example is the spline surfaces (e.g. Bézier surfaces, B-splines/NURBS) with control points on a regular grid ((Yang, et al., 2006) and (Liu, et al., 2005)).

Similar to depth map construction, other types of surface descriptors can also be extracted on a regular grid to constitute 2D maps such as normal maps, curvature maps, shape index maps, etc.

The above-mentioned representations are illustrated in **Figure 1**. Their more detailed descriptions can be found in (Fabry, et al., 2010) together with some others that are not yet reported in current face recognition literature, but considered to be promising.

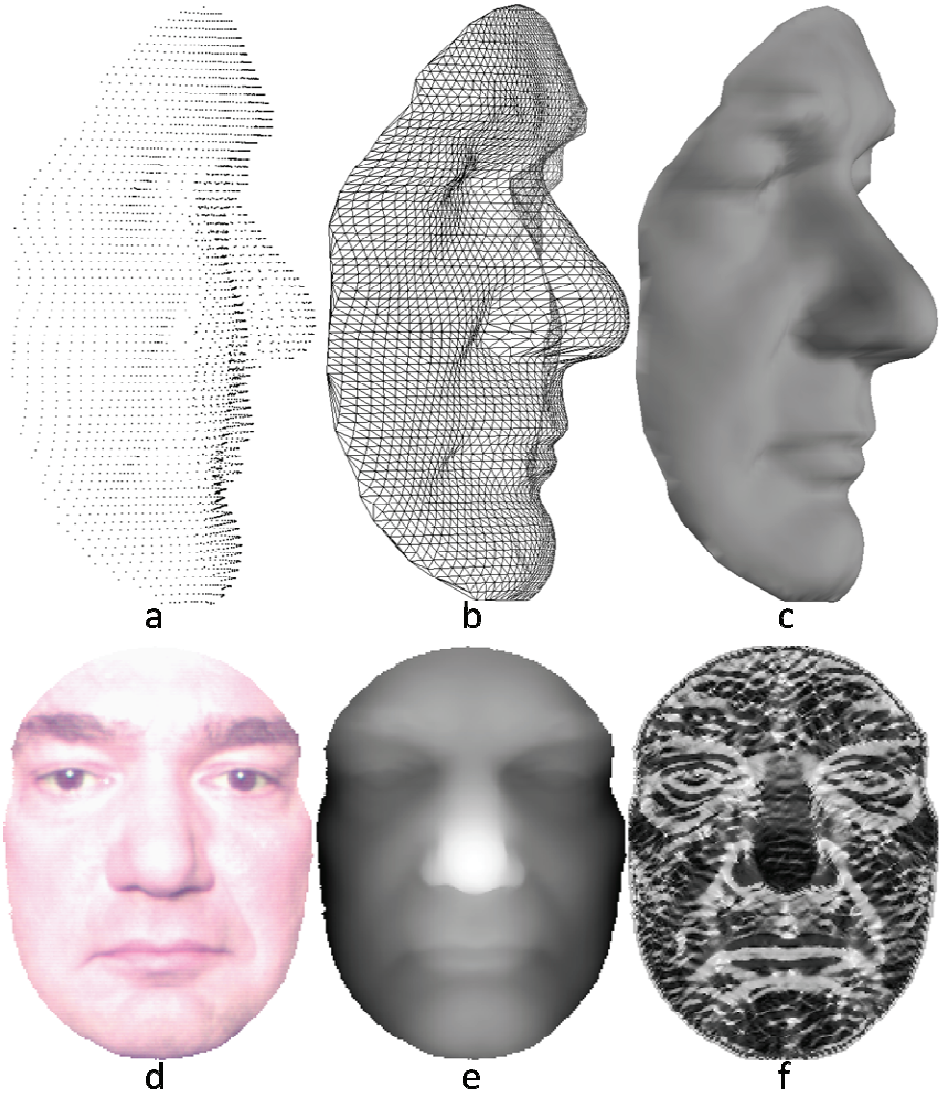


Figure 1: Some representations are illustrated on a 3D face sample: a) Point cloud b) Triangular mesh c) 3D rendered as shaded model d) Texture map e) Depth map f) Shape index map

4. 3D Face Recognition: A Survey

Each having their own strengths and weaknesses, many different approaches have been developed to utilize 3D shape information in face recognition. In this survey, we will study them under two different categories.

The first class of approaches relies on using only 3D shape information. 3D model representations or feature vectors derived from them are compared for recognition purposes. We will associate 3D modules of "multi-modal" approaches with this class as well. Since facial appearance is a combination of texture and shape, utilization of both components is adopted by many studies, labeled as "multi-modal". For further reading, we can refer the reader to two very good examples (Chang, et al., 2003) and (Husken, et al., 2005), in which strategies and benefits of 2D+3D fusion have been investigated thoroughly.

In the second class of approaches, 3D face models are utilized to assist 2D face recognition. Either obtained by 3D scanners or reconstructed using single or multiple 2D images, 3D shape information is proven useful to overcome common problems in 2D face recognition, such as illumination, pose and expression variations.

Additionally, multi-algorithm systems (like (Huang, et al., 2010)) and fusion method evaluations (you can refer to (Gokberk, et al., 2006) for a comparative study) is left out of scope the of this survey

4.1. 3D Shape-Based Face Recognition

Whatever approach is adopted, methods that utilize 3D data in face recognition typically follow several main steps: Pre-processing of the 3D data, alignment of faces (registration) and recognition using 3D shape. Each of these steps will be reviewed in detail in the following subsections.

In addition, we will dedicate a separate part for 3D approaches proposed to deal with non-rigid structure of the face. Mostly associated with facial expressions, this behavior of the 3D face data is proven detrimental (Phillips, et al., 2005) for the recognition performances. Partially owing to the increasing availability of large 3D face databases embodying facial expressions, a large variety of strategies have emerged to deal with this challenge. The main characteristics and properties of those approaches will be summarized in the last subsection of this chapter.

4.1.1. Pre-processing

The raw 3D data obtained from most scanners needs to be processed before any further operation as it contains non-facial points (regions like shoulders and clothing), spikes, holes and noise.

In order to speed up the subsequent steps, firstly the non-facial regions included in the raw scanner output like hair, clothing and shoulders are removed. Also referred as face

cropping (as in 2D case), this is often accomplished using landmark points like nose tip and outer eye corners. For example, (Mian, et al., 2006) detect nose tip and crop the face via a sphere centered at the nose tip and (McCool, et al., 2007) utilize the eyes to extract the face in the range map as a square region of interest. It is also possible to simply apply a depth thresholding (assuming that face is the closest region) and then to remove non-facial regions via connected-component analysis.

Spikes are outlier errors in the data that can be caused by specular regions like eyes or glossy regions like nose tip or forehead. Through smoothness assumption, they can be detected as being far from their neighbors and removed leaving additional holes behind.

Holes are essentially areas of missing data which can be introduced by less reflective regions that cannot be captured by the sensors, like eyebrows, moustache or hair. In order to patch the holes, missing points can be restored by various interpolation techniques and/or using facial symmetry. Linear interpolation can be sufficient for small holes; on the other hand, (Mian, et al., 2007) have shown that bicubic interpolation can give better results. For larger holes, facial symmetry is often utilized by simply copying the corresponding symmetric vertices from the other side of the face (Malassiotis, et al., 2004). Additionally, model-based approaches are also proposed in which a model is morphed until it gives the best fit for the missing points (Kakadiaris, et al., 2007).

Another type of error in 3D data is noise which can be introduced by imaging conditions like illumination, surface texture, as well as by the optical/mechanical components of the scanner like lens or CCD. Low pass and median filters can be used to remove the noise. Smoothing using spline is also possible (Bronstein, et al., 2003).

Regularization of the face representations, mentioned in the previous section "Surface Representations for 3D Face", can also be included in the preprocessing step. This operation facilitates the surface comparison and can be realized by remeshing (Xu, et al.,

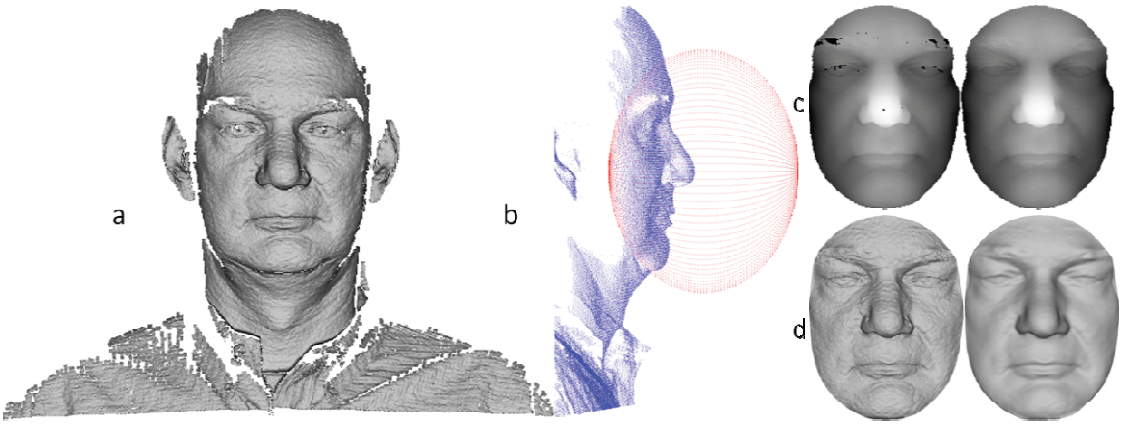


Figure 2: Some pre-processing operations are applied on a sample model: a) The raw output of the 3D scanner includes spikes, holes and noise. b) A sphere of radius 100 mm and centered 10mm away from the nose tip is utilized to crop the area of interest. c) Cropped range maps before and after hole-filling operation. d) Cropped range maps before and after smoothing operation.

2004) and (Irfanoglu, et al., 2004) or voxel discretization (Achermann, et al., 1997). A chain of pre-processing steps is illustrated on a sample in **Figure 2**.

4.1.2. Alignment

Once the clean 3D face is obtained, the next problem to deal with is the facial pose differences. In order to make a comparison possible, face models need to be aligned, i.e. the translation and rotation that optimally aligns the two point sets should be computed. This is a critical part which heavily influences the overall performance of 3D face recognition systems.

The translation issue can be solved by colliding faces' center of mass (Bronstein, et al., 2003) or nose tip (Hesher, et al., 2003). There also exist approaches which detect and utilize multiple points to coarsely deal with the rotation issue. (Ben Amor, et al., 2005) utilize more than two corresponded points to calculate rotation, translation and scale. (Lu, et al., 2004) automatically detect three feature points (inner and outer corners of an eye and the nose tip) using shape indices and compute a rigid transform using a least square fitting between the triangles formed by the two sets of detected feature points. In a different approach, (Achermann, et al., 2000) fit a plane on the facial surface to be aligned with the image plane of the camera.

Alternatively or additionally, Iterative Closest Point (ICP) algorithm (Besl, et al., 1992) can be employed for which no point correspondences are required to be known; however, a reasonably good initialization is important to avoid local minima. This most crucial disadvantage of ICP results from its sensitivity to initialization. A relatively good starting point is required in order to avoid trapping into a non-optimal local minimum. For this purpose, a coarse alignment prior to ICP is proposed by many methods (e.g. (Lu, et al., 2004) and (Maurer, et al., 2005).

In ICP, firstly an alignment estimate is computed by defining correspondences as the closest points from one surface to another. Next, a rigid transformation is realized according to the estimated alignment parameters and point correspondences are updated. These iterations are repeated until the observed improvement is less than a threshold. After convergence, the ICP algorithm also approximates the volume difference between two surfaces as the sum of residual distances between final corresponding points. It is used as the similarity score in many studies (For the examples please refer to the next section).

Another important method that better aligns the facial models by warping them (non-rigid registration) is the Thin Plate Spline algorithm (Bookstein, 1989). Generating a smoothly interpolated mapping between two sets of landmark points, TPS method establishes a dense point-to-point correspondence excellently. Still, it has two main disadvantages: Firstly, TPS highly depends on the landmarks and it does not take into account surface intensity information. For this reason, a good match is not very likely where no landmark information is available. Secondly, it results in a non-rigid deformation which in the end may degrade the recognition performances. There exists a

trade-off between the two disadvantages. As the number of landmarks rises for better alignment, the discriminatory information loss becomes larger, causing success rates to go down.

Lastly, as an alternative approach, instead of aligning faces, alignment invariant data can be extracted, such as distances, angles and areas between facial points. As an example, (Gupta, et al., 2007) employ geodesic and Euclidean distances between 25 facial fiducial points as features and select most discriminatory ones using stepwise linear discriminant analysis (LDA). On the other hand, (Fabry, et al., 2008) present a 3D face recognition method based on point cloud kernel correlation which only requires the actual point cloud of the facial scan without any point correspondence or prior knowledge.

Automatic Localization of Facial Points in 3D

Most cropping and alignment techniques rely on one or more facial feature points, like eye corners, nose tip, lip corners etc. The correct localization of these points is critical for the rest of the system to work properly. Many 3D face recognition studies utilize manually selected points on the assumption that there already exist numerous solutions for their automatic detection. Whereas, some researchers propose their own automatic landmarking methods within the 3D face recognition scheme. For this reason, this field of research can also be included in the pre-processing of the 3D faces.

(Yacoob, et al., 1993) identify convexity and concavity points and group them into components via qualitative reasoning. In (Colbry, et al., 2004), feature points are located based on local shape index information and a search algorithm is utilized to find the best candidate points.

In (Dibeklioglu, et al., 2008), a 3D facial landmarking algorithm is proposed which relies on accurate statistical modeling of facial features and can generalize to different acquisition conditions. On the other hand, (Zhou, et al., 2010) present a 3D Active Shape Model (3DASM) -based system to automatically locate facial feature points from different views. 3DASM is trained on synthetic data, generated using a 3D morphable model.

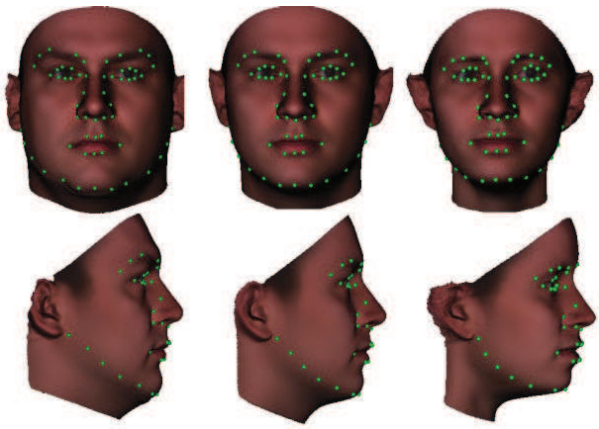


Figure 3: 3DMM and 58 manually selected landmarks. Middle: the average face model. Left and right: The first component of shape parameter is changed (Zhou, et al., 2010).

4.1.3. Recognition

There exists a range of methods that relies on only 3D shape information for face recognition. In this section, they will be discussed under several main categories:

Curvature Map Segmentation

The earliest 3D face recognition approaches are based on curvature information. Curvature describes the degree of bending at a particular point and it is closely related to the second derivative of the curve defining the surface. Maximum (k_{max}) and minimum (k_{min}) curvatures, referred as the principal curvatures, are defined by the plane orthogonal to the surface and measure the maximum and minimum bending (positive or negative) of the surface at one point, respectively. Additionally, combinations of these metrics are widely used, such as mean curvature (($k_{max}+k_{min}$)/2), Gaussian curvature ($k_{max}\cdot k_{min}$) and shape index ($2/\Pi\cdot\tan^{-1}((k_{max}+k_{min})/(k_{max}-k_{min}))$). One limitation of purely curvature-based approaches is that they are sensitive to size variations between faces.

(Lee, et al., 1990) utilize curvature values to segment the range images. Based on the sign of the mean and Gaussian curvatures, convex regions are detected and represented by their Extended Gaussian Images (EGI). Finally, match scores are obtained by correlating EGIs of the face pairs. In (Tanaka, et al., 1998), the same method is tested using a spherical correlation. Moreover, (Wong, et al., 2004) propose to find an optimal transformation for EGI representations based on genetic optimization.

Likewise, (Gordon, 1991) begins with segmenting the range maps by sign of mean and Gaussian curvature into four regions: concave, convex, and two types of saddle. Next, the faces are transformed to the standard position using location of the bridge of the nose, the base of the nose and the outside corners of the eyes that are detected automatically and the volume difference between aligned surfaces are computed as a measure of their similarity.

In (Moreno, et al., 2003), after segmentation of faces based on curvatures, 86 features are extracted using areas of regions, distances and angles between mass centers of regions, average Gaussian and mean curvatures of the regions, etc. and sorted according to their discriminatory power computed using Fisher coefficient and first 35 is selected to represent faces for matching.

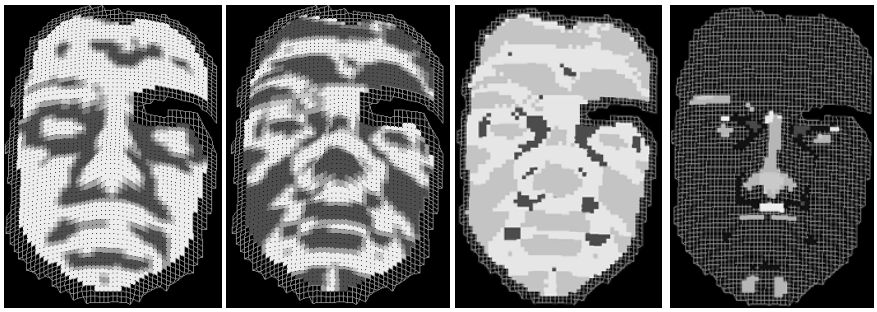


Figure 4: Sign of (a) median and (b) Gaussian curvatures; (c) point HK classification and (d) HK classification after curvature thresholding (Moreno, et al., 2003).

Contour / Profile Curves

Especially, in the earlier studies, due to severe computation and storage requirements, utilization of the whole facial data was not considered feasible. Since then, many facial curve-based methods have been proposed which extract profile or contour curves from the 3D geometry of the face and use them for matching.

Contour curves are closed, non-intersecting curves that can be defined according to various criteria. Isodepth, isoradius and isogeodesic curves are three popular examples. On the other hand, profile curves have a starting and an end point and they can be extracted in infinitely many different ways. Vertical and horizontal profiles through nose tip are widely utilized.

As one of the first studies, in (Cartoux, et al., 1989), the profile plane of a face is found using its principal curvatures and facial symmetry. The extracted profile curves are aligned and compared for recognition.

(Nagamine, et al., 1992) normalize the face position and extract various vertical and horizontal profiles and isoradius curves using five automatically located feature points. According to their results, vertical profile curves that pass through the central region of the face achieve the best recognition rates. Later in (Beumier, et al., 2000), in addition to central profiles, mean lateral profiles are also compared and the fusion of both methods provides a clear advantage compared to each one individually.

(Lee, et al., 2003) perform 3D face recognition by extracting contour lines from some depth value and comparing them after being resampled and stored consecutively as feature vectors. Similarly, (Samir, et al., 2006) represent facial shapes via isodepth contours which they refer as level curves. The distances between two level curves are computed using a different geometrical approach that adopts geodesic lengths as a shape metric. A large set of profile and contour curves and their hybrid combinations are evaluated in (Ter Haar, et al., 2008) using a generic distance measure for fair comparison.

Especially after being verified as invariant to facial expression in (Bronstein, et al., 2007), the geodesic distance methods to measure distances between points on the facial surface gained popularity. This property was extended to facial curve extraction as well. (Feng, et al., 2007) present an Euclidean integral invariant signature obtained by

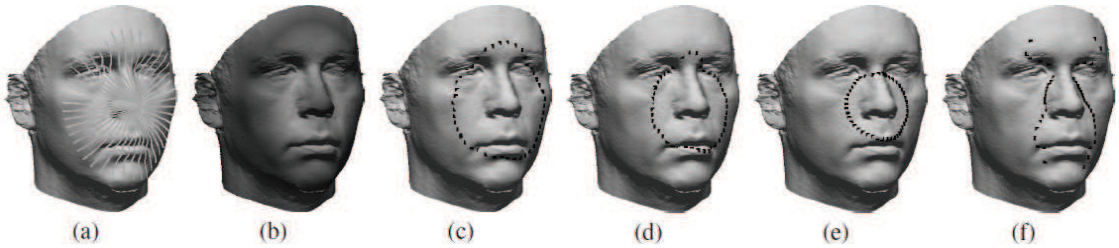


Figure 5: (a) Profile curves of shortest geodesic paths (b) Geodesic distances to the nose tip are colored (c) Isogeodesic contour curve (d) Contour consisting of samples with the same curve distance to the nose tip (e) Isoradius contour curve (f) Isodepth contour curve (Ter Haar, et al., 2008)

transforming the facial isogeodesic curves. (Jahanbin, et al., 2008) examine both isodepth and isogeodesic curves by encoding their characteristics with shape descriptors like convexity, compactness, circular variance and polar Euclidean distance from the nose tip.

2.5D Approaches

Representation of 3D facial shapes with depth (or range) maps has a great advantage of applicability to many existing 2D techniques, with mostly utilized ones being subspace analysis and Hidden Markov Model (HMM) based methods.

In (Achermann, et al., 1997), both eigenface and HMM methods are extended to range images and report 100% recognition rate on a dataset of 24 subjects. With these promising results, many other similar approaches appeared in the literature.

(Hesher, et al., 2003) utilize principal component analysis (PCA) and independent component analysis (ICA) to reduce the dimensionality of the depth maps and compare the resulting feature vector to the gallery templates. Then, they perform classification using the nearest neighbor criterion on the Euclidean metric. In (Tsalakanidou, et al., 2004), a multi-modal method is presented for which two Embedded Hidden Markov Model (EHMM) classifiers, one for intensity and one for depth, are trained and utilized for recognition. (McCool, et al., 2010) examine the applicability of two HMM topologies (HMM 1D and HMM 2D) on depth maps and conduct a comparative study with a GMM system. They show that HMM 2D system can greatly outperform its GMM equivalent.

In (Srivastava, et al., 2006), an optimal component analysis is applied on the training set of range images to find a subspace that maximizes nearest-neighbor classifier performance. The optimal subspace is searched based on a stochastic gradient algorithm on Grassmann manifold and the experimental results show that it performs better than standard component analyses such as PCA and ICA.

Another 2D approach that has recently gained popularity recently is based on Local Binary Pattern (LBP) which is proven to be very efficient despite its simplicity. This approach is applied to surface normal maps in (Li, et al., 2011) and enhanced by learning based spatial weights.

Besides, some researchers explore different ways of shape representations in 2D. In (Bronstein, et al., 2003), the authors propose an isometric transformation that can be applied both on 2D and 2.5D images to better cope with facial expression variations. After obtaining a bending-invariant canonical form of the facial surface, they utilize eigen-decomposition coefficients and Euclidean distance to recognize faces. Additionally, (Pan, et al., 2005) present a recognition method based on flattened depth maps which preserve the intrinsic geometric properties of facial surfaces and eigenface application. The flattening of the facial surface is achieved by one-to-one mapping to an isomorphic planar space and PCA is applied on the obtained maps.

An annotated deformable model approach is proposed by (Passalis, et al., 2005) which utilizes a face model (AFM) with anthropometric landmarks annotated and a continuous

global UV parameterization applied on it. After fitting the AFM on the face model under analysis, a deformation image (referred as geometry image) is obtained using its UV parameterization which retains all the shape information and is suitable to be processed with 2D-based methods. Wavelet analysis is adopted to derive descriptive and compact biometric signatures from the geometry images in (Ocegueda, et al., 2011) and (Passalis, et al., 2011).

Different construction approaches for depth maps were mentioned in Section 3. Moreover in this part , we can add (Xu, et al., 2004) in which first a regular 3D mesh model is built to represent the face shape and then PCA is utilized by regarding Z-coordinates of the mesh as the intensity values. Also in (Russ, et al., 2006) a dense correspondence is achieved using a generic reference face model, enabling 3D PCA.

Finally, we would like to note a thorough study (Heseltine, et al., 2008) which analyze a variety of surface feature maps such horizontal/vertical gradients and curvatures in addition to depth, within the Fishersurface technique. A combination of multi-feature subspace components was extracted empirically on a validation set and the tests for this system prove that better results can be achieved compared to each feature individually.

Comparing Point Clouds

Point cloud representation of faces is not very easy to work with and the performance of recognition algorithms based on this representation highly depends on the registration step. The faces are required to be very well-aligned to measure the distance between them correctly.

(Achermann, et al., 2000) propose to utilize Hausdorff metric to calculate the distance between two facial point clouds by discretizing them into voxels. The Hausdorff distance is popular in computer vision applications since it is insensitive to small perturbations and tolerates for small positional errors. In (Pan, et al., 2003), Hausdorff distance is adopted as the closeness evaluation function for both fine alignment and 3D model comparison. (Koudelka, et al., 2005) employ Hausdorff distance as a pre-screener that selects most likely matches from a large gallery for further consideration.

Alternatively, point-set differences between two facial point clouds can be computed as a similarity metric. However, this requires a dense correspondence between surfaces. As

Figure 6: The average surface (left) and first five Fishersurfaces (right), in which lighter and darker regions indicate greater positive and negative variance, respectively, while mid-grey pixels indicates zero variance (Heseltine, et al., 2008).

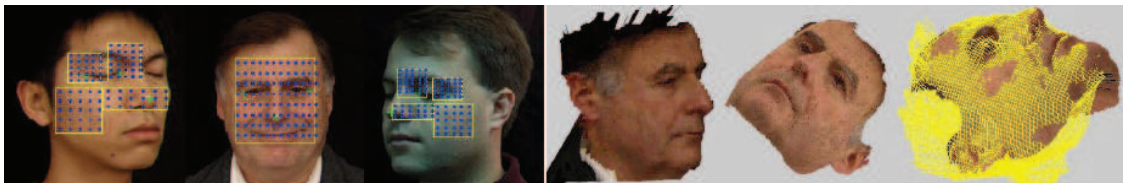


Figure 7: (Left) Control point (~100) selection for a left profile, frontal and right profile scan (Right) Matching result after fine alignment of the 2.5D test scan (wireframe) to the 3D model (textured) (Lu, et al., 2004).

mentioned in the previous section, ICP and TPS are two common techniques for registration and they are also widely used in point-cloud based recognizers.

In (Medioni, et al., 2003), ICP algorithm is used to align and extract the minimum distance between two faces for comparison. Similarly, in (Lu, et al., 2004) an ICP-based registration process is proposed to generate a similarity metric. Later in (Lu, et al., 2005), this surface matching system is integrated with an appearance-based component. In another example which combines 3D and texture information, (Papatheodorou, et al., 2004) the ICP algorithm is extended to 4D by including the intensity of each point. By minimizing the 4D Euclidean distances between points, the optimal rotation and translation is estimated.

(Pan, et al., 2005) combine facial profile and surface matching for 3D recognition. Again, ICP is utilized for registration of surfaces, but different from previously mentioned methods, a point-based statistical discriminative model is constructed for each subject in a similar fashion to Fisher's linear discriminant and a weighted distance is calculated as the matching score.

In (Wang, et al., 2006), partial ICP is utilized to match two surfaces for which the worst n% of the point pairs are rejected during computations for implicit and dynamic extraction of rigid parts of the facial surface.

In some studies, researchers investigate the point distances obtained by ICP to further discriminate inter-class and intra-class variations. (Maurer, et al., 2005) register faces to a reference mesh and compute a difference map which captures the distances between the two point clouds. Rather than considering the entire distance image, they compute statistics associated with it to be classified as correct or false match. Similarly in (Cook, et al., 2004), Gaussian Mixture Models is used to statistically model the distance distributions and differentiate between intra- and extra-personal comparison.

Among TPS-based approaches we can cite the work of (Irfanoglu, et al., 2004) in which after ICP registration and automatic landmarking of sparse points, point-to-point dense correspondence is established with TPS and point set difference is computed as the sum of the Euclidean distances between corresponding points.

Furthermore, in (Lu, et al., 2006), in addition to rigid matching distance calculated by using ICP, a non-rigid registration difference is measured after aligning faces via TPS. It is defined as the root mean square distance between deformed point set after TPS and its

counterpart in the other surface. Two matching distances are then combined according to the sum rule. This fusion is shown to improve matching accuracies, especially in the presence of facial expressions.

Feature-based Recognition

Holistic algorithms mentioned before such as point-set differences or PCA on depth-maps use global features and they are limited by normalization/registration requirements and sensitive to distorted or missing data. Accordingly, numerous feature-based approaches for 3D face recognition have been developed.

In (Chua, et al., 2000), authors present a 3D face recognition algorithm based on point signatures. After being aligned using point signature matching and ICP, expression-robust regions are extracted and used to create a model library. For recognition, the similarity with a test model is first assessed by number of matching point signatures and final decision is based on the ICP registration error. This work is later extended by integrating texture information in (Wang, et al., 2001).

In (Mian, et al., 2007), a feature-based algorithm is proposed based on repeatedly identifiable person-specific keypoints and pose invariant features extracted using a local coordinate basis at each of them. Recognition is achieved via a graph based feature matching algorithm. (Gupta, et al., 2007) employ 300 3D Euclidean and 300 geodesic distances between all pairs of 25 anthropometric facial fiducial points as features and reduce their dimension using LDA. Those features are shown to perform significantly better than the same approach with arbitrarily selected regularly spaced points.

(Maes, et al., 2010) adapt Scale Invariant Feature Transform (SIFT) algorithm for 3D faces to be utilized in pose normalization and recognition. The extracted keypoints are matched according to their local descriptors using the number of matches – Bag of Features approach. On the other hand, (Li, et al., 2011) present a novel local shape descriptor to be utilized in a SIFT-like matching process. They detect salient points and calculate weighted statistical distributions of multiple order surface differential quantities within their neighborhood. Finally, surface matching is achieved by comparing the set of local descriptors.

(Wang, et al., 2010) use three kinds of local features extracted from Signed Shape Difference Maps (SSDM) computed between two aligned 3D faces for shape comparison: Haar-like features, Gabor features and LBP. The most discriminative features are selected by boosting, three Collective Shape Difference Classifiers (CSDC) are trained and their results are fused.

Feature-based approaches are often employed to handle cases where just some parts of the 3D face models are available. In (Berretti, et al., 2011), SIFT keypoints are extracted from the depth image and the facial curves connecting pairs of keypoints are matched to determine the similarity together with the keypoints themselves. Similarly, in (Smeets, et al., 2011), features are located and described using SIFT. Additionally, facial symmetry is

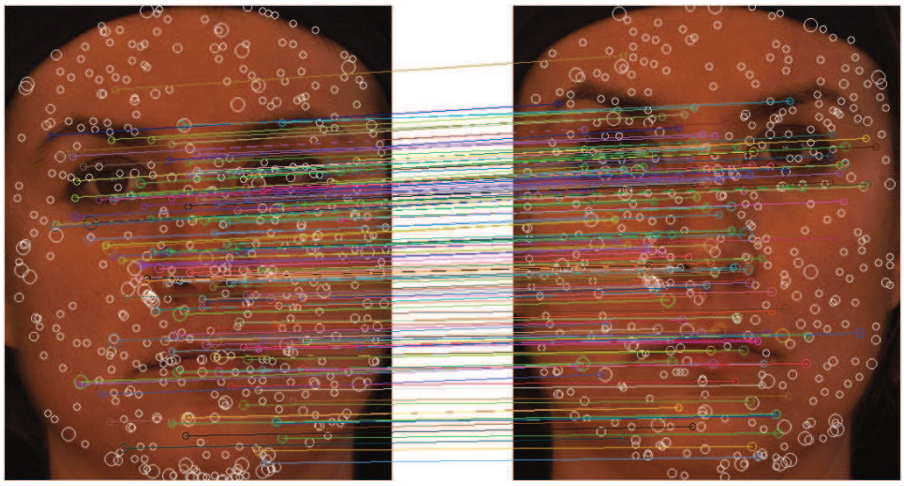


Figure 8: Two pictures of textures belonging to 3D face scans from the same person with detected meshSIFT points and detected matches (Maes, et al., 2010).

exploited to increase the overlap between two face scans by mirroring one of them. The experimental results demonstrate the accuracy of the approach in case of missing data.

4.1.4. Robustness against Facial Surface Deformations

3D shape data is advantageous to texture data being invariant to head pose and illumination changes. However, there are other factors that can also change the 3D facial surface such as expressions, occlusion and aging.

Also mentioned in the previous sections, the first kind of methods that seek robustness against such variations utilizes the geometric information that is invariant to isometric deformations like bending, particularly geodesic distances. This approach can be analyzed in two different classes. While the first class uses geodesic measurements to obtain an expression-invariant representation, the second class tries and extracts discriminatory features based on geodesic distances.

Among the first class, we can cite (Mpiperis, et al., 2007) in which a geodesic polar parameterization of the facial surface is presented. In this way, an expression-invariant representation is claimed to be used successfully for 3D face recognition in the presence of expressions, proving consistency with the assumption that intrinsic surface attributes, namely geodesics, do not change under isometric deformations.

Likewise in (Bronstein, et al., 2007), facial expressions are assumed as isometries of the facial surface and intrinsic metric structure of the face is represented in its canonical form which is approximately invariant to facial expressions.

(Berretti, et al., 2006) obtain a compact expression-invariant representation face by extracting isogeodesic stripes from the facial surface and structuring them in the form of an attributed relational graph. In a more recent work from the same group (Berretti, et al., 2010); an extensive and comparative performance evaluation is presented.

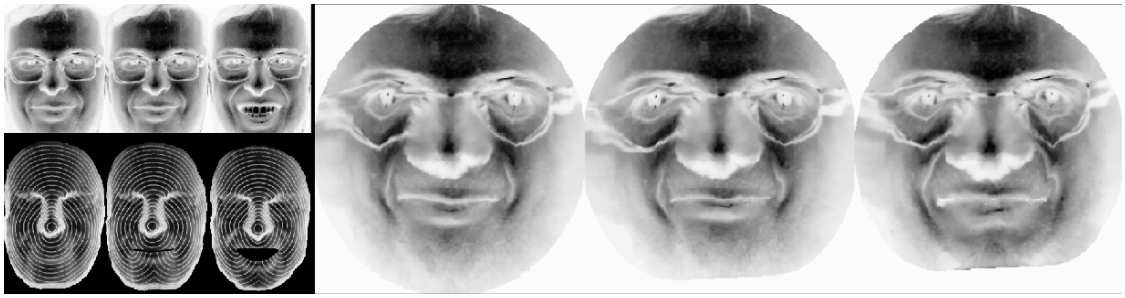


Figure 9: (Left) Original images and surfaces for which the geodesic circles are given. (Right) The obtained geodesic images (Mpiperis, et al., 2007)

Among the second class, (Li, et al., 2007) regularly remesh facial surfaces and extract multiple intrinsic geometric attributes (angles, geodesic distances, curvatures, etc.). Expression-insensitivity is further improved by element-wise weights obtained through training which give higher importance to stable elements. Previously mentioned as a feature-based method, work of (Gupta, et al., 2007) can also be included here.

The main advantage of this first kind of methods is that one sample per subject is sufficient to cope with a wide range of expressions. Nonetheless, in the case of strong expressions where the facial surface is not only bent but also stretched or split (like open mouth), the assumption of invariance for geodesics does not hold.

In addition to approaches mentioned above, some researchers propose to analyze facial deformation in order to understand and handle the non-rigid variations on facial surface. In this way, they aim to learn or model the non-rigid behavior of expressions and either to discriminate or to eliminate its effects, relying on the assumption that expressions are transferrable from one individual or group of individuals to another.

In (Lu, et al., 2005), rigid and non-rigid registration of the face models are achieved via ICP and TPS, respectively. For non-rigid registration, three point sets are defined: Two control point sets for TPS and a validation set to calculate alignment error. After the two registration procedures are applied, displacement vectors for all validation points are extracted and classified as intra or inter-subject using a nonlinear classifier (SVM).

In (Mpiperis, et al., 2008), a symmetric bilinear model is utilized to decouple the impact of identity and expression on the facial surface. Bilinear decomposition is applied on the extracted parameters by fitting a deformable model to each face. Finally, the probe and galley images are forced to have a common expression using the computed expression control vector and then matched. The authors extend their work by adopting an asymmetrical model to be able to capture wider range of expressions in (Mpiperis, et al., 2008) and test their approach also for recognition of expressions.

Similarly, (Li, et al., 2010) also attempt to differentiate the expression deformations from the interpersonal differences. A normal face is deformed to fit target faces and the optimized parameters that minimize the matching error are regarded as feature vectors. For training, feature vectors are extracted from face model pairs of the same identity but

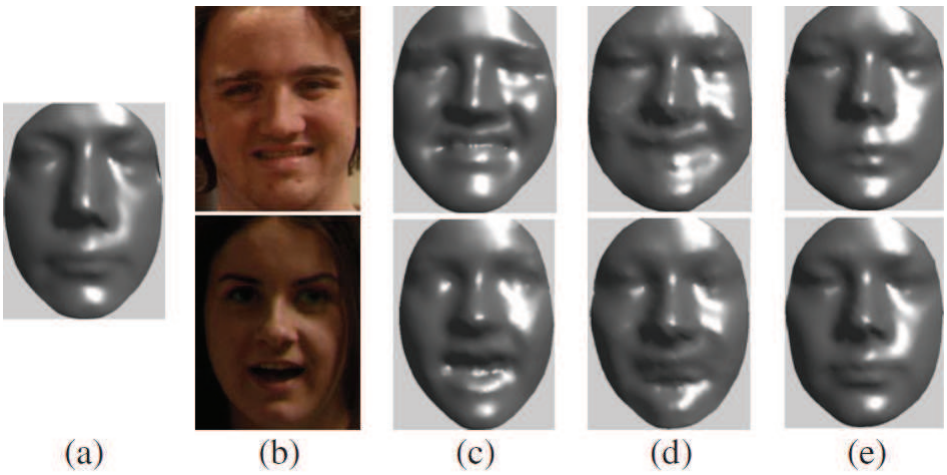


Figure 10: Examples of the face deformation. (a) The normal face. (b) 2D face. (c) Corresponding 3D face. (d) Synthesized face with expression. (e) Synthesized normalized face with expression being removed (Li, et al., 2010).

with different expressions and pairs of different identities but with no expression (neutral) as intra-class and inter-class deformations, respectively. These deformations are later linearly combined to synthesize a probe face with certain expression and inter-class coefficients are utilized for recognition.

In (Wang, et al., 2007), a guidance-based constraint deformation model (GCD) is presented which is utilized to cope with the expression variations by deforming a non-neutral face model towards its neutral state under certain constraints. With a similar aim to "neutralize" the probe face models, (Amberg, et al., 2008) differentiate two possible deformations by fitting an identity/ expression separated 3DMM and (Pan, et al., 2010) model expression residue from samples and then use the inferred expression residue from the expressional test scans to recover their neutral version. Lastly, in (Al-Osaimi, et al., 2008), expression deformations are learned in a PCA subspace built from shape residues of training data. In the recognition stage, the residues extracted in the same way as the training set, are minimized using PCA subspace and utilized for similarity score computation.

Exceptionally in (Lu, et al., 2006), instead of removing the present expression in the probe, the authors try to infuse it on the whole gallery.

One big disadvantage of this second kind of methods is that they require additional data for training. No matter how large the utilized training set is, still the range of possible expressions are so wide that full comprehension is difficult to achieve. Additionally, the validity of transferring expressions between individuals is far from being indisputable.

One last type of deformation-robust approaches that we will cite here is the region-based methods which analyze the 3D face by dividing it into multiple components and applying fusion at score-, rank- or abstract-level. Robustness against facial deformations is

achieved by simply concentrating on regions that change the least with varying expressions and/or occlusions. Yet, numerous different techniques have been proposed.

The first class of such methods divides facial surface into multiple regions, often for better alignment and fuse the regional match scores equally. For instance in (Alyuz, et al., 2008), average regional models are constructed and utilized for region-based alignment via ICP. The recognition performances of individual face regions are reported together with combined performances using various fusion methods.

In (Faltemier, et al., 2008), a committee of small regions are independently matched using ICP and the match scores are combined one by one, in the order of individual recognition performances by different fusion techniques. Accordingly, 28 of those regions are selected and fused with a modified version of the Borda count method.

Similarly, (Spreeuwers, 2011) presents a 3D face recognition system based on the fusion of 30 dependent region classifiers that use PCA-LDA for feature extraction and the likelihood ratio as the similarity score. The obtained regional scores are combined by majority voting.

The second class of methods among approaches of this type relies on the a priori assumption that some parts are less affected by expressions, like nose or eye regions. In (Chang, et al., 2005), Adaptive Rigid Multi-region Selection algorithm is introduced which matches both automatically and manually extracted multiple facial regions that are focused around nose. Comparative results with PCA and ICP based algorithms show better performance against expression variations.

In (Cook, et al., 2006), after applying 18 Log-Gabor filters, partial range images (excluding mouth region) is decomposed into 25 overlapping regions by 3 scales of filter, resulting in 75 semi-independent observations. PCA is applied for dimension reduction and Mahalanobis Cosine distance is adopted as the similarity metric. Finally, linear Support Vector Machines trained to maximize the margin between the client and impostor scores are utilized to combine all scores into a single value.

In (Mian, et al., 2005), the effects of facial expressions and facial hair are avoided by performing recognition only on the basis of eyes-forehead and nose regions. Those two regions are aligned using ICP and the final matching distances are fused using the product rule. Later in (Mian, et al., 2007), the authors additionally incorporate the texture information and implement a rejection classifier to eliminate a large number of candidates using a SIFT-based approach.

(Faltemier, et al., 2006) sample 7 regions around the nose for matching by ICP and the obtained match scores are combined with empirically determined regional thresholds which allow the regions to vote and maximize the verification rate. With the addition of the minimum root mean square distance value among all regions, 8 matching scores are fused by voting. (Smeets, et al., 2010) fuse their above-mentioned isometric deformation modeling approach with an ICP-based system applied on more rigid parts of the face (again around nose).

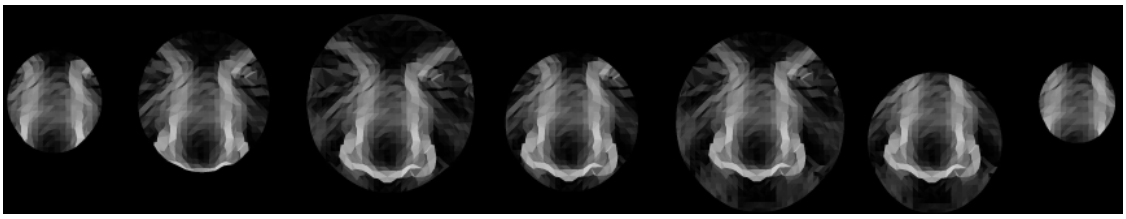


Figure 11: Example facial regions selected to be matched via ICP (Faltemier, et al., 2006)

(Queirolo, et al., 2010) analyze face in four regions: circular and elliptical regions around nose, forehead and the entire face. They utilize Surface Interpenetration Measure (SIM) for similarity metric and report both individual and combined performances.

In the third and last class of region-based approaches, the discriminatory power of each facial region is assessed and utilized in the fusion stage. For example in (Gunlu, et al., 2010), the features independently extracted from each sub-region using DCT and each region is assigned as the proportion of inter-variance to intra-variance of the depth of each pixel in the range map.

In (McKeon, et al., 2010), 3D Fisherface approach is employed for an ensemble of 22 different regions. Sequential Forward Search (SFS) is performed to select the most discriminating regions.

In (Ben Amor, et al., 2008) a study on the deformability and elasticity of facial regions is presented based on an anatomical analysis. Measuring facial deformations on five people, different regions are sorted by their rigidity. Finally, regional match scores obtained by applying ICP are fused using the empirically measured rigidity levels as weights.

Lastly, in (Li, et al., 2011), an LBP-based approach is presented for which the 3D facial surfaces are divided into patches and the patch weights are learned based on individual rank-one recognition rates.

4.2. 3D Shape Assisted 2D Face Recognition

Possible solutions for illumination and view point invariance in 2D face recognition are limited due to the 3D nature of the problem. By incorporating the 3D shape data of the face, researchers aim to improve 2D recognition performances in the presence of such variations. On the other hand, acquisition of facial models using 3D scanners can be problematic in the operational mode, especially under the scenario of uncooperative persons to be recognized. Mainly due to these two factors, the idea of 3D shape assisted 2D face recognition emerged, for which 3D acquisition can be realized asymmetrically only during the enrollment phase or 3D shape can be reconstructed based on the captured 2D images.

As one of the earliest works, (Chang, et al., 1997) gives an interesting approach to the use of 3D and 2D images. Taking only range maps as the gallery images, facial surface normals are decomposed into three components and their weighted sums are utilized to

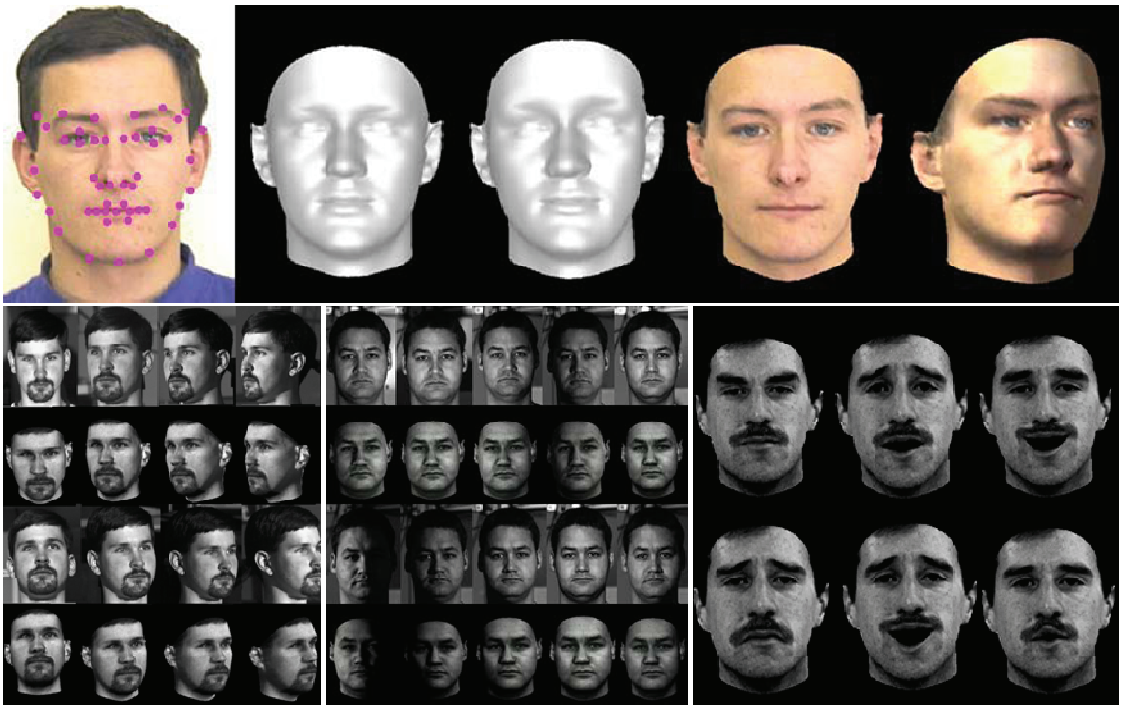


Figure 12: An example study (Hu, et al., 2004) is illustrated. Detected feature points and 3D reconstruction result is given in the top row. At the bottom, the synthesized pose and illumination images are compared with their originals and some generated expressions are presented.

simulate 2D intensity images with different illuminations. The proposed recognition is reported to be less sensitive to the illumination variance. Later in (Huang, et al., 2010), a similar approach is adopted in the sense of 3D-2D face matching. Exploiting the Canonical Correlation Analysis (CCA) to learn the mapping between range and texture LBP faces, similarity scores are computed and fused with the pure 2D counterpart.

In many approaches of this kind, the 3D face model is utilized to generate synthetic images of the enrolled subjects with different poses, expressions or under different illumination conditions, in order to span their all possible appearances. The probe images in 2D are then compared with the synthetically augmented gallery for better match.

(Lu, et al., 2006) develop a multi-modal system in which 3D models in the gallery are used to synthesize new appearance samples with pose and illumination variations for discriminant subspace analysis in order to improve the performance of the 2D modality. In a similar way, again a multi-model system is proposed by (Tsalakanidou, et al., 2004) which generates synthetic views depicting various head poses and illumination conditions and uses them to train Embedded Hidden Markov Models.

Methods based-on 2D images as their unique modality try and extract the 3D data from intensity. (Zhao, et al., 2000) propose a shape-from-shading (SFS) based method to generate synthetic facial images under different rotations and illuminations. In (Hu, et al., 2004), a personalized 3D face is reconstructed from a single frontal face image with neutral expression and normal illumination using 83 automatically located feature points

and different images are synthesized with variant pose, illumination and expression. A similar scheme is also presented by (Lu, et al., 2004) where a 3D generic face model is aligned onto a given frontal image using 115 feature vertices.

In an earlier study, (Lee, et al., 2003) present a combination of an edge model and color region model for face recognition after synthetic images with varying pose are created via a deformable 3D model. In (Huang, et al., 2003), a 3D morphable model is used to generate 3D face models from three input images of the enrolled subjects. Similar to previously mentioned studies, the generated 3D models are utilized to augment the 2D training set with synthetic images rendered under varying pose and illumination conditions.

Lately, a study in which 3D model reconstruction is achieved by applying the 3D Generic Elastic Model approach is published by (Prabhu, et al., 2011). Instead of enlarging the training set, they choose to estimate the pose of the test query and render the constructed 3D models at different poses within a limited search space about the estimated pose. In another recent study, (Park, et al., 2010) address the problem of facial aging by using a 3D age modeling technique to compensate for age variations. A shape and texture aging pattern is simulated on a given 2D image via reduced morphable model-based 3D model fitting process.

Another type of methodology in the reverse direction tries to remove variations in the probe images by pose normalization and/or illumination compensation. (Kittler, et al., 2005) propose a system in which 3D models of the users are acquired during enrollment and utilized for illumination correction by separating the effect of light and albedo in the 2D test images.

Similarly, in (Malassiotis, et al., 2004), a multi-modal system is presented for which the 3D component of the face is utilized to compensate pose and illumination in its 2D counterpart. After estimating the pose using the nose tip and the facial symmetry line in the range image, it is normalized to be frontal. Next, scene illumination is recovered from the depth and color image pair and the input image is re-lighted. Two compensations are reported to improve the 2D face recognition performance.

Differently from these mentioned methods, one last class of approaches relies on 3D morphable models (3DMM) and uses 3D data as an intermediate step for 2D image matching. Again a generic 3D face model is morphed to match 2D images, but instead of synthesizing new facial images under different conditions, model coefficients obtained after fitting are used for recognition as feature vectors to be compared.

In (Blanz, et al., 2003), a morphable model that is based on a vector space representation of faces is constructed from 3D scans of 100 males and 100 females. After dense correspondences are established between the scans, Principal Component Analysis (PCA) is performed on the shape and texture vectors resulting in two orthogonal bases formed by 90 eigenvectors. During the fitting process, the shape and texture coefficients together with illumination parameters are estimated iteratively to bring the morphable

model as close as possible to the query image. Finally, two faces are compared by the set of coefficients that represent shape and texture using Mahalanobis distance.

In (Amberg, et al., 2007), an optimal step non-rigid ICP registration is proposed to replace the general optimization and this approach is proven to be more robust against missing data. More recently, (Bustard, et al., 2010) suggest extending 3DMM approach to ears which are mostly excluded due to being noisy.

Lastly, numerous surveys on 3D (assisted) face recognition exist that can be referred for further information, such as ((Bowyer, et al., 2004), (Kittler, et al., 2005), (Akarun, et al., 2005), (Scheenstra, et al., 2005), (Bowyer, et al., 2006), (Abate, et al., 2007) and (Smeets, et al., 2010)).

5. Challenges and Trends

In the final section of this chapter, we will discuss current challenges to 3D face recognition and trends in the related research directions.

The utilization of 3D data in the field of face recognition is obviously proven very successful and useful. On the other hand, it still bears some problematic sides, most crucial of them being 3D sensors.

Shortcomings of 3D Sensing Technology

The quality of 3D scanners has improved a lot today but their employment for face recognition or surveillance applications is not very straightforward. Mainly 6 properties are listed to make practical applications possible (Bowyer, et al., 2006):

- A small acquisition time to minimize artifact due to subject motion and to require less cooperation
- A large depth of field with stable depth accuracy
- Robust operation under "a range of normal conditions"
- Harmlessness for eyes
- Dense depth sampling
- High depth resolution

For the time being, there exist numerous tradeoffs between these items.

First tradeoff is between acquisition quality and time. Passive stereo systems have the advantage of capturing the 3D scene in shorter times comparable to 2D cameras, but their sampling rate and depth accuracy is mostly not very sufficient. Whereas, active sensors require longer acquisition time but they can output more accurate results.

Second tradeoff is between harmlessness for eyes and robust depth of field for active sensors. Coherent lights like lasers are smaller and brighter compared to non-coherent

ones. A coherent light beam can be directed and scanned with precision but it bears potential eye safety issues. On the other hand, non-coherent projections cannot produce comparable acquisition accuracy and depth of field.

Hybrid passive stereo and structured light systems also decrease the time required for acquisition, but they often have troubles in regions with facial hair. Additionally, as it is the case with all systems that rely on stereo correspondence, they have sparse sampling of points.

One last important point about 2D sensors is their sensitivity to illumination. Albeit the intrinsic illumination independence of 3D data, the acquisition can be greatly affected by the illumination conditions. For stereo-based 3D sensors which take one or more standard 2d images, light levels of too high or too low intensity can degrade acquisition quality.

3D Face Datasets

Availability of standard and public datasets is an important aspect for researchers to be able to benchmark their algorithms and compare their results. The early studies on 3D face recognition suffer from lack of appropriate datasets with high quality face models of sufficient number and demographic variety of people. But today, many 3D facial databases exist with different specifications and size that we will briefly catalog chronologically.

- 3D-RMA Database: Published in 1998 (DB01, 1998), this earliest example of 3D face dataset includes facial models of 120 people acquired by a structured light system in two sessions, with different orientations of the head. The texture images are not made available for privacy reasons.

- York 3D Face Database: A wide variety of conditions, such as pose, expression and occlusion is captured in this database which consists of over 5000 3D face models of approximately 350 people (DB02, 2003).

- GavabDB: Consisting of 3D images of 61 different individuals, GavabDB (DB03, 2004) provides systematic variations of pose and facial expression. 9 images per person are captured: 2 frontal views and neutral expression, 2 x-rotated (±30º, looking up and down) views with neutral expression, 2 y-rotated (±90º, left and right profiles) views with neutral expression and 3 frontal gesture images, laugh, smile and a random gesture chosen by the user.

- FRGC v1: This database contains over 900 2D and 3D images of 275 persons with frontal view and no expressions (DB04, 2004).

- FRGC v2: FRGC v2 contains over 4000 frontal scans of 466 persons with different expressions. An advantage of this database which was published within a challenge is that a well-organized infrastructure is also provided by the Biometric Experimentation Environment (BEE), an XML based framework to describe and document computational experiments (DB04, 2004).

- BJUT-3D: BJUT-3D (DB05, 2005) is a 3D face database including 500 Chinese subjects (250 male and 250 female). The acquisitions are done with 4 different expressions and no accessories like glasses. 3D face models are acquired by the CyberWare 3D scanner and preprocessed (holes filled, the facial surface smoothed and cropped).

- FRAV3D: This database (DB06, 2006) contains multiple scans of 105 volunteers (81 men/24 women) obtained by Minolta Vivid-700 red laser light-stripe triangulation range finder. Each scan is different from the rest in the sense of expression variations, illumination and viewing conditions.

- CASIA 3D Face Database: This database (DB07, 2007) consists of 4324 scans of 123 persons scanned by a non-contact 3D digitizer Minolta Vivid 910. Pose, expression and illumination variations are captured both separately and in combinations.

- Bosphorus: Intended for research on 3D and 2D human face processing tasks such as face/expression recognition, facial action detection and 3D face reconstruction, Bosphorus database (DB08, 2008) has a rich repertoire of expression (up to 35 different expressions per subject), pose (13 rotations) and occlusion variations. Additionally, it includes 24 labeled facial landmarks per sample.

- Texas 3DFRD: The Texas 3D Face Recognition (DB09, 2010) database of 1149 2D and 3D facial images from 105 adult subjects is constructed using a stereo-based system. Both raw and preprocessed versions of the images are provided together with locations of 25 anthropometric facial fiducial points.

- NTU-CSP: Acquired using Minolta Vivid 910, 3D scans of 80 different persons are captured under different facial expressions and head pose variations and preprocessed (noise and outliers removal). 16 samples per person are collected: 4 frontal views with neutral expression, 2 frontal views with random expression, 2 upward views, 2 downward views, 2 rightward views and 2 leftward views (DB10, 2010).

- UMB: The University of Milano Bicocca (UMB) 3D face database is a collection of 3D and 2D images of 143 subjects (98 men, 45 women) with 4 facial expressions (neutral, smiling, bored, angry) and various occlusions (such as hair, eyeglasses, scarves, etc.). For the samples with occlusions, the masks representing the visible part of the faces are also available (DB11, 2011).

- 3D TEC: 3D Twins Expression Challenge (3D TEC) dataset consists of 3D face models of 107 pairs of identical twins with one neutral and one smiling expression (DB12, 2011). The face scans were acquired by Minolta Vivid 910. More details about the dataset can be found in (Phillips, et al., 2011)

Table 1: The general characteristics of the aforementioned databases in chronological order.

	Database	Number of subjects	Number of samples per subject	Total number of samples	Available variations	Number of available landmarks	Texture
1998	3D RMA	120	6	720	Pose	-	No
2003	York	~350	15	~5000	Pose, expression, distance, occlusion	-	Yes
2004	GavabDB	61	9	549	Pose, expression	-	No
2004	FRGC v1	275	1-8	943	-	-	Yes
2004	FRGC v2	466	1-22	4007	Expression	-	Yes
2005	BJUT-3D	500	4	2000	Expression	-	Yes
2006	FRAV 3D	106	16	1696	Pose, expression, illumination	-	Yes
2007	CASIA	123	37-38	4624	Pose, expression, illumination	-	Yes
2008	Bosphorus	105	31-54	4666	Pose, expression, occlusion	24	Yes
2010	Texas 3D	105	1-72	1149	Expression	25	Yes
2010	NTU-CSP	80	16	1280	Pose, expression	-	Yes
2011	UMB	143	~10	1473	Expression, occlusion	7	Yes
2011	3D TEC	214	2	428	Expression	-	Yes

Trends

Although expression and occlusion variations are studied to a great extent, these issues are still very far away from being closed. 3D face recognition algorithms need to be more tolerant of real-world scenarios that in addition to expressions also include many other factors like pose variations, facial hair and accessories.

Even though the 3D shape of the face is invariant of the viewing angle, this problem can be interpreted as missing data challenge since some parts of the probe face may not be available due to self-occlusion. Additionally, the impact of facial accessories like eyeglasses or lip gloss could not yet arouse the interest as strongly as facial expressions or pose variations.

Moreover, new challenges for 3D face recognition are emerging each day with the most recent ones being age variations and identical twin phenomenon. Especially with the introduction of the new 3D TEC dataset, it is reasonable to expect an orientation towards research on identical twin-robust recognition. On the other hand, researching on age-invariant 3D face recognition still carries the problem of database absence. 3DMM's are proposed to model aging for robustness agains age variations by (Park, et al., 2010) in the context of 3D assisted 2D face recognition.

Another challenge to be addressed is 3D recognition at a distance. Today, the solution for this issue is mostly searched focusing on the acquisition stage of the process and aim at high quality sensing at longer ranges.

Before concluding this section, Kinect deserves to be mentioned as one last point with the novelties it brought into 3D sensing. With its affordable price, reliability and speed, it has already become an important 3D sensor for a wide range of applications from computer vision to HCI and face recognition domain has no reason to be any different.

6. Conclusion

In this chapter, we presented an in-depth description for the current state of 3D face recognition. Starting with a short introduction to general biometrics, 3D imaging techniques and 3D surface representation methods were described in detail. A complete survey on the place of 3D data in face recognition was given which reviews major approaches in the related fields. Finally, possible challenges and trends in research directions were discussed

Utilization of 3D data for recognizing human faces has drawn a strong and constant interest for a long time now and there is no doubt that it will maintain its popularity in the future.

Chapter III. Pre-processing and Automatic Landmarking for 3D Face

1. Introduction

As previously discussed in Section 4.1.1, outputs of 3D sensors mostly need to be pre-processed before being utilized for recognition. The raw scans often include non-facial parts like shoulders or clothing and are imperfect with spikes, holes and noisy surface. An example raw scan is given in **Figure 13**.

Furthermore, automatic landmarking is also included in this chapter, since many cropping and alignment techniques require on one or more facial feature points, like eye corners, nose tip, lip corners etc. Hence, it can be regarded as a part of the preprocessing stage.

In the following sections, we firstly present the pre-processing approach that will be used throughout the rest of the work. Next, our automatic landmarking algorithm is introduced in detail for which experimental results are also provided.

2. Preprocessing

The erroneous 3D scanner outputs are preprocessed mainly for three purposes:

- To extract the face region
- To eliminate spikes and holes on the facial mesh
- To smooth the 3D surface

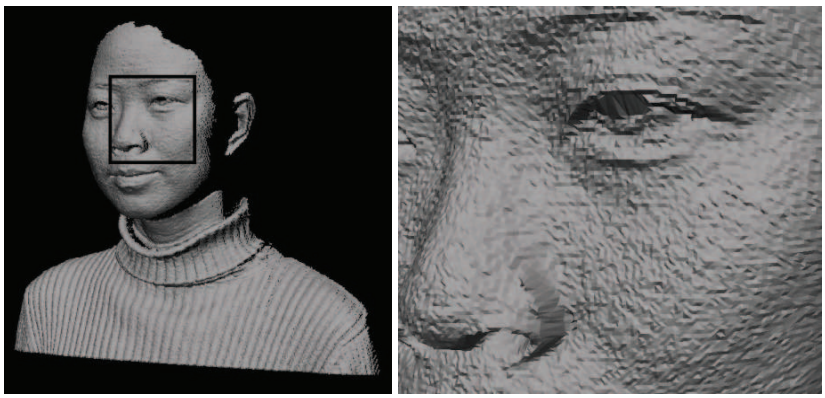


Figure 13: Sample raw scan from FRGC v1 database and a detail

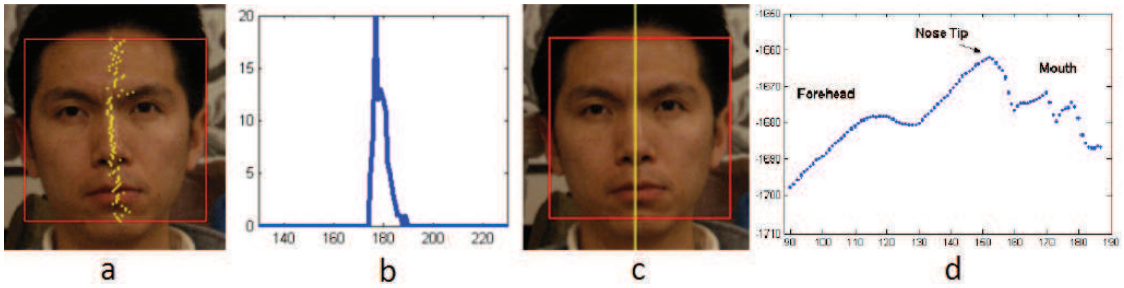


Figure 14: (a) The positions of maximum z value for each row is marked with yellow. (b) The histogram created by counting marked columns (c) The detected vertical midline (d) Nose tip is selected as the peak of the vertical profile (Lu, et al., 2005).

Each of these steps will be explained in detail in the following subsections.

2.1. Facial Region Cropping

In order to crop the facial region, first we coarsely locate the nose tip as proposed in (Lu, et al., 2005). In the range map, for each row, the position with the maximum z value is found. Then for each column, the number of these positions is counted and a histogram is created. The peak of this histogram is chosen as the column for the position of the vertical midline, and the maximum point of this contour is identified as the nose tip. This method is illustrated in **Figure 14**.

Finally, using a sphere of radius 100mm and centered 10mm away from the nose tip in +z direction, the facial surface is cropped (see **Figure 2**).

2.2. Removing Spikes and Holes

In order to be removed, spikes have to be detected first. To this end, we analyze the neighboring pixels to decide whether a pixel value in the range map indicates a spike or not. The decision mechanism is rather simple. If the pixel value differs largely from more than the half of its valid neighbors in 5x5 kernels, it is treated as a spike and its value is changed to the mean of these neighboring pixels.

For filling the holes, we employ a linear interpolation based method. The missing data is interpolated by fitting a straight line between the neighboring vertices with known x, y and z coordinates. Once the cropped and cleaned facial surface is obtained, denoising (also referred as smoothing) is performed.

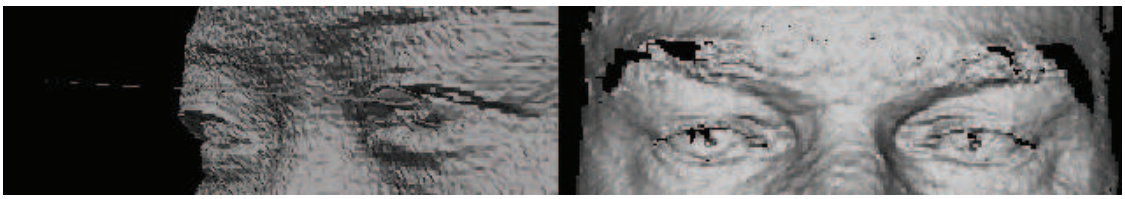


Figure 15: Examples of spikes and holes that occur more frequently around eye and eyebrow regions

2.3. Smoothing

Even with most reliable scanners, surface noise is inevitable on the acquired 3D models and it must be removed before processing further. However, while estimating the noise-free models, in addition to smoothness, another concern is to preserve the features in the surface so that the shape descriptiveness is unharmed for further applications.

If no preventive measure is taken, details of the facial surface (e.g. sharp features like corners) are often blurred. In order to avoid this, a non-iterative feature-preserving filtering technique – bilateral filtering is applied with some modifications. Proven to be a powerful yet simple scheme for edge preserving smoothing in 2D, this technique has also been successfully extended to 3D by (Jones, et al., 2003).

Bilateral filter is distinguished from Gaussian blur by its additional "range" term. In this case, not only the distances between positions matter, but also the variation of the intensities are also taken into account in order to keep the features from derogation. Consequently, the edges where high intensity differences occur are preserved.

In 2D, the intensity values are a function of position values, whereas in 3D the position in fact, is the signal itself. Hence, the modification of bilateral filter to be applied to 3D data is not very straightforward. Like the most image processing algorithms that are extended to surfaces, normal information at each point of the surface can be used to form an intensity space like in images. This approach can be formulated as follows:

$$P_e = BF(P) = \frac{1}{w_p}\sum_{s\in V} G_{\sigma_s}(||P - S||)F_{\sigma_r}(|\left|N_p - N_s\right||)S \qquad \text{Equation 1}$$

In this equation, P_e is the point to be estimated and S are the points in the voxel V, which are going to contribute to the calculation of denoised position of P. $G\sigma_s$ is the spatial weight function, which is taken as a one-dimensional Gaussian filter. Hence, the weight decreases with the spatial distance to P. $F\sigma_r$ is the range weight function, which is again a one dimensional Gaussian function that decreases the influence of pixels S which have a normal value, N_s, much different from N_p. This means, if S is very distant to P, the drastic decline in $G\sigma_s$ will attenuate the weight of S. Similarly, if the difference between the normals of S and P is high; the falloff in $F\sigma_r$ will make S less effective on the estimation. Without the range weight function, this filter simply applies a Gaussian blur. The range function is what provides us with the feature preserving behavior. Finally, W_p is the normalization factor calculated by adding all weights for each point.

Two important parameters in this formula are σ_s and σ_r, that have big influence on their corresponding weights. σ_s and σ_r are the standard deviations in space and intensity (range in 3D) difference, respectively. In this study, σ_s is taken proportional to the model size, in other words it is taken to be 2% of the largest distance that is between the points on each end of the diagonal. On the other hand, σ_r is taken proportional to normal variations throughout the point cloud. That is, it is calculated by taking the average of the values in the gradient of the normal image.

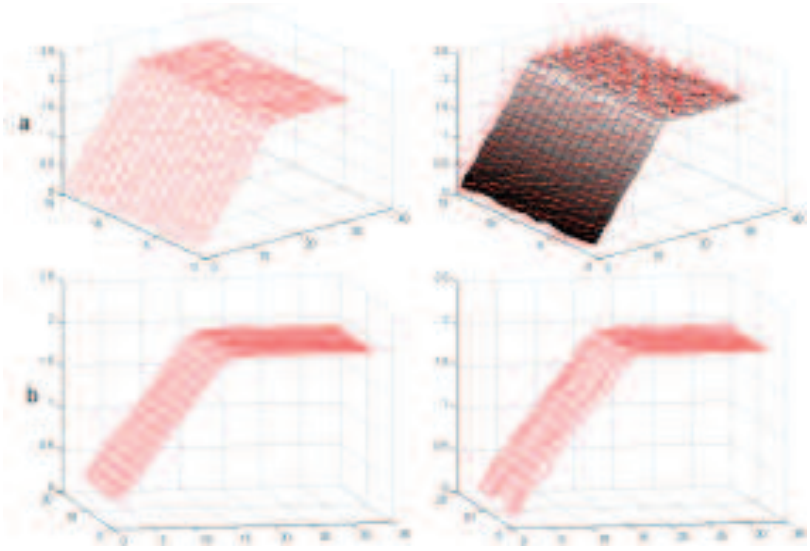


Figure 16: (a) Artificially created noisy surface and normal vectors at each vertex. (b) Surfaces smoothed with Gaussian and bilateral filters, respectively.

In **Figure 16**, the advantage of bilateral filtering is illustrated on a synthetic 3D surface with an additional random noise. The results of the proposed bilateral filtering on 3D facial data are quite satisfactory as shown in **Figure 17** where the noisy raw scan data and the same model after bilateral smoothing are given for comparison. The edge details around the eye are well preserved while the facial surface is smoothed. In our algorithm, the bilateral filtering is applied in 3x3 depth-pixel neighborhood and the normals at each point is approximated as the average normal of all neighboring triangles.

3. Automatic Landmarking

Facial feature points are one of the most important clues for many computer vision applications such as face normalization, registration and model-based human face coding. Hence, automating the extraction of these points would have a wide range of usage. In this chapter, we will present an algorithm to detect a subset of Facial Definition Parameters (FDPs) defined in MPEG-4 automatically by utilizing both 2D and 3D face data.

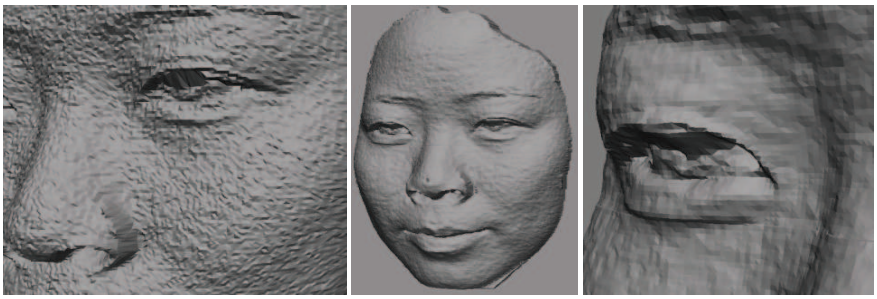


Figure 17: The noisy and the cleaned surface with a detail around eye region.

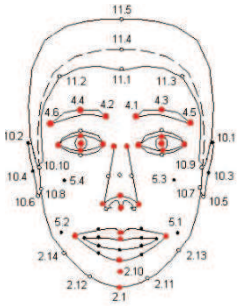


Figure 18: Facial Definition Parameters defined in MPEG-4 FA and the chosen subset to be automatically located (in red)

3.1. Introduction

Considering that subject cooperation is required during the enrollment, we base our system on the restrictive assumption of a well-controlled acquisition environment in which subjects are registered under following conditions:

- In frontal pose
- Without any occlusions and prominent facial expressions

According the next steps in our plans, we aim to extract a subset of MPEG-4 Facial Definition Parameters (FDPs) to be utilized for the alignment of the faces with the animatable generic model (see Chapter IV). For the extraction of the points, 2D and/or 3D data is used according to the distinctive information they carry in that particular facial region.

A method for detection of anchor points using 2.5D face images is proposed in (Colbry, et al., 2005), in which the decision is based on local shape characteristics with parameters trained on sample scans. This approach is improved with a more robust nose tip extraction algorithm in (Lu, et al., 2005) where additionally, the cornerness response from the intensity map is utilized to determine the positions of the corners of the eyes and the mouth. This study is extended further in (Lu, et al., 2006) by handling pose variance in 3D facial scans. It is achieved by estimating the nose tip based on a nose profile model. In a similar way, feature points are described by Gabor filter responses in the 2D domain and point signature in the 3D domain in (Wang, et al., 2001).

Integrating range and texture information for face processing is proven to be advantageous in studies like (Husken, et al., 2005) and (Lu, et al., 2005). An algorithm that combines 2D and 3D registered data to accurately detect the facial feature points is proposed in (Boehnen, et al., 2005), where firstly skin regions are segmented and eye and mouth candidates are determined by processing formerly produced eye and mouth maps.

In our approach, firstly, a facial midline (vertical profile) analysis is done and 9 fiducial points on that midline is detected. Based on that information; face is broken into sub-regions. This facilitates the coarse localization of eyebrow, eye, nose and lip. After that, further analysis is done in the extracted sub-regions to detect the points of interest. In the

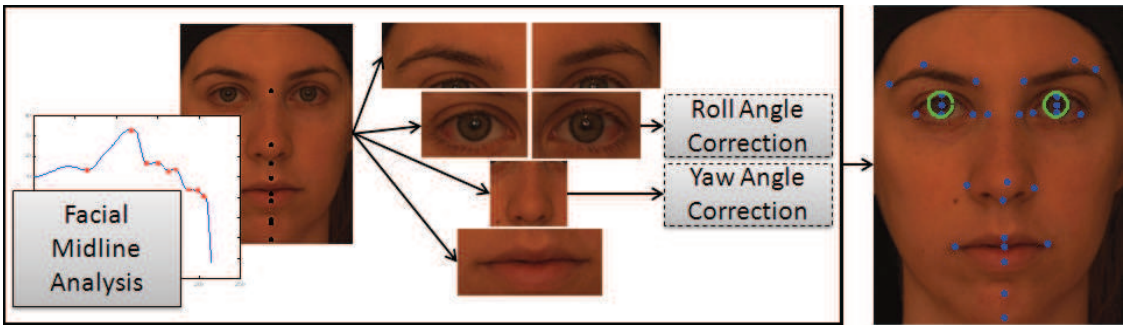


Figure 19: Flow chart for the proposed system

proposed algorithm, for those regions with non-informative texture (like nose) the 3D data is analyzed. On the other hand, for the regions with noisy surface and/or distinctive color information 2D data is utilized. As a result, 29 facial interest points are detected in total, consisting of 3 points for each eyebrow, 5 points for each eye and 6 points for the nose, 5 points for the lip and finally 2 points for the chin (**Figure 19**).

3.2. Vertical Profile Analysis

The analysis done on the vertical profile constitutes the backbone of the whole system. It starts with the extraction of the facial midline and for this purpose; the method proposed in (Lu, et al., 2005) is adopted. For each row, the position with the maximum z value is found and then for each column, the number of these positions is counted so that a histogram is created. The peak of this histogram is chosen as the column for the position of the vertical midline. The vertical profile is extracted and smoothed by mean

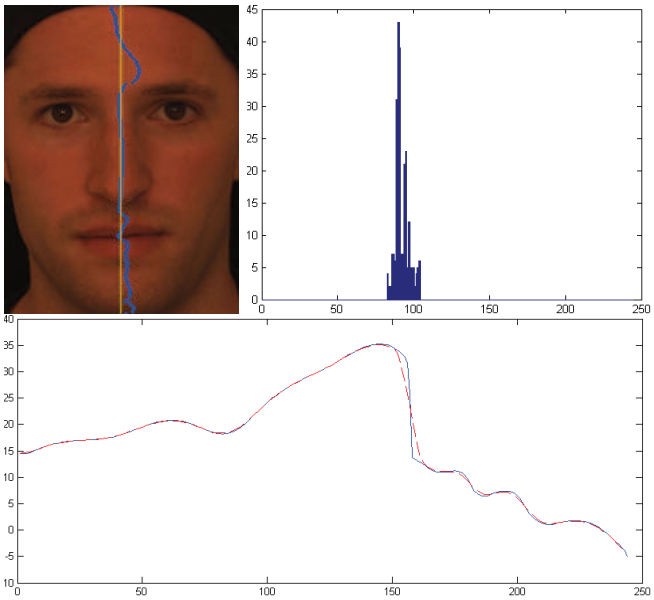


Figure 20: The positions with maximum z value for each row on the example face are marked with blue dots and the chosen column is marked with yellow. The histogram and the extracted profile before and after smoothing are given.

filtering. This procedure is illustrated in **Figure 20**.

The highest point in the profile curve is detected as the nose tip. Since this knowledge of nose tip position lets us to look for the eyes in the upper half of the face, before going any further, coarse iris detection is applied in that region.

For coarse iris extraction, the non-skin region is found by removing the pixels with the most frequent (Cb, Cr) values present in the half-image, using YCbCr space in the 2D image. Edge maps are constructed for the non-skin region using Canny edge detector by iteratively adjusting the threshold until a descriptive edge map is obtained.

Subsequently, Hough transform is applied to the edge map to detect circles. For each detected circle, an overlapping score is calculated by the ratio of the detected portion of the circle to the whole circle parameter. After grouping the detected iris candidates as right and left iris according to their positions, among the compatible pairs, the one with the maximum total overlapping score is chosen to be the two irises.

After the positions of the two iris centers are found, the 2D and 3D images of the face are rotated in order to align to iris centers on the same horizontal line. Thereby, our assumption for vertical profile is better assured.

Subsequent to the roll angle correction, the profile analysis is repeated, and this time in more detail. The vertical midline of the face is again extracted and smoothed as explained before.

On the profile curve, there are bulges (forehead, nose, upper and lower lips and chin) and bores in between. Even though, those shapes do not fully expose the location of the facial interest points, they can be highly informative. For this reason, the peaks and nadirs in the profile curve are found with the help of the zero-crossings of the curve's first and second derivatives (differences between adjacent elements of the mid-line profile).

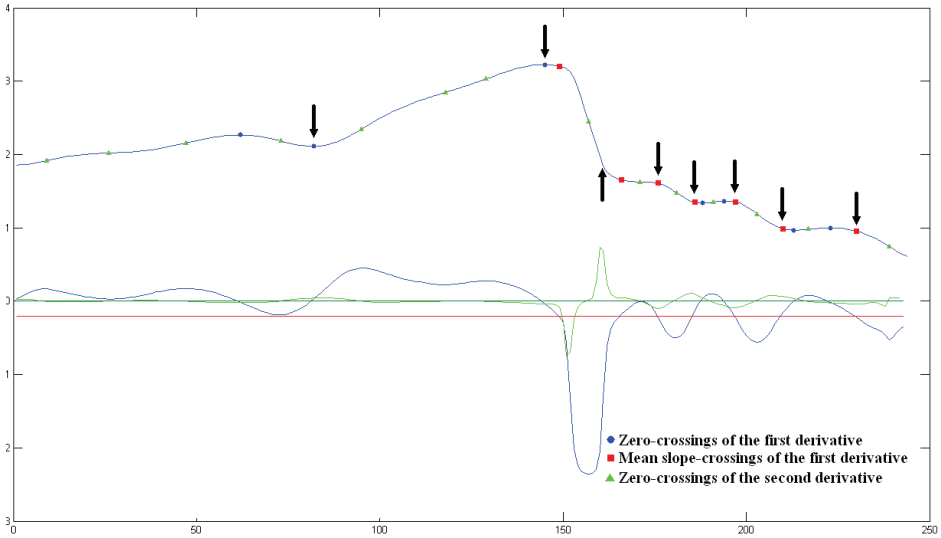


Figure 21: A sample profile curve and its first (blue) and second (green) derivative curves. The arrows show the five detected interest points among the candidates.

Firstly, the highest point on the curve is detected as nose tip. The first zero-crossing of the first derivative of the curve that is closest to the nose tip from left is labeled as the starting point of the nose (the nose bridge). Next, the point with the maximum second derivative value to the right of the nose tip is labeled as the end point of the nose (the bottom of the nose). Afterwards, the lower edge of the face is found as the point whose second derivative exceeds a threshold (0.7). Thus, three points on the nose (the nose bridge, nose tip and the bottom) and the border of the chin is detected.

For the points between the end of the nose and the border of the chin, zero-crossings for both first and second derivative are calculated. Since usually, due to the spherical shape of the facial surface, this region is inclined, the "mean slope" is calculated as the difference between the end of the nose and the bottom of the face. Assuming that slope of the curve in that region is biased as that mean value, another set of "zero-crossings" are calculated by subtracting the calculated mean slope from the first derivatives.

The bulges and bores are created by the subsequent increase and decrease in the profile curve so each peak (or nadir) is bounded by two zero-crossings of the second derivative. Considering this fact, an interest point is detected as the zero-crossing of the first derivative between two "boundaries" if it exists. Otherwise, the "mean slope-crossing" is taken into account. If both are not found, the value is calculated as the mid-point of the boundaries. As a result, points on the lips and on the chin are coarsely extracted. The curves for the profile and the first and the second derivatives, lines for zero and the mean slope and their intersection points are depicted in **Figure 21**.

Now those landmarks are known, the face can be broken into more meaningful sub-images for locating or refining the points of interest.

3.3. Extraction of Facial Interest Points

After analyzing the vertical midline of the face, regions of interest for each eye and eyebrow, for nose and lips are extracted. The method for each part will be explained in detail in the following subsections.

3.3.1. Eyes

The 3D surface around the eyes tends noisy because of the reflective properties of the sclera and the pupil and the eyelashes. On the other hand, its texture carries highly descriptive information about the shape of the eye. For that reason, 2D data is utilized to detect the interest points around the eyes, namely the iris center, the inner and outer eye corners and the upper and the lower borders of the iris.

The non-skin region is found by removing the pixels with the most frequent (C_b, C_r) values present in the image, using YC_bC_r space. For this purpose, firstly the histogram is calculated for distribution analysis. Even though the face image is cropped into its upper half where the eyes are located, still the skin pixels constitute the majority. Taking the histogram into account, a threshold is set according to the maximum count and the image size. Afterwards, the pixels with higher value than this threshold is eliminated as skin

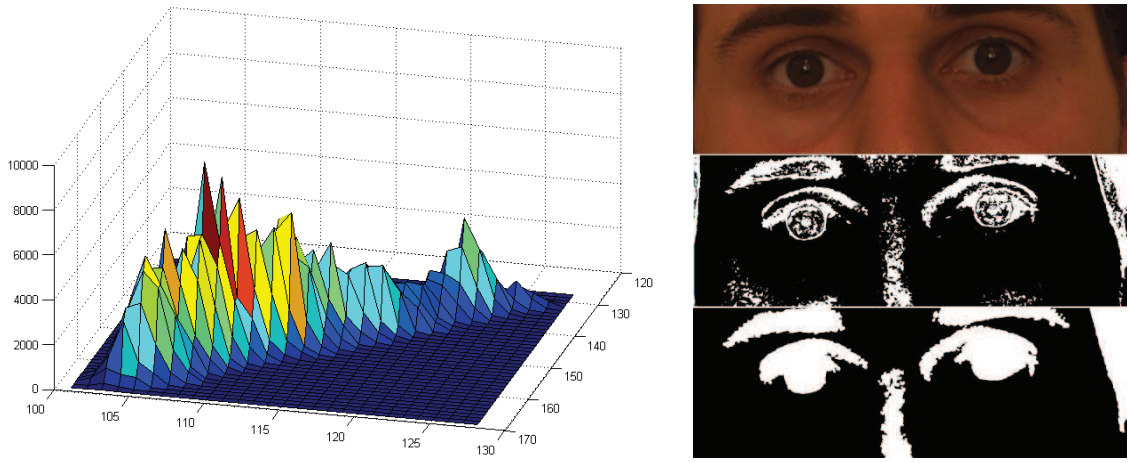


Figure 22: A (C_b, C_r) histogram and the resulting mask after thresholding. As you can see, the non-skin region which includes eyes is clearly separated from the rest.

pixels. Lastly, the small islands in the obtained binary mask are removed. In **Figure 22**, an example set of images is given to demonstrate the process.

Since the algorithm proposed is stepwise, the iris detection results affect the performance of the rest. Hence, this eye region extraction part is added to the system as a supportive module to improve the iris extraction by removing other possible circular edges as much as possible. In this module, the removal of the skin pixels is applied to suppress possible edges out of the eye region. A perfect segmentation is not necessary and it is out of our scope.

Iris Extraction

After obtaining the eye regions, firstly edge maps are constructed by Canny edge detector. The drawback of this edge detection method is that it requires a good adjustment of the threshold. In order to overcome this issue, we propose to use the edge detector iteratively, by tuning the threshold parameter until a descriptive edge map is obtained. Here, we define the "descriptiveness" of an edge map by the number of the edge pixels detected with respected to cropped eye image size and the number of circles that can be detected using these edges.

Afterwards, Hough transform is applied to the edge map to detect circles. For each detected circle, an overlapping score is calculated by the ratio of the detected portion of the circle to the whole circle parameter. In other words, it is the ratio of number of pixels detected as iris and the number of pixels circumscribed by the circle whose parameters are found via Hough transform. For circle detection, minimal and maximal radius values are defined to speed up the process.

Subsequently, the detected circles (iris candidates) are grouped into two classes according to their position: right side and left side. Then, for all possible pairs of right and left circles, following metrics are assessed and utilized to eliminate incompatible iris pairs generated:

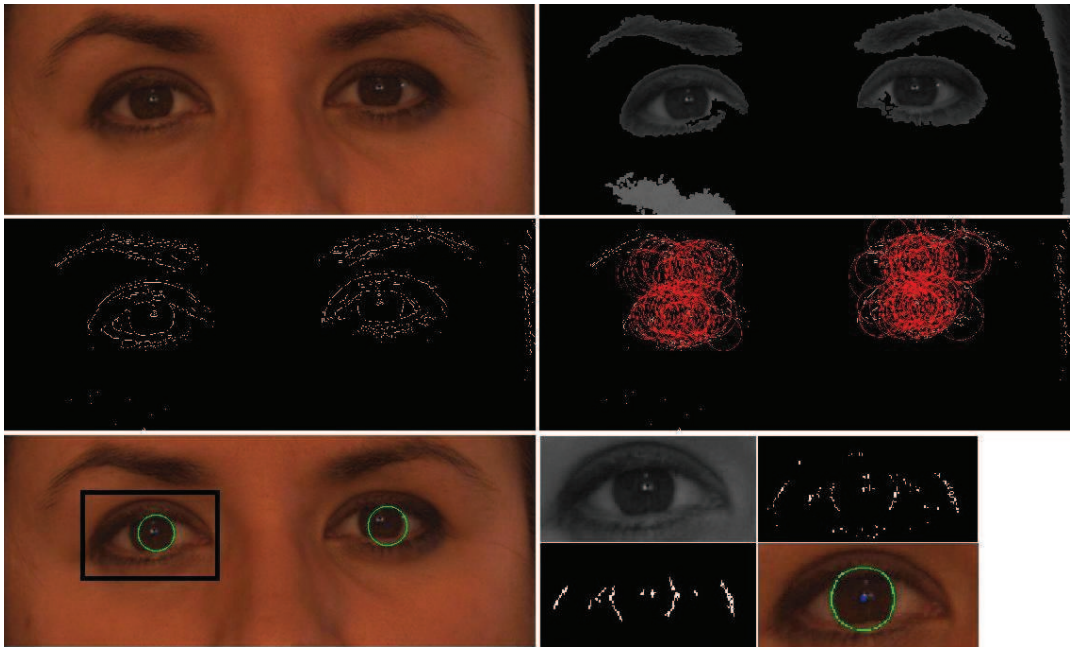


Figure 23: From left to right, top to bottom: a. Input image b. Masked image after skin region removal c. Detected edges d. Detected circles e. New eye region window f. Refining of the iris position and radius after detecting best circle to detect vertical edges

- Vertical distance of the centers: It should be less than a threshold.
- Horizontal distance of the centers: It should be larger than a threshold.
- Difference between radiuses: It should be less than a threshold.

Among the well-matched pairs, the one with maximum total overlapping score is chosen to be the two irises. In **Figure 23**, the procedure to roughly obtain the iris positions and dimensions is depicted with examples.

Once the approximate positions of the irises are obtained, rectangular windows centered at the detected iris centers are extracted and analyzed separately. Firstly, an averaging filter is applied with a rectangular kernel, where the noise and horizontal edges are suppressed and vertical edges are preserved to some point. Then, the vertical edges are detected by using the Sobel operator. As also explained in (Boehnen, et al., 2005), the upper and the lower parts of the iris border are mostly occluded by eyelids. This leads to incorrect hints for circle detection (**Figure 24**). Therefore, only the vertical edges are detected. Then, the obtained edge map is cleaned with the help of morphological operations where only the connected components which are larger than a threshold, are preserved. Since edges detected by the Sobel operator are often broken, vertical dilation is applied before the removal of small islands.

Using this edge image, similar to the previous approach, circles are again detected by using the Hough transform method. The circle with the maximum score provides us the center and the radius of the iris.

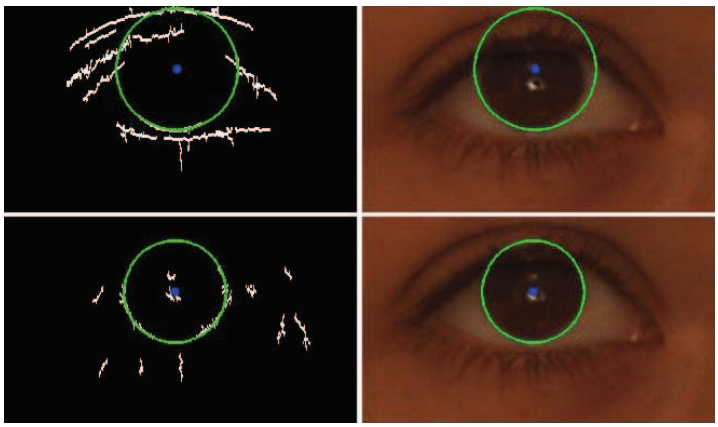


Figure 24: The positive effect of using vertical edges only can be observed when the two detected iris circles are compared

Eye Corners Extraction

For this part, the eye images are further cropped since now the accurate iris centers and radiuses are known. In this approach, firstly the eyelids contours are aimed to be detected which can be used to determine the eye corners. For this purpose, the edges created by the eyelids are searched for. The edge detection is done in two ways:

- On the color segmented image
- On the grayscale image

Firstly, the color eye image is segmented into 3 regions: dark regions like iris and eye lashes, skin regions and sclera (white part of the eye ball). In this segmentation, at the beginning the input eye image is coarsely represented using 10 colors. For this coarse representation, spatial information from a Histogram based windowing process is used (Chen, et al., 2002). Next, k-means is used to cluster the coarse image data. The cluster centroid locations are initialized with the mean value of the 70 manually collected sample colors for each region.

After clustering, the resulting segmented image is convolved with horizontal and

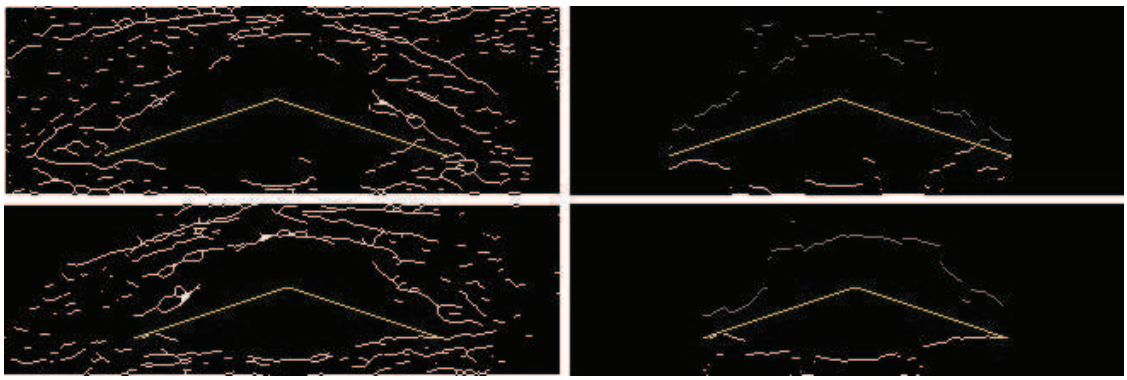


Figure 25: Two example edge maps before and after the edge pixel elimination method is applied

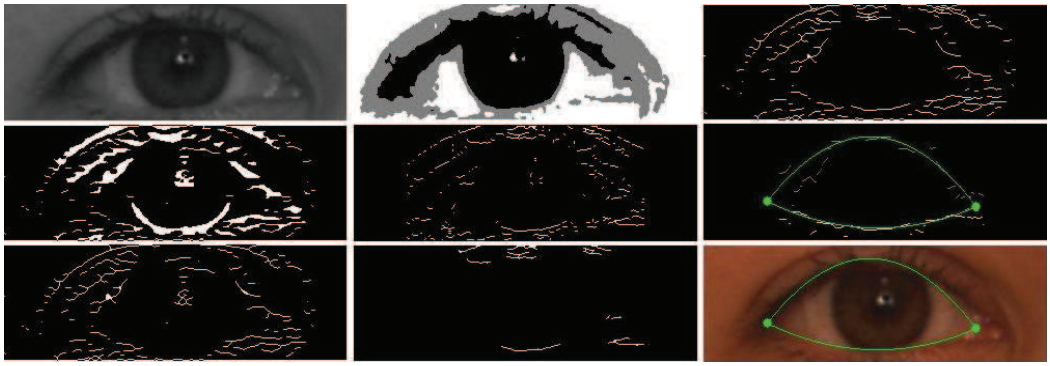


Figure 26: Each column from left to right: a. Edge detection using gray-scale image and removal of the edges close to the iris contour b. Detection of edges with less than 45° using segmented image and removal of small sections c. Fusing the two edge maps and curve fitting after eliminating edges that are not related to eyelids.

vertical Sobel operators to detect the corresponding edges and for each edge its angle is calculated. In view of the fact that eyelids are mostly closer to horizontal, only the edges with less than 45 degrees are taken into account. Additionally, similar to the processing in section 3 for iris extraction, horizontal dilation is applied to connect broken edges and then small sections are removed.

In a similar manner to the segmented ones, in this part gray-scale eye images are processed to detect the eyelids. Horizontal edges are detected again using Sobel operators. Since the edges are not as well-defined, small parts of the iris border are also detected as horizontal. In order to solve this, the edges detected in the close neighborhood of the previously detected iris contour are removed. Lastly, morphological thinning operation is applied on the resulting edge map.

After the two detected edge maps are superimposed, the following method is applied to remove outliers:

Observing that the eye corners are mostly located lower than the iris center, two lines are created, which are imagined to be approximately connecting the iris center and the corners. The slope of both lines is empirically chosen to be 1/3. Afterwards, only the closest edges that are below and above these lines are labeled as upper and lower eyelids. This method is illustrated in **Figure 25.**

In the final step, 2nd and 3rd degree polynomials are fitted for lower and upper eyelids edges respectively, in a least squares sense. The fitting is repeated once more with only the edges close to the first estimation, to further remove the outliers that still exist. The inner (near the nose bridge) and outer eye corners are determined as the intersection points of the two fitted polynomial. In **Figure 26**, a set of sample images is given to illustrate each step of this section.

3.3.2. Nose

Contrary to the case with the eyes, nose region is extremely distinctive in surface but quite plain in texture. For this reason, we choose to proceed in 3D.

Initially, the yaw angle of the face is corrected in 3D. In order to achieve this correction, the horizontal curve passing through the previously found nose tip is examined. Ideally, the area under this curve should be equally partitioned by a vertical line passing through its maximum (assuming the nose is symmetrical) or at least the difference should be minimal. With this approach, the curve is iteratively rotated to minimize this difference between the two areas under the curve, divided by the newly calculated maximum point as the new nose tip. Once the angle is determined, the whole surface is rotated, so that a "more frontal" face is obtained.

After this adjustment, the minimum Gaussian curvature is calculated for each point in the nose region. Then, edge detection is applied on the image in which the intensity is taken as the minimum curvature. . Those edges reveal the position of the interest points on both sides of the nose tip (**Figure 27**). Since they are not clearly defined, for the right and left upper edges of the nose, the points are assumed to be on the same horizontal line as the lower iris bounds and with a z value higher than ¼ of the nose tip.

3.3.3. Eyebrows

In many studies to detect and track eyebrows ((Singh, 2004), (Hara, et al., 1997) and (Nikolaidis, et al., 1997)), template matching is utilized. The eyebrows are extracted with the help of edge map and according to a template. In a similar manner, in (Hammal, et al., 2004) the segmentation of the eyebrows involves fitting parametric models accurately based on the luminance gradient and in (Kapoor, et al., 2002), a template for which the

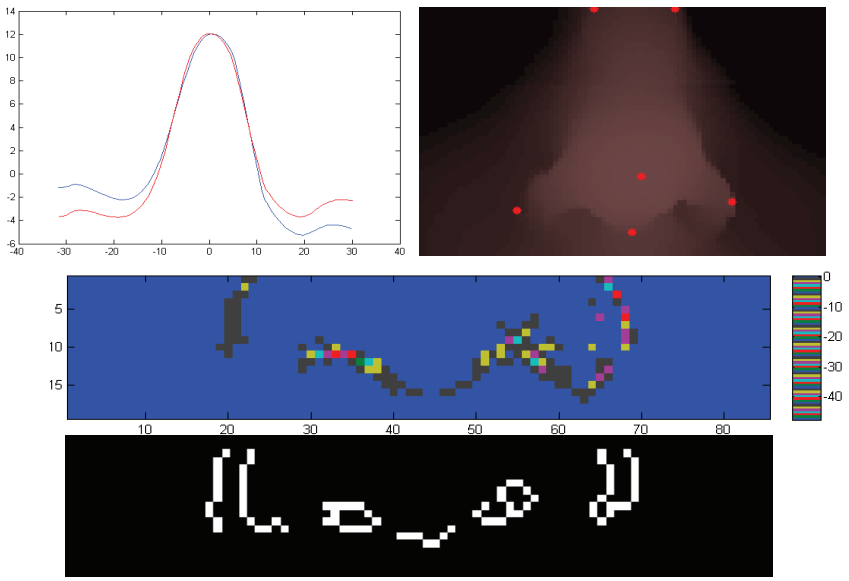


Figure 27: The horizontal profile of the nose tip before (blue) and after correction (red) is given along with the depth map of the nose region with detected points marked. The other two graphs show the minimum curvature and the corresponding edge map.

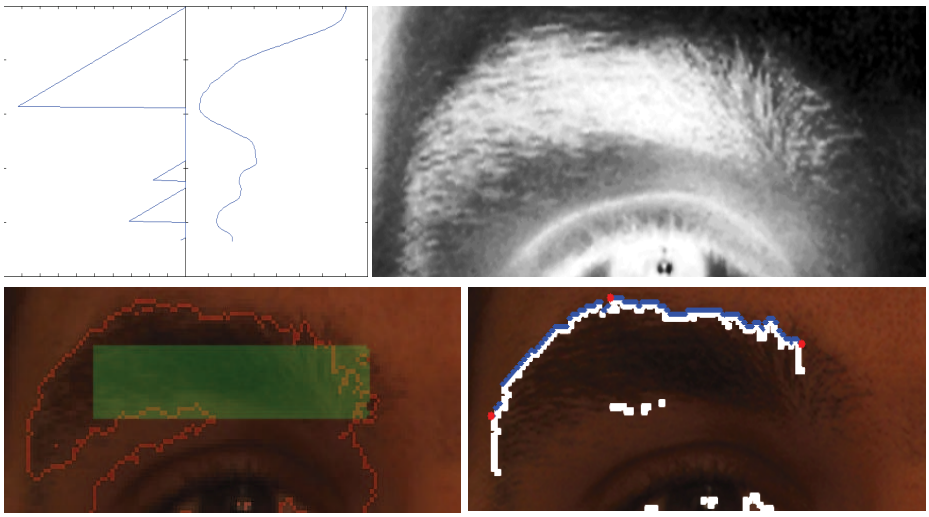


Figure 28: In the top row, vertical projection curve and its accumulation plot are given for the red channel of the eyebrow image whose histogram is equalized. In the bottom row, the initial mask (green) and the resultant segmentation is presented with the final contour merged with the edge map (white), the noise-free section after polynomial fitting (blue) and the points detected (red).

parameters are recovered by PCA analysis is utilized to represent the detailed shape information of eyebrows. On the other hand, there are also approaches like (Chen, et al., 2002) where active contour model (snake) method is applied to extract the accurate eyebrow contour.

In our system, due to the fact that for eyebrow regions, the foreground and background are statistically different from each other but homogeneous within, region based (without edges) active contours approach is adopted. For this approach, initialization is a crucial step for accurate results. This problem is overcome by selecting the minimum in vertical projection of the smoothed grayscale images. For this purpose, histogram-equalized red channel of the eyebrow image is vertically projected (integral projection) and after the obtained curve is smoothed, an accumulation plot is created in which successive points with a negative derivative are counted. As a result, an approximate position for the eyebrow is found by taking the minimum value of the accumulation plot. The initial mask is created accordingly and the eyebrows are segmented based on the method proposed in (Chan, et al., 2001). After the contour is obtained, only the upper boundary is extracted and horizontal edge information, obtained by using Canny operator is added to it, in order to improve the robustness against false contour sections. Finally, a 3rd degree polynomial is fitted to remove the outliers and the first, the highest and the last points on the curve are taken as points of interest. Detailed example eyebrow segmentation is given in **Figure 28**.

3.3.4. Lips

Although the lip region possesses similar properties with the eyebrow region, in the sense that they are darker regions on a plain background, the color difference between

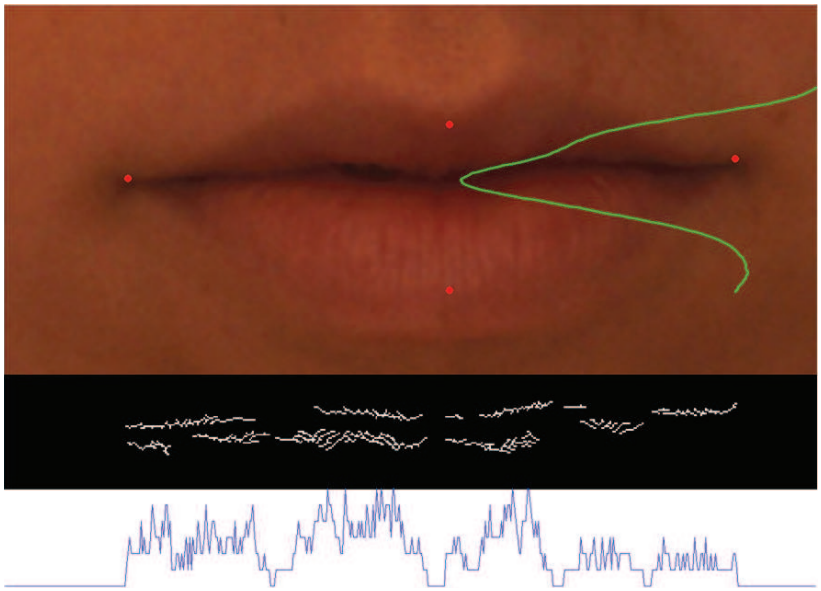


Figure 29: A lip image is given with the detected points. The green line shows the horizontal projection result. At the bottom, the calculated edge map and its vertical projection is given.

skin and lips is not always as distinctive as the difference between the skin and the eyebrows. For this reason, we need a more appropriate approach than region based active contours.

Numerous studies are proposed to extract the lip contour based on deformable (Liew, et al., 2000) or active contour models (Liu, et al., 2010), working only with the 2D image. On the contrary, this is not the case with our system and thanks to the previous analysis on the vertical midline of the face; we have good estimates of three interest points on the lip.

Since we work on faces with neutral expressions, the mouth is assumed to be closed. A closed mouth always yields to a darker line between the two lips and based on this knowledge, the contact point of the lips is refined by applying a vertical projection analysis similar to that is suggested for eyebrows. Since that line should be represented with a nadir point in the projection, the row with a zero derivative and a minimum value is taken.

After obtaining the middle contact point for the lips, a narrow horizontal window is taken around that point to further analyze the dark line in between and to use that information to find the two endpoints of the lips. Horizontal edges are detected by iteratively decreasing the threshold until enough edge pixels are found and by horizontally projecting this edge map, the left and right corners of the lips are detected (**Figure 29**).

The upper and lower edges are not modified (they are located by the vertical profile analysis) unless there is a mistake due to the presence of moustache. In the case that a moustache is detected (**Figure 30**), the dark pixels in the gray-scale image are masked out

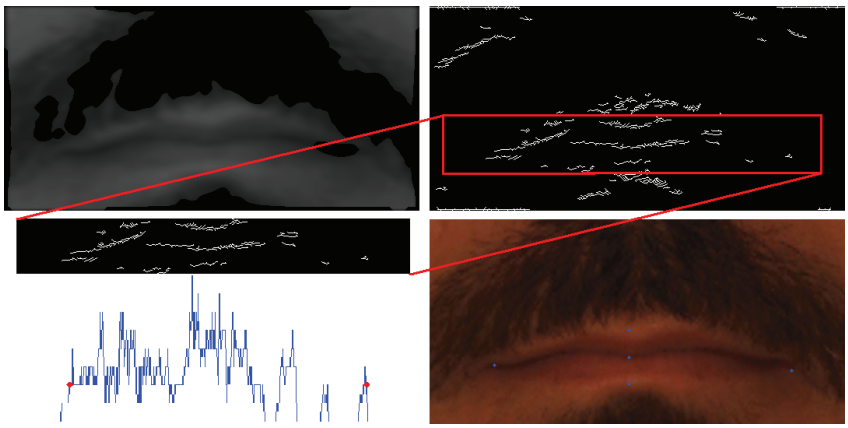


Figure 30: The obtained mask by thresholding removes the edges that the moustache creates and good estimates for the both corners of the lips are obtained.

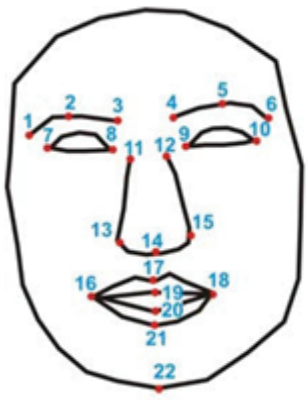
and the lip points' detection is executed in the remaining part. The moustache masking is achieved by intensity thresholding in the highly blurred gray-scale image.

3.4. Tests and Results

In order to evaluate the performance of the proposed system, Bosphorus database [22] is utilized. This database contains 3D face scans and corresponding 2D color images, together with 2D and 3D coordinates of 24 labeled facial landmarks (**Figure 31**). Most of those landmarks coincide with the ones we detect. For the rest of them, the ground truth is created manually. Since the purpose of the proposed system is defined as to landmark neutral and frontal faces, the neutral set is utilized, which consists of 299 frontal shots (2D+3D) of 105 different subjects. The results are presented separately for each region in the following subsections.

3.4.1. Eyes

For the eye corners, manually noted landmarks in the database are utilized. The error is taken to be the ratio of the Euclidean distance between the detected and the manually



Labels of the 24 landmarks	
1. Outer left eye brow	2. Middle of the left eye brow
3. Inner left eye brow	4. Inner right eye brow
5. Middle of the right eye brow	6. Outer right eye brow
7. Outer left eye corner	8. Inner left eye corner
9. Inner right eye corner	10. Outer right eye corner
11. Nose saddle left	12. Nose saddle right
13. Left nose peak	14. Nose tip
15. Right nose peak	16. Left mouth corner
17. Upper lip outer middle	18. Right mouth corner
19. Upper lip inner middle	20. Lower lip inner middle
21. Lower lip outer middle	22. Chin middle

Figure 31: The manually labeled points in the Bosphorus database

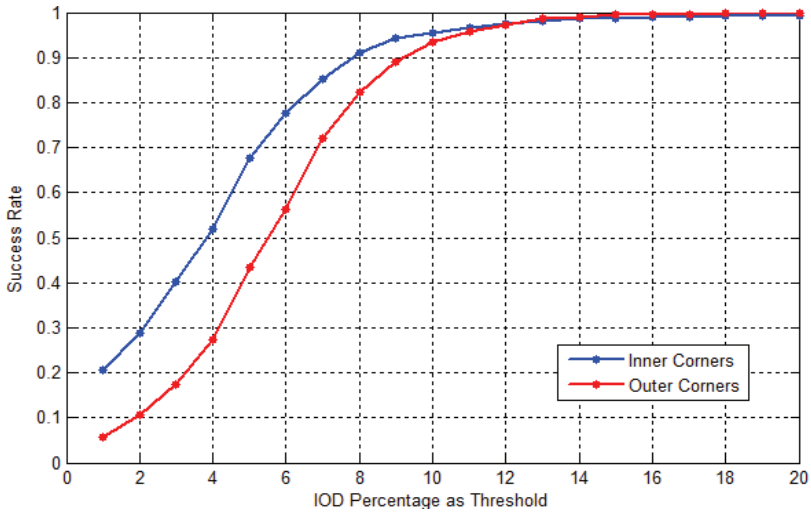


Figure 32: Eye corner detection rates for different error thresholds.

labeled corner to the interocular distance (IOD). For the error threshold being 10% of IOD, 95.48% and 93.48% success rates are achieved for inner and outer eye corners, respectively. The results show that the algorithm performs better for the inner corners. This is because closer to the inner eye corners, the eyelid contours are more prominent than the ones around the outer eye corners. The error threshold vs. success rate graph is given in **Figure 32**. In (Celiktutan, et al., 2008), 98.5% success rate is reported for inner eye corners in neutral images. This result is achieved by fusing 4 several feature modalities.

The iris locations are not reported in the database and hence, it is labeled additionally. In the following table, the success rate at 8% IOD threshold is reported with two other existing algorithms.

Table 2: Success rates for iris localization compared with two other methods

Method	Threshold	Success rates
(Guan, 2007)	-	94.82%
(Kuo, et al., 2005)	-	94%
Our method	8%	100%

3.4.2. Nose

Similar to eye corners, the ground truth files in the database are used for comparison for the interest points around the nose. The error is again calculated as the Euclidean distance between the detected points and the ground truth, divided by the IOD distance. In **Figure 33**, the success rates versus the error threshold for 5 points (except the one at the bottom of the nose which is not included in the ground truth files) are given.

The first two points (nose saddle left and right) have the highest errors. When analyzed further, the reason is found to be in the definition of those interest points. Since in our

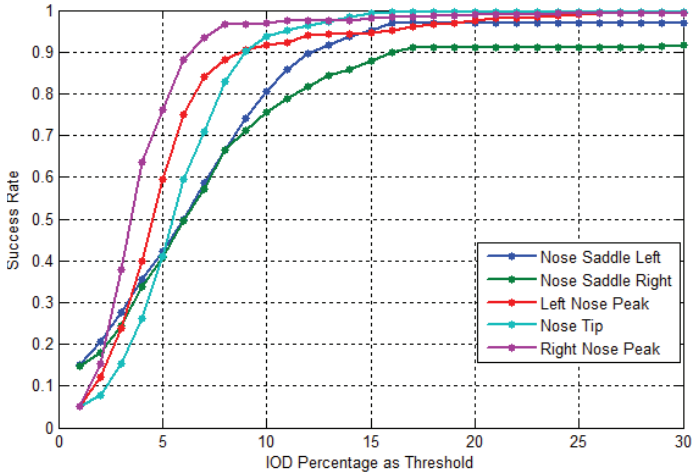


Figure 33: Nose point detection rates for different error thresholds.

approach those points are taken to be at the same horizontal line as the lower iris bound and with a z value higher than ¼ of the nose tip, the error represents the deviation of manual decisions from this definition The success rates at 10% IOD threshold is given in **Table 3**.

For the point at the bottom of the nose, since its location is not always obvious in 2D images, the evaluation is done on the 3D vertical profile lines for each face. It is observed that 100% success rate is achieved with 2 pixels error threshold and for 94% of the faces, the bottom of the nose labeled impeccably.

Table 3: Success rates for the points around the nose when the success threshold is taken as 10% of IOD.

Point (as labeled in Figure 31)	Success Rates
Point #11	80.60%
Point #12	75.59%
Point #13	91.64%
Point #14	93.65%
Point #15	96.99%

3.4.3. Eyebrows

The difficulty in evaluating the detection of points around eyebrows stems from the ambiguity in their definition. To clarify the extent of the subjectivity, a test was conducted in which two persons are asked to manually label the points of interest and the differences between the detected and labeled points are analyzed together with the difference between two sets of labeled points (3 for each eyebrow: left, middle and right). In total, results for 210 eyebrows are compared. The success rates vs. error threshold plots for the three

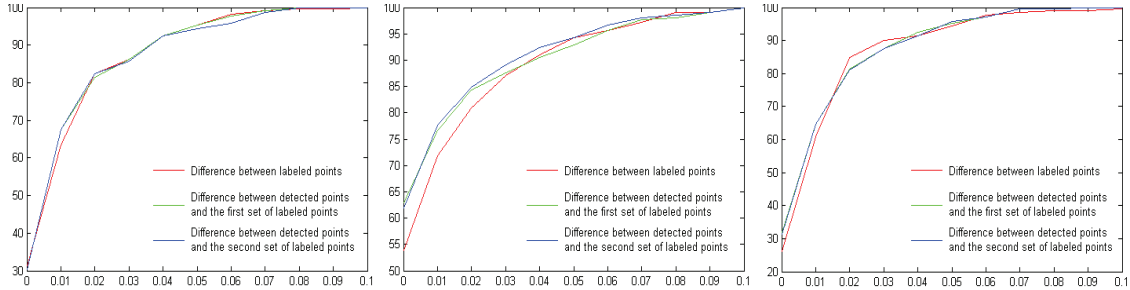


Figure 34: Rates vs. error threshold plots for the three points

points are given separately in **Figure 34**, which include error between two labeled sets and between the detected points and the two sets of ground truth points.

As shown in the plots, the detection error follows a close path to the subjective error. The error axes in the graphs represent the Euclidean distance between points divided by the diagonal length of the cropped eyebrow image. For the final evaluation, again the rate of the distance difference to the IOD length is utilized. The obtained results for 10% threshold are given in **Table 4**. Additionally in **Figure 35**, the plots for the 6 points (left and right eyebrows separately) are given.

Table 4: Detection rates for the outer, middle and inner points on the eyebrows when the success threshold is taken as 10% of IOD.

Points	Success Rates
Outer	77.09%
Middle	61.04%
Inner	77.43%

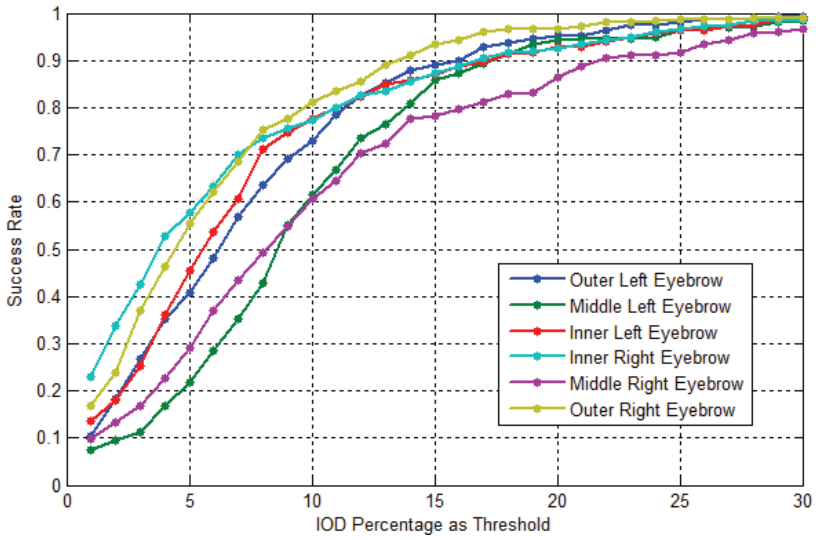


Figure 35: Eyebrow points detection rates for different error thresholds.

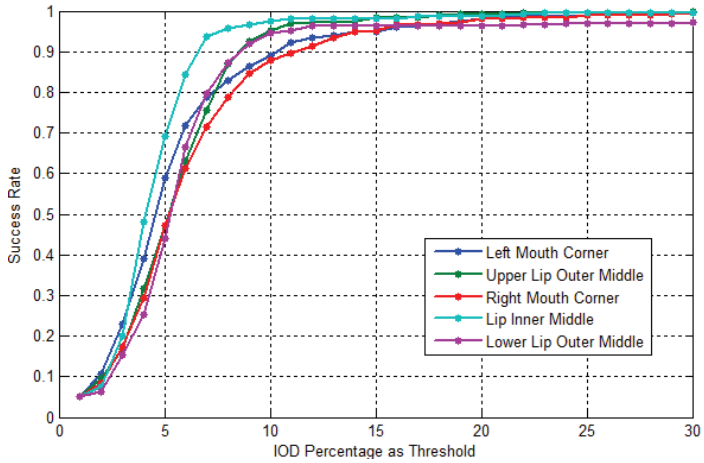


Figure 36: Lip points detection rates for different error thresholds.

3.4.4. Lips

The landmark files supplied with the database are used to evaluate the detection of interest points in the lip region. The success rates at error threshold of 10% IOD for each of the 5 points are given in **Table 5**.

Since the mouths are closed in the test set, the "upper lip inner middle" point is chosen as the middle point on the contact line of the lips. Again in (Celiktutan, et al., 2008), 92.6% success rate is reported for mouth corners in neutral images. For the same points, we reach 88.46%.

Table 5: Error mean and its standard deviation for each lip point

Point (as labeled in Figure 31)	Success Rates
#16	88.96%
#17	95.32%
#18	87.96%
#19-20	97.66%
#21	94.65%

3.4.5. Chin

For the two points on the chin, the lower one is included in the database landmark files (point #22). For this point, the success rate is found to be 73.58% with the error threshold of 10% IOD. On the other hand, for the upper point, evaluation is done in the vertical profile plot in a similar manner as the bottom of the nose. It is seen that for 91.43% of the faces the point is detected within 3 pixels error threshold and 97.14% success rate is achieved when this threshold is taken as zero (perfectly detected). The errors are mostly due to the presence of the beard on the chin.

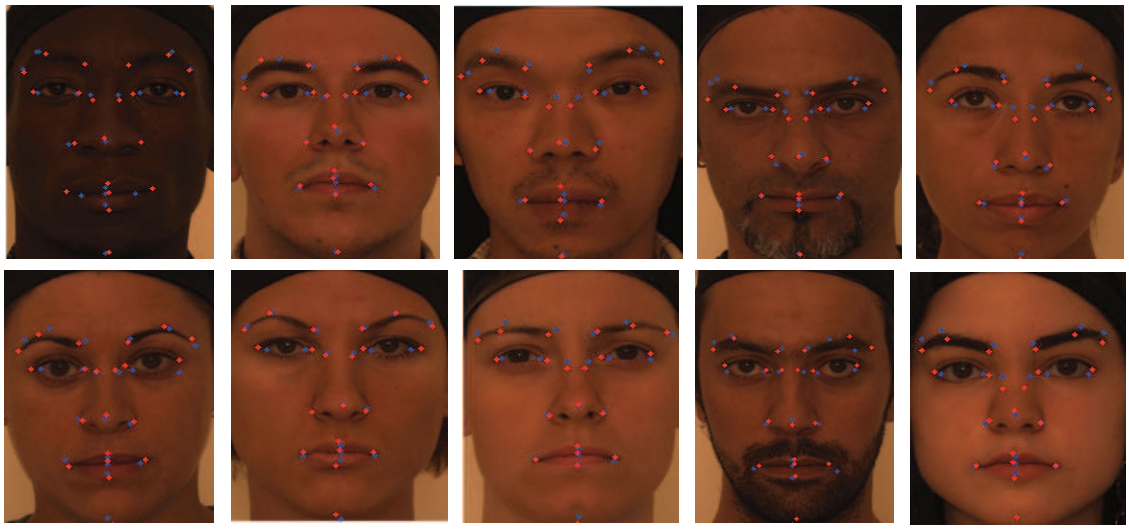


Figure 37: Some examples of the resultant detected points (blue) and the labeled points (red)

Some example faces with detected and labeled points are given in **Figure 37**.

3.5. Conclusion

In this chapter, a method to find 29 interest points on frontal and neutral faces is presented, in which both 2D and 3D images of the faces are utilized. With the help of an extensive vertical profile analysis, the face is broken down to different sub-regions for eyes, eyebrows, nose and lips. Next, according to the region characteristics either 2D color images, either 3D scan data or both are analyzed in order to detect the points of interest. The algorithm is tested on the Bosphorus database that includes 299 scans with registered 3D scans and color images. As a result, it is observed that accurate results can be achieved for each one of the points.

Chapter IV. Asymmetrical Approach: Learning the Expression Variations

One of the most critical sources of variation in the face recognition problem is facial expressions, especially in the common case where only a single sample per person is available for enrollment. Methods that improve the accuracy in the presence of such variations are still required for a reliable person-authorization system. In this chapter, we address this problem with an analysis-by-synthesis based scheme, in which a number of synthetic face images with different expressions are produced. For this purpose, an animatable 3D model is generated for each user based on 17 landmark points (for which an automatic extraction method is presented in the previous chapter). The contribution of these images to the recognition performance is evaluated with a PCA-based implementation.

1. Introduction

As 3D sensing technologies advance and the acquisition devices become more accurate and less expensive, the utilization of range data instead of / together with color data broadens. Consequently, in FRVT 2006 (Phillips, et al., 2009), an order-of-magnitude improvement in recognition performance was achieved over the FRVT 2002 from 3D face images, with FRR of 0.01 at a FAR of 0.001. Similar results were also achieved with high resolution still images under controlled illumination. However, for the uncontrolled scenario, performances again drop.

Even though, 3D face recognition has a better potential than its 2D counterpart, in practice it is not very straightforward to obtain an accurate 3D image. Systems that extract shape information from 2D images, e.g. passive stereo approach, rely on the knowledge of extrinsic parameters of the scene and intrinsic parameters of the camera to obtain a certain degree of accuracy. On the other hand, with active sensors like laser scanners, a scan can take several seconds, requires the subject to remain still during the process and furthermore reconstruction of depth is allowed at short range. While these aspects are inconsequential in the well-controlled enrollment phase, they complicate the capturing of the probe image. By taking this fact into consideration, we believe that a system, in which a combined enrollment is realized by both 2D and 3D information whereas the target images are still 2D images, is the optimal strategy to fuse advantageous features of the two modalities.

With the assumption of this asymmetrical scenario, in this study, we address the problem of expressions in 2D probe images. Our aim is to facilitate recognition by

simulating facial expressions using 3D models for each subject. With regard to the causes of the intra-class variations, synthesis of facial images under various pose and illumination conditions using 3D face models is straightforward since these variations are external. However, this does not hold for expressions which alter the facial surface characteristics. In order to obtain realistic facial expression simulations, we propose an automatic procedure to generate MPEG-4 compliant animatable face models from the 2.5D facial scans (range images) of the enrolled subjects based on a set of automatically detected feature points. Using a facial animation engine, different expressions are simulated for each person and the synthesized images are used as additional samples for the recognition task. It is important to emphasize that synthetic sample augmentation is carried out during enrollment only once for each subject.

1.1. Existing Work on Robustness against Expression Variations in 2D Face Recognition

Due to the inconveniences in 3D acquisition, 2D image-based face recognition still remains the most practicable in real-world applications. Numerous solutions for the problem of recognition in the presence of facial deformations, due to expressions, have been proposed. These can be classified as appearance-based and model-based approaches (Lu, 2003).

In appearance-based methods, face recognition is treated as a pattern recognition problem, where the face image is considered as a high-dimensional vector. Several dimension reduction techniques are proposed such as Principal Component Analysis (PCA) (Turk, et al., 1991), Linear Discriminant Analysis (LDA) (Swets, et al., 1996) and Independent Component Analysis (ICA) (Bartlett, et al., 1998) in which a face vector is projected to different basis vectors and the decision is made based on the distance between projection coefficients. Those methods are extended by creating subsets of images through masking the regions where significant modifications are expected to occur and using them to build different face projection spaces (Tarres, et al., 2005). Additionally, in (Bing, et al., 2005), PCA is applied on the motion vectors domain, caused by the motion of the facial features due to facial expressions and two spaces are used for the reconstruction of the test images. However, the small sample problem which is commonly faced in real-world situations, may give rise to faulty system design in appearance-based recognition procedures (Martinez, et al., 2001).

In model-based methods, a face model is constructed based on the prior knowledge of human face. By localizing feature points on a face, the constructed model is fitted to probe face images and this fitting process produces parameters to be used as feature vectors. Two of the most significant examples are Elastic Bunch Graph Matching (EGBM) (Wiskott, et al., 1997) and Active Appearance Model (AAM) (Edwards, et al., 1998).

In EBGM, Gabor wavelet coefficients are proposed to be obtained at fiducial points, detected on the face. This approach removes most of the variability in images due to

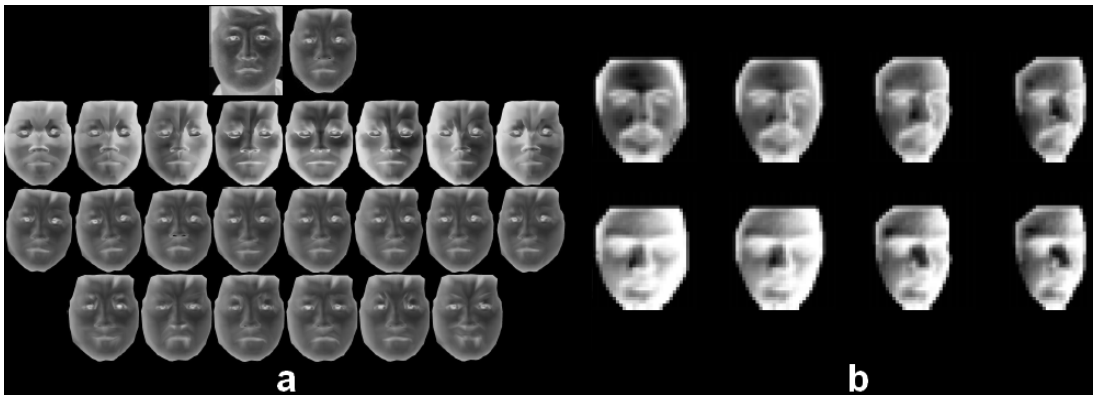


Figure 38: Synthesis examples: (a) input intensity image and accordingly synthesized face images under 8 different lighting conditions, 8 different pose variants and 6 different expressions (Lu, et al., 2004) (b) Images rotated (left to right) by angles 5°, 10°, 25°, 35°; (top to bottom) illuminated under conditions where ($\alpha = 0°$) and ($\alpha = 30°$, $\tau = 120°$) (Zhao, et al., 2000)

illumination conditions; nevertheless it is not robust to strong deformations on the facial surface. Similarly, in (Ersi, et al., 2006), another local feature based method is proposed in which a set of feature points with highest deviations from the expectation is automatically extracted by using statistical Local Feature Analysis (LFA). Subsequently, each point is described by a sequence of local histograms captured from the Gabor responses at various frequencies and orientations around the feature point.

In AAM, shape and texture information are utilized to compute a 2D morphable model. Model parameters which bring the morphable model as close as possible to the target face image are calculated and used as feature vectors. Later, Blanz et al. (Blanz, et al., 2003) proposed a more advanced 3D morphable face model that represents the shape and the texture of a face via its parameters, which are obtained by optimizing intrinsic and extrinsic parameters, using 3D computer graphics technology.

The adopted approach to overcome facial variations in this study is face synthesis, for which synthetic sample images are generated for training under varying conditions. Previously, in (Zhao, et al., 2000), a view synthesis method based on shape-from-shading is proposed. The images that are synthesized at different views and illuminations are fed into an LDA-based system. Similarly, in (Lu, et al., 2004), a 3D generic face model is aligned onto a frontal face image and synthetic face images are obtained under different light, pose and expression conditions to build an affine subspace for each subject.

As shown in **Figure 38**, examples of synthetically generated faces are defective. In the first example (Lu, et al., 2004), although a large number of (115) manually labeled feature vertices are used to iteratively adapt the generic face model; still the facial data is observed to be highly degraded.

In these examples, 3D shape assistance is acquired by obtaining 3D models of face using the texture information (Zhao, et al., 2000) or by utilizing a general shape model instead of a personalized 3D model (Lu, et al., 2004). Employing a combined enrollment of 2D and 3D data would help to eliminate faulty modeling for the face.

This potential scenario was also proposed during the Face Recognition Grand Challenge (FRGC), where enrolled images are 3D and the target images are still 2D images (Phillips, et al., 2005) but no baseline was provided. Reinforcing this approach, in their analysis (Husken, et al., 2005), Husken et al. deduce that combining both modalities on different algorithmic levels to compensate for malfunctions of each of them separately is a promising approach.

Accordingly, in our study, acquired 2D and 3D face information are blended together in multiple steps of enrollment. Based on the assumption of a controlled environment, the subjects are enrolled with a neutral face and under ambient light, by both 2D and 3D sensors. In the pre-processing step, the 3D surface is cleaned of its holes and spikes and smoothed. Then, positions of 17 landmarks are automatically located by utilizing both modalities (i.e. shape and texture) according to their inherent characteristics in different regions of the face, as explained in the previous chapter. Finally, based on the detected feature points, an MPEG-4 compliant animatable model is produced by warping an animatable generic model using Thin Plate Spline (TPS) method.

The main contribution of this study revolves around this last stage, in which the rigid facial surface with no semantics is transformed in to a highly realistic animatable model.

Once this model is obtained, it is animated using a compatible animation engine for various expressions. The efficacy of the generated synthetic face images are evaluated on a PCA-based recognition system.

2. Proposed System

In the proposed system, the enrollment is done in both 2D and 3D for each subject under a controlled environment – frontal face images with a neutral expression under an ambient illumination. The obtained 3D shape of the facial surface together with the registered texture is preprocessed, firstly to extract the face region.

On the extracted facial surface, scanner-induced holes and spikes are cleaned and a bilateral smoothing filter is employed to remove white noise while preserving the edges. After the complete, noise-free face model (3D+2D) is obtained, 17 feature points are automatically detected using both modalities according to the regional properties of the face (Chapter III). These detected points are then utilized to warp a generic animatable face model to completely transform it into the target face. The generic model with manually labeled 71 MPEG-4 points is suitable to simulate facial actions and expressions via an animation engine that is in accordance with MPEG-4 Face and Body Animation (FBA) specification.

Finally, in order to simulate the facial expressions on the obtained animatable model, an animation engine, called visage|life™ (Visage) is utilized. Multiple expression-infused face images are generated for each subject to enhance face recognition performance. The whole system is illustrated in **Figure 39**.

3. Constructing the Animatable Face Models

In order to construct an animatable face model for each enrolled subject, a mesh warping algorithm based on the findings in (Tena, et al., 2006) is proposed. A generic face model, with holes for the eyes and an open mouth is strongly deformed to fit the facial models in the database, using the TPS method. 17 automatically detected points, together with the rest of the FDP points for an MPEG-4 compliant animation is marked manually on the generic face (**Figure 39**).

MPEG-4 specifications and the mathematical background of the TPS method will be briefly explained before going into details about the proposed animatable face construction method.

3.1. MPEG-4 Specifications and Facial Animation Object Profiles

MPEG-4 is an ISO/IEC standard developed by Moving Picture Experts Group which is a result of efforts of hundreds of researchers and engineers from all over the world. Mainly defining a system for decoding audiovisual objects, MPEG-4 also includes a definition for the coded representation of animatable synthetic heads. In other words, independent of the model, it enables coding of graphics models and compressed transmission of related animation parameters.

The facial animation object profiles defined under MPEG-4 often are classified under three groups ((Lavagetto, et al., 1999), (Kuriakin, et al., 2001) and (Pandzic, et al., 2003)):

- Simple facial animation object profile: The decoder receives only the animation information and the encoder has no knowledge of the model to be animated.
- Calibration facial animation object profile: The decoder also receives information on the shape of the face and calibrates a proprietary model accordingly prior to animation.
- Predictable facial animation object profile: The full model description is

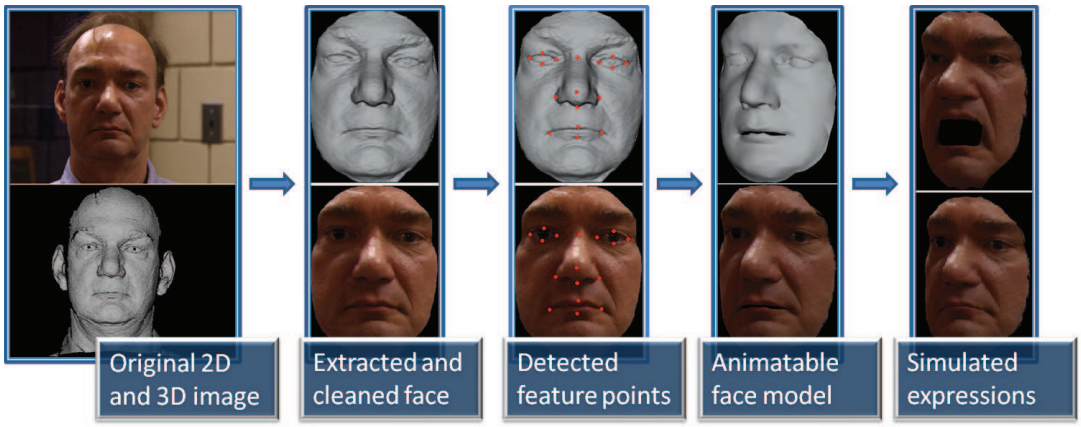


Figure 39: The enrollment procedure is illustrated on an example.

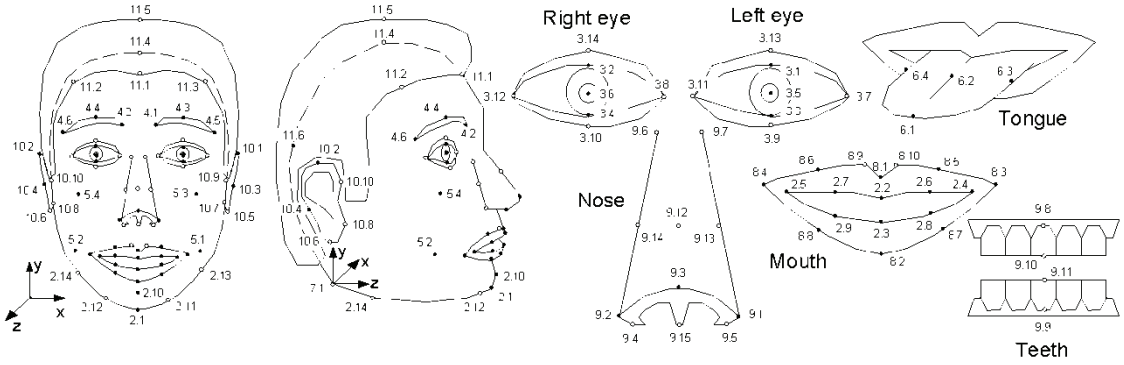


Figure 40: MPEG-4 Facial Definition Parameters

transmitted. The encoder is capable of completely predicting the animation produced by the decoder.

The profile most conformable to our approach is the second one: calibration facial animation object profile, since we are aiming to calibrate the "generic" model according to the samples in our database. Our system generates an animatable model by using 17 of 71 MPEG-4 specified face definition parameters (FDP) which are annotated automatically. The rest of the points are only marked on the generic model for animation.

In **Figure 40**, the positions of the MPEG-4 FDP points are given. Most of these points are necessary for an MPEG-4 compliant animation system, except for the ones on the tongue, the teeth or the ears, depending on the animation tool structure.

3.2. Thin Plate Spline Warping

As the name indicates, the TPS method is based on a physical analogy to how a thin sheet of metal bends under a force exerted on the constraint points. The TPS method was made popular by Fred L. Bookstein in 1989 in the context of biomedical image analysis (Bookstein, 1989).

For the 3D surfaces S and T, and a set of corresponding points (point pairs) on each surface, P_i and M_i respectively, the TPS algorithm computes an interpolation function f(x,y) to compute T', which approximates T by warping S:

$$T' = \{(x', y', z')\ st. \forall\ (x, y, z) \in S, x' = x, y' = y, z' = z + f(x, y)\} \quad \text{Equation 2}$$

$$f(x, y) = a_1 + a_x x + a_y y + \sum w_i U(|P_i - (x, y)|) \quad \text{Equation 3}$$

with U(.), the kernel function, expressed as:

$$U(r) = r^2 ln(r^2), r = \sqrt{x^2 + y^2} \quad \text{Equation 4}$$

In the interpolation function f(x,y), the w_i, $i \in \{1,2,...n\}$ are the weights. As given in Equation 2, the interpolation function consists of two distinct parts: An affine part ($a_1+a_x x+a_y y$) which accounts for the affine transformation necessary for the surface to match the constraint points and a warping part ($\Sigma w_i U(|P_i-(x,y)|)$).

3.3. The Method

TPS is commonly used to establish registration in non-rigidly deformed surface patches, like two different facial surfaces (Lu, et al., 2005). The deformation of the registered models is minimal since only few point pairs are utilized. As more control points are added to the TPS warping, the amount of deformation increases and the face becomes more and more similar to the target surface. By exploiting this fact, in this study we propose to strongly deform a common face to fit target faces in the gallery.

On the other hand, before applying the TPS warping, we have to make sure that the target face and the generic face models are aligned correctly. Each step of the process to obtain the animatable model is detailed in the following subsections.

3.3.1. Rescaling and alignment:

In the first step, the generic model is rescaled in three directions:

- X: The x-distances between the two outer eye corners of the generic model and target are equalized.
- Y: The y-distances between the outer left eye corner and the lower mid-point of the lips are equalized.
- Z: The z-distances between the outer left eye corner and the nose tip are equalized.

Next, a rigid transformation is calculated based on the 17 corresponding point pairs. Using the two sets of landmarks, the best fit mapping is computed in a least squares sense, where the squared distance between the point sets is minimized. The calculated transformation is applied on the target face. This step corrects the existing deviations from the frontal pose of the enrolled face (if any) while better aligning the two surfaces.

3.3.2. Coarse Warping

The generic model is warped coarsely for better alignment to the target face. By taking the 17 feature point pairs as the source and target landmarks, a non-linear warp is calculated, which moves any point on the mesh around a source landmark at the same location with the corresponding target landmark. The points in between are interpolated smoothly using the Thin Plate Spline algorithm.

3.3.3. Fine Warping

At this stage, the two surfaces are very well aligned and hence, we assume for all points on the generic model, the corresponding pair on the target model is the one that is closest. Based on this assumption, for every other vertex on the generic face, the matching vertex on the target face is found and used in TPS calculation. This way, half of the points on the generic model conform into the target face counterparts and maintain their smoothness.

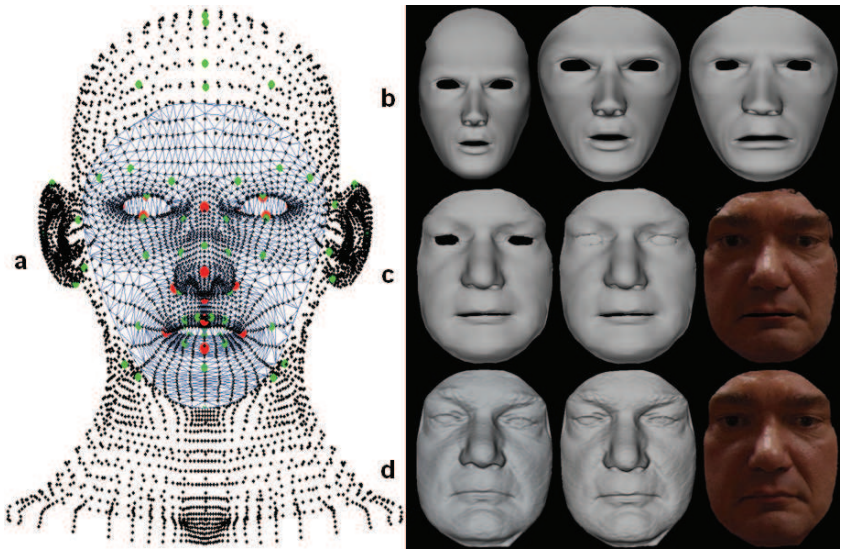


Figure 41: a) The point cloud for the generic head model is given with MPEG-4 points marked. Red points are the 17 feature points that are also located on the target faces. Only the manually cropped face region (in blue) of this generic model is utilized for warping. b) Generic face before the warping procedure, after anisotropic scaling and coarse warping (using 17 point pairs only). c) Generic face after fine warping (using half of the points on the generic face), after creating the eyeballs and after being textured. d) The target face, before alignment, after alignment without and with texture.

Finally, two spheres for the two eyes are created based on the 4 feature points detected around each eye and the texture is copied onto the obtained animatable model. The proposed method for animatable model generation is illustrated on a sample model in **Figure 41**.

4. Performance Evaluation

In order to evaluate the proposed system two different databases are utilized: FRGC (Phillips, et al., 2005) and Bosphorus (Savran, et al., 2008) 3D face databases.

Contrary to the scenario at hand, in the FRGC database, the images are taken in uncontrolled illumination conditions. For this reason, automatic landmarking is tested only on the Bosphorus database for which the subjects were recorded in a highly controlled environment and a 1000W halogen lamp was used in a dark room to obtain homogeneous lighting for good quality texture images.

The 3D data for FRGC experiments is divided into two partitions for training and testing. The training set consists of 943 3D scans and controlled and uncontrolled still images, whereas the validation partition contains 4007 3D and 2D images collected from 466 individuals with different facial expressions. On the other hand, in the Bosphorus database, there are 105 subjects in various poses, expressions and occlusion conditions, with the total number of 4666 face scans.

Figure 42: The 12 simulated expressions on a sample face

4.1. Experiments with FRGC Database

In the validation set of the FRGC database, 466 persons are recorded with different facial expressions. According to the proposed system, the enrollment process is realized with a single neutral 3D+2D image for each person. Hence, the persons with at least one neutral image in the database are selected, resulting in a gallery of 400 persons and the rest of the images that belong to the enrolled people (3522 images) are taken as the probe set.

Firstly, all images in the training, gallery and probe sets are preprocessed by applying cropping, geometrical normalization (64x80 pixels) and histogram equalization. 3 key tecniques are adopted in order to observe the effect of gallery augmentation: PCA, LDA (Belhumeur, et al., 1997) and LBP (Ojala, et al., 2002).

2D subset of the FRGC v1.0 is used to train PCA/LDA space for both FRGC and Bosphorus experiments. Considering the cumulative sum of the eigenvalues, the most significant 97 eigenvectors (90%) are used for representation for PCA. Finally, the recognition is achieved by the minimum cosine distance between the projected test and gallery images in PCA and LDA methods.

Table 6: Recognition rates for PCA, LDA and LBP methods with the original and the augmented galleries

Method	VR	EER	IR
PCA (o)	30,80%	18,00%	44,72%
PCA (s)	35,48%	16,91%	50,54%
LDA (o)	40,43%	15,99%	58,14%
LDA (s)	43,98%	13,73%	63,41%
LBP (o)	60,32%	10,68%	75,08%
LBP (s)	66,86%	8,70%	82,23%

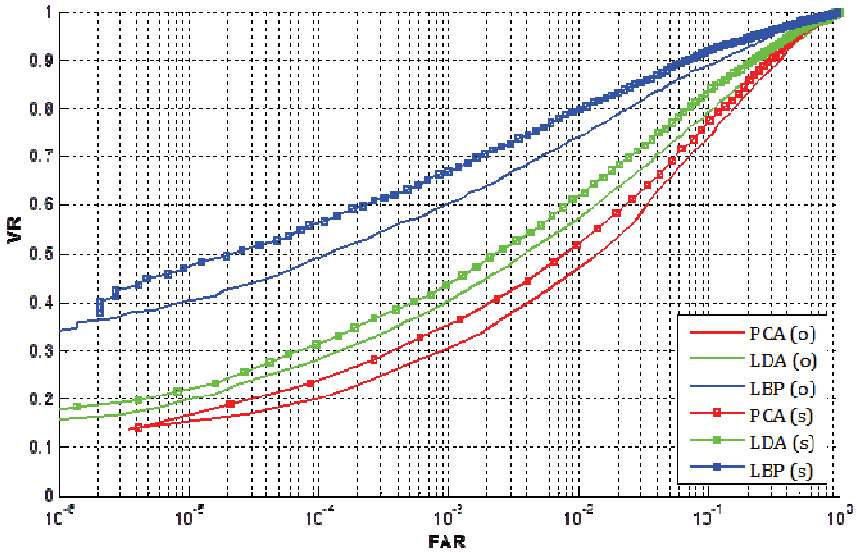


Figure 43: ROC curves for PCA, LDA and LBP methods applied using FRGC v2.0

Next, 3D face data is cropped and processed as detailed in Section III and in order to produce the synthetic face images with expressions, an animatable model for each enrolled person is constructed based on 17 manually labeled landmarks by Szepycki et al (Szeptycki, et al., 2009). The obtained faces are animated for 12 different expressions (pre-defined in the animation tool) and stored in the simulation gallery after going through the same preprocessing steps as detailed previously. An example is given in **Figure 42**.

The recognition results for the entire database using 3 different approaches with (s) and without (o) using simulated images are given in **Figure 43**. Additionally in **Table 6**, verification rates (VR) at 0.001 FAR is given with equal error rates (EER) and rank-1 identification rates (IR).

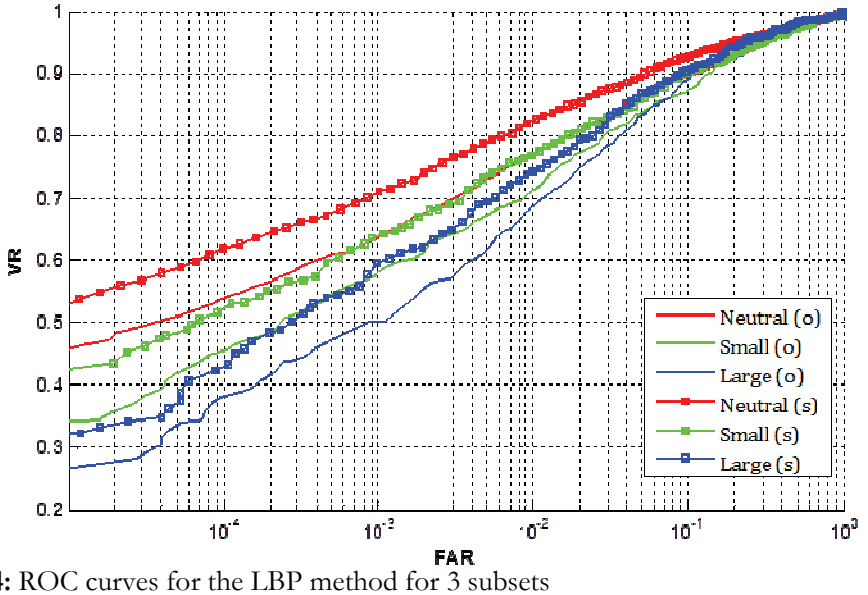


Figure 44: ROC curves for the LBP method for 3 subsets

For further analysis, three categories of facial expressions are defined: Neutral, small and large. (Maurer, et al., 2005) This approach is adopted because for recognition the amount of the shape change is more important than its type. Small expressions include moderate smiles and talking gestures. On the other hand, large contains unnatural expressions like blown cheeks. This classification results in a neutral probe set of size 2085 (60%), small probe set of size 780 (20%) and large probe set of size 742 (20%). In **Figure 44**, verification results for each bin is presented and more detailed breakdown of the results is given in **Table 7**

Table 7: Recognition rates for LBP method with the original and the augmented galleries for each subset

Method	VR	EER	IR
Neutral (o)	63,79%	10,19%	75,93%
Neutral (s)	71,01%	7,99%	83,56%
Small (o)	58,09%	11,52%	75,50%
Small (s)	64,04%	9,67%	80,93%
Large (o)	50,32%	10,34%	71,45%
Large (s)	59,74%	9,51%	78,63%

4.2. Experiments with Bosphorus Database

Among 4666 facial scans in the database, one neutral face scan for each individual is chosen as the single gallery sample. For testing, the emotions subset of the database which includes posed expressions of the six basic emotions (anger, disgust, fear, happiness, sadness and surprise) and the unused neutral scans compose a test set of 647 images. Since the facial surfaces are already cropped and cleaned, only smoothing is applied.

4.2.1. Automatic Landmarking

For each of the 105 subjects, 17 feature points are located automatically (**Figure 41**). The automatic landmarking method is explained in detail in Chapter III.

4.2.2. Face Identification

To observe the effects of the automatic landmarking errors, identification tests are conducted with both automatically detected and manually marked feature points. For 105 persons, animatable models are generated and 12 expressions (same as FRGC) are simulated to create the synthetic gallery images. All images in both gallery and probe sets undergo the same preprocessing steps as in the FRGC experiments.

Using a single neutral image per person, the rank-1 identification rate with PCA is found to be 64.14%. With the addition of synthetic gallery images, generated by using the manually marked feature points, the rate rises to 70.94%. On the other hand, synthetic

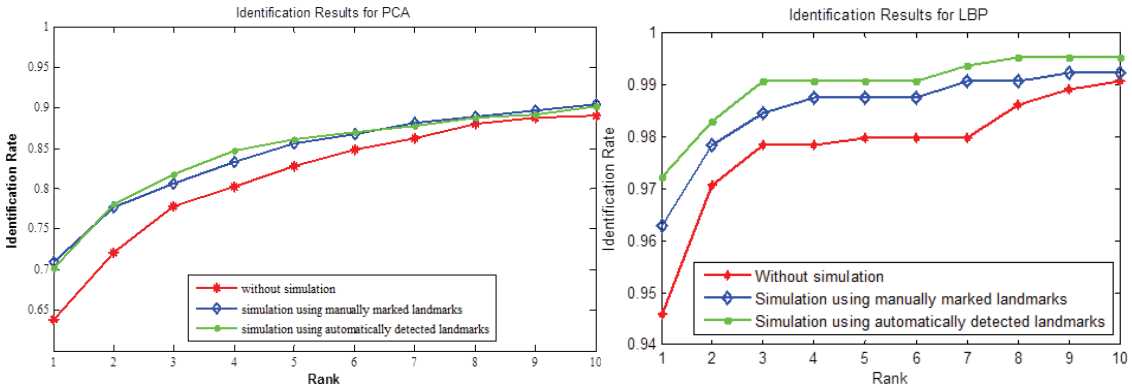


Figure 45: Recognition rates from rank-1 to rank-10 with PCA and LBP for the three experiments: without simulation, with simulation using manually marked landmarks and with simulation using automatically detected landmarks

gallery images that are created based on automatically detected landmarks increase the rank-1 identification rate to 70.14%.

Furthermore, the rank-1 identification rate with LBP method is observed to be higher with the points that are detected automatically when compared to the ones that are manually marked. The graph for recognition rates from rank-1 up to rank-10 for both algorithms are is given in **Figure 45**.

These results are very impressive in the sense that despite the errors in the automatic detection of landmarks, using these points for simulated image generation leads to a comparable amelioration in results with the ones created using manually marked points. The reason behind this result lies in the fine warping step in the animatable model generation process. At this stage, the landmarks are left aside and the correspondences are created by pairing every second point on the generic model to the closest vertex in the face scan. Hence, the errors in coarse warping due to the inaccuracies in automatic landmarking are partially corrected. Several examples for the faces that are corrected in the fine warping stage are given in **Figure 46** with the resulting simulation results.

5. Conclusions

Based on the assumption of a fully-controlled environment for enrollment, a face recognition framework is proposed in which the widely-encountered single sample problem for identification of faces with expressions is targeted by augmenting the dataset with synthesized images. Several expressions are simulated for each enrolled person on an animatable model which is specifically generated based on the 3D face scan of that subject.

For the animatable model generation, a generic model for which MPEG-4 FDPs are located manually is utilized. Based on only 17 common points on both the generic and the target models, the generic model is first coarsely warped using the TPS method. Then,

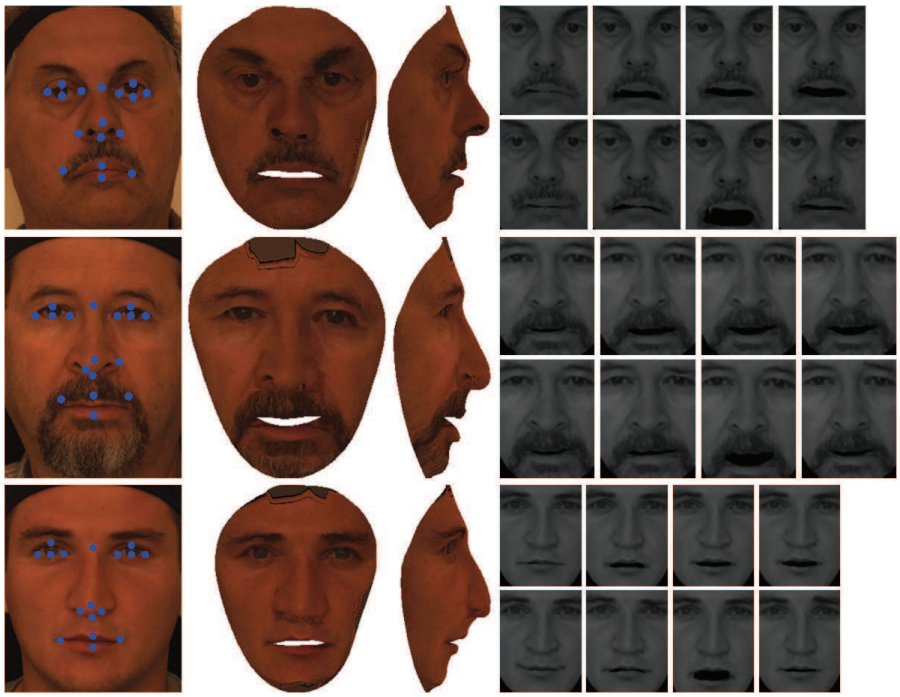


Figure 46: Three examples with high errors in automatic detection of landmark points, the resulting animatable model after coarse warping (frontal and profile) and final images with simulated expressions after fine warping

assuming the surfaces are close enough, new and denser point correspondences are formed by pairs with minimum distance and fine warping is applied. Finally, the texture is copied.

A sub-procedure on automatic detection of those 17 landmarks is presented utilizing both 2D and 3D facial data.

For the simulation of facial expressions on the generated models, an animation engine, called visage|life™ is utilized. The facial images with expressions constitute a synthetic gallery, of which the contribution to the face recognition performance is evaluated on a PCA-based implementation.

The experiments are conducted on two large and well-accepted databases; FRGC and Bosphorus 3D face database. The experiment results reveal that introduction of realistically synthesized face images with expressions improves the performance of the identification system. Additionally, based on the evaluations on Bosphorus database, the error introduced by the proposed automatic landmarking algorithm does not have an adverse effect on the success rates thanks to the corrective property of the warping phase.

Chapter V. Discriminative Properties of Warping Parameters

Due to the advances in the acquisition systems, three-dimensional (3D) facial shape information has been increasingly used for human face recognition. In order to be able to compare different facial surfaces, various registration approaches have been proposed, including Thin Plate Spline (TPS) based algorithms. In this paper, instead of adopting the TPS for registration purposes, we analyze the discriminative properties of the parameters obtained by deforming a generic face model onto target faces using TPS. The warping parameters (WP) that describe the non-global and non-linear transformations and represent the deviations from the common geometric structure are given to the classifier for face recognition. The descriptiveness of those vectors is analyzed on the FRGC database where total of 4569 3D face models are utilized. In spite of its low complexity compared to other proposed approaches, this method yields promising accuracy rates.

1. Introduction

On account of the recent advances in 3D acquisition tools, the decrease in their prices and the higher computational power of computers, 3D facial shape information has become a practical modality in face recognition systems. For most of the 3D face recognition methods which use local similarity metrics, the success rates are highly dependent on the registration step, where dense correspondences are established among faces, in order to be able to make a comparison.

The Iterative Closest Point (ICP) is the most frequently employed registration algorithm ((Pan, et al., 2005), (Irfanoglu, et al., 2004) and (Lu, et al., 2004)) for the alignment of the 3D facial surfaces. By regarding them as free-form curved surfaces, ICP minimizes the distance between two point clouds while preserving the rigidity. However, initialization before starting the iterations is a crucial problem, since the ICP algorithm converges to a local minimum monotonically.

A surface matching framework that takes into account both rigid and non-rigid variations is proposed in (Lu, et al., 2005), where ICP is extended by applying a Thin Plate Spline (TPS) warping algorithm, which can establish registration in non-rigidly deformed surface patches. After originally proposed in (Bookstein, 1989) by Bookstein, this elegant mathematical framework has been extensively utilized to define the deformation of 3D surfaces ((Hu, et al., 2009), (Schneider, et al., 2008) and (Hu, et al., 2009)).

In (Gokberk, et al., 2006), the aforementioned two registration techniques are compared through 3D face recognition via some of the most commonly used features

such as point clouds, facial profiles, curvature-based features, depth maps and surface normals. The experiment results indicate that warping the faces in 3D causes some losses in the discriminative shape information. The deformation of the probe models, even though it is minimal, is proven to be disadvantageous to the recognition performance.

As more and more control points are used for the TPS warping, the amount of deformation increases and the face becomes more and more similar to the target surface.

By exploiting this fact, in this study we propose to utilize warping parameters as a biometric signature for 3D face recognition. Instead of minimally deforming the faces to a common target, a common (i.e. generic) face is strongly deformed to fit target faces. In this case, there exists a tradeoff between the number of control points (and hence the accurate approximation of the facial surface) and the alignment between faces. If all points on the generic model are used as landmarks, it completely deforms into the face. However, in this case the registration data is lost.

1.1. Related Work

In (Ju, et al., 2001), Ju et al., proposes to fit a generic human animation model to the 3D images of real people in order to achieve highly realistic human animation models. Correspondences are established for segmented surface conformations in two major steps: global mapping and local deformation. Based on this work, Mao et al. (Mao, et al., 2004) later develop a similar technique to construct dense correspondences for 3D facial analysis in which a generic model is mapped onto a 3D surface of a face model as opposed to forming correspondences between the raw 3D scans directly. As is the case in (Ju, et al., 2001), global mapping is achieved by Radial Basis Functions (RBF) based on manually selected landmark pairs. On the other hand, for the local deformation, as an extension to the previous work, a similarity measure is proposed for finding the most similar point on the raw scanned data for each vertex of the generic model. The similarity measure is calculated by using the surface distance, normal and curvature. Afterward, the generic model is deformed based on the selected reliable point pairs. The test results are reported to give better results than (Ju, et al., 2001).

However, in a more recent work (Tena, et al., 2006) that presents a large scale evaluation of the face registration method proposed in (Mao, et al., 2004), it is shown that the curvature shape index does not help to find the matching of compatible features, but the matching of compatible shapes, and hence degrades the quality of the fit. It is also stated that, given the variability of the face features, it is difficult to discriminate outliers from a large face variation. For this reason, better results are achieved by removing the reliable point selection from the algorithm.

1.2. Proposed System

In this study, we propose a system which employs parameters that are extracted from a mesh warping algorithm based on the findings in (Tena, et al., 2006), as biometric signatures. A generic head model (that is also used in the previous chapter), with holes for

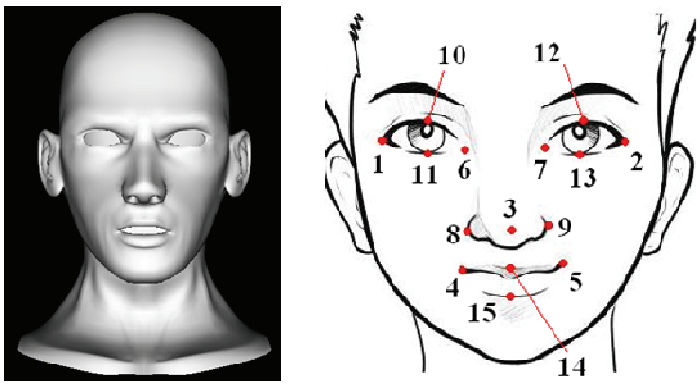


Figure 47: The generic head model and the manually annotated landmarks

the eyes and an open mouth, (shown in **Figure 47**) is strongly deformed to fit the facial models in the database, using the TPS method. The approach is similar to but much simpler than the work presented in (Kakadiaris, et al., 2007), in the sense that an annotated model of the human face is fitted to each individual data. However, our aim is to analyze the discriminative properties of the warping parameters obtained during fitting (which have mostly been overlooked as superfluous sediment in registration applications), rather than utilizing the output of the fitting process.

When a generic model is deformed to fit an individual face in the database, an approximate representation of the facial surface is obtained. If we think of the generic model as the common geometric structure of all faces, these parameters represent the deviations from the conventional. Therefore, they may be claimed to possess concentrated information about the facial shape.

In the next section, the proposed feature extraction method based on TPS warping is explained in detail. After that, an extensive analysis on the descriptiveness of the TPS parameters is presented with the experimental setup and the achieved results.

2. Feature Extraction

The generic model is a symmetrical head model which does not give an impression of any gender. Based on the manually labeled landmarks provided by Szepycki et al. (Szeptycki, et al., 2009) on both target faces and the generic model (**Figure 47**), firstly an alignment is applied. Then the local deformations are replicated using the TPS method.

2.1. Thin-Plate Splines

In this section, for the sake of completeness, we will briefly give a short explanation on the mathematical background of TPS again.

For the 3D surfaces S and T, and a set of corresponding points on each surface, Pi and Mi respectively, the TPS algorithm computes an interpolation function f(x,y) to compute T', which approximates T by warping S:

$$T' = \{(x', y', z')\ st. \forall\ (x, y, z) \in S, x' = x, y' = y, z' = z + f(x, y)\} \quad \text{Equation 5}$$

$$f(x, y) = a_1 + a_x x + a_y y + \sum w_i U(|P_i - (x, y)|) \quad \text{Equation 6}$$

with U(.), the kernel function, expressed as:

$$U(r) = r^2 ln(r^2), r = \sqrt{x^2 + y^2} \quad \text{Equation 7}$$

Given in (6), the interpolation function f(x,y) includes the warping coefficients: w_i, $i \in \{1,2,...n\}$ to be utilized. It consists of two distinct parts. The affine part ($a_1+a_x x+a_y y$) which accounts for the affine transformation necessary for the surface to match the constraint points and a warping part ($\Sigma w_i U(|P_i-(x,y)|)$).

2.2. Rescaling and Alignment

Before warping the generic model, a linear transformation is computed in a least square sense (LSS), based on two sets of landmarks (Huang, et al., 1986). The best fit mapping is calculated by minimizing the squared distance between the point sets. Then, the obtained transformation that includes rotation, translation and isotropic scaling is applied onto the generic model, aligning it with the subject's face.

Due to the difficulties and accuracy problems in automatic extraction of facial interest points, less number of landmarks to be used is always more desirable. An in-depth analysis is presented in the experiments section, demonstrating the trade-off between number of points used and the recognition rate. An additional ICP alignment becomes necessary if the number of points is too few. It is shown that ICP stabilizes the recognition rate independent of the number of points used.

2.3. Warping

After the generic model is rescaled and aligned, it is ready to be warped to compensate for the remaining local non-rigid transformations.

In the related works mentioned before, (Ju, et al., 2001), (Mao, et al., 2004) and (Tena, et al., 2006), on which this study is based, establishing the dense correspondence is achieved in two stages. In the first stage, called global mapping, two sets of landmarks that consist of small number of points are perfectly aligned using TPS interpolation, namely the generic model is warped coarsely. In the second stage, for each vertex on the generic face, the most "similar" vertex on the target scan is found within a search radius of 10mm. However, different from these cited works, our aim is to represent facial surfaces with the generic model and a set of parameters. Hence, a two-step warping, in which a different intermediate state is achieved for each facial surface after global warping stage, is not suitable for our case. Warping results after the first step will not be correlated and hence will not be comparable. For this reason, we proceed with a single-step TPS warping.

Similarity between vertices on generic (G) and test (T) models is determined by the weighted sum of different metrics (4), where D is the distance between them, N is the

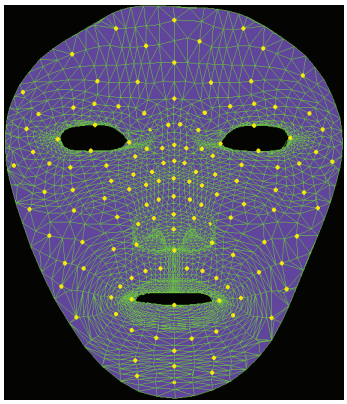


Figure 48: 140 points marked on the generic model

angle between surface normals at those points and C is the difference between curvature shape indices.

$$S = -\alpha D - \beta\left(\frac{arccos(N_G \cdot N_T)}{\pi}\right) - \gamma(C_G - C_T) \qquad \text{Equation 8}$$

The vertex pairs that are more "similar" than a threshold are used in the energy minimization process that warps the generic model to the target face.

However, after being evaluated on the FRGC database, the curvature shape index is proven to have a detrimental effect on the results (Tena, et al., 2006). Taking into account also that the surface normals on the relatively noisy part of the face like eyes and nose are less reliable, we return back to the initially proposed matching criteria, which was based solely on distance (α=1, β=0, γ=0).

Additionally, in (Tena, et al., 2006), similarity (in our case distance) limitation is also removed from the algorithm since better fits can be obtained without it.

To construct point pairs, 140 points are manually selected on the generic face (**Figure 48**) and they are coupled with the closest vertices in the target surface after alignment. Using those point pairs, TPS warping is applied to the generic head model. The function f(x,y) given in (Equation 6) performs interpolation in z direction. Since the generic face is rescaled and aligned beforehand, the affine part is ignored and the weights w_i are taken into account. When we transpose the formula for the other two directions, as a result we obtain the warping vector: [(w_{1x},w_{1y},w_{1z}), (w_{2x},w_{2y},w_{2z})… (w_{140x},w_{140y},w_{140z})]. It is a dense representation of facial surfaces, since the bulk of the data is stored in the generic model.

In this study, we investigate the performance of these warping parameters as a feature vector to be used in a biometric system.

2.4. Distance Metrics

Feature vectors for each face are structured as a 140x3 matrix, in which for each control point there is a 1x3 warping vector. In order to measure the distance between facial surfaces, the cosine of the angle and the Euclidean distance between these two

warping vectors are calculated. This results in two 140x1 distance vectors for the compared face pair.

Two distance vectors are calculated for each face pair in the gallery. By measuring the central tendency of these vectors, two distance values are assigned. In order to avoid the sensitivity to the outliers, (e.g. mean value can be easily corrupted with a small number of extreme values.) trimmed mean approach is adopted during the experiments. Additionally, for each control point, N closest persons are assessed and for each person in the gallery, a weight is computed based on its occurrence count in those possible solution sets. The two distance values are fused by addition after being normalized to [0, 1] range.

The proposed feature extraction scheme is presented and the warping process is illustrated on a sample, in **Figure 49**. Additionally, the frontal and profile 2D/3D images of the original sample face and the generic model at different stages of the algorithm are given for comparative visualization.

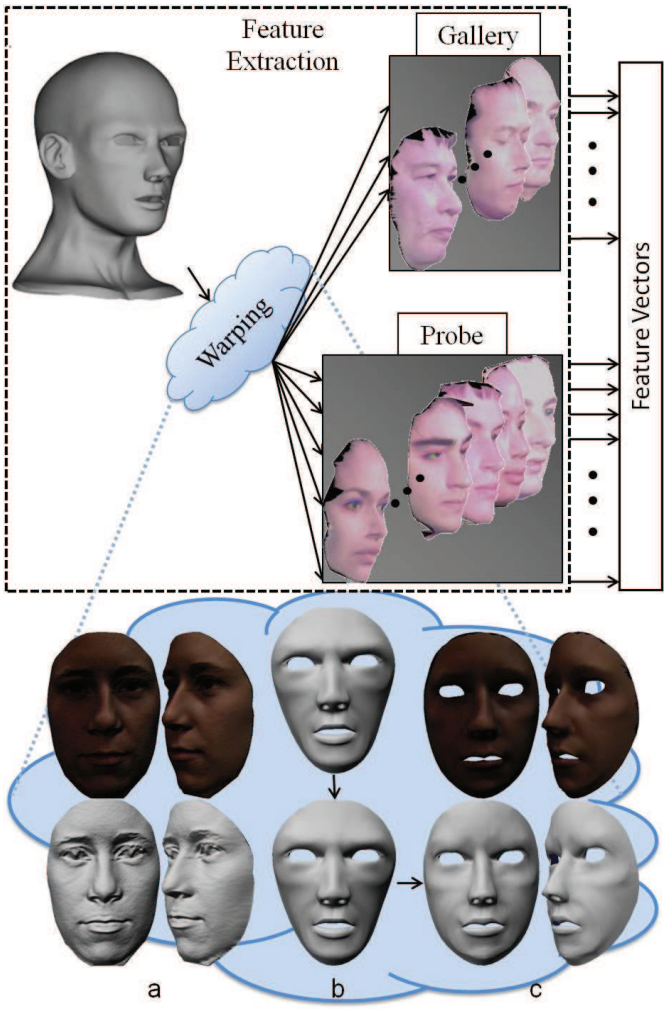


Figure 49: The proposed feature extraction scheme and an illustration on a sample model: (a)The target model with and without texture (b) generic model before and after alignment (c) generic model after warping with and without texture

3. Experiments and Analysis

In order to test the discriminative properties of the extracted feature vectors, we worked on both versions of the FRGC database (Phillips, et al., 2005) where the first version is used as the training partition to analyze different techniques and to obtain optimal parameters.

In FRGCv1, the persons with at least 2 images constitute a gallery of 198 classes with a single scan per person. The remaining 668 face scans are appointed to the probe set. The most important property of this version of FRGC is that all facial scans are with a neutral appearance.

For FRGCv2, in which facial models can be either with or without an expression, the gallery is composed of persons with multiple scans in the database with at least one of them being neutral. For 401 subjects that fulfill this condition, there are 2164 neutral and 1359 non-neutral test scans in the probe set.

At the beginning, the face scans in the whole database are preprocessed as explained previously in the related chapter. After that, the generic model is aligned to each face and finally, TPS warping is applied, resulting in feature vectors of size 140x3 for all face models.

The evaluation is mainly based on three performance metrics: rank-1 identification rate (IR), verification rate with false acceptance rate (FAR) at 0.1% (VR) and equal error rate (EER).

3.1. Experiments with FRGCv1

When comparing two faces, based on 140 control points, the cosine of the angle and the Euclidean distance between two warping vectors are calculated. In order to obtain a single distance metric from all control points, trimmed mean approach is adopted, where a portion of the values on both extremities are ignored.

The recognition tests are conducted both with and without an additional ICP alignment and the results is given in **Figure 50** for Euclidean (TM_{euc}) and Cosine (TM_{cos}) distances and their fusion by sum rule. With the additional ICP alignment, the success rates become independent of the number of landmarks and in fact, the best result (95.81%) is achieved using only 3 points (**m2**). On the other hand, without ICP, the recognition rates increase with more points. The best rate is 96.41% with 15 points (**m1**).

These results are obtained with non-normalized distances and without using weights (C). With the addition of weights and normalization (Equation 9), mentioned in Section 2.4, the accuracy ameliorates. Details are given in **Table 8**.

$$D(i) = \frac{TM_{cos}(i)}{C(i)} + \frac{TM_{euc}(i)}{C(i)} \qquad \text{Equation 9}$$

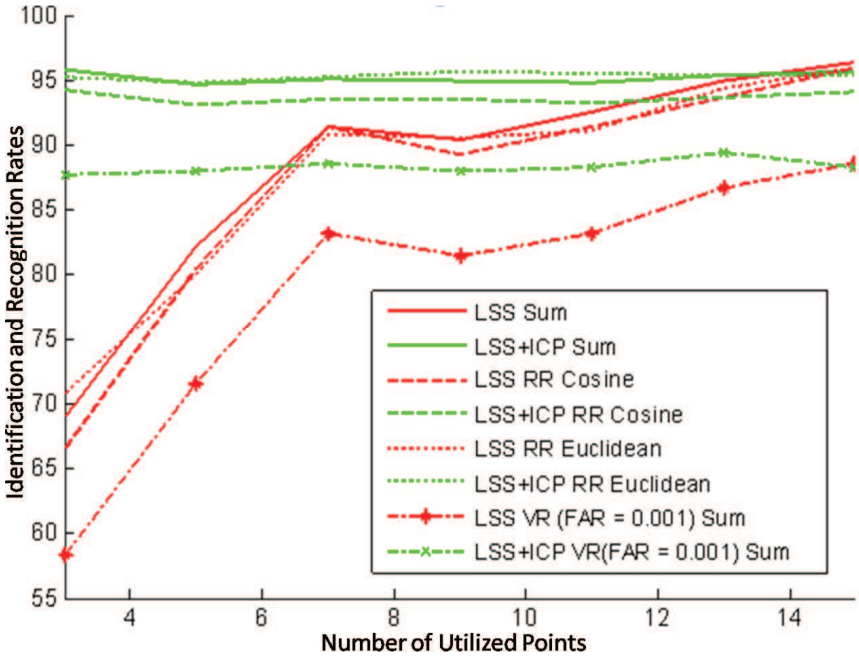


Figure 50: The identification and verification rates with increasing number of landmarks used for alignment

Table 8: The identification and verification (at 0.001 FAR) rates and EER values for FRGCv1 by adding Cosine and Euclidean distances; with and without weighting and normalization

	IR	VR	EE
m1	96.41	88.62	0.021
m2	95.81	87.72	0.030
m1 with weights + normalization	97.30	99.10	0.005
m2 with weights +	97.31	98.65	0.003

In a similar approach (Cadavid, et al., 2009), after transforming a generic model to each face model, adaboosted geodesic distances are calculated additionally for face recognition. The rank-1 identification rate on FRGCv1 is reported as 95.0%. With our much less complex method, we reach 97.31%.

The experiments on FRGCv2 are realized with the parameters that optimize the rank-1 identification rate on FRGCv1. For trimmed mean method, the portion to be ignored is taken as 10% and for the calculation of the coefficients, the possible solution sets are created by rank N, where N equals to 20% of the gallery size.

3.2. Experiments with FRGCv2

After the tests and analyses on FRGCv1, the resulting system is applied to the second version. For this experiment, three categories of facial expression are defined: Neutral, small and large (Maurer, et al., 2005). This approach is adopted because the real concern

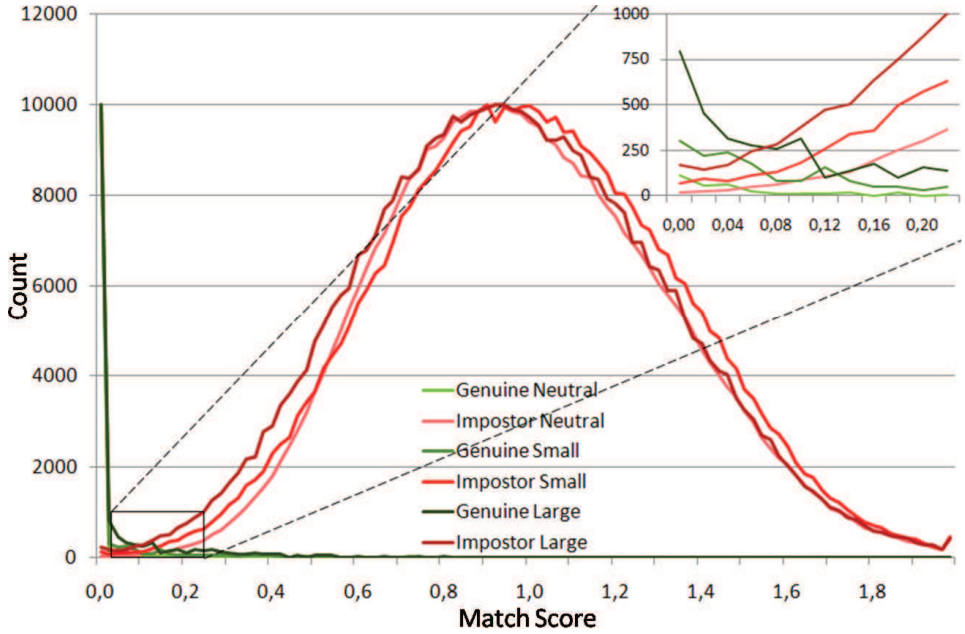


Figure 51: Impostor and genuine score distributions for all comparisons in 3 subsets: neutral, small and large

in recognition is more likely the amount of the shape change rather than its type. Small expressions include moderate smiles and talking gestures. On the other hand, large contains unnatural expressions like blown cheeks. This classification results in a neutral probe set of size 2053 (~60%), small probe set of size 747 (~20%) and large probe set of size 723 (~20%).

In **Figure 51**, genuine and impostor score distributions for all-vs.-all comparison in FRGCv2 is given for 3 subsets. As the figure implies, a decline occurs in the accuracy as stronger expressions are introduced into the probe sets. This is mainly due to the control points around the regions that are sensitive to facial expressions.

Performances of the two approaches: the first one with 15 points and without ICP (m1) and the second one with 3 points and ICP (m2) are compared in **Table 9**. The results obtained for both identification and verification (at 0.001 FAR) are given with EER values. As expected, the impact of the expression variations is also visible on the success rates

Table 9: The identification and verification (at 0.001 FAR) rates and EER values for the whole FRGCv2 database (A) and for three subsets: Neutral (N), small (S) and large (L).

	IR m1	VR m1	EER	IR m2	VR	EER
N	97.13	99.37	0.003	92.40	95.37	0.015
S	85.81	91.70	0.018	84.74	88.22	0.031
L	71.92	76.35	0.033	72.75	75.93	0.050
A	89.55	94.52	0.015	86.74	90.92	0.026

3.3. Robustness to Degradation

For the third set of experiments, a gallery of 410 persons is constructed with a single neutral face scan per individual. For the probe set, one test scan is selected randomly from the FRGCv2 database which can be with or without expression and 9 degraded versions of this probe set are created: 3 decimated versions (by 2, 4 and 8), 3 noisy versions (Gaussian noise added to z, with standard deviation of 0.2, 0.4 and 0.8mm) and 3 other versions with 1, 2 or 3 holes on the facial surface (**Figure 52**). The identification test is conducted on those sets to evaluate the robustness of the extracted feature vector against various degradations (using m1). The results are shown in Table 10. The identification rate for the original probe set is obtained as 89.05%.

Table 10: Rank-1 identification rates for different probe sets with various degradations

Type	Rate	Type	Rate	Type	Rate
Decimation 2	86.86%	Noise 0.2	84.43%	1 hole	88.57%
Decimation 4	81.75%	Noise 0.4	77.13%	2 holes	88.08%
Decimation 6	72.02%	Noise 0.8	61.56%	3 holes	87.11%

According to the results, the extracted warping parameters can still perform well with slightly low sample rate or mild noise. But as the degradation becomes stronger, the accuracy begins to drop for both decimation and noise, whereas it does not get affected strongly by the introduction of holes on the surfaces.

4. Conclusion

This study presents an evaluation on the distinctive properties of warping parameters that are obtained by aligning a generic model and deforming it to face samples using TPS algorithm.

The alignment procedure is analyzed using different number of landmarks and with/without an additional ICP module. ICP is observed to stabilize the recognition rates as the number of points of interest decreases.

The evaluation is performed on FRGC database with total number of 4596 face scans. A series of tests is conducted on FRGCv1 to analyze different alignment approaches and to obtain optimal parameters for the trimmed mean method and the computation of weights.

FRGCv2 is utilized in order to evaluate the robustness of the extracted feature vectors against facial expressions and degradations; such as noise, holes and decimation. The warping parameters are shown to be robust against moderate decimation and noise and

Figure 52: Degradation illustrated on a sample face scan

the presence of holes on the facial surface. On the other hand, facial expression variations are observed to be highly detrimental.

In the next chapters, we will explore the quality of the facial surface patches to obtain a reliability score for one point or groups of control points, aiming robustness against expression variations.

Chapter VI. Regional Reliability Estimation for 3D Face

3D shape data for face recognition is advantageous to its 2D counterpart for being invariant to illumination and pose. However, expression variations and occlusions still remain as major challenges since the shape distortions hinder accurate matching. Numerous algorithms developed to overcome this problem mainly propose region-based approaches, where similarity scores are calculated separately by local regional matchers and fused for recognition. In this chapter, we present a regional confidence score assessment scheme that estimates the expression or occlusion induced distortions in different facial regions. Thereby, reliability scores are obtained which can be used in fusion step for recognition. For 7 regions of face, primitive shape distributions are extracted and the surface quality is measured automatically by an Iterative Closest Point (ICP) based method. Using these measurements, an Artificial Neural Network (ANN) is trained and utilized to estimate regional reliability scores. Experiments have been conducted on FRGC v2 3D face database and results demonstrate a high accuracy in surface quality estimation.

1. Introduction

3D face recognition becomes more popular each day as it offers superiority over 2D face recognition by being intrinsically robust against illumination and pose variations. However, despite many algorithms developed, the deterioration in the recognition accuracy due to intra subject deformations that are introduced by facial expressions and occlusion still needs to be handled.

Approaches aimed to solve these issues are seemed to be gathered around two main directions: extraction of distortion-invariant shape descriptors and region-based matching.

Among the methods to extract distortion-insensitive features, (Smeets, et al., 2009) proposes the geodesic distance matrix as a representation of 3D face and claims that the set of largest singular values is an expression-robust descriptor. In (Xiaoxing, et al., 2009), expression-insensitive low-level geometric features are ranked by their *confidence* which is calculated in a fashion similar to Fisher's Linear Discriminant. An expression-insensitive descriptor is formed by the higher-ranked features.

Region-based face recognition methods can be again subdivided into two classes. The first class relies on the facial regions that are least affected by the distortions. For example in (Cheng, et al., 2008), (Kyong, et al., 2005) and (Gunlu, et al., 2010), a relative robust

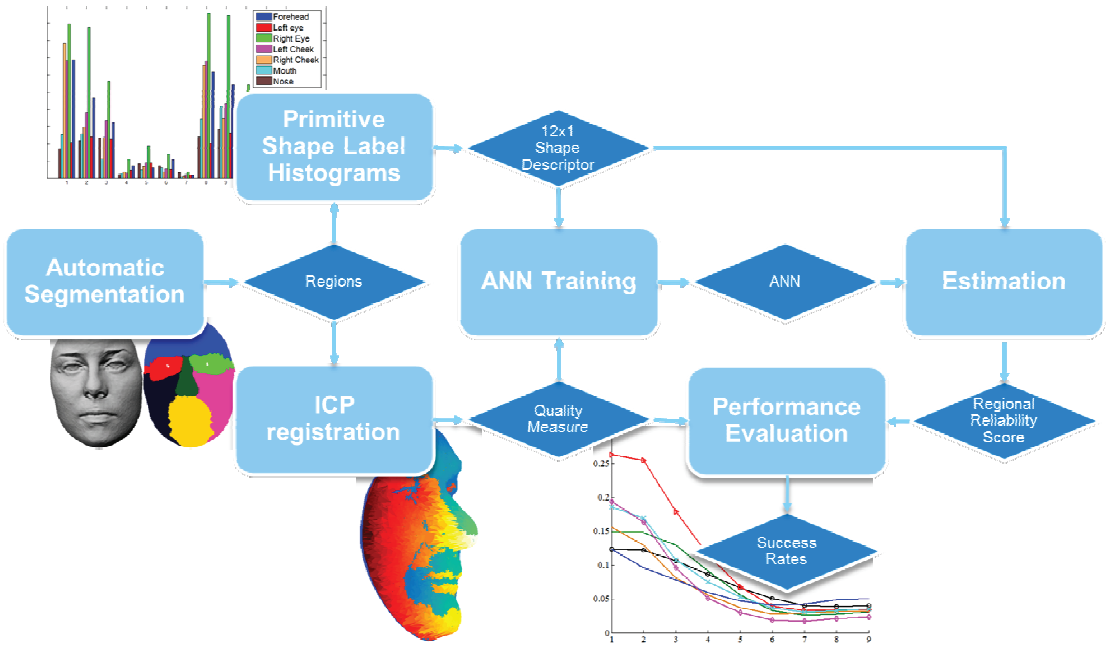


Figure 53: Flowchart of the proposed regional reliability estimation method and its evaluation

area is selected around the nose and eyes. In (Ben Amor, et al., 2008), a preliminary analysis is proposed to weigh the regions according to their stability by studying the anatomical face. The second class separately matches all regions of face and makes a decision based on the fusion of all scores. In (Wei-Yang, et al., 2007), Linear Discriminant Analysis is utilized for optimal linear fusion of scores from 10 different pre-defined regions. In (Alyuz, et al., 2008), face segments are registered individually to specific Average Region Models and point-set differences of all regions are combined for decision-making. Among many different fusion techniques evaluated, the product rule is concluded to be performing best.

In this chapter, we present a scheme to evaluate different regions of face according to the present distortions due to expressions and occlusions. First, the face is divided into 7 components (forehead, left eye, right eye, left cheek, right cheek, mouth, and nose) and for each component; every vertex is labeled as one of the 12 primitive shape categories. The distributions of these labels in each region are taken as shape descriptors to determine the quality. Here, quality refers to presence of occlusions or expressions, rather than mesh resolution, noisiness, etc. Next, for every face model, each region is registered to its neutral and clean equivalent which belongs to the same person, using the ICP algorithm (Besl, et al., 1992). Accepting ICP registration errors as the approximations of volumetric differences between two regions and hence, as the deviations from the undistorted versions, these errors are taken as metrics of imperfection. Finally, an ANN is trained to map between primitive shape distributions (input set) and ICP registration errors (target set).

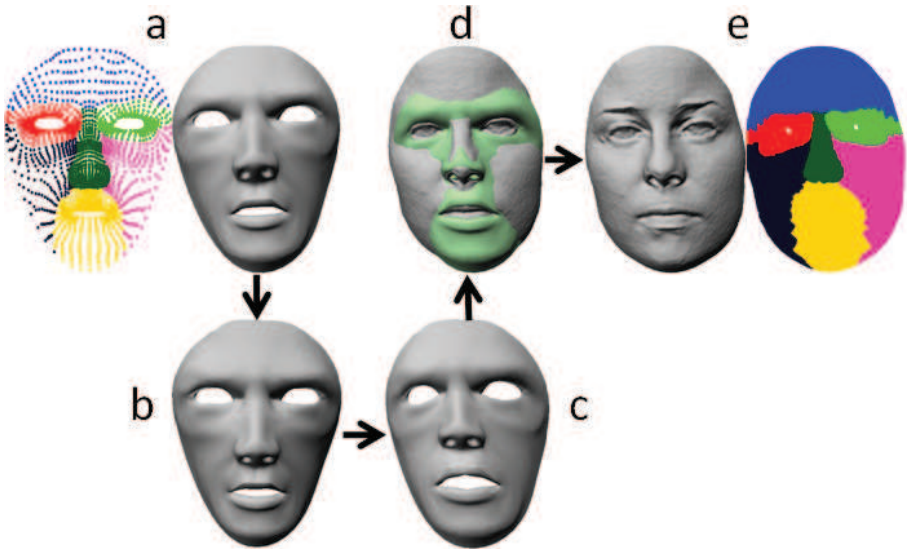


Figure 54: (a) The generic face and segmented regions (b) after alignment (c) after TPS warping (d) resultant generic face and the model to be segmented superimposed (e) the model to be segmented and the obtained regions

2. Automatic Segmentation

For the region-based analysis, automatic segmentation of the face is crucial. For this purpose, a generic neutral 3D template which is previously segmented into 7 regions is utilized. After matching it to a model in the database, the segmentation of the model is achieved by mapping the region labels.

In order to establish dense correspondence between the face model under analysis and the generic face, 3 manually-marked fiducial points are utilized: outer corners of the eyes and the nose tip. There exist many algorithms capable of automatically detecting these points and additionally an automatic landmarking system is proposed in the previous chapters.

First, the generic face is rigidly aligned to the inspected model. A linear transformation is computed in a least square sense, based on two sets of landmarks (Huang, et al., 1986). The best fit mapping is calculated by minimizing the squared distance between the point sets. Then, the obtained transformation that includes rotation, translation and isotropic scaling is applied onto the generic model.

Next, the alignment is further improved for non-rigid deformations by applying TPS warping onto the generic model still using the same 3 point pairs. Finally, since the two surfaces are well-aligned, the closest point on the generic face is found for each vertex on the face to be segmented and its region label is copied. The proposed method is illustrated in **Figure 54.**

3. Primitive Shape Labeling

Utilization of primitive shape distributions within different facial regions is suggested in (Wang, et al., 2006) for recognition of facial expressions. In this study, we use similar features not to recognize expressions but to assess the quality of the face segments.

To determine the primitive shape class of each vertex, firstly, gradient magnitudes for all vertices on the facial triangular mesh are computed. For this purpose, initially normals of neighboring triangles are calculated for each point. Then, based on the distances between the vertex and the centroids of the triangles, the weighted average of these normal vectors is taken and assigned as the normal of that vertex. Finally, for a vertex with a normal vector of $n_p = (a,b,c)^T$, the gradient magnitude is calculated as:

$$G = \sqrt{\left[\left[-\frac{a}{c}\right]^2 + \left[-\frac{b}{c}\right]^2\right]} \qquad \text{Equation 10}$$

After that, in order to compute the maximum and minimum curvatures, for each vertex on the facial surface, the adjacent points are found and transformed to a local coordinate system, where that vertex becomes the origin and its normal becomes the unit vector along the positive z axis. To fit a smooth second order polynomial patch on (X, Y, Z), where X, Y and Z are column vectors of x, y, z coordinates of the transformed adjacent points, a least-square solution that minimizes the residual error is computed for the following equation:

$$A \times \left[\frac{1}{2}X^2 \; XY \; \frac{1}{2}Y^2 \; X^3 \; X^2Y \; XY^2 \; Y^3\right] = Z \qquad \text{Equation 11}$$

Lastly, the eigenvalues of the Weingarten matrix W = [A(1) A(2); A(2) A(3)] are found. The first eigenvalue (the one with the largest absolute value) gives us the amount of greatest curvature whereas the other one gives the least. Therefore, the first and the second eigenvalues are taken as maximum and minimum curvatures, respectively.

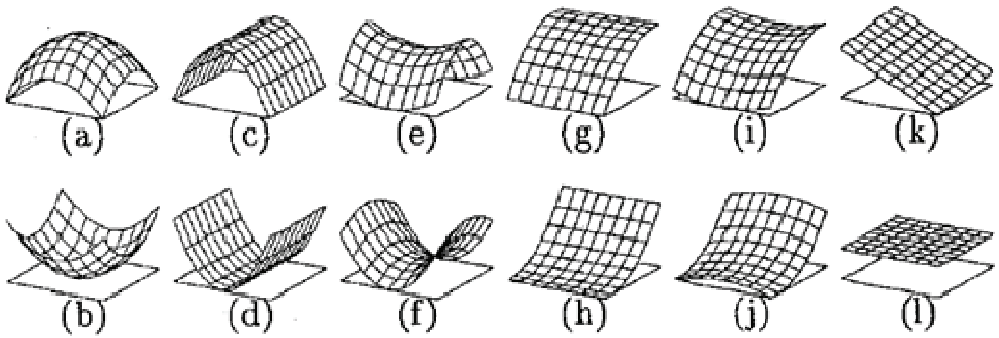


Figure 55: Topographic labels for the center vertex in each example: (a) peak; (b) pit; (c) ridge; (d) ravine; (e) ridge saddle; (f) ravine saddle; (9) convex hill; (h) concave hill; (i) convex saddle hill; (j) concave saddle hill; (k) slope hill; and (I) flat. (Trier, et al., 1995)

Once the gradient magnitude (G – the steepness of the surface) and maximum and minimum curvatures (k_1 and k_2 – maximum and minimum degrees of bending, respectively) for a vertex are obtained, they are utilized to classify it as one of the twelve primitive shapes shown in **Figure 55**, according to the rules given in **Table 11**.

Table 11: Classification rules for the vertices according to magnitudes of their gradients and principal curvatures

		G = 0	G ≠ 0
$k_1 = 0$	$k_2 = 0$	flat	slope hill
$k_1 = 0$	$k_2 < 0$	ridge	convex hill
$k_1 = 0$	$k_2 > 0$	ravine	concave hill
$k_1 < 0$	$k_2 < 0$	peak	convex hill
$k_1 < 0$	$k_2 > 0$	ravine saddle	concave saddle hill
$k_1 > 0$	$k_2 < 0$	ridge saddle	convex saddle hill
$k_1 > 0$	$k_2 > 0$	pit	concave hill

The decision for the gradient and principal curvature magnitudes to be equal to zero is made based on a threshold $\varepsilon = 0.001 \cdot s$, where s is the average distance from each vertex to the mean of whole vertices. If the magnitudes of the gradients and curvatures are smaller than this threshold, they are accepted as zero.

Finally, histogram distributions of primitive shapes for each region are calculated, resulting in 7 shape descriptors of size [12x1].

4. Surface Quality Measurements

In this part of the study, an automatic system is developed to measure surface qualities of the facial regions. The outputs of this system together with the corresponding primitive shape distributions are utilized to train an ANN and to evaluate the final scheme.

To this end, each region of all facial scans of a subject is registered to the corresponding region of a neutral (without expression) and clean (without occlusion) scan of the same person (reference model). The final registration error obtained is accepted to measure the distortions on the surface, in view of the fact that it gives us the deviations from an undistorted equivalent.

In order to have more accurate registrations, firstly, the faces are aligned rigidly based on 3 manually-marked fiducial point pairs: outer corners of the eyes and nose tip (same points mentioned previously in Section 2 – Automatic Segmentation).

Subsequently, the faces are registered using ICP algorithm which iteratively adjusts the transformation needed to minimize the distance between two point clouds until the

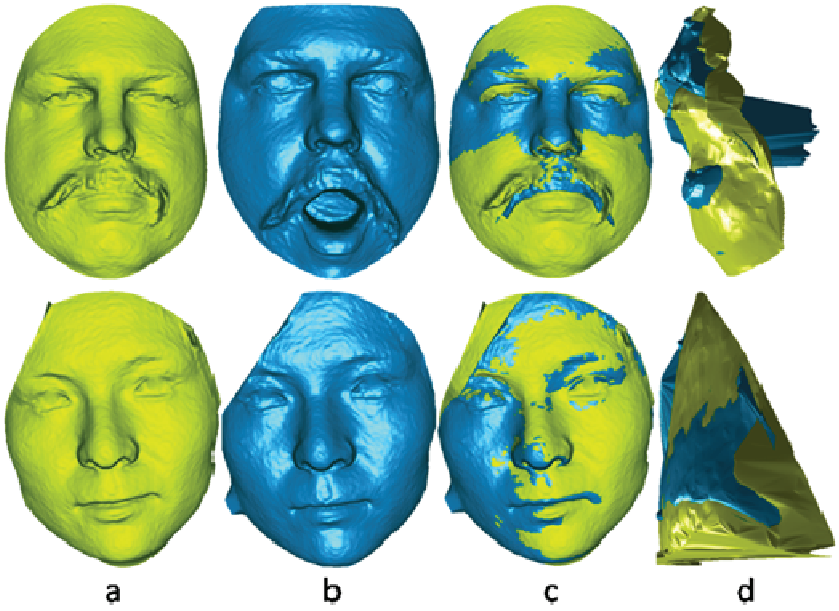


Figure 56: Examples for a bad and a good quality region: (a) neutral and clean reference model; (b) models to be evaluated; (c) model pairs after initial registration; (d) close-up to mouth to region for the 1st example and to forehead region for the 2nd example. The scores computed are 7.67 and 0.96, respectively.

improvement is stabilized. The final distance obtained is adopted as the surface quality metric where a smaller value indicates a more reliable region. Two example regions of a bad and a good quality are given in **Figure 56**.

As the baseline method, the system proposed in the previous chapter is employed which adopts warping parameters extracted from a TPS warping as biometric signatures. The experimental results given in Chapter V – Section 5 have shown that this baseline algorithm is vulnerable against facial expression variations. With our proposed approach of regional confidence estimation, we aim to compensate the disadvantages introduced by such mesh deformations.

5. Experiments

The proposed surface quality assessment scheme is tested on FRGC v2 database (Phillips, et al., 2005). Since a neutral and clean model is required for each person to measure the regional surface qualities automatically, the subjects without a reference model are eliminated. This results in a gallery set of 343 subjects and a probe set of 3123 face models.

For the experiments, the whole sample set is divided into 2 blocks: the first 50% of the samples for training and validation (1562), and the other 50% for testing (1561).

Firstly, the face is automatically segmented and the primitive shape histograms are computed for 7 regions of all face models. For each point on every region, a label is

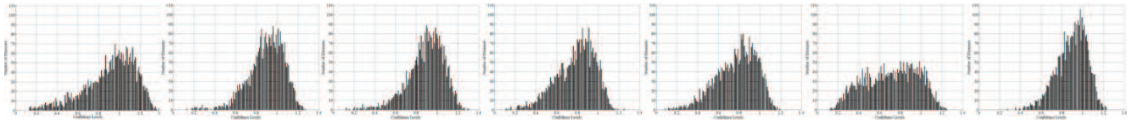


Figure 57: Distributions of confidence levels for each region: forehead, left eye, right eye, left cheek, right cheek, mouth and nose

assigned based on its gradient and principal curvatures. Then, those labels are counted to form a feature vector of size [12x1].

Next, by ICP-based registration of each facial component to the corresponding neutral and clean sample from the same person, regional qualities are measured. The distributions of the obtained quality measurements for different regions are given in **Figure 57**. In the FRGC v2 database, the facial expressions in the non-neutral faces affect mostly the mouth and the cheek regions. Another common distortion is due to hair in the forehead region. For this reason, larger deviations can be seen in the corresponding histograms. The most scarcely distorted part of the face is observed to be the nose region.

It is clear from the given histograms that the distributions of the quality scores are not uniform. In order to supply the training input of the ANNs with data as equally distributed as possible, the quality measurements in the training set is binned into 10 equally spaced containers for each region. The adjacent bins are merged until the minimum number of elements in each container is larger than 20. After that, equal number of samples which is the size of the smallest container, are collected from each bin for training. The obtained partition for the training, validation and test sets are given in **Table 12**.

Subsequently, 7 ANNs to map between primitive shape distributions (12 inputs) and quality measurements (1 output) with a hidden layer of 20 neurons are constructed and trained according to Levenberg-Marquardt optimization until the mean square error (MSE) of the validation samples stops decreasing.

Table 12: Size of the training, validation and testing sets for each region

	Forehead	Left Eye	Right Eye	Left Cheek	Right Cheek	Mouth	Nose
Training	280	216	280	224	235	380	280
Validation	1282	1346	1282	1338	1327	1182	1282
Test	1561	1561	1561	1561	1561	1561	1561

The resulting ANNs are utilized to estimate the regional confidence levels for the test set. The obtained MSE for all regions with the corresponding standard deviations are presented in **Table 13**. Moreover, for the purpose of having a reference point to evaluate the results, minimum and maximum quality scores that are automatically measured; in other words, worst and best registration errors for each region of all face models are provided.

Table 13: Average MSE for regional quality estimations and their standard deviations and minimum scores measured for each region

	Forehead	Left Eye	Right Eye	Left Cheek	Right Cheek	Mouth	Nose
MSE	0.0405	0.0312	0.0345	0.0318	0.0352	0.0511	0.0236
STD	0.0626	0.0463	0.0582	0.0462	0.0524	0.0620	0.0318
Min.	0.1371	0.1305	0.1261	0.1095	0.0896	0.1236	0.2750
Max	1.3852	1.2746	1.3081	1.2557	1.1904	1.2458	1.2277

Additionally, the accuracies of the ANNs with respect to the magnitude of distortions present on the surface are analyzed. MSE values for different bins with increasing quality scores are calculated and illustrated in **Figure 58**. As can be seen, the performance tends to decrease at large distortions; however the accuracy still remains very high. The least accurate estimations are done for the mouth region. This is mainly because the possible deformations for mouth are much more diverse than the other regions.

Once the regional confidence levels are obtained, region-based feature vectors - WP are extracted for comparison using the proposed baseline algorithm.

A generic face is aligned rigidly to the inspected face model and TPS transformation is computed using 140 point pairs: points that are manually marked once on the generic face and their closest neighbor on the face model under analysis. The obtained warping vector is adopted as the biometric signature of the face model.

The test set includes 1561 scans (probe) for 191 subjects (gallery). To identify a probe model, its WP vector is compared with those of all gallery models.

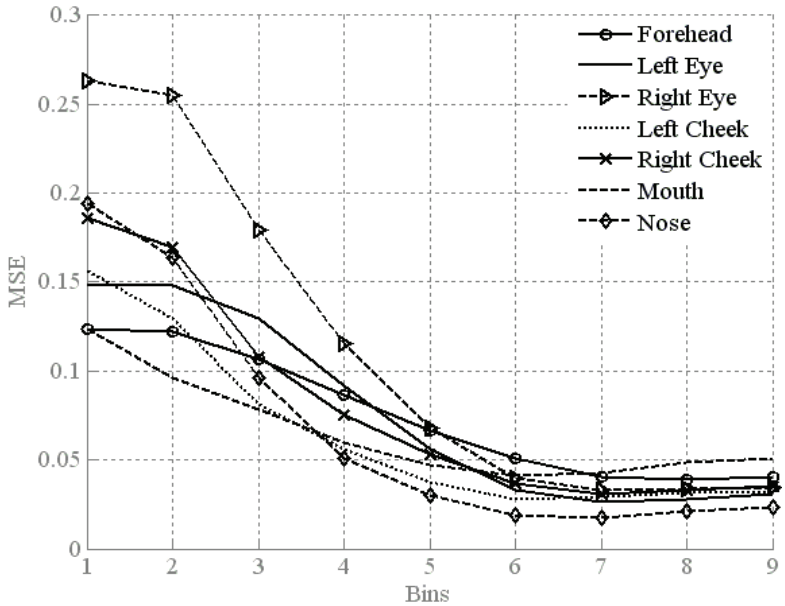


Figure 58: MSEs and success rates calculated separately for different bins, where each bin consists of regions with a quality measurement less than a score threshold.

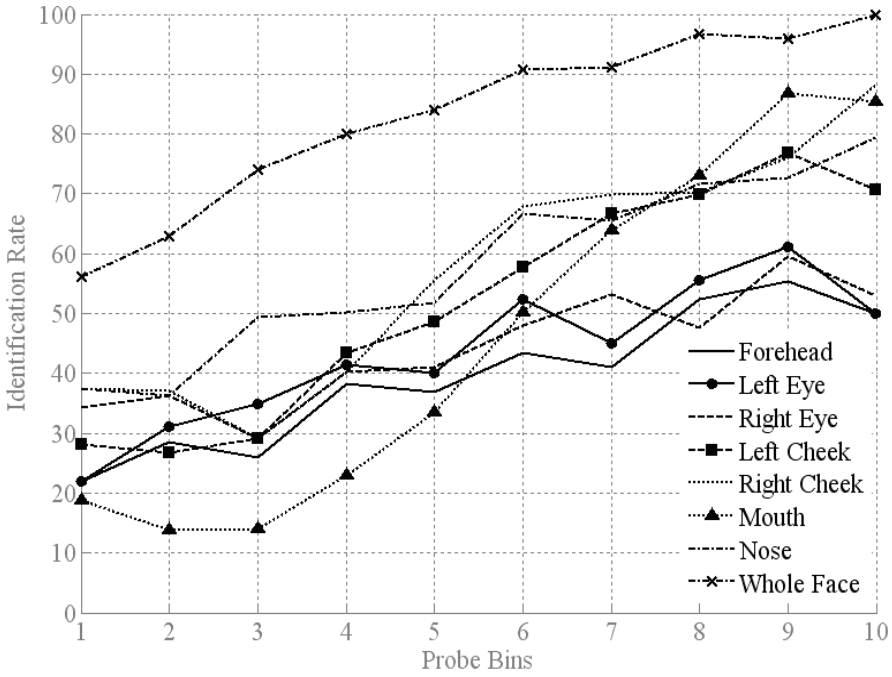


Figure 59: The recognition rates of different regions separately and the whole face for 10 containers of probe images with increasing quality.

The WP vectors are structured as a 140x3 matrices, in which for each control point there is a 1x3 warping vector. In order to compare two models, the Euclidean distance between all warping vector pairs are calculated, resulting in a 140x1 distance vector. The central tendencies of these vectors are measured by adopting the trimmed mean approach in order to avoid outliers. By ignoring 10% from both extremities, average distance is calculated and assigned as the dissimilarity score between the two compared models. The identification is done by choosing the gallery scan with minimum dissimilarity score.

For evaluation, the test set is again divided into bins according to present distortions on the facial surface. For this purpose, the minimum quality scores among all regions for every probe face are taken and mapped between 0 and 1, by min-max normalization. Then, the probe set is separated into 10 equally spaced containers. The recognition rates for each region separately and the whole face is given in **Figure 59** for all bins.

The results show that the extracted feature vectors are highly susceptible to facial distortions. Even though the recognition rate is around 90-100% for the bins with least distorted models, as more irregularities are introduced, the success rate declines down to 56.25%.

In order to test the fusion of 7 regional classifiers, two approaches are investigated:

5.1. Fusion: Using Confidence Scores as Weights

Estimated confidence scores are utilized as weights with two different fusion methods: at the score level - sum rule and one at the rank level - the Borda count.

In the first technique, the combined matching score is computed as the weighted sum of the dissimilarity scores of the individual matchers. As typically required by score level fusion schemes, the dissimilarity scores are normalized before addition. Assume that s_{ij} is the normalized dissimilarity score between a probe model and the j^{th} subject in the gallery by the i^{th} region. The final score S_j is computed as:

$$S_j = \sum_{i=1}^{7} w_i s_{ij} \qquad \text{Equation 12}$$

For the baseline method the weights, w_i, are assumed equal. For the proposed approach, the surface qualities are incorporated by assigning the estimated regional confidence levels to w_i. Finally, the identity decision is made by choosing the subject with the lowest S value. Results for both approaches are given in **Figure 60**. For the bin with the highest distortions, the identification rate is increased from 53.13% to 71.82% (the gain is 35.18%) with the inclusion of the reliability levels. As the 3d face quality improves, the gain decelerates as expected. For the second bin, the improvement is 3.45% and for the third, it is 3.17%.

For fusion at rank level, the Borda count method is adopted. For this purpose, once the dissimilarity scores from 7 regional matchers are obtained, they are arranged in ascending order to form a ranking list among the gallery $_{subjects}$. Assume that for a probe model, r_{ij} is the rank assigned to the enrolled user j by the region i. The final score R_j is computed as:

$$R_j = \sum_{i=1}^{7} w_i r_{ij} \qquad (7)$$

Again, two different identification results are obtained: with equal weights and with the estimated confidence levels. The success rates are compared in **Figure 60**. Overall, the Borda count method is observed to perform less accurately than the sum rule. Nevertheless, the amelioration in the identification rates is more clearly visible in the first three bins. 12.50%, 6.04% and 10.76% increase is achieved for the first, second and third bins, respectively, with the reliability levels included as weights.

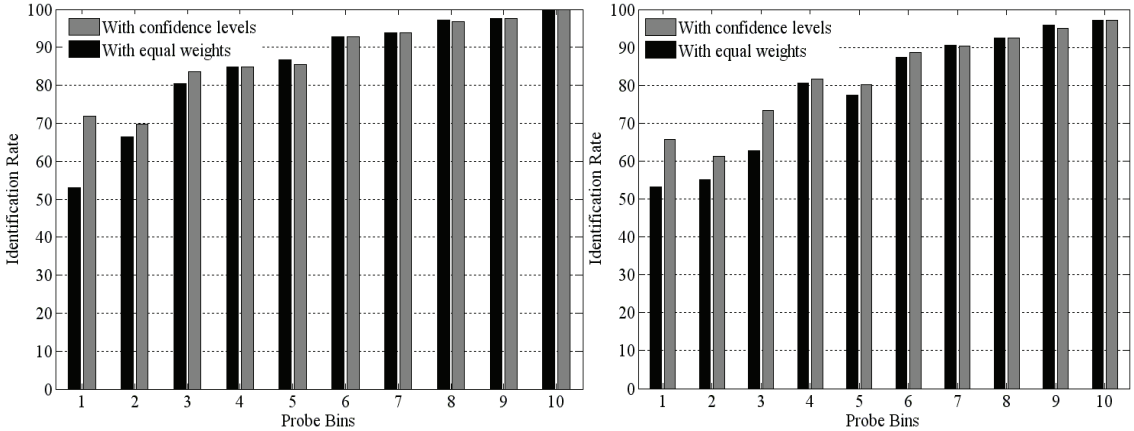


Figure 60: The recognition rates of the sum rule and Borda count based fusion approaches: using confidence levels as weights and assuming equal weights.

5.2. Fusion: Probabilistic Conversions

The performance of a biometric system employing a single trait is constrained by several factors: feature overlaps, noisy inputs, sensor failures, etc. (Ross, et al., 2006) In order to overcome these issues, the use of multiple biometric sources was proposed in (Most, 2003) and has been thoroughly investigated since then, especially on the domain of score normalization and fusion ((Verlinde, et al., 1999) and (Jain, et al., 2005)).

In this section, prior to fusion, we propose to obtain posteriori probabilities, P(genuine|s), of being genuine given the regional matching scores (s). However, as expressed in (Jain, et al., 2005), the outputs of individual classifiers are better combined directly without being converted into probabilities in the absence of confidence measures (c) which assess the nature of the input samples. By taking this into consideration, we obtain P(genuine|s) by computing the conditional densities P(s|genuine) and P(s|impostor) based on the regional confidence scores estimated via our proposed method.

5.2.1. Probability Fusion at Match Score Level

We can represent the fusion of regional match scores as the classifier combination problem formulated in (Kittler, et al., 1998): A face is to be assigned to one of the *m* possible classes in the gallery based on distinct measurements (match scores - s_i) obtained from its R sub-regions. In our approach, this is simplified by converting the class assignment to a binary decision which answers if the match score is "*genuine*", i.e. it is computed between two samples of the same subject or "*impostor*". These two classes are modeled by the probability density functions $p(s_i|\text{genuine})$ and $p(s_i|\text{impostor})$ and their a priori probabilities of occurrence are assumed to be equal. Under this assumption, using the Bayes theorem, P(genuine|s1,s2,...,sR) is expressed as:

$$P(\text{genuine}|s_1,s_2,\ldots,s_R)=\frac{p(s_1,s_2,\ldots,s_R|\text{genuine})}{p(s_1,s_2,\ldots,s_R)} \quad \text{Equation 13}$$

where

$$p(s_1,s_2,..,s_R) = p(s_1,s_2,..,s_R|\text{genuine}) + p(s_1,s_2,..,s_R|\text{impostor}) \quad \text{Equation 14}$$

As suggested in the same study (Kittler, et al., 1998), a posteriori probability of being genuine given R measurements is expressed in terms of decision support computations and possible dependence on joint probability density functions is ignored. This assumption of conditional statistical independence gives way to combining the a posteriori probabilities obtained from different regions by means of a product rule. However, sum approximation of the product rule, despite its foundation on highly unrealistic assumptions, is proven to be more resilient to estimation errors and hence, is adopted in this study (Equation 15).

$$P(\text{genuine}|s_1,s_2,\ldots,s_R) \approx \sum_{i=1}^{R} \frac{p(s_i|\text{genuine})}{p(s_i|\text{genuine})+p(s_i|\text{impostor})} \quad \text{Equation 15}$$

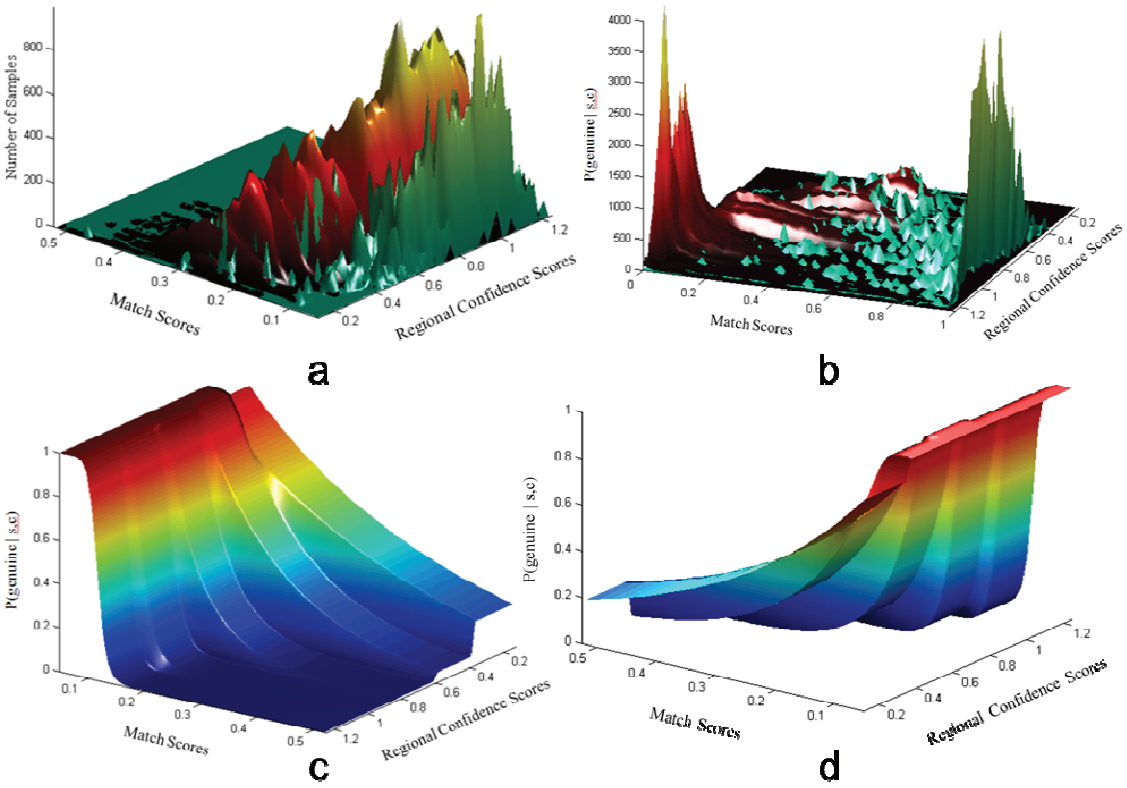


Figure 61: For the mouth region: (a) Histogram of raw match scores with respect to the regional confidence scores for mouth region: impostor scores in red and genuine scores in green (b) the match score distributions after converting them to probabilities (c & d) Calculated a posteriori probabilities for genuineness for 5 confidence level bins from to different view angles

5.2.2. Conditional Density Estimation

As formulated in Equation 15, conditional probability density functions, p(s|genuine) and p(s|impostor) are essential to compute the posterior probabilities of genuineness.

In (Snelick, et al., 2003), the authors utilize mean and variance of genuine and impostor scores in the training set and assume a normal distribution for the conditional densities. However, this assumption may not be true in many cases. For this reason, (Jain, et al., 2005) propose the use of the Parzen window based non-parametric density estimation method and obtain actual conditional densities of genuine and impostor scores.

In our work, the Parzen window is replaced with a smooth Gaussian kernel function in order to avoid discontinuities in the estimations and to take the sample point distances to the estimation point into consideration. The bandwidth is empirically set to 0.03.

Additionally, considering that as the match score increases the probability of being genuine should also increase, a monotonic sigmoidal function (Equation 16) is fit on the kernel density estimation via nonlinear regression.

$$f(x)=\frac{1}{1+e^{a*(b-x)^c}} \qquad \text{Equation 16}$$

The conditional probability densities are estimated with and without considering regional confidence scores for comparison reasons. In case, where the estimated confidence levels (c) are taken into account, P(genuine|s,c) is approximated by dividing confidence scores in several bins and estimating p(s|genuine) and p(s|impostor) for each bin separately. Impostor and genuine score distributions differ greatly for different levels of region quality. With our approach, better estimates are achieved with discontinuous density functions in c-axis.

5.2.3. Evaluation

The evaluation of the proposed fusion is done on the same datasets used in Section 5.1. For each region, conditional distributions are estimated after the confidence scores are binned into several equally spaced containers. The conditional probability densities are estimated for different number of bins for comparison: 1(no bins), 2, 3, 4 and 5. Taking mouth as the exemplar region, the distribution of match scores according to the regional confidence scores, the estimated posterior probabilities P(genuine|s,c) using 5 bins and the match score distributions after being converted into probabilities are given in **Figure 61**.

Verification and identification performances are obtained for the proposed fusion scheme with 1, 2, 3, 4 and 5 bins using the sum rule. Success rates of the original match scores are calculated by the same fusion technique. Additionally, in order to provide equal advantage to the original scores, the regional confidence levels are incorporated via weighted sum combination:

$$S = \frac{\sum_{r=1}^{7} s_r * c_r}{\sum_{r=1}^{7} c_r} \qquad \text{Equation 17}$$

The verification rates at 0.001 FAR, the equal error rates and the rank-1 identification rates for all experiments are given in **Table 14**. Moreover, in **Figure 62**, the verification and identification performances of the fusion techniques are evaluated with receiver operation characteristics (ROC) and cumulative match characteristics (CMC), respectively.

Table 14: Comparative results for verification and identification tests before and after probability conversion of match scores

	method	VR	EER	IR
original	sum	63.39%	0.109	90.33%
original	w. sum	68.22%	0.088	91.04%
probability	1 bin	66.83%	0.072	89.76%
probability	2 bins	71.38%	0.049	91.61%
probability	3 bins	72.66%	0.047	92.13%
probability	4 bins	74.08%	0.046	91.93%
probability	5 bins	73.15%	0.045	92.00%

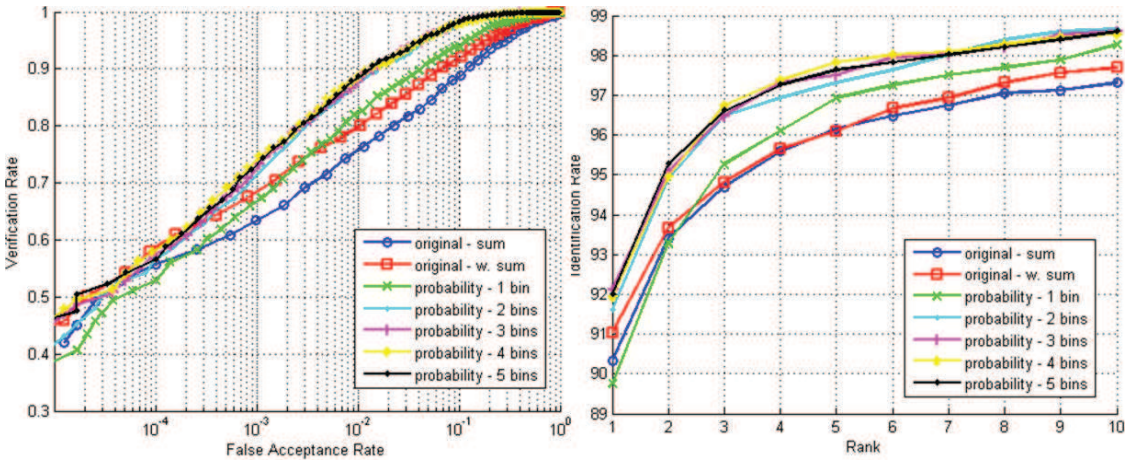


Figure 62: ROC and CMC curves for different fusion methods

The results show that utilization of confidence scores for regional classifiers is more advantageous with the probabilistic fusion approach than employing the raw scores directly. The performances improve as the number of the bins increase but tend to converge after 3 bins. This is mainly due to scarcity of samples and hence erroneous conditional density estimations as the bins get smaller.

6. Conclusion

In this study, a region-based 3D face recognition framework with an additional regional confidence level assessment scheme is presented. After segmenting the face into 7 components automatically, primitive shape histograms are extracted and utilized to estimate regional distortion levels due to expressions and occlusions. The estimations are made by 7 ANNs that are trained separately for each region, using the automatically measured surface qualities. For this purpose, each region is registered to its neutral and clean equivalent of the same subject using ICP and the registration errors are accepted as metrics of imperfection. Finally, a region-based baseline algorithm is adopted for identification in which a feature vector is extracted for matching by warping the facial scans to a generic model using TPS.

Extensive experiments are conducted on the FRGC v2 database. Firstly, the regional confidence level assessment technique is tested. The results reveal that the surface qualities of different facial regions can be estimated with high accuracies. Average MSE at 0.035 is achieved among 7 regions where the regional quality measurements lie in the range of 0.1-1.4.

Next, the estimated confidence levels are included in the fusion step of the identification procedure. For the probe models with high distortions, 35.18% and 23.53% gains are obtained compared to the utilization of equal weights, for the sum rule and the Borda count methods, respectively.

Finally, we incorporated regional confidence scores to the fusion process via the probabilistic framework developed by (Kittler, et al., 1998) through confidence-score based conditional density estimations. Experimental results show that this approach can make better use of the estimated confidence levels compared to utilization of raw scores as weights.

Chapter VII. Impact of Regional Alterations on Face Recognition

Numerous major challenges in face recognition, such as pose, illumination, expression and aging, have been investigated extensively. All those variations modify the texture and/or the shape of the face in a similar manner for different individuals. However, studies on alterations applied on face via plastic surgery or prosthetic make-up which can be in countless different ways and amounts, are still very limited. In this chapter, we analyze how such changes on nose region affect the face recognition performances of several key techniques. For this purpose, a simulated face database is prepared using FRGC v1.0 in which nose in each sample is replaced with another randomly chosen one. Since this is a 3D database, the impact analysis is not limited to only 2D, which is one of the novelties of this study. Performance comparisons of three 2D and four 3D algorithms are provided. In addition, differently from previous works, baseline results for the original database are also reported. Hence, the impact which is purely due to the applied nose alterations can be measured. The experimental results indicate that with the introduction of alterations both modalities lose precision, especially 3D.

1. Introduction

Plastic surgery is considered to be a relatively new challenge in face recognition when compared to pose, expression or illumination variations. With the increasing number of people resorting to plastic surgery for correction of feature defects, cosmetic reasons or even law evasion, it becomes of interest for the biometric community to investigate and prevent the impact of facial alterations on recognition performances. Yet, very few studies exist which address this problem.

An evolutionary granular approach is proposed in (Bhatt, et al., 2011) for matching a post-surgery face image with a pre-surgery face image and 15% improvement in identification performance is reported. Furthermore, two new methods, FARO and FACE, based on fractals and a localized version of correlation index, respectively, are implemented in (De Marsico, et al., 2011) which claims that the performance of these two algorithms compare favorably against standard face recognition methods such as PCA and LDA in case of plastic surgery changes. Singh et al. adopted the near set theory to classify facial images that have previously undergone some feature modifications in (Ojala, et al., 2002).

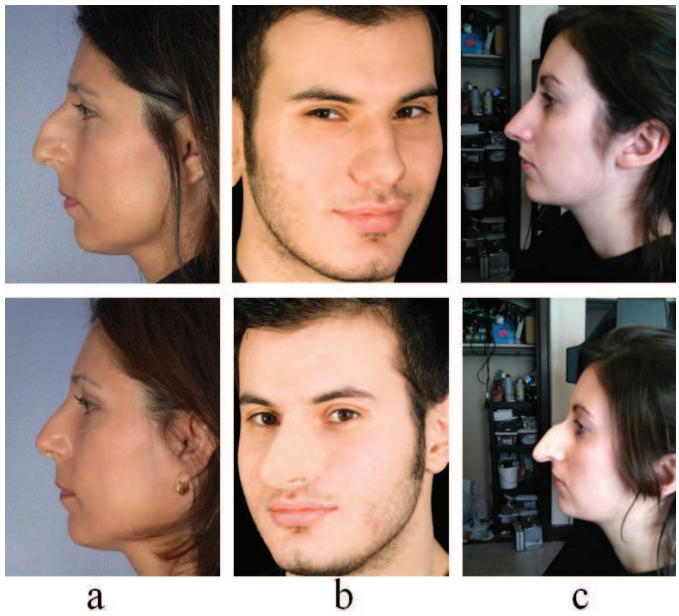


Figure 63: Examples of nose alterations with before (upper row) and after (lower row) photos: (a) plastic surgery (Blogger, 2010) (b) latex appliance (CinemaSecrets) (c) makeup using wax

In this study, we focus on the nose modifications and analyze their effects on success rates of different face recognition methods. According to the statistics published by The American Society for Aesthetic Plastic Surgery in 2010 (ASAPS, 2010), nose reshaping (rhinoplasty) is the second most common surgical procedure on face after cosmetic eyelid surgery (blepharoplasty).

On the other hand, plastic surgery is just one of many ways to change the appearance of the nose. For example, latex/silicone-based prosthetic appliances can be simply purchased as off-the shelf products. Alternatively, makeup using wax or putty can also alter the nose shape very easily. Three nose alteration examples for the three aforementioned methods are given in **Figure 63**.

1.1. Related Work

To the best of our knowledge, the impact of facial alterations, specifically due to plastic surgeries, on face recognition was first analyzed in (Singh, et al., 2009) where the effect of plastic surgery is evaluated on six recognition algorithms. The database used consisted of 506 subjects with 2 images: before and after the plastic surgeries. Later, this work was extended in (Singh, et al., 2010) by augmenting the database up to 900 subjects and additionally including a different non-surgery database for performance comparison. The results showed that the evaluated appearance, feature and texture-based algorithms were unable to effectively mitigate the decline caused by plastic surgery procedures.

Three shortcomings of these studies, which will be addressed throughout this chapter, can be identified as follows:

Figure 64: Examples of facial hair, expression and makeup variations on the facial images between before (upper row) and after (lower row) plastic surgery procedure

- Due to the fact that a single image is provided before the plastic surgery procedure, a non-surgery vs. non-surgery recognition experiment had to be conducted on a separate database with different subjects. Unfortunately, for face recognition algorithms, the accuracy can vary widely depending on the difficulty of the database. Hence, an authentic comparison is not possible.

- In the plastic surgery database, the before and after images differ not only as a result of the procedure, but also due to expressions, makeup and facial hair variations (**Figure 64**). This leads to an additional decrease in the performances which clouds the true measurement of the plastic surgery effect.

- Since this is an image database, the analyses are restricted to 2D. However, 3D face recognition gains a rising popularity as it offers superiority over to its 2D counterpart by being intrinsically robust against illumination and pose variations. For this reason, the impact of the facial alterations on 3D algorithms should also be investigated.

In this study, these limitations are eliminated by creating a synthetic database using FRGC v1.0 (Phillips, et al., 2005) for which nose regions are randomly exchanged between subjects. In this way, a 2D+3D database is obtained for nose alterations and since the conditions and the subjects are identical for the original and the simulated databases, measuring the exact impact of the applied changes is possible.

2. Simulating Nose Alterations

The nose region can be altered in many ways using plastic surgery, prosthetic appliances or makeup and it can be made bigger, smaller, wider or thinner. In order to simulate these changes and preserve the authenticity of the facial shape, noses in the

database are replaced by randomly chosen ones from different subjects. For this purpose, a Thin Plate Spline (TPS) based method is implemented.

A metamorphosis technique for 3D plastic surgery simulation was proposed in (Lee, et al., 1999), where three morphing operations: augmentation, cutting and lacerating were simulated. Later in (Rabi, et al., 2006), an automatic virtual plastic surgery system was presented which similarly to our approach, replaced an individual's facial features with corresponding features of another individual and fused the replaced features with the original face, but only in 2D. In a more recent work (Bottino, et al., 2010), effective patient-specific improvements for facial attractiveness are automatically suggested using 3D scans of the patients and the results are simulated by merging the target feature of the most similar face in the 3D database of attractive faces with the patient's face.

In this study, beautification of the faces is not a concern. What we aim is to change nasal regions in the database as realistically as possible and create nose variations for all subjects. Therefore, a target list is randomly generated to transfer noses (from another person) for each sample in the database.

For this purpose, firstly, nose regions of all facial scans are automatically segmented in a similar manner to (Kakadiaris, et al., 2007) where an annotated generic face model is deformed to fit the target models and the annotations are transferred.

Next, nose deformations are applied using TPS method (Bookstein, 1989) in 3D. Prior to warping, the target model is aligned with the source model using 4 of 5 landmark points around the nose (**Figure 65**), excluding the nose tip. A linear transformation that includes rotation, translation and isotropic scaling is computed in a least square sense, based on the two sets of landmarks and applied onto the source model. Subsequently, using all 5 point pairs a coarse TPS approximation is computed.

In the final step, for one-fifth of the vertices on the target nose, the closest vertices on the source nose are found and coupled to be utilized in a second and denser TPS warping, which results in the source nose completely transforming into the target nose. The proposed method is illustrated in **Figure 65**.

The original database FRGC v1.0 consists of 943 multimodal samples from 275 subjects and the simulated 3D database is of the same size.

For simulated samples in 2D, the synthesized 3D models with the corresponding original texture mapped on. However, due to some mismatches in 2D and 3D samples in the original database, 39 samples had to be removed, leaving us with a database of 904 samples from 268 subjects.

In order to evaluate the visual plausibility of the created database, an online survey was conducted, for which the participants were asked to classify the randomly displayed facial images (with or without texture) as original or simulated. According to a total number of 81 participations, success rate is found to be 60.68% for the images displayed with texture. For the other ones, the performances deteriorate as expected (58.77%) since the texture gives a better hint about originality. Being very close to the average performance of a

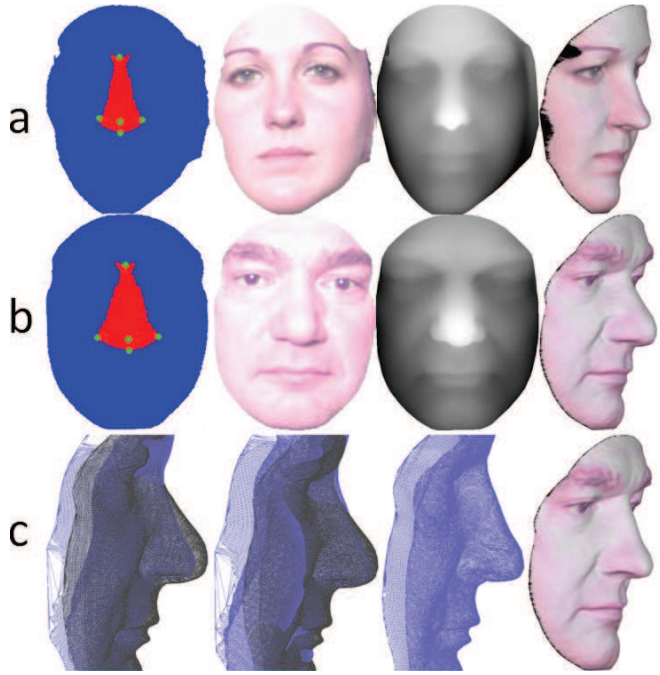


Figure 65: From left to right: (a) Nose region with landmark points, color map, depth map and profile view for the target model (b) Same images for the source model (c) Two models superimposed before and after alignment, resulting mesh after warping and profile view for the synthesized model

random classifier (50%), this result indicates very low distinguishability, and hence a highly realistic look for the simulated noses.

For the sake of clarity, the original databases in 2D and 3D will be referred as DB-o2 and DB-o3, while the simulated nose alteration databases will be referred as DB-s2 and DB-s3, in the rest of this chapter.

3. Experimental Evaluation

The effect of the applied nose alterations on face recognition performances are evaluated with three different scenarios in both 2D and 3D which are determined according to the study of (Singh, et al., 2010) for comparison purposes.

Initially, all four databases, DB-o2, DB-o3, DB-s2 and DB-s3 are partitioned in non-overlapping training and testing datasets. This is done by randomly selecting 40% of the subjects and assigning their samples to the training set, while the rest is used for testing. The partitioning is repeated 10 times and verification and identification performances are computed over these 10 trials.

For verification tests, the Receiver Operating Characteristic (ROC) curves which plot Verification Rates (VR) as a function of False Acceptance Rates (FAR) are reported together with the verification rates at 0.001 FAR.

For identification tests, the first sample of every individual in the test set is used as gallery and the rest as probes. The rank-1 recognition rates are reported.

- **Experiment 1 – Performance on the original database**: It is important to compute the performances on the original datasets in terms of having a baseline. In this way, the impact of the applied changes can be measured accurately. For this purpose, 2D and 3D algorithms are evaluated on DB-o where the similarities are calculated between each original image pair.

- **Experiment 2 – Performance on the simulated database:** In this scenario, the similarity scores between every DB-o and DB-s sample pairs are calculated and used to evaluate recognition performances. For the training set, for each subject selected, half of the corresponding images are taken from DB-o and the rest from DB-s. This experiment is identical to Experiment 1, except the probe images are now replaced by their modified versions.

- **Experiment 3 - Performance on the simulated database with training on an external database:** Face recognition algorithms are usually trained using different databases. Therefore in this scenario, Experiment 2 is repeated, but the training partition is composed of samples from an external database, namely Texas 3D Face Recognition Database (Gupta, et al., 2010). Briefly, the Texas 3D Face Recognition (Texas 3DFRD) database is a collection of 1149 pairs of facial color and range images of 105 subjects. In order to obtain a training set of similar subject and sample numbers as in experiments 1 and 2, a subset of Texas 3DFRD is compiled with 350 color and range images of 103 subjects, without expressions.

3.1. Impact on 2D Face Recognition

Three key methods are chosen to be evaluated for 2D face recognition: Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) (Belhumeur, et al., 1997) and Circular Local Binary Pattern (CLBP) (Ojala, et al., 2002). PCA and LDA are appearance-based approaches which are widely used for dimensionality reduction and feature extraction. On the other hand, CLBP is a texture-based algorithm for describing local structures.

The plastic surgery database in (Singh, et al., 2010) has images after several types of surgery operations such as forehead surgery, ear surgery, eyelid surgery, nose surgery, skin resurfacing, face lift etc. However, since the results are reported for each surgical operation separately, we can compare our results with the reported rank-1 identification accuracies for nose surgery (rhinoplasty) as shown in **Table 15**.

According to this comparison, it is observed that even the evaluated databases are completely different, very similar identification results are achieved with both PCA and LDA. Observing such consistent results with a real plastic surgery database indicates high accuracy for our synthetic database.

Table 15: Rank-1 identification accuracies for 2D FR algorithms for Experiment 1, 2 and 3

Algorithm	Exp. 1	Exp. 2	Exp. 3	Exp.3 [2]
PCA	40.24%	30.02%	24.74%	23.1%
LDA	64.74%	51.56%	28.94%	24.1%
CLBP	92.90%	88.52%	88.52%	44.8%

However, this is not the case for CLBP. Very different identification rates are obtained, mainly due to two main reasons: The significant variance between the pre-surgery and post-surgery images in (Smeets, et al., 2010) (as shown in **Figure 64**) and the fact that in our case, the only variation is due to the nose alteration and hence the change in the image texture is minimal.

Table 16: Verification rates at 0.001 FAR for 2D FR algorithms for Experiment 1, 2 and 3

Algorithm	Exp. 1	Exp.2	Exp. 3
PCA	27.50%	21.18%	11.66%
LDA	50.69%	40.11%	15.30%
CLBP	81.51%	71.72%	71.72%

The verification rates at 0.001 FAR and the ROC curves for all three algorithms are given in **Table 16** and **Figure 66**, respectively. For CLBP, the rates are identical for Experiments 2 and 3, since no training is required.

According to the results in **Table 15** and **Table 16**, best performance is obtained using CLBP method for both identification and verification with a marked difference. This shows that being a texture based method, CLBP is much more appropriate than appearance based methods, PCA and LDA in case of nose alterations. With CLBP, the relative difference between the results of Experiment 1 and 2 for identification and verification are found as 4.71% and 12.01%, respectively.

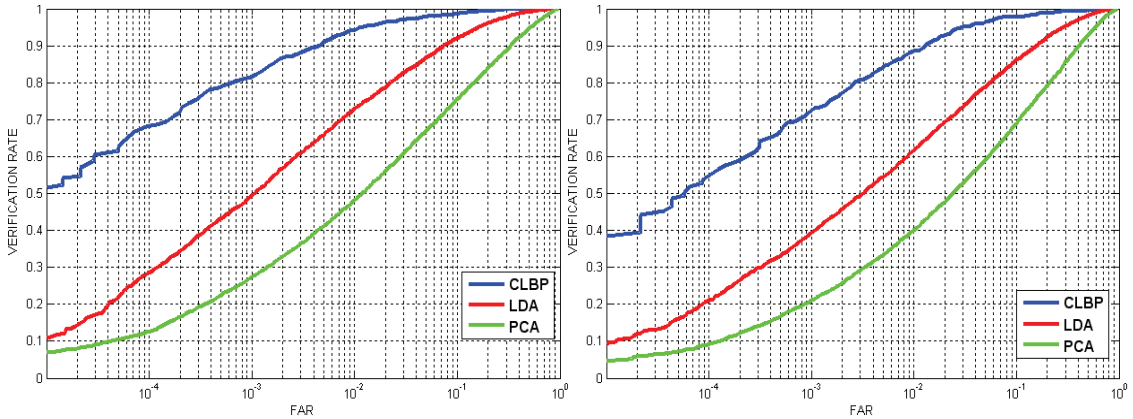


Figure 66: Verification rates for all 2D FR algorithms by Experiment 1 (left) and Experiment 2 (right)

Robustness of LDA is observed to be higher than PCA, with ~20% decrease for both verification and identification scenarios. Whereas, PCA suffers 25.40% and 22.98% loss in identification and verification accuracies, respectively.

Utilization of an external database worsens the results even further for both identification and verification experiments.

3.2. Impact on 3D Face Recognition

For the evaluation of 3d face recognition systems, 4 algorithms are selected where the facial surfaces are represented as depth maps or point clouds.

Depth maps can be involved in most of the existing 2D techniques, including subspace methods. In this part of the study, similar to the 2D evaluations, PCA and LDA are selected to be evaluated.

Additionally, two different approaches are implemented for 3d face recognition using point clouds. In this representation, faces are required to be registered prior to similarity measurements.

For this reason in the first technique, the faces are aligned with a generic face model using 3 landmark points (2 outer eye corners and the nose tip) and then the depth values are regularly sampled. The similarity (in this case distance) between two faces is obtained by averaging the z-distances of all vertices. In this way, the volume difference (VD) between two facial surfaces is approximated.

For the second approach, we implemented the method presented in Chapter V which employs the TPS warping parameters as biometric signatures.

The achieved rank-1 identification rates and the verification rates at 0.001 FAR by all 3D algorithms on databases DB-o3 and DB-s3 are given in **Table 17** and **Table 18**. As is the case with CLBP, since the two point cloud methods do not require any training, rates for Experiments 2 and 3 are the same.

For identification, the best performing and most robust algorithm is observed to be

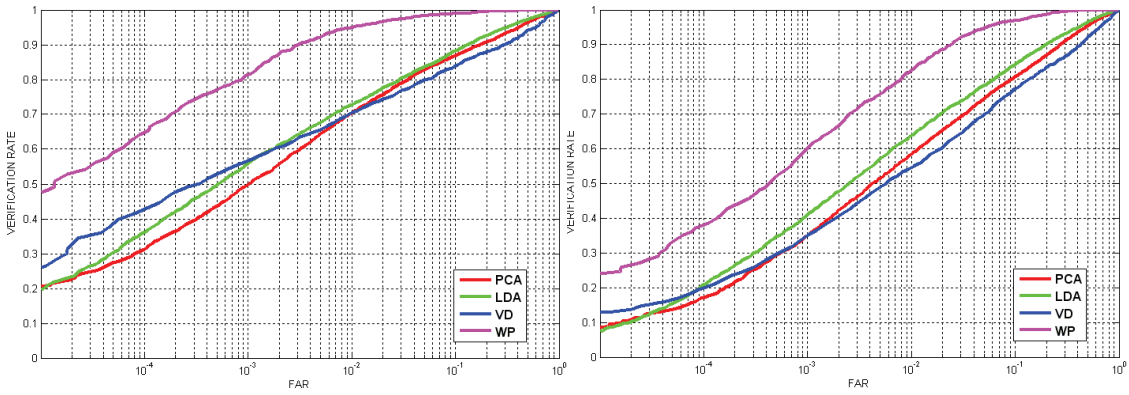


Figure 67: Verification rates for all 3D FR algorithms by Experiment 1 (left) and Experiment 2 (right)

WP, followed by LDA on range images. Both PCA and VD suffer a drastic decline (~25%) when nose alterations are introduced.

Table 17: Rank-1 identification accuracies for 2D FR algorithms for Experiment 1, 2 and 3

Algorithm	Exp. 1	Exp. 2	Exp. 3
PCA	64.11%	48.40%	33.96%
LDA	68.47%	58.15%	42.03%
VD	68.26%	51.95%	51.95%
WP	94.46%	86.64%	86.64%

Likewise, analysis concerning the verification rates reveals that LDA and WP are least affected from nose alterations. However in verification, deteriorations are much more visible for all four methods.

Table 18: Verification rates at 0.001 FAR for 3D FR algorithms for Experiment 1, 2 and 3

Algorithm	Exp. 1	Exp. 2	Exp. 3
PCA	49.85%	35.22%	17.42%
LDA	56.67%	42.18%	17.74%
VD	56.97%	35.23%	35.23%
WP	81.18%	60.79%	60.79%

Similar to the case observed in 2D experiments, utilization of an external database has a negative effect on the recognition accuracies. Algorithms have better performances when they are trained on both pre- and post-alteration images.

4. Conclusion

As means of altering facial shape proliferate, its impact on recognition performances becomes crucial to measure and prevent. Today, more and more people undergo plastic surgeries (From 2009-2010, there was almost a 9% increase in the total number of cosmetic surgical procedures and since 1997, there has been over 155% increase in the total number of cosmetic procedures (ASAPS, 2010).) not only for medical reasons but also to improve their appearance or even to hide their true identity. Easy-to-use appliances and makeup products are within reach of everyone who seeks ways to evade recognition.

In this study, a synthetic nose alteration database is obtained for which the nose of every subject in FRGC v1.0 is transfigured by replacing it with another randomly selected one. It is utilized to evaluate the performances of face recognition algorithms in presence of nose alterations.

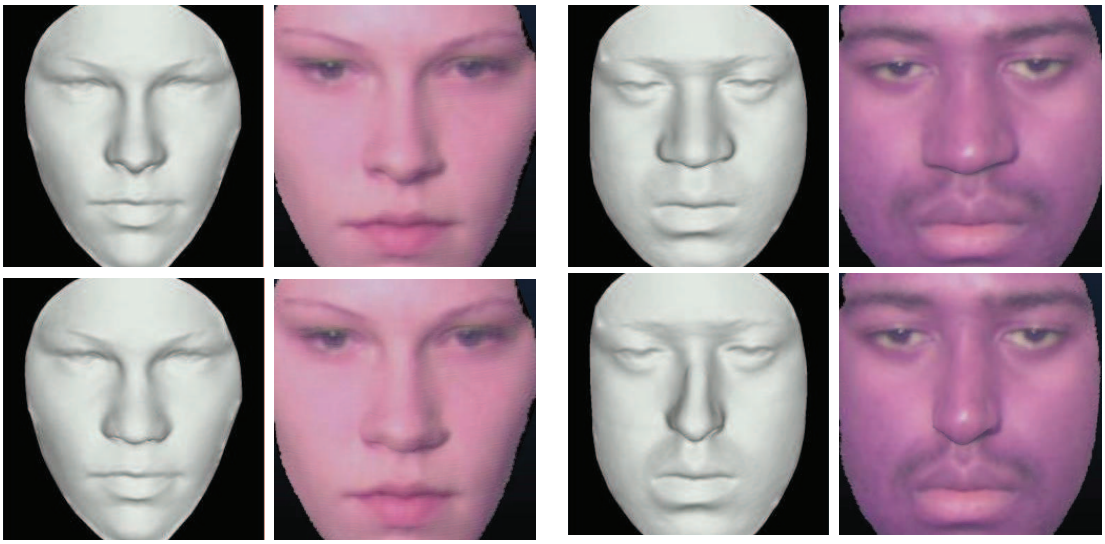


Figure 68: Two examples of nose alterations with and without textures (upper row: originals lower row: altered)

The novelty of this contribution is that the analyses are not restricted to 2D images. Thanks to the nature of the simulated database, the effect of the applied modifications can be determined also in 3D. Additionally, since it is possible to measure the original performances on FRGC v1.0, an authentic comparison between pre- and post-alteration performances can be provided, which is a significant advantage of this study when compared to the previous ones.

The results reveal that the evaluated algorithms are not robust to the variations caused by nose alterations, especially for the purpose of verification. Furthermore, comparing verification performances of 2D and 3D algorithms show that 3D is much more vulnerable against the nose variations. On the other hand, robustness in identification is observed to be more method dependent then modality.

Robust face recognition algorithms are necessary to mitigate the effects of facial modifications. Our future research direction is to develop such face recognition methods. Additionally, we would like to measure the efficiency of nose alterations for face spoofing purposes.

Chapter VIII. Conclusions

1. Summary

In this thesis, we investigated the influence and advantages of 3D data in face recognition domain. With the fast advancement in 3D acquisition technologies, the accessibility and usability of range data has greatly improved, expanding its field of application in many directions and the topic of face recognition is only one of them. Motivated by many successful applications launched previously, we explored methods to incorporate 3D modality into identification/verification process of human faces.

Firstly, we examined the state of the art on utilization of 3D data in face recognition and presented a thoroughly structured review. The role of 3D is explored under two different classes approach: Recognition of faces by matching 3D facial data and exploitation of 3D knowledge for recognition of faces from their 2D images.

After briefly describing the pre-processing methods adopted throughout the thesis, in Chapter III, an automatic landmarking system is proposed. Based on the assumption of a well-controlled acquisition environment during enrollment, points of interest were detected either using 2D images, either 3D scan data or both, according to the regional characteristics.

In Chapter IV, we proposed a system, in which a combined enrollment is realized by both 2D and 3D information whereas the target images are still 2D images and claimed to be an optimal strategy to fuse advantageous features of the two modalities. With the employment of this asymmetrical scenario, we addressed the problem of expressions in 2D probe images. We tried and facilitated 2D recognition by simulating facial expressions using 3D models for each subject. Realistic facial expression simulations were achieved with automatically generated MPEG-4 compliant animatable face models of the enrolled subjects. A TPS-based algorithm was developed for animatable model generation and gallery images were augmented with synthesized images of different expressions. The tests conducted on a PCA-based baseline method revealed increased robustness against expression variations.

Next, the discriminative properties of TPS warping parameters were investigated as detailed in Chapter V. Often employed in surface registration stage, TPS generates those parameters during interpolation computations and they are mostly overlooked as superfluous sediment. However in our studies, they were proven to possess concentrated information about the facial shape when TPS is applied densely. Instead of minimally deforming the faces to a common target as is the case for registration applications, a common (i.e. generic) face was strongly deformed to fit target faces and the warping

parameters were used as biometric signatures. Extensive tests with FRGCv2 showed that the warping parameters were robust against moderate decimation and noise and the presence of holes on the facial surface. On the other hand, the facial expressions were observed to be highly detrimental for the recognition performances.

Stimulated by these results, in Chapter VI, we proposed a regional confidence level estimation algorithm for facial surface to be used in cooperation with region-based 3D face recognition methods. Based on local primitive shape distributions, confidence scores were found by ANNs for 7 non-overlapping segments of face. We analyzed two different approaches to incorporate the estimated confidence levels in fusion of regional match scores: employing them as weights and including them in probability density function calculations of match scores. Both approaches were proven to improve the robustness of our baseline algorithm (presented in Chapter V). On the other hand, the comparative study revealed that the latter achieved higher accuracies.

Finally in Chapter VII, a relatively more recent field of research was explored. The impact of partial modifications on the facial surface, nose region in particular, on recognition performances was analyzed. To this end, a synthetic nose alteration database was generated for which the alterations are simulated by replacing nose of each sample with another randomly chosen one. The experiments were conducted with several key methods for both 2D and 3D recognition and the results showed that the success rates deteriorated significantly with the simulated database.

2. Future Work

As we concluded our survey in Chapter II, utilization of 3D data for recognizing human faces has drawn a strong and constant interest and there is no doubt that it will maintain its popularity in the future. We can list two main reasons for that: Firstly, 3D face is a promising modality for biometric authentication and it has the potential to offer novel solutions and to bring new ideas for existing problems in this domain. Secondly, recognition of faces from their 3D shape still suffers from numerous major challenges, such as surface deformations due to expressions or aging, missing data due self-occlusion or lack of quality in 3D acquisition at a distance. Consequently, there exist many possible directions to be explored and many possible fields to be improved. The studies conducted within the context of this thesis can also be expanded in many ways.

To start with, simulated expressions in the asymmetrical approach proposed in Chapter IV, could be analyzed to measure their similarity to real expressions. In the light of these evaluations, one could try to understand shortcomings of the animatable model or the animation engine and generate more realistic, hence more helpful, synthetic images to achieve higher robustness.

In Chapter V, the warping parameters were extracted for arbitrary regularly spaced facial points. The impact of the number and location of those control points on recognition performances could be further investigated. Similarly, the role of the selected

generic model in the system accuracy could be assessed. Additionally, based on the extracted parameters, one could compute the discriminatory power of each control point individually and use this information to fuse point-wise calculated match scores.

For the estimation of the regional confidence levels in Chapter VI, we would like to extend the experiments by including Bosphorus database. In this way, it would be possible to utilize two databases as training and test sets interchangeably, enabling us to perform training with higher number of samples and to make comparisons with other approaches tested with the same experimental setup. Additionally, one could evaluate various face segmentations of different numbers and shapes for confidence estimations.

Finally, regarding to our work on partial facial alterations and their impact on face recognition, we would like to investigate the possibility of spoofing by changing the nose region only as explained in Chapter VII. With our experiments, it has become evident that there is a need for methods that are robust to such changes. New techniques could be developed in this direction to avoid the decline in the success rates.

3. Publications

The work presented in this dissertation resulted in the following refereed conference and journal articles:

Book Chapter

- Velardo Carmelo, Dugelay Jean-Luc, Daniel Lionel, Dantcheva Antitza, Erdogmus Nesli, Kose Neslihan, Min Rui, Zhao Xuran, "Introduction to biometry", Multimedia Image and Video Processing, CRC Press; 2012; ISBN:978-1439830864

Journal

- Erdogmus Nesli, Dugelay Jean-Luc, "3D Assisted Face Recognition: Learning the Expression Variations," submitted to IEEE Transactions on Multimedia – *under review*

- Dugelay Jean-Luc, Erdogmus Nesli, Kose Neslihan, "3D face recognition: Where we are", IEEE COMSOC MMTC E-Letter, Vol 6, N°11, November 2011

Conference

- Erdogmus Nesli, Dugelay Jean-Luc, "Probabilistic fusion of regional scores in 3d face recognition," IEEE International Conference on Image Processing, ICIP 2012, Orlando, Florida, USA – *under review*.

- Erdogmus Nesli, Dugelay Jean-Luc, "Impact analysis of nose alterations on 2d and 3d face recognition," IEEE International Conference on Multimedia and Expo, ICME 2012, Melbourne, Australia – *under review*.

- Erdogmus Nesli, Dugelay Jean-Luc, "3d face recognition with regional confidence levels," IEEE International Conference on Multimedia and Expo, ICME 2012, Melbourne, Australia – under review.

- Erdogmus Nesli, Dugelay Jean-Luc, "On Discriminative Properties of TPS Warping Parameters for 3D Face Recognition," International Conference on Informatics, Electronics & Vision, ICIEV 2012, Dhaka, Bangladesh.

- Ben Amor, Boulbaba; Drira, Hassen; Daoudi, Mohamed; Ardabilian, Mohsen; Ben Soltana, Wael; Lemaire, Pierre; Chen, Liming; Erdogmus, Nesli; Dugelay, Jean-Luc; Colineau, Joseph, "3D face recognition: A robust multi-matcher approach to data degradations," 5th IAPR International Conference on Biometrics, March 29-April 1, ICB 2012, New Delhi, India.

- Erdogmus Nesli, Dugelay Jean-Luc, "Regional confidence score assessment for 3D face," IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2012, Kyoto, Japan.

- Erdogmus Nesli, Dugelay Jean-Luc, "Classification of facial expressions by sparse coding," 19th IEEE Signal Processing and Communications Applications Conference, SIU 2011 (in Turkish)

- Erdogmus Nesli, Dugelay Jean-Luc, "Automatic extraction of facial interest points based on 2D and 3D data," Electronic Imaging Conference on 3D Image Processing (3DIP) and Applications, SPIE 2011

- Dantcheva Antitza, Erdogmus Nesli, Dugelay Jean-Luc, "On the reliability of eye color as a soft biometric trait," IEEE Workshop on Applications of Computer Vision, WACV 2011

- Erdogmus Nesli, Dugelay Jean-Luc, "An efficient iris and eye corners extraction method," Structural, Syntactic, and Statistical Pattern Recognition - Joint IAPR International Workshop, SSPR and SPR 2010

- Erdogmus Nesli, Etheve Rémy, Dugelay Jean-Luc, "TPS Yöntemiyle MPEG-4 Uyumlu Canlandırılabilir 3B Yüz Modelleri Elde Etme - Obtaining MPEG-4 Compliant Animatable 3D Face Models by Using TPS Method," Sinyal Isleme ve Iletisim Uygulamalari Kurultayi SIU, 2010 (in Turkish)

- Erdogmus Nesli, Etheve Rémy, Dugelay Jean-Luc, "Realistic and animatable face models for expression simulations in 3D," SPIE 2010, Electronic Imaging Conference on 3D Image Processing (3DIP) and Applications, January 17-21, 2010, San Jose, California

- Bozkurt N., Halici U., Ulusoy I., Akagündüz E., "Yüz Tarayıcısı Verilerinin İyileştirilmesi için 3B Veri İşlem – 3D Data Processing for Enhancement of Face Scanner Data," Sinyal Isleme ve Iletisim Uygulamalari Kurultayi SIU, 2009 (in Turkish)

Résumé Etendu en Français

1. Introduction

Vu l'importante et l'augmentation de l'investissement dans les applications de sécurité, un sujet particulier, parmi ceux appartenant aux domaines de la reconnaissance faciale, a certainement attiré beaucoup plus d'attention que les autres, il s'agit de la biométrie- reconnaissance des humains basée sur leurs traits physiques et/ou comportementaux. Ces derniers comportent, entre autres, un élément qui peut être traité "favori" pour posséder la combinaison parfaite entre accessibilité et fiabilité: le visage. Dans cette thèse de doctorat, nous étudions principalement la reconnaissance 3D qui représente une modalité de reconnaissance faciale relativement récente. Bien qu'elle soit plus avantageuse que son homologue 2D en étant intrinsèquement invariante aux changements de poses ainsi qu'à l'éclairage, la reconnaissance faciale 3D se heurte encore à des défis majeurs comme l'inadéquation de l'acquisition ou encore les déformations de la surface faciale dues aux expressions ou aux occultations. Compte tenu des faits précédemment cités, une étude minutieuse est menée sur ce thème, y compris le prétraitement des données faciales 3D, le repérage automatique, l'expression faciale et les simulations d'altération en 3D, l'extraction des traits faciaux en 3D et l'analyse 3D de la surface régionale.

1.1. Motivation

L'authentification biométrique permet d'obtenir des systèmes de gestion d'identité aussi fiables qu'efficaces en exploitant les caractéristiques physiques et comportementales des sujets qui sont permanents, universels et facile à accéder. Il existe de nombreux systèmes biométriques, dont chacun est soumis à ses propres limitations, qui utilisent une variété de caractéristiques humaines telles que la voix, l'iris, le visage, l'empreinte digitale, la manière de marcher ou bouger ou l'ADN. La "supériorité" de ces traits ne constitue plus un concept réaliste dès qu'elle se sépare du scénario de l'application. Les contraintes et les exigences du système doivent être prises en compte, mais aussi l'objet de l'utilisation et le contexte qui comprend des facteurs techniques, sociaux et éthiques (Introna, et al., 2009). Par exemple, alors que les empreintes digitales représentent la forme biométrique la plus répandue d'un point de vue commercial (Abate, et al., 2007), un fait principalement dû à une longue histoire avec la médecine légale, cette technique requiert une coopération importante du sujet. Pareillement, la reconnaissance par l'iris, quoique très précise, nécessite quant à elle une participation active des personnes, sans omettre le fait qu'elle soit extrêmement dépendante de la qualité d'image.

La reconnaissance faciale offre une identification à des distances relativement grandes pour des sujets non-coopératifs et non-informées. Hormis l'intérêt qu'elle a attité depuis des décennies et les progrès importants réalisés, la reconnaissance faciale doit encore faire face au problème des variations "intra-classe" liées aux différents facteurs présents dans les scénarios du monde réel comme l'éclairage, la pose, l'expression, l'occultation et l'âge.

Selon FRVT (Face Recognition Vendor Test, évaluation marchand de reconnaissance faciale en Anglais), on a démontré que l'utilisation de l'intensité 2D ou des images en couleurs donne un taux de reconnaissance supérieur à 90% dans des conditions contrôlées (Phillips, et al., 2003). Néanmoins, en introduisant les variations susmentionnées, les performances ont détériorées. Les résultats obtenus ont motivé l'accélération des études sur des modalités alternatives, particulièrement le visage tridimensionnel, puisqu'il semble de par sa nature, être une solution logique pour éviter les problèmes de pose et d'éclairage.

Par conséquent, en FRVT 2006 (Phillips, et al., 2009), une amélioration significative concernant les performances de reconnaissance a été accomplie sur FRVT 2002 à partir des images faciales en 3D, avec un FRR de 0.01 et un FAR de 0.001. Des résultats similaires ont été également obtenus en haute résolution mais toujours sous un éclairage contrôlé. Toutefois, pour le scénario contraire, où l'éclairage était non-contrôlé, les performances ont à nouveau baissé.

Dans les travaux de thèse, nous recherchons les contributions possibles de l'information 3D pour l'obtention de meilleures performances. La première motivation de notre travail est alors d'explorer les méthodes qui permettent d'incorporer des caractéristiques avantageuses de la modalité 3D dans le processus d'identification/ vérification tout en évitant des inconvénients. Un des principaux défis liés au visage 3D est certainement son acquisition. Même avec le progrès important achevé lors des quelques dernières décennies, elle n'est toujours pas aussi immédiate que son équivalent 2D pour obtenir une image 3D précise. La plupart des systèmes d'acquisition de haute précision (des scanners laser "dits tridimensionnels" par exemple) demandent une collaboration de l'utilisateur qui peu prendre plusieurs secondes. D'autres part, les systèmes qui extraient les informations de la forme à partir des images bidimensionnelles (stéréo passive par exemple) peuvent faire des mesures instantanées; cependant, pour atteindre un certain degré de précision, ces systèmes-là se reposent fortement sur la connaissance des paramètres extrinsèques de la scène et intrinsèques de la caméra. Prenant ces faits en considération, nous proposons un système asymétrique qui inclut une combinaison d'informations 2D et d'informations 3D tandis que les images de tests restent en 2D. Via cette approche, nous abordons le problème des variations liées aux expressions faciales en reconnaissance 2D en augmentant synthétiquement la taille de la galerie. Pour ce faire, des simulations réalistes de diverses expressions pour les sujets enregistrées sont obtenues par la dérivation des modèles compatibles MPEG-4 animable depuis leurs données faciales 3D numériques dans la galerie. Générer un modèle animable consiste à déformer un modèle annoté pour le transformer en un sujet qui sera animé, par la suite, par l'intermédiaire de TPS.

Dans l'étape qui suit, nous étudions les propriétés discriminantes des paramètres TPS obtenus durant le processus de déformation du modèle générique. TPS est le plus souvent utilisé en tant qu'algorithme de recalage qui établit des correspondances denses entre les modèles du visage; ses paramètres ont fréquemment été considérés comme redondants. Cependant, si le modèle générique est totalement déformé pour l'adapter au visage cible au lieu d'une déformation minimale à l'image de celle effectuée pour le recalage, les paramètres de déformation prennent un sens nouveau. En l'occurrence, tant qu'elles représentent des déviations depuis un modèle commun, elles peuvent être utilisées pour traiter des informations denses concernant la forme faciale.

Notre analyse des paramètres de déformation prouve qu'ils constituent des signatures biométriques solides lorsque les conditions d'acquisitions à l'enregistrement et à la vérification sont les mêmes. Toutefois, ces paramètres ne sont pas robustes face aux variations de l'expression. Cette approche suppose que le visage est rigide d'où l'impossibilité de gérer ces changements. Récemment, la recherche dans le domaine 3D est principalement orientée vers le problème des variations de l'expression; néanmoins, il est encore reconnu comme étant l'un des défis les plus difficiles rencontrés en reconnaissance faciale 3D (Smeets, et al., 2010). Ainsi, nous élargissons notre recherche à travers ce sujet-là.

Comme elles sont des distorsions faciales locales, on propose souvent de traiter les variations d'expression et d'occultation en se servant des méthodes dites "par régions" qui suggèrent la division du visage en parties multiples pour une comparaison plus robuste. Les scores de similarité calculés à partir des toutes les régions sont souvent fusionnés eu utilisant une métrique de similarité pondérée qui accorde plus d'importance aux régions les plus statiques, (Ben Amor, et al., 2008), (Smeets, et al., 2010) par exemple.

Dans notre travail, au lieu de faire des hypothèses en ce qui concerne la fiabilité (à laquelle le terme "confiance" fait référence tout au long du texte) des différentes régions faciales, nous proposons de l'estimer en nous basant sur les caractéristiques de la surface. Utilisant un ensemble d'apprentissage de modèles de visages 3D avec différentes expressions et occultations, un Réseau Neuronal Artificiel est formé pour chaque région afin d'évaluer la présence de telles distorsions depuis la distribution des catégories de la forme primitive des sommets (Due Trier, et al., 1995). Le calcul initial des scores de fiabilité est réalisé en recalant chaque région à la région neutre et nette équivalente et qui appartient à la même personne, au moyen de l'algorithme ICP (Iterative Closet Point) et le calcul des distances résiduelles entre les points mutuellement correspondants. Une fois que les taux de confiance régionale sont obtenus, ils sont intégrés dans la fusion des scores de correspondances régionales. Trois différentes approches de fusion sont évaluées et comparées pour ce schéma: la loi des Sommes, la méthode Borda et les approches probabilistes. Les résultats expérimentaux montrent une robustesse améliorée face aux distorsions d'un réseau induites par l'expression et/ ou occultation.

Malencontreusement, les variations qui modifient la texture et la forme du visage ne sont pas limitées aux expressions et aux occultations. Il y a d'autres sources de variations

comme le vieillissement, l'indice de masse corporelle (IMC) ou encore la chirurgie plastique. En particulier, en devenant plus avancés et plus abordables, la chirurgie plastique introduit de nouveaux défis à propos des futurs conceptives de systèmes de reconnaissance faciale (Singh, et al., 2010), néanmoins, il existe seulement quelques études qui s'intéressent à ces sujets.

Dans cette dissertation, nous abordons ce problème avec une perspective plus vaste comme les altérations faciales, puisque la chirurgie plastique n'est qu'une méthode parmi d'autres pour changer l'apparence du visage. Nous choisissons de nous concentrer sur la région du nez comme elle est fréquemment envisagée en tant que région robuste face aux variations, et par ailleurs, puisqu'elle possède la plus haute importance dans les approches "par régions" précédemment citées. Afin d'analyser l'impact des altérations du nez, une base de données synthétique est préparée dans laquelle le nez de chaque personne est remplacé par un autre aléatoirement choisi. Ensuite, des comparaisons de performance de plusieurs techniques de reconnaissance faciale sont réalisées.

1.2. Contributions originales:

Les contributions originales de cette thèse peuvent être résumées en quatre points: animation réaliste 3D d'expressions faciales, utilisation des paramètres TPS comme signatures biométriques, estimation de la fiabilité régionale pour un visage 3D et base de données simulée pour altération du nez.

Apprendre ou modéliser un comportement non rigide d'expressions a été proposé dans de nombreuses études qui avaient pour but, dans la plupart des cas, l'élimination des expressions des images tests (Wang, et al., 2007) ou l'ajout d'expressions à la galerie d'images (Lu, et al., 2006) pour une correspondance 3D robuste. Cette approche comporte deux inconvénients: le besoin de données additionnelles pour entrainer apporta le système à une large gamme d'expressions possibles et la validité du transfert d'expressions. Obtenir des modèles animables spécifiques au client pour le système proposé est simple et exécuté seulement une fois durant l'enregistrement. En outre, la simulation des expressions est très directe avec les moteurs d'animation grâce à la conformité par MPEG-4.

L'analyse des paramètres TPS est une autre contribution pour leurs propriétés discriminantes. On a prouvé que ces caractéristiques, fréquemment négligées en tant qu'un résidu de l'étape du recalage, sont extrêmement informatives en dépit de la faible complexité de leur extraction et leur petite dimension.

De plus, un schéma est présenté pour évaluer les différentes régions du visage selon la présence des distorsions pour la fusion des scores de correspondance régionale au lieu des hypothèses de stabilité. On propose d'utiliser l'erreur résiduelle du recalage ICP entre deux échantillons (dont un est l'échantillon de référence) appartenant à la même personne comme indicateur de la fiabilité d'une région.

Finalement, un domaine de la reconnaissance faciale 3D qui n'a encore jamais été exploré est étudié où des altérations faciales sont simulées pour la région du nez et leurs impacts sur les performances de la reconnaissance est déterminés.

1.3. Plan:

Dans ce résumé, la dissertation est brièvement expliquée à travers le plan qui suit:

- Dans la section 2, nous fournissons une introduction à la reconnaissance faciale 3D, passant en revue le spectre complet de la technologie du traitement facial en 3D, l'état de l'art des travaux connexes, les défis actuels ainsi que les orientations possibles à venir.
- Dans la section 3, nous présentons notre prétraitement et l'approche du repérage automatique pour un visage 3D. Cette approche est utilisée à plusieurs reprises dans le reste de cette étude.
- Dans la section 4, nous introduisons notre scénario asymétrique dans lequel un enregistrement combiné est réalisé par, à la fois, les informations 2D et les informations 3D alors que les images cibles restant en 2D. Notre méthode pour générer des modèles spécifiques aux clients, animables, et conformes au format MPEG-4 est détaillée et l'effet positif de l'augmentation de la galerie sur les performances de la reconnaissance est établi.
- Dans la section 5, les propriétés discriminantes de la déformation et les paramètres TPS sont analysés.
- Dans la section 6, nous proposons une méthode pour estimer le score de fiabilité régionale pour les modèles de visages en 3D. Plusieurs, méthodes pour la fusion des scores de correspondance basées sur des estimations plutôt que sur des hypothèses sont évaluées.
- Dans la section 7, l'impact des altérations du nez sur les performances de la reconnaissance est examiné. Notre méthode pour simuler les altérations du nez est expliquée en détails et les résultats expérimentaux sont fournis pour montrer l'effet des modifications appliquées.
- Dans la section 8, les conclusions sont données.

2. Reconnaissance faciale en 3D:

Alors que les exigences en identification humaine connaissent des changements importants avec un développement et une mobilité démographique incroyablement rapides, la technologie "biométrie" est considérée comme une solution prometteuse.

La biométrie est le domaine de la reconnaissance automatique d'une personne basée sur un ou plusieurs traits intrinsèques associés à lui/elle. Ces traits peuvent être physiques

(visage, iris, empreinte digitale...) ou comportementaux (démarche, mouvement des lèvres, signature,...).

La reconnaissance faciale - la caractéristique biométrique la plus utilisée chez les humains, s'est placée au premier plan en fournissant un bon compromis entre efficacité et accessibilité. En plus d'être non invasive et naturelle, elle offre le potentiel d'une reconnaissance à distance, en n'exigeant ni la coopération ni la connaissance du sujet. Avec la croissance de la population, les systèmes de reconnaissance faciale ont atteint des taux de reconnaissance plus haut que 80% dans des situations difficiles (FRVT 2002 - (Phillips, et al., 2003)) en dépassant même la performance humaine; en particulièr dans le cas de galeries très grandes.

Toutefois, le FRVT 2002 a clairement montré que les scénarios, en monde réel, où les images faciales peuvent comporter un large éventail de variations, les performances se dégradent très rapidement. Parmi ces variations, on peut citer principalement l'éclairage, la pose, l'expression faciale, l'occultation et l'âge. Bien que de nombreuses méthodes prometteuses -dont chacune comporte ses propres limitations- puissent être listées, on a encore besoin d'efforts significatifs pour arriver au but recherché: une reconnaissance faciale robuste aux variations. Ces problèmes ont motivé l'approche de la reconnaissance faciale en 3D qui se diffère fondamentalement de celle en 2D, car l'acquisition faciale en 3D est -d'une certaine façon- invariante aux changements d'éclairage, et par ailleurs, les variations de pose peuvent être plus facilement maîtrisées par la rotation des visages en 3D.

2.1. Reconnaissance faciale en 3D: une étude

Bien que cela fasse approximativement quatre décennies depuis la réalisation des plus anciens travaux en reconnaissance faciale 3D (Harmon, 1973), ce domaine n'a commencé a vraiment attiré une attention considérable que depuis dix ans environ. Cela est principalement dû à l'évolution remarquable des technologies d'acquisition et des ordinateurs qui ont permis le développement de systèmes de reconnaissance faciale en 3D beaucoup plus efficaces d'apparaitre. Ces développements ont mené également à l'apparition d'une multitude de bases de données publiques compréhensives de visages en 3D qui ont largement influencé et inspiré la recherche.

Ayant chacun ces avantages et ces inconvénients, plusieurs approches ont été développées pour utiliser les informations en 3D en reconnaissance faciale. Dans cette étude, nous les examinerons selon deux catégories distinctes.

La première classe d'approches repose sur l'utilisation des informations 3D uniquement. Les représentations d'un modèle 3D ou les vecteurs de caractéristique dérivés à partir d'eux sont comparés à des fins de reconnaissance. En outre, nous associerons les modules 3D des approches multimodales à cette classe. Tant que l'apparence du visage est une combinaison de texture et de forme, l'utilisation de ces deux composantes est adoptée par plusieurs études, étiquetée "multimodale". Pour des lectures supplémentaires, nous pouvons orienter le lecteur vers deux très bons exemples: (Chang,

et al., 2003) et (Husken, et al., 2005) dans lesquels les stratégies et les bénéfices de la fusion 2D+3D ont été profondément examinées.

Dans la deuxième classe d'approches, les modèles de visages 3D sont utilisés pour assister à la reconnaissance faciale en 2D. Qu'elles soient obtenues par des scanners 3D ou reconstruites par l'intermédiaire de simples ou multiples images en 2D, les informations des formes 3D sont utiles pour surmonter les problèmes habituellement rencontrés dans la reconnaissance faciale en 2D, comme l'éclairage, de la pose et les variations de l'expression.

Par ailleurs, les systèmes multi-algorithme (comme (Huang, et al., 2010) et les évaluations de la méthode de fusion (voir (Gokberk, et al., 2006) pour une étude comparative) n'ont pas été inclus dans cette étude.

2.1.1. Reconnaissance faciale basée sur la forme 3D:

Quelque soit l'approche adoptée, les méthodes qui utilisent les données 3D dans la reconnaissance faciale suivent typiquement trois étapes principales: prétraitement des données 3D, alignement des visages (recalage) et reconnaissance par la forme 3D.

Les données 3D brutes obtenues la majorité des scanners ont besoin d'être traitées avant d'effectuer n'importe quelle autre opération car elles contiennent des points non-faciaux (régions comme épaules et vêtements), pointes, trous et bruit numérique. Afin d'accélérer les étapes susmentionnées, on élimine en premier lieu les régions non-faciales incluses dans les données brutes acquises en sortie du scanner comme les cheveux, les vêtements et les épaules. Aussi appelé "rognage facial" (comme dans le cas 2D), cette opération est souvent accomplie au moyen de points de repérage comme la pointe du nez et les (cantus) externes. Par exemple, (Mian, et al., 2006) détectent la pointe du nez qui est par la suite considérée comme le centre de la sphère à travers laquelle le visage est rogné.

Les pointes sont des erreurs aberrantes dans les données pouvant être causées par les régions spéculaires (yeux) ou luisantes (pointe du nez, front...). Avec une hypothèse de régularité, elles peuvent être détectées comme étant loin de leurs voisins et supprimées, laissant des trous derrière. Les trous sont essentiellement des régions de données manquantes qui peuvent être introduites par des régions moins réfléchissantes dont la saisie par les capteurs est impossible, telles que les yeux, les moustaches ou les cheveux. L'interpolation linéaire peut être suffisante pour des trous minuscules; d'autre part (Mian, et al., 2007) ont montré que l'interpolation bi-cubique offraient de meilleurs résultats. Pour les plus grands trous, la symétrie faciale est fréquemment utilisée en copiant simplement les sommets symétriques correspondant depuis l'autre côté du visage (Malassiotis, et al., 2004).

Dès que le visage 3D net est obtenu, un autre problème à résoudre se pose: les différences d'une pose faciale. Afin de rendre possible une comparaison, les modèles du visage doivent être alignés, c'est à dire, que la translation et la rotation nécessaires pour un alignement optimal de deux ensembles de points doivent être calculés. Pour la translation, on peut résoudre le problème en réalisant une collision des centres de masses des visages

(Bronstein, et al., 2003) ou bien ceux des pointes de nez (Hesher, et al., 2003). Alternativement ou par addition, l'algorithme Iterative Closest Point (ICP) (Besl, et al., 1992) peut être employé, pour lequel, aucun point de correspondance ne nécessite d'être connu; toutefois, une initialisation raisonnablement faible est importante pour éviter un minimum local. L'algorithme Thin Plate Spline (TPS) constitue une autre méthode importante d'alignement des modèles faciaux par déformation (Bookstein, 1989). En générant une correspondance interpolée entre deux ensembles de points de repérage, la méthode TPS établit habilement une correspondance dense point à point. Enfin, en tant qu'approche alternative, et au lieu d'aligner les visages, des données invariantes peuvent être extraites telles que les distances, les angles et les régions entre les points faciaux. Par exemple, (Gupta, et al., 2007) emploient les distances Géodésiques et Euclidienne entre 25 points faciaux de réfèrence comme caractéristiques et sélectionnent les plus discriminants en utilisant l'analyse discriminante linéaire (LDA).

La plupart des techniques d'alignement et de rognage dépendent d'au moins un point facial caractéristique, comme une pointe de nez, un coin de lèvre. La localisation correcte de ces points est critique pour le bon fonctionnement du reste du système. Dans (Dibeklioglu, et al., 2008), un algorithme de repérage facial en 3D est proposé. Il est caractérisé par la capacité de se généraliser aux différentes conditions d'acquisition et repose sur la modélisation des traits faciaux. D'ailleurs, (Zhou, et al., 2010) présentent un système basé sur un Active Shape Model en 3D (3DASM) pour localiser automatiquement les points des traits faciaux à partir de différentes vues. 3DASM est propre aux données synthétiques, générées par l'intermédiaire d'un modèle 3D malléable.

Une fois le prétraitement et l'alignement faits, il sollicite une des méthodes qui se basent uniquement sur les informations de la forme 3D pour la reconnaissance faciale:

- **Segmentation en carte de courbure:** Une courbure décrit le degré de pliage en un certain point et elle est étroitement liées à la dérivée seconde de la courbe qui définit la surface. Dans (Moreno, et al., 2003), après une segmentation des visages basée sur les courbures, 86 traits sont extraits en utilisant des zones de régions, des distances et des angles entre les centres de masse des régions, la moyenne de Gauss et la courbure moyenne des régions, etc. puis triés selon leurs pouvoirs discriminants calculés par l'intermédiaire du coefficient de Fisher. Ensuite, 35 d'entre eux sont sélectionnés pour représenter les visages pour une correspondance.

- **Courbes faciales:** Les courbes faciales sont des courbes fermées, qui ne se croisent pas et qui peuvent être définies selon divers critères. Les courbes de niveaux, les courbes iso-géodésiques et les courbes radiales sont trois exemples populaires. D'autre part, les courbes de profil ont un point de début et un point de fin, et peuvent être extraites via une infinité de manières différentes. Les profils horizontaux et les profils verticaux à travers la pointe du nez sont largement utilisés. (Lee, et al., 2003) élaborent une reconnaissance faciale 3D on extrayant des lignes de contour en fonction d'une certaine valeur de profondeur, puis en comparant ces mêmes lignes après qu'elles soient ré-échantillonnées et consécutivement stockées comme vecteurs de

caractéristiques. D'une façon similaire, (Samir, et al., 2006) représentent, les formes faciales par "courbes de niveaux".

- **Approches 2.5D:** La représentation des formes faciales tridimensionnelles par une carte de disparité possède un énorme avantage d'applicabilité à plusieurs techniques bidimensionnelles, notamment celles basées sur le Modèle de Markov Caché (MMC) et l'analyse a sous-espaces (McCool, et al., 2010) examinent l'applicabilité de deux topologies MMC (MMC 1D et MMC 2D) sur des cartes de disparité et conduisent une étude comparative avec un système GMM. Ils montrent que le MMC 2D peut devenir considérablement plus performant qu'un système GMM équivalent. Une autre approche 2D qui a récemment gagné beaucoup de popularité est sans doute celle basée sur les Motifs Binaires Locaux (connus sous LBP, leur abréviation en Anglais), dont la grande efficacité a été prouvée malgré leur simplicité. Cette approche est appliquée aux cartes à surfaces normales dans (Li, et al., 2011) et améliorée par des pondérations spatiales estimées par un algorithme d'apprentissage.

- **Comparaison des nuages de points:** Ce n'est pas facile de travailler avec une représentation de visages par nuages de points, et de plus, la performance des algorithmes de reconnaissance basés sur cette représentation dépend fortement de l'étape de recalage. Il est nécessaire que les visages soient très bien alignés pour mesurer correctement la distance qui les sépare. Dans (Pan, et al., 2003), la distance de Hausdorff est adoptée en tant que fonction d'évaluation de proximité pour à la fois l'alignement fin et même pour la comparaison de modèles 3D. Dans (Lu, et al., 2004) un processus de recalage basé sur ICP est proposé pour générer une mesure de similarité. Dans (Lu, et al., 2006), en plus de la distance rigide de correspondance calculée par ICP, une différence non-rigide de recalage est mesurée après un alignement de visage via TPS.

- **Reconnaissance basée sur traits:** Les algorithmes globaux déjà mentionnés tels que les différences point-ensemble ou l'ACP sur cartes de disparité utilisent des caractéristiques globales. Ils sont limités par les exigences de normalisation/recalage et sensible aux données déformées ou manquantes. Dans (Mian, et al., 2007), un algorithme basé sur des caractéristiques est proposé; il est fondé sur des points-clés spécifiques à une personne qui sont identifiables à maintes reprises et des traits invariants à la pose extraits en se servant d'une base locale de coordonnées pour chacun. La reconnaissance est accomplie via un algorithme de correspondance de traits reposant sur un graphe. (Wang, et al., 2010) se servent des traits locaux extraits d'une SSDM calculée entre deux visages 3D alignés pour une comparaison de forme: les traits comme-Haar, les traits Gabor et LBP. Les traits les plus discriminants sont sélectionnés par un algorithme de boosting, trois "Collective Shape Difference Classifiers" (CSDC) sont entraînés et leurs résultats sont fusionnés.

Les données 3D sont avantageuses puisqu'elles sont invariantes à la pose de la tête et aux changements d'éclairage. Toutefois, il y a d'autres facteurs qui peuvent affecter la surface faciale 3D comme les expressions, l'occultation et le vieillissement. Les approches

qui visent une robustesse contre de telles variations peuvent être analysées en deux classes distinctes. Alors que la première classe utilise des mesures géodésiques pour obtenir une représentation invariante de l'expression, la deuxième classe tente d'extraire des traits discriminants basés sur des distances géodésiques.

Dans le cadre de la première classe, on peut citer (Mpiperis, et al., 2007) dans laquelle un paramétrage polaire géodésique de la surface faciale est présenté. Ainsi, une représentation "expression-invariante" est demandée pour être utilisée pour une reconnaissance faciale 3D avec succès en la présence d'expressions, tout en prouvant la pertinence associée à l'hypothése que les attributs d'une surface intrinsèque, en particulier les attributs géodésiques, ne changent pas sous déformations isométriques. (Berretti, et al., 2006) obtiennent une représentation faciale "expression-invariante" compacte par extraction des rayures iso-géodésiques à partir des surfaces faciales, qui sont ensuite, structurées dans une forme de graphe relationnel attribué.

Parmi la deuxième classe, (Li, et al., 2007) remaillent régulièrement, les surfaces faciales et extraient des nombreux attributs géométriques intrinsèques (angles, distances géodésiques, courbures, etc.). La relation expression/insensibilité est davantage améliorée par des pondérations des éléments qui accordent une importance plus haute aux éléments stables.

En plus des approches susmentionnées, quelques chercheurs proposent d'analyser des déformations faciales afin de comprendre et contrôler les variations non-rigides de la surface faciale. De cette façon, ils tentent d'apprendre ou à modéliser le comportement non-rigide des expressions pour soit différencier soit éliminer ses effets, en se basant sur la supposition que les expressions sont transférables d'un individu (ou d'un groupe d'individus) à un autre.

Dans (Lu, et al., 2005), l'enregistrement rigide et non-rigide des modèles faciaux est faite via ICP et TPS, respectivement. Pour l'enregistrement non-rigide, trois ensembles de points sont définis: deux ensembles de contrôle pour TPS et un ensemble de validation pour calculer l'erreur d'alignement. Après l'application des deux procédures de recalage, les vecteurs de déplacement sont extraits pour tout point de validation, puis classés en vecteurs intra-sujets ou inter-sujets par l'intermédiaire du classificateur non linéaire SVM.

De façon similaire, (Li, et al., 2010) tentent eux aussi de différencier les déformations d'expression d'après les différences interpersonnelles. Un visage normal est déformé pour qu'il soit adapté aux visages cibles et les paramètres optimisés qui minimisent les erreurs de correspondance sont considérées comme vecteurs caractéristiques. Pour l'entraînement, les vecteurs caractéristiques sont extraits, comme déformations intra-classe et interclasse, des paires des modèles faciaux ayant la même identité mais pas les mêmes expressions et des paires de différentes identités mais sans expressions (neutres), dans cet ordre. Ces déformations sont plus tard combinées linéairement pour synthétiser un visage test avec une certaine expression, et les coefficients interclasses sont utilisés pour la reconnaissance.

Exceptionnellement dans (Lu, et al., 2006), les auteurs essayent d'introduire l'expression présente dans les données test à travers toute la galerie au lieu de la supprimer.

Un énorme inconvénient avec cette deuxième catégorie de méthodes est le fait qu'elle exige des données supplémentaires pour l'entrainement. Quel que soit la taille de l'ensemble d'apprentissage, l'intervalle d'expressions possible est si large qu'une compréhension exhaustive est difficilement réalisable. En outre, la validité du transfert d'expressions entre individus est loin d'être incontestable.

Le dernier type d'approches robustes aux déformations, que nous citons ici est la méthode basée sur ces régions; elle analyse le visage 3D en le divisant en plusieurs composants et en appliquant une fusion au niveau score, rang ou abstrait. Par exemple, dans (Alyuz, et al., 2008), les modèles régionaux moyens sont construits et utilisés pour un alignement basé sur région via ICP. Les performances de reconnaissance des régions faciales de chaque individu sont rapportées ainsi que les performances combinées utilisant des méthodes de fusion différentes. D'une manière similaire, (Spreeuwers, 2011) présente un système de reconnaissance faciale 3D basé sur la fusion de 30 classificateurs qui utilisent PCA-LDA pour l'extraction de traits et le rapport de probabilité comme score de similarité. Les scores régionaux sont combinés par "majorité qualifiée".

Dans (Mian, et al., 2005), les effets des expression faciales et barbe/moustache sont évités en accomplissant une reconnaissance basée sur les régions des yeux-front et nez. Ces deux régions sont alignées en utilisant l'ICP et les distances de correspondance finale sont fusionnées en utilisant la règle du produit. Dans (McKeon, et al., 2010), l'approche "fisherface" est employée pour un ensemble de 22 régions diverses. Le SFS (Sequential Forward Search) est effectué pour sélectionner les régions les plus discriminantes.

2.1.2. Reconnaissance faciale 2D assistée par la forme 3D:

Des solutions possibles pour une invariance à l'éclairage et à la pose dans la reconnaissance faciale 2D sont limitées à cause de la nature tridimensionnelle du problème. En incorporant les données de la forme 3D du visage, les chercheurs ciblent l'amélioration des performances de la reconnaissance 2D en la présence de telles variations. D'autre part, l'acquisition de modèles faciaux par l'intermédiaire des scanners peut être problématique en mode opérationnel, en particulier dans le scénario ou des individus peu coopératifs doivent être reconnus. Principalement en raison de ces facteurs, l'idée de la reconnaissance 2D assistée par la forme 3D a émergé, et consiste à reconstruire la forme 3D en se basant sur les images 2D capturées, ou bien en acquérant de données 3D asymétriquement durant la phase d'inscription.

Selon plusieurs approches, le modèle facial 3D est utilisé pour générer des images synthétiques des sujets inscrits avec différentes poses ou expressions, ou dans des conditions d'éclairage non similaires, afin de couvrir toutes leurs apparences éventuelles. Les images tests en 2D sont ensuite comparées à la galerie synthétiquement augmentée pour une meilleure correspondance.

(Lu, et al., 2006) développent un système multimodal dans lequel les modèles 3D présents dans la galerie sont utilisés pour synthétiser les échantillons de chaque nouvelle apparence avec des variations de pose et d'éclairage pour une analyse discriminante à sous-espace visant une amélioration de la performance de la modalité 2D.

Dernièrement, une étude dans laquelle un modèle de reconstruction 3D est établi en appliquant l'approche "3D Generic Elastic Model" est publiée par (Prabhu, et al., 2011). Au lieu d'élargir l'ensemble d'apprentissage, ils ont choisi d'estimer la pose du test et de générer les modèles 3D construits pour différentes poses au sein d'un espace de recherche limité autour de la pose estimée.

Un autre type de méthodologie dans la direction inverse tente de se débarrasser des variations présentes dans les images de test par une normalisation de pose et/ou une compensation d'éclairage. (Kittler, et al., 2005) proposent un système dont les modèles 3D des sujets sont acquis durant l'enregistrement et utilisés pour la correction de l'éclairage en séparant les effets de lumière et l'albédo dans les images 2D du test.

Une dernière classe d'approches diffère des deux classes précédemment citées ; celle-ci se repose sur des modèles 3D malléables (3DMM) et utilise des données 3D dans une étape intermédiaire de correspondance d'images 2D. À nouveau, un modèle facial 3D est transformé pour qu'il corresponde aux images 2D, mais, au lieu de synthétiser de nouvelles images faciales sous des conditions différentes, les coefficients du même modèle obtenus après correspondance sont utilisés comme vecteurs caractéristiques à comparer mutuellement.

Dans (Blanz, et al., 2003), un modèle malléable basé sur une représentation des visages dans un espace vectoriel est construit à partir des images numérisées de 100 sujets de sexe masculin et 100 sujets de sexe féminin. Après que des correspondances denses sont établies entre les images numérisées, une Analyse en Composantes Principales (ACP) est effectuée sur les vecteurs de formes et de textures résultant en deux bases orthogonales formées par 90 vecteurs propres. Durant le processus d'adaptation, les coefficients de forme et de texture ainsi que les paramètres d'éclairage sont estimés itérativement pour approcher le modèle malléable le plus près possible de l'image de test. Finalement, deux visages sont comparés à travers l'ensemble de coefficients qui représente la forme et la texture en utilisant la distance de Mahalanobis.

2.2. Défis et Tendances:

L'utilisation des données 3D dans le domaine de la reconnaissance faciale est évidemment très utile et fructueuse. D'autre part, cela comporte encore quelques côtés problématiques, dont le plus crucial est "les détecteurs 3D".

Défaut de la technologie de détection 3D:

La qualité des scanners 3D s'est remarquablement améliorée au fil des dernières années, cependant, leur emploi pour la reconnaissance faciale ou pour les applications de

surveillance n'est pas immédiat. Principalement, 6 propriétés peuvent être énumérée pour rendre possible les applications pratiques (Bowyer, et al., 2006):

- Temps d'acquisition minimal pour réduire l'artefact dû au mouvement du sujet et, par conséquent, pour requérir moins de coopération;
- Profondeur de champ importante avec une précision de profondeur stable;
- Opération robuste sous une "gamme de conditions normales";
- Innocuité par les yeux;
- Échantillonnage de profondeur dense;
- Profondeur à haute résolution.

Actuellement, de nombreux compromis entre ces points existent.

Le premier compromis est entre la qualité et le temps d'acquisition. Les systèmes stéréos passifs ont l'avantage de capturer les scènes 3D en un temps inférieur à celui pris par les cameras 2D, mais le taux de leur échantillonnage et la précision de leur profondeur ne sont pas souvent suffisants. Tandis que, les détecteurs actifs nécessitent un temps d'acquisition plus long mais ils peuvent fournir des résultats plus précis.

Le deuxième compromis réside entre l'innocuité pour les yeux et la profondeur robuste du champ pour les détecteurs actifs. Les lumières cohérentes telles que les lasers sont moins larges et plus éclatantes que celles non cohérentes. Un faisceau lumineux cohérent peut être dirigé et numérisé avec précision, toutefois, il comporte des dangers potentiels pour les yeux. En revanche, les projections non cohérentes ne peuvent pas fournir une telle précision d'acquisition ni une telle profondeur de champ.

Les systèmes hybrides (stéréo passif et lumière structurée) minimisent aussi le temps requis pour l'acquisition, mais ils ont souvent des difficultés avec les régions présentant de la barbe et/ou moustache. Entre autre, comme c'est le cas pour tout système dépendant de la correspondance stéréo, ils ont un échantillonnage de points clairsemé.

Un dernier point important à propos des détecteurs 2D est leur sensibilité à l'éclairage. Bien que les données 3D soient indépendantes de l'éclairage intrinsèque, l'acquisition peut être énormément affectée par les conditions d'éclairage. Pour les détecteurs basés sur la stéréo qui prend une ou plusieurs images standards 2D, les niveaux de lumière de très haute ou très basse intensité peuvent altérer la qualité d'acquisition.

Tendances:

Malgré le fait que les variations d'expression et d'occultation sont étudiées exhaustivement, elles demeurent très loin d'être closes. Les algorithmes de reconnaissance faciale 3D ont besoin d'être plus tolérants aux scénarios du monde réel qui englobent, en plus des expressions, de nombreux facteurs différents comme les variations de pose, barbe ou moustaches et les accessoires.

Bien que la forme faciale 3D soit invariante à l'angle de vision, le problème peut être interprété comme un défi de données manquantes puisque quelques parties du visage de test peuvent être indisponibles à cause de l'occultation. Par ailleurs, l'impact des accessoires faciaux tels que les lunettes et le rouge à lèvres n'arrivent pas encore à faire surgir un intérêt similaire à celui généré par l'impact des expressions faciales ou des variations de pose.

De plus, de nouveaux défis face à la reconnaissance faciale 3D vrais chaque jour, dont les plus récents sont le vieillissement et le phénomène de jumeaux. Surtout avec l'introduction de la nouvelle base de données 3D TEC, il est raisonnable de s'attendre à une orientation de la recherche vers une reconnaissance faciale robuste de vrais jumeaux. D'un autre côté, la recherche visant une reconnaissance 3D invariante au vieillissement rencontre encore un problème sérieux : l'absence d'une base de données. Les 3DMMs sont proposés par (Park, et al., 2010), dans le contexte de la reconnaissance faciale 2D assistée en 3D, pour modéliser le vieillissement afin d'obtenir une robustesse contre les variations d'âge .

Un autre challenge à affronter est la reconnaissance 3D à distance. Aujourd'hui, la solution pour ce problème est principalement recherchée en se concentrant sur le processus d'acquisition ciblant une détection de haute qualité à larges gammes.

Avant de conclure cette section, Kinect mérite d'être cité comme un dernier point avec les nouveautés qu'il a apportées à la détection 3D. Avec son coût abordable, sa fiabilité et sa rapidité, il est déjà devenu un détecteur 3D important pour une large gamme d'applications à partir d'une vision HCI par ordinateur et le domaine de la reconnaissance faciale n'a aucune raison d'être différent.

3. Prétraitement et annotation automatique d'un visage 3D:

Les sorties des détecteurs 3D ont souvent besoin d'être prétraitées avant qu'elles ne soient utilisées pour la reconnaissance. Les images numérisées initiales incluent fréquemment des parties non faciales comme les épaules ou les vêtements et sont imparfaites à cause de la présence de bruit, trous et pointes. En outre, le repérage automatique peut aussi être inclus dans l'étape de prétraitement, puisque plusieurs rognages et alignements techniques exigent un ou plusieurs points caractéristiques faciaux tels que les coins des yeux, la pointe du nez, les coins des lèvres etc.

3.1. Recadrage de régions faciales:

Afin de recadrer une région faciale, on définit grossièrement, avant tout, la pointe du nez comme proposé dans (Lu, et al., 2005). Sur la carte de repérage, pour chaque ligne, la position avec la valeur maximale de z est trouvée. Puis, pour chaque colonne, le nombre de ces positions est calculé et un histogramme est créé. Le sommet de cet histogramme

est choisi pour être la colonne de la position de la ligne médiane verticale, et le point maximum de ce contour est identifié en étant la pointe du nez. Finalement, on utilisant une sphère de 100 mm de rayon centrée à 10 mm de la pointe du nez dans la direction +z, la surface faciale est recadrée.

3.2. Suppression des pointes et des trous:

Pour les supprimer, les pointes doivent être tout d'abord détectées. A cette fin, nous analysons les pixels avoisinants pour décider si une valeur de pixel sur la carte de repérage indique ou non la présence d'une pointe. Pour remplir les trous, nous employons une méthode basée sur l'interpolation linéaire. Les données manquantes sont interpolées en ajustant une ligne droite entre les sommets avoisinants avec des coordonnées x, y et z connues.

3.3. Lissage

Pour obtenir un lissage tout en préservant les traits anguleux tels que les coins, une technique de filtrage non itérative que préserve les traits (filtrage bilatéral) est appliquée avec quelques modifications. Un filtre bilatéral se distingue du filtre "Flou Gaussien" par son terme additionnel de gamme. Dans ce cas-là, non seulement les distances entre les positions comptent, mais également la variation des intensités sont prises en compte afin de protéger les traits contre toute sorte d'atténuation. Cela peut être formulé comme suit:

$$P_e = BF(P) = \frac{1}{w_p}\sum_{s\in V} G_{\sigma_s}(||P - S||)F_{\sigma_r}(|\left|N_p - N_s\right||)S \qquad \text{Équation 18}$$

Dans l'équation ci-dessus, P_e est le point à estimer et S sont les points appartenant au voxel V, qui vont contribuer au calcul de la position débruitée de P. $G\sigma_s$ est la fonction de pondération spatiale prise comme un filtre Gaussien unidimensionnel. Par conséquent, la pondération décroit avec la distance spatiale à P. $F\sigma_r$ est la longueur de la fonction de pondération, qui est à son tour une fonction Gaussien unidimensionnelle qui fait baisser l'influence des pixels S ayant une valeur normale, N_s, très différente de N_p.

3.4. Repérage automatique:

Les points des traits faciaux sont très importants pour beaucoup d'applications de vision par ordinateur comme la normalisation des visages, le recalage et le codage des visages humains basé sur des modèles. En considérant que la coopération du sujet est requise durant l'enregistrement, notre système de repérage automatique repose sur l'hypothèse restrictive d'un environnement d'acquisition bien contrôlé dans lequel les sujets sont inscrits sous les conditions suivantes:

- En pose frontale;
- Sans aucune occlusion ni expression importante.

Dans notre approche, premièrement, une analyse de la ligne médiane faciale (profil vertical) est faite et 9 points de repère sur cette ligne sont détectés. En se basant sur ces informations, le visage est découpé en sous-régions. Cela facilite la localisation grossière des sourcils, des yeux, du nez et des lèvres. Ensuite, une analyse supplémentaire est faite dans les sous-régions extraites pour détecter les points d'intérêt.

3.4.1. Analyse du profil vertical:

L'analyse du profil vertical commence par l'extraction de la ligne médiane faciale, et, à cette fin, la méthode proposée dans (Lu, et al., 2005) est utilisée. Pour toute ligne, la position ayant la valeur maximale de z est trouvée, puis, pour chaque colonne, le nombre de ces positions est calculé pour qu'un histogramme soit créé. Le sommet de cet histogramme est choisi pour la position de la ligne médiane verticale. Le point le plus haut de la courbe du profil est détecté comme étant la pointe du nez. Les iris sont localisés grossièrement en haut du visage par rapport à la pointe du nez. Ensuite, les images 2D et 3D du visage subissent une rotation afin d'aligner horizontalement les centres des iris et l'analyse du profil est répété.

Sur la courbe du profil, il existe des bosses (front, nez, lèvres et menton) et des trous entre elles. Les sommets et les nadirs de la courbe du profil sont trouvés à l'aide des points d'intersections des courbes représentant la première et la deuxième dérivées avec l'axe des zéros (les différences entre les éléments adjacents de la ligne médiane). Désormais, ces points sont connus, le visage peut être divisé en sous-images plus significatives pour localiser ou raffiner les points d'intérêt.

3.4.2. Extraction des points d'intérêt faciaux

Yeux:

Les données 2D sont utilisées pour détecter les points d'intérêt autour des yeux, notamment les centres des iris, les coins intérieurs et extérieurs des yeux et les limites supérieures et inférieures des iris. La région "sans peau" est retrouvée en supprimant les pixels avec les valeurs (C_b, C_r) les plus fréquemment présentes dans l'image, en utilisant l'espace YC_bC_r et les petites îles dans le masque binaire obtenu sont éliminées.

Après l'obtention des régions oculaires, des contours de pixels sont tout d'abord construits en utilisant le filtre de Canny (dit aussi détecteur de Canny). Ensuite, la transformée de Hough est appliquée sur les contours de pixels pour détecter les cercles.

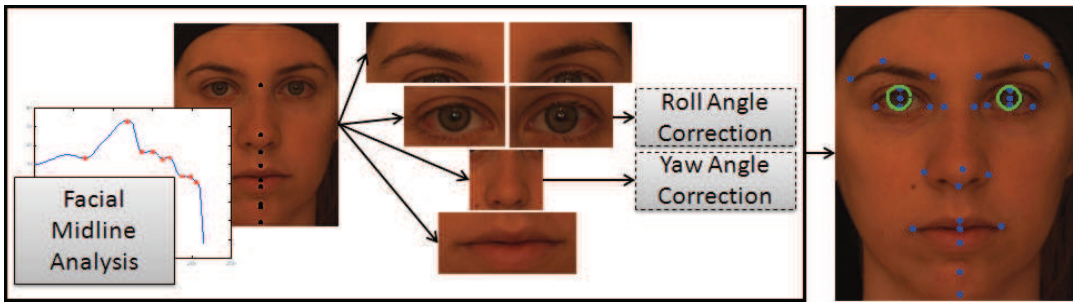


Figure 69: Organigramme du système proposé

Pour chaque cercle détecté, un score de chevauchement est calculé comme étant le rapport des portions détectées du cercle par rapport au périmètre du cercle complet.

Par la suite, les cercles détectés (candidats iris) sont regroupés en deux classes selon leurs positions: côté droit ou côté gauche. Puis, pour toute paire possible (cercle droit, cercle gauche), les mesures qui suivent sont évaluées et utilisées pour éliminer les paires incompatibles:

- La distance verticale des centres: elle ne doit pas dépasser un certain seuil;
- La distance horizontale des centres: elle doit être supérieure à un certain seuil;
- La différence entre les rayons: elle ne doit pas dépasser un certain seuil.

Parmi les paires bien adaptées, celle présentant le score de chevauchement le plus grand est élu pour être les iris. Leurs emplacements sont raffinés dans de plus petites fenêtres situées aux centres des iris détectés.

Pour les coins des yeux, les arêtes horizontales crées par les paupières sont recherchées à la fois sur l'image segmentée en couleurs ainsi que sur l'image en échelle de gris. Après que les deux cartes de contours de pixels soient superposées, des polynômes du deuxième et du troisième ordre sont ajustés pour les arêtes des paupières inférieures et supérieures respectivement, en utilisant la méthode des moindres carrés. Les coins des yeux intérieurs (près de l'arête du nez) sont déterminés comme les points d'intersection des deux polynômes ajustés.

Nez

Les données 3D sont utilisées pour détecter les point3s du nez. La courbure de Gauss minimale est calculée pour tout point appartenant à la région du nez. Ensuite, la détection des arêtes est appliquée à l'image dans laquelle l'intensité est considérée comme la courbure minimale. Ces arêtes révèlent la position des points d'intérêt sur les deux côtés de la pointe du nez.

Sourcils

Dans notre système, une approche de contours actifs basée sur les régions (sans arêtes) est adoptée pour les sourcils. Après l'obtention du contour, seulement la limite supérieure est extraite, ainsi que les informations de l'arête horizontale, obtenues par l'intermédiaire du détecteur de Canny, y sont ajoutées afin d'améliorer la robustesse contre les sections à contours erronés. Finalement, un polynôme du 3ème degré est ajusté pour supprimer les points aberrants, avant que les premiers, plus haut et dernier points soient sélectionnés comme points d'intérêt.

Lèvres

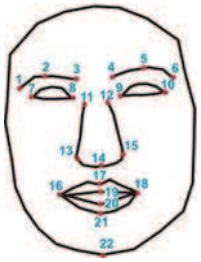
Puisque nous travaillons sur des visages avec des expressions neutres, la bouche est supposée être fermée. Une bouche fermée engendre toujours une ligne plus foncée entre les deux lèvres. En se basant sur cet acquis, le point de contact des lèvres est raffiné en appliquant une analyse de projection verticale. Dans une fenêtre étroite prise autour de ce point-là, les arêtes horizontales sont détectées et en projetant horizontalement ce contour

de pixels, les coins droit et gauche des lèvres sont détectés. Les arêtes supérieures et inférieures sont localisées au moyen de l'analyse de profil vertical.

3.4.3. Tests et Résultats

Afin d'évaluer la performance du système proposé, la base de données Bosphorus [22] est utilisée. L'erreur est prise pour être le rapport de la distance Euclidienne entre le coin manuellement étiqueté et le point détecté par la distance interoculaire (DIO); les taux de succès dans la Table 19, pour tout point détecté, sont donnés pour un seuil de 10% de DIO.

Table 19: Les taux de réussite associés à la détection des points de repérage faciaux.



Points	Taux de Réussite	Points	Taux de Réussite	Points	Taux de Réussite
Points 1 et 6	77.09%	Point 12	75.59%	Point 18	87.96%
Points 2 et 5	61.04%	Point 13	91.64%	Points 19 et 20	97.66%
Points 3 et 4	77.43%	Point 14	93.65%	Point 21	94.65%
Points 7 et 10	93.48%	Point 15	96.99%	Point 22	73.58%
Points 8 et 9	95.48%	Point 16	88.96%	Centres d'iris	100%
Point 11	80.60%	Point 17	95.32%		

3.4.4. Approche Asymétrique: Apprendre les Variations d'Expressions:

Une des sources de variation les plus critiques dans la reconnaissance faciale est constituée par les expressions faciales, en particulier dans le cas où un seul échantillon par personne est disponible à l'enrolment. Nous abordons ce problème avec un schéma basé sur une analyse par synthèse dans lequel un certain nombre d'images faciales synthétiques avec différentes expressions sont produites. Pour ce faire, un modèle animable 3D est généré pour chaque utilisateur en utilisant 17 points de repérage pour lesquelles une méthode d'extraction automatique est précédemment présentée.

3.5. Système proposé:

Pour le système proposé, l'enregistrement est fait en 2D et en 3D pour chaque personne sous un environnement contrôlé (images frontales du visage avec expression neutre sous un éclairage ambiant). La forme 3D obtenue de la surface faciale ainsi que la texture inscrite sont prétraitées, pour extraire les régions faciale puis pour éliminer les pointes, les trous et le bruit. 17 points caractéristiques sont automatiquement détectés en utilisant les deux modalités en fonction des propriétés régionales du visage et utilisés pour modifier un modèle facial animable générique pour le transformer complètement en un visage cible. Le modèle générique avec 71 points MPEG-4 manuellement étiquetés est approprié pour simuler des mouvements et des expressions faciales via un logiciel d'animation appelé visage|life™.

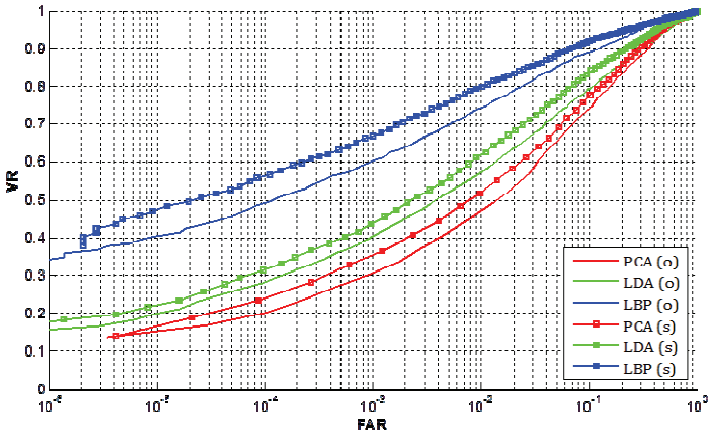


Figure 70: ROC courbes pour les méthodes ACP, LDA et LBP avec FRGC v2

3.6. Construction des Modèles Faciaux Animatable:

Afin de construire un modèle faciale animable pour chaque sujet inscrit, un algorithme de déformation basé sur les conclusions de (Tena, et al., 2006) est proposé. Un modèle générique facial, avec des trous pour les yeux et une bouche ouverte est considérablement déformé pour qu'il s'adapte aux modèles faciaux dans la base de données, en utilisant la méthode TPS. 17 point automatiquement détectés, ainsi que le reste des points FDP pour une animation MPEG-4 conforme, sont marqués manuellement sur le visage générique.

Avant que la déformation TPS ne soit appliquée, nous devons nous assurer que le visage cible et le visage générique est correctement aligné. Premièrement, l'échelle du modèle générique est modifiée dans les trois directions et une transformation rigide est calculée à partir des 17 paires de points correspondantes.

Ensuite, le modèle générique est grossièrement déformé pour un meilleur alignement au visage cible. En prenant les 17 paires de points caractéristiques comme étant les repères de la source et de la cible, une déformation non linéaire est calculée. Cette déformation déplace chaque point du réseau facial se situant autour d'un repère source vers la même position mais avec le repère cible qui correspond. Les points dans l'intervalle sont doucement interpolés utilisant l'algorithme TPS.

Enfin, pour tout autre sommet sur le visage générique, le sommet qui correspond sur le visage cible est trouvé et utilisé dans le calcul TPS. Ainsi, la moitié des points sur le modèle générique se conforment à leurs équivalents au visage cible et maintiennent leur régularité.

3.7. Evaluation de Performance

Afin d'évaluer le système proposé, deux différentes bases de données sont utilisées: FRGC (Phillips, et al., 2005) et Bosphorus (Savran, et al., 2008). Contrairement au scénario qu'on a sous la main, dans la base de données FRGC, les images sont prises sous des conditions d'éclairage non contrôlées. Pour cette raison, le repérage automatique est testé seulement dans la base de données Bosphorus.

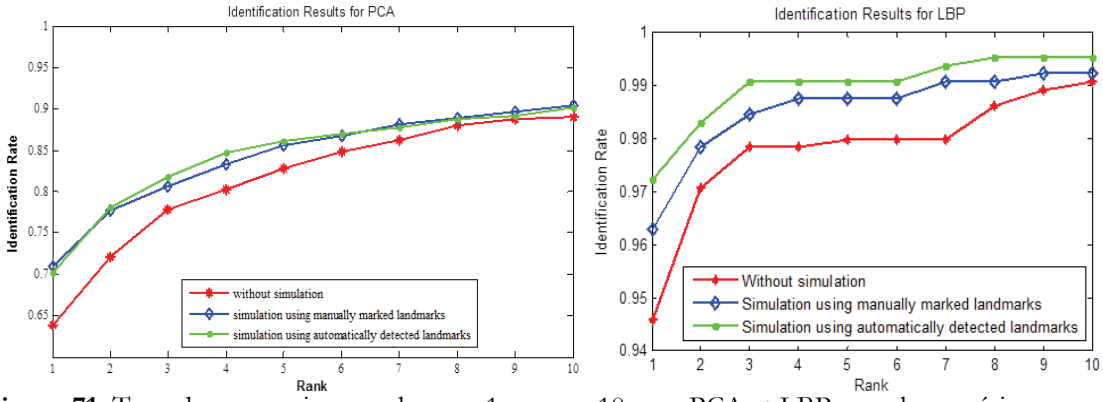


Figure 71: Taux de reconnaissance du rang-1 au rang 10 avec PCA et LBP pour les expériences

3.7.1. Expériences avec la Base de Données FRGC

Une fois qui un modèle animable pour chaque personne enregistrée est construit en se basant sur 17 repères manuellement étiquetés par Szepycki et al (Szeptycki, et al., 2009), il est animé pour 12 expressions distinctes (prédéfinies dans l'outil d'animation) et est stocké dans la galerie de simulation.

Les images, les ensembles de galeries et de tests sont tous prétraités en appliquant un rognage, une normalisation géométrique (64x80 pixels) et une égalisation d'histogramme. Trois techniques-clés sont adoptées pour observer l'effet de l'augmentation d'une des galeries: ACP, LDA (Belhumeur, et al., 1997) et LBP (Ojala, et al., 2002) (Figure 70)

3.7.2. Expériences avec la Base de Données Bosphorus

Pour observer les effets des erreurs du repérage automatique, les tests d'identification sont conduits avec à la fois des points caractéristiques automatiquement et manuellement marqués. Les modèles animable sont générés et les même 12 expressions sont simulées pour créer la galerie synthétique d'images.

Les taux de reconnaissance pour ACP et LBP sont représentés à la Figure 71. Ces résultats sont très impressionnants étant donné que, malgré les erreurs tiées à la détection automatique des repères, l'utilisation de ces points pour la génération d'une image simulée mène à une amélioration des résultats comparables a cette obtenue avec des points manuellement marqués.

4. Propriétés Discriminantes Des Paramètres de Déformation

Dans cette étude, au lieu d'adopter un algorithme TPS pour enrichir l'enregistrement, nous analysons les propriétés discriminantes des paramètres obtenus par la déformation d'un modèle générique facial vers des visages cibles utilisant le TPS. Les paramètres de déformation (WP) qui décrivent les transformations non globales et non linéaires et qui représentent les déviations de la structure géométrique générale sont transférés au

classifieur pour une reconnaissance faciale. Le caractère descriptif de ces vecteurs est analysé sur FRGC.

4.1. Système Proposé

Nous proposons un système qui utilise des paramètres extraits d'un algorithme de déformation basé sur les conclusions de (Tena, et al., 2006), en tant que signatures biométriques. Un modèle générique de la tête (également utilisé dans le chapitre précédent), avec des trous pour les yeux et une bouche ouverte considérablement déformée pour qu'elle s'adapte aux modèles faciaux présents dans la base de données. Notre objectif est d'analyser les propriétés discriminantes des paramètres de déformation obtenus durant l'adaptation.

4.2. Extraction des Traits

Conformément à une démarche similaire à celle suivie dans la section précédente, l'échelle du modèle générique de la tête est initialement modifiée et le modèle est aligné selon le modèle facial a analyser. Puis, une déformation TPS à une seule étape est réalisée. Pour les surfaces 3D S et T, et pour l'ensemble des points correspondants sur chaque surface, Pi et Mi respectivement, l'algorithme TPS calcule la fonction l'interpolation f(x,y) pour estimer T', qui s'approche de T par déformation de S:

$$T' = \{(x', y', z')\ st. \forall\ (x, y, z) \in S, x' = x, y' = y, z' = z + f(x, y)\} \quad \text{Équation 19}$$

$$f(x, y) = a_1 + a_x x + a_y y + \sum w_i U(|P_i - (x, y)|) \quad \text{Équation 20}$$

Avec la fonction de noyau U(.) exprimée comme suit:

$$U(r) = r^2 ln(r^2), r = \sqrt{x^2 + y^2} \quad \text{Équation 21}$$

Donnée par (6), la fonction d'interpolation f(x,y) inclut les coefficients de déformation w_i, $i \in \{1,2,...n\}$ à utiliser.

En utilisant 140 points manuellement sélectionnés sur le visage générique et couplés avec les sommets les plus proches dans la surface cible après alignement, la déformation TPS est appliquée au modèle générique de la tête. La fonction f(x,y) donnée dans l'équation 20 exécute l'interpolation dans la direction z. Etant donné que l'échelle du visage générique est changée et le visage générique est aligné d'avance, la partie raffinée est ignorée et les pondérations w_i sont prises en considération. Quand nous transposons la formule pour les deux directions qui restent, nous obtenons comme résultat le vecteur de déformation: $[(w_{1x},w_{1y},w_{1z}), (w_{2x},w_{2y},w_{2z}) \ldots (w_{140x},w_{140y},w_{140z})]$. C'est une représentation dense des surfaces faciales, puisque les données sont stockées dans le modèle générique.

4.3. Mesures de Distance

Afin de mesurer la distance entre surfaces faciales, le cosinus de l'angle et la distance Euclidienne entre les deux vecteurs de déformation correspondants sont calculés. Cela a

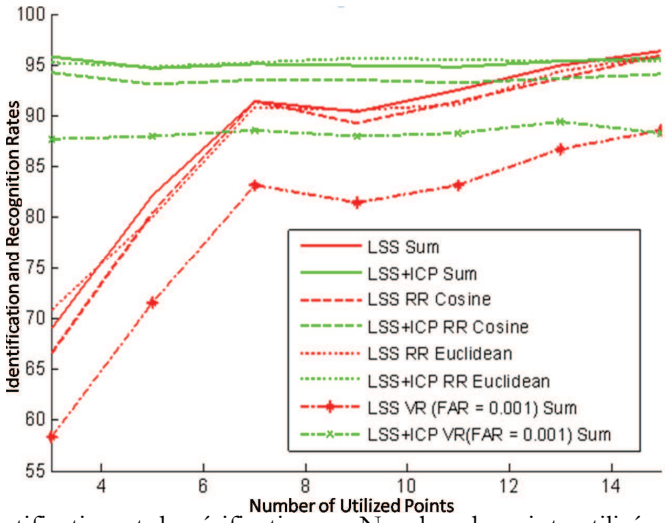


Figure 72: Taux d'identification et de vérification vs. Nombre de points utilisés pour alignement

pour résultat deux vecteurs de distance 140×1 pour la paire de visage comparée. En mesurant la tendance centrale de ces vecteurs par la méthode de la moyenne tronquée, deux valeurs de distance sont obtenues. Entre autres, pour chaque point de contrôle, les N personnes les plus proches sont évaluées, et, pour chaque personne de la galerie, une pondération est calculée en fonction de son nombre d'occurrences dans les possibles ensembles de solutions.

4.4. Expériences et Analyses

Pour tester les propriétés discriminantes des vecteurs caractéristiques extraits, nous avons travaillé sur les deux versions de la base de données FRGC (Phillips, et al., 2005) où la première version est utilisée comme étant la partition d'entrainement pour analyser les différentes techniques et obtenir les paramètres optimaux.

Pour la version 1 de FRGC, les tests de reconnaissance sont conduits avec et sans alignement additionnel ICP et les résultats sont fournis à la Figure 72 pour la distance Euclidienne (TM_{euc}) et Cosinusoïdale (TM_{cos}) ainsi que leur fusion obtenue par la règle de la somme. Avec l'alignement additionnel ICP, les taux de succès deviennent indépendants du nombre de repéres, et en fait, le meilleur résultat (95.81%) est obtenu en utilisant trois point seulement (**m2**). D'autre part, sans ICP, les taux de reconnaissance augmentent avec le nombre de points. Le meilleur taux a été enregistré avec 15 points (**m1**), et il valait 96.41%. En ajoutant les pondérations et la normalisation, la précision s'améliore.

$$D(i) = \frac{TM_{cos}(i)}{C(i)} + \frac{TM_{euc}(i)}{C(i)} \qquad \text{Équation 22}$$

Les expériences sur la version 2 de FRGC sont réalisées avec les paramètres qui optimisent l'identification du rang-1 et le taux de la version 1 de FRGC. Pour la méthode de la moyenne tronquée, la portion à ignorer est prise comme 10%, et pour le calcul des coefficients, les ensembles possibles de solutions sont créés par le rang N, où N vaut 20% de la taille de la galerie.

Table 20: Les taux d'identification et de vérification (à 0.001 FAR) et les valeurs EER pour FRGCv1 m1 et m2, sans et avec pondérations et normalisation.

	IR	VR	EER
m1	96.41%	88.62%	0.021
m2	95.81%	87.72%	0.030
m1 avec pondérations+ normalisation	97.30%	99.10%	0.005
m2 avec pondérations + normalisation	97.31%	98.65%	0.003

Les performances des deux approches: la premiére, avec 15 points et sans ICP (m1), et la deuxième, avec 3 points et avec ICP (m2) sont comparés dans la Table 20.

Table 21: Les taux d'identification et de vérification (à 0.001 FAR) et les valeurs EER pour toute la base de données (A) de FRGCv2 et pour trois sous-ensembles: Neutre (N), petit (S) et grand (L).

	IR m1	VR m1	EER m1	IR m2	VR m2	EER m2
N	97.13%	99.37%	0.003	92.40%	95.37%	0.015
S	85.81%	91.70%	0.018	84.74%	88.22%	0.031
L	71.92%	76.35%	0.033	72.75%	75.93%	0.050
A	89.55%	94.52%	0.015	86.74%	90.92%	0.026

5. Estimation de Fiabilité Régionale pour Visage 3D

Les données 3D pour reconnaissance faciale sont avantageuses par rapport à leurs équivalentes 2D étant donné qu'elles sont invariantes à l'éclairage et à la pose. Néanmoins, les variations d'expressions et les occultations demeurent considérées comme des défis majeurs puisque les distorsions de forme entravent une correspondance précise. De nombreux algorithmes ont été développés pour surmonter ce problème en proposant, principalement, des approches basées régions suivant lesquelles les scores de similarité sont séparément calculés par des méthodes de correspondance (locales régionales) et fusionnés pour la reconnaissance. Dans ce chapitre, nous présentons un schéma d'évaluation de scores de fiabilité régionale qui estime les distorsions induites par les expressions ou les occultations dans différentes régions faciales. Ainsi, des scores de fiabilité sont obtenus, qui peuvent ensuite être utilisés lors de l'étape de fusion pour la reconnaissance.

5.1. Segmentation Automatique

Pour une segmentation automatique du visage, un modèle 3D neutre générique, précédemment segmenté en 7 régions, est utilisé. Après l'avoir adapté à un modèle appartenant à la base de données, la segmentation du modèle est faite par la projection topographique des labels des régions.

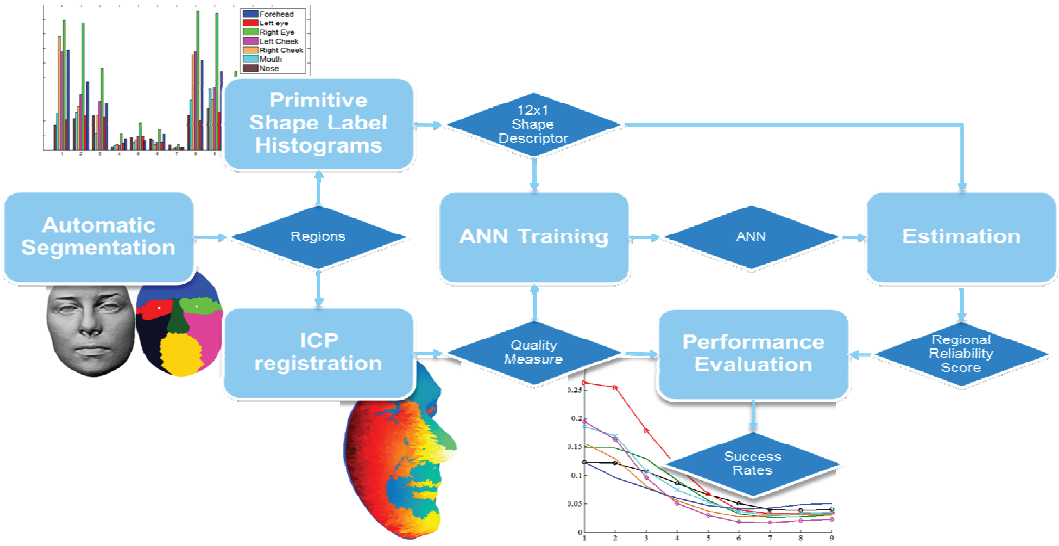


Figure 73: Organigramme de la méthode proposée pour l'estimation de la fiabilité régionale et son évaluation

En vue d'établir une correspondance dense entre les modèles faciaux à analyser et le visage générique, 3 points de référence manuellement marqués sont utilisés : les coins extérieurs des yeux et la pointes du nez. Tout d'abord, le visage générique est rigidement aligné au modèle inspecté. Une transformation linéaire est calculée selon la méthode des moindres carrés, en se basant sur deux ensembles de repérage. Ensuite, l'alignement est davantage amélioré pour des déformations non rigides en appliquant une déformation TPS sur le modèle générique et toujours en utilisant les mêmes 3 points. Enfin, puisque les deux surfaces sont bien alignées, le point le plus proche sur le visage générique est trouvé pour chaque sommet du visage afin d'être segmenté et l'étiquette de sa région est copiée.

5.2. Étiquetage de la Forme Primitive

L'utilisation de distributions de forme primitive dans les différentes régions faciales est suggérée dans (Wang, et al., 2006) pour la reconnaissance des expressions faciales. Dans cette étude, nous utilisons des traits similaires, non pas pour reconnaître les expressions mais pour évaluer la qualité des segments faciaux.

Pour déterminer la classe de la forme primitive de chaque sommet, tout d'abord, les valeurs des gradients de tous les sommets sur le réseau triangulaire du visage sont calculées. Ensuite, la courbure maximale et la courbure minimale sont obtenues pour chaque sommet sur la surface en ajustant un polynôme du second ordre aux points adjacents. Dès que la magnitude du gradient (G - la pente de la surface) et les courbures maximale et minimale (k_1 et k_2 – degrés de courbure maximum et minimum, dans cet ordre) pour un sommet sont obtenues, elles sont utilisées pour classer le sommet en question comme une des douze formes primitives dont les distributions d'histogramme sont calculées pour chaque régions, entraînant 7 descripteurs de forme, de taille [12x1] chacun.

5.3. Mesures de la Qualité de Surface

Dans cette partie de l'étude, un système automatique est développé pour mesurer les qualités de surface des régions faciales. Les sorties de ce système ainsi que les distributions de forme primitive qui correspondent sont utilisées pour instruire un ANN et pour évaluer le schéma final.

Pour ce faire, chaque partie appartenant à toute image faciale numérisée d'un sujet est recalée à la partie, qui correspond, de l'image numérisée neutre (sans expression) et nette (sans occlusion) du même sujet (modèle de référence) en utilisant la méthode ICP. L'erreur obtenue après le recalage final est acceptée afin de mesurer les déformations (distorsions) de la surface. Cette mesure nous livrera les déviations d'une surface équivalente non déformée.

5.4. Expériences

Le schéma proposé pour l'évaluation de la qualité de surface est testé sur la base de données FRGC v2 (Phillips, et al., 2005). Etant donné qu'un modèle net et neutre est requis pour chaque personne pour mesurer automatiquement les qualités d'une surface régionale, les sujets sans modèle de référence sont éliminés. Pour les expériences, la totalité de l'ensemble d'échantillons est divisée en 2 blocs: la première moitié (50%) des échantillons pour l'entrainement et la validation (1562) et le reste (1561) pour le test. Le système proposé au chapitre précédent qui adopte les paramètres de déformation extraits d'une déformation TPS en tant que signatures biométriques est employé comme méthode de base.

7 ANNs pour établir des correspondances entre les distributions de forme primitive (12 entrées) et les mesures de qualité (1 sortie) avec une couche cachée de 20 neurones sont construits et entrainés selon l'optimisation de Levenberg-Marquardt jusqu'à ce que l'erreur quadratique moyenne (MSE) des échantillons de validation cesse de croître. Les ANNs qui en résultent sont utilisés pour estimer les niveaux de confiance régionale pour l'ensemble de test. La MSE obtenue pour toutes les régions et la déviation standard correspondante sont présentées à la Table 22.

Table 22: MSE moyenne pour les estimations de la qualité régionale, leurs déviations standards et les scores minimaux mesurés pour chaque région

	Front	Œil (G)	Œil (D)	Joue (G)	Joue (D)	Bouche	Nez
MSE	0.0405	0.0312	0.0345	0.0318	0.0352	0.0511	0.0236
STD	0.0626	0.0463	0.0582	0.0462	0.0524	0.0620	0.0318
Min.	0.1371	0.1305	0.1261	0.1095	0.0896	0.1236	0.2750
Max	1.3852	1.2746	1.3081	1.2557	1.1904	1.2458	1.2277

Pour une évaluation, l'ensemble de test est divisé en plusieurs «boites» pour présenter les distorsions sur la surface faciale. Pour cela, les scores minimaux de qualité parmi toutes

les régions pour chaque visage de test sont pris et des projections sont faites entre 0 et 1, par une normalisation min-max. Puis, l'ensemble de test est fragmenté en 10 boîtes uniformément espacées. Afin de tester la combinaison des 7 classificateurs régionaux, deux approches sont examinées:

5.4.1. Fusion: Utilisation des Scores de Confiance en tant que Pondérations:

Les scores de confiance estimés sont utilisés comme des pondérations avec deux méthodes différentes de fusion : au niveau "score" - la règle de somme et au niveau "classe"- la méthode de Borda. Les résultats pour les approches sont illustrés à la Figure 74. En général, la méthode de Borda fournit des résultats moins précis que ceux obtenus par la règle de la somme. Toutefois, l'amélioration dans les taux d'identification est plus clairement visible pour les trois premières boites. Des augmentations de 12.50%, 6.04% et 10.76% sont obtenues pour, respectivement, les premier, deuxième et troisième boites, avec les niveaux de fiabilité utilisés comme pondérations.

5.4.2. Fusion: Conversions Probabilistes

Dans cette section, préalablement à la fusion, nous proposons d'obtenir des probabilités a posteriori P(véritable|s), d'avoir un échantillon authentique sachant les scores de correspondance régionale (s).

Cependant, comme exprimé dans (Jain, et al., 2005) les sorties des classificateurs individuels sont meilleures lorsqu'elles sont combinées directement sans être converties en probabilités en l'absence des mesures de confiance (c) qui évaluent la nature des échantillons de l'entrée. En prenant cela en considération, nous obtenons P(véritable|s) en calculant les densités conditionnelles P(s| véritable) et P(s|imposteur) basées sur les scores de confiance régionale estimés selon notre méthode proposée.

Chaque visage est attribué à une des classes possibles *m* présentes dans la galerie et en fonction des mesures distinctes (scores de correspondances - s_i) obtenues depuis leurs sous-régions R. Avec notre approche, cela est simplifié en convertissant l'attribution de la classe en une décision binaire qui répondra si le score de correspondance est "*véritable*", c'est à dire, calculé entre deux échantillons du même sujet ou "*imposteur*". Ces deux classes sont modélisées par les densités de probabilités dont les occurrences sont supposées égales. Afin de tenir compte des niveaux de confiance (c) estimés, une valeur approximative de P(véritable|s,c) est obtenue en divisant les scores de confiance en plusieurs boites et en estimant p(s|véritable) et p(s|imposteur) pour chaque boite, séparément.

Pour les expériences, les densités de probabilités conditionnelles sont estimées pour différents nombres de boites pour une comparaison : 1 (pas de boites), 2, 3, 4 et 5. Les performances de vérification et d'identification sont obtenues pour le schéma de fusion proposé avec 1, 2, 3, 4 et 5 boites en se servant de la même règle.

Les taux de succès des scores de correspondance originaux sont calculés par l'intermédiaire de la même technique de fusion. Les taux de vérification (VR) à 0.001

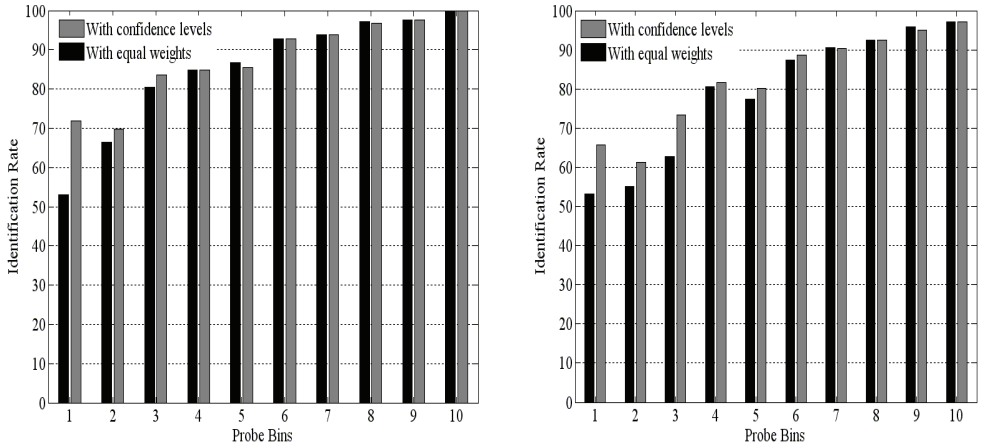


Figure 74: Taux de reconnaissance des approches basées sur la fusion pour la règle de somme et la méthode de Borda: En utilisant les niveaux de fiabilité comme pondérations et consideront des pondérations égales.

FAR, les taux erreurs d'égales (EER) et ceux d'identification (IR) rang-1 pour toutes les expériences sont fournis à la Table 23. Les résultats montrent que l'utilisation des scores de confiance pour les classificateurs régionaux est plus avantageuse avec l'approche de fusion probabiliste qu'avec l'emploi direct des scores bruts.

Table 23: Résultats comparatifs pour des tests de vérification et d'identification avant et après conversion probabiliste des scores de correspondance.

	méthode	VR	EER	IR
original	somme	63.39%	0.109	90.33%
original	w. somme	68.22%	0.088	91.04%
probabilité	1 boite	66.83%	0.072	89.76%
probabilité	2 boites	71.38%	0.049	91.61%
probabilité	3 boites	72.66%	0.047	92.13%
probabilité	4 boites	74.08%	0.046	91.93%
probabilité	5 boites	73.15%	0.045	92.00%

6. Impact des Altérations Régionales sur la Reconnaissance Faciale

La reconnaissance faciale fait face à de nombreux défis majeurs, tels que les variations de pose, d'éclairage, d'expression et le vieillissement, qui ont été considérablement examinés. Toutes ces variations modifient la texture et/ou la forme du visage d'une manière similaire pour des individus différents. Toutefois, les études sur les altérations, appliquées aux visages par chirurgie plastique ou maquillage prosthétique et pouvant être effectuées par de nombreuses méthodes différentes et en une multitude de quantités distinctes, restent toujours très limitées. Dans cette étude, nous analysons l'influence de

telles altérations, affectant la région du nez, sur les performances de plusieurs techniques fondamentales de reconnaissance faciale.

6.1. Travaux Connexes

A notre connaissance, l'impact des altérations faciales, spécifiquement dues aux chirurgies plastiques, sur la reconnaissance faciale a été analysé pour la première fois dans (Singh, et al., 2009) où l'effet de la chirurgie plastique est évalué en considérément six algorithmes de reconnaissance. La base de données utilisée comporte 506 sujets avec 2 images chacun : avant et après chirurgie. Plus tard, ce travail a été étendu dans (Singh, et al., 2010) en augmentant la base de données pour qu'elle couvre 900 sujets, mais aussi en incluant une base de données différente "sans-chirurgie" pour une comparaison de performance. Trois défauts de ces études, ultérieurement discutés au cours de ce chapitre, peuvent être identifiés comme suit:

- Etant donné qu'une seule image est fournie avant la chirurgie, une expérience de reconnaissance sans-chirurgie vs sans-chirurgie doit être conduite sur une base de données indépendante avec des sujets différents.
- Les images d'avant et d'après (chirurgie) appartenant à la première base de données ne dépendent pas tout simplement de la procédure de chirurgie, mais en plus des variations des expressions, du maquillage et barbe ou moustache. Par conséquent, elles ne peuvent pas être traitées comme des résultats de chirurgie uniquement.
- Puisqu'il s'agit d'une base de données d'images, les analyses sont limitées en 2D. Cependant, la reconnaissance faciale 3D ne cesse de gagner en popularité, étant donné qu'elle a l'avantage par rapport à son équivalant 2D d'être intrinsèquement plus robuste contre les variations de pose et d'éclairage.

6.2. Simulations des Altérations du nez

La région du nez peut être altérée de différentes façons par chirurgie plastique, applications prosthétiques ou maquillage et elle peut être transformée en plus grande, plus petite, plus large ou plus étroite. Afin de simuler ces changements et préserver l'authenticité de la forme faciale, les nez dans la base de données sont remplacés par d'autres aléatoirement choisis à partir d'autres sujets. Pour ce faire, une méthode basée sur TPS est implémentée.

Premièrement, les régions du nez de toutes les images faciales numérisées sont automatiquement segmentées d'une manière similaire à celle de (Kakadiaris, et al., 2007), où un modèle facial générique annoté est déformé pour qu'il s'adapte aux modèles cibles et les annotations sont transférées. Ensuite, les déformations nasales sont appliquées en utilisant une méthode TPS en 3D (Bookstein, 1989). Préalablement à la déformation, le modèle cible est aligné avec le modèle source utilisant 4 des 5 points de repérage se trouvant autours du nez. Une transformation linéaire incluant une rotation, une translation et une dilatation isotrope est tout d'abord calculée en utilisant la méthode des

moindres carrés en se basant sur les deux ensembles de repérage, puis appliquée au modèle source. Ensuite, une approximation TPS grossière est calculée pour toutes les paires à 5 points. Enfin, lors de la dernière étape, pour un cinquième des sommets sur le nez cible, les sommets les plus proches sur le nez source sont sélectionnés et couplés afin d'être utilisés dans une seconde et plus dense déformation TPS, ce qui donne un nez source complètement transformé en un nez cible.

Afin d'évaluer la plausibilité visuelle des bases de données créées, une enquête en ligne a été menée, qui demandent aux participants de classer les images faciales affichées au hasard (avec ou sans texture) en "images originales" ou "images simulées". Pour un total de 81 participants, le taux de réussite s'élevait à 60.68% pour les images affichées avec texture. Pour les images restantes, la performance s'est dégradée comme prévu (58.77%) étant donné que la texture donne une meilleure information sur l'originalité. Ce résultat, qui est très près de la performance moyenne du classifier aléatoire (50%), indique une très basse perceptibilité, et par conséquent, un air extrêmement réaliste des nez simulés.

6.3. Évaluation Expérimentale

L'effet des altérations nasales appliquées sur les performances de la reconnaissance faciale est évalué avec trois différents scénarios 2D et trois autres en 3D. Ces scénarios sont déterminés selon l'étude de (Singh, et al., 2010) pour des fins de comparaison.

Initialement, les quatre bases de données: DB-o2, DB-o3, DB-s2 and DB-s3 sont partitionnées en deux ensembles de données sans chevauchement pour respectivement l'entrainement et les tests. Cela est réalisé en sélectionnant arbitrairement 40% des sujets et en attribuant leurs échantillons à l'ensemble d'entrainement, tandis que le reste est attribué à l'ensemble de test. Le partitionnement est répété à 10 reprises et les performances de vérification et d'identification sont calculées à chaque fois.

- **Expérience 1 - Performance sur la base de données originale:** Les algorithmes 2D et 3D sont évalués sur DB-o où les similarités sont calculées entre les images de chaque paire originale.

- **Expérience 2 - Performance sur la base de données simulée:** Dans ce scénario, les scores de similarité entre chaque paires d'échantillons (DB-o - DB-s) sont calculés et utilisés pour évaluer les performances de reconnaissance. Pour l'ensemble d'entrainement, pour chaque sujet sélectionné, la moitié des images correspondantes sont prises de DB-o et le reste d'une DB-s.

- **Expérience 3 - Performance sur la base de données simulée avec entrainement sur une base de données externe:** Les algorithmes de la reconnaissance faciale sont habituellement entrainée en utilisant des bases de données différentes. Ainsi, dans ce scénario, l'expérience 2 est répétée, mais la partition de l'entrainement est composée d'échantillons venant d'une base de données externe, notamment "Texas 3D Recognition Database" (Gupta, et al., 2010).

6.3.1. Impact sur la Reconnaissance Faciale 2D

Trois méthodes-clés sont choisies pour être évaluées dans le cadre de la reconnaissance faciale 2D: Analyse en Composantes Principales (ACP), Analyse Discriminante Linéaire (LDA) (Belhumeur, et al., 1997) et Motif Binaire Local Circulaire (CLBP) (Ojala, et al., 2002). Les résultats sont fournis dans la Table 24.

Table 24: Précisions d'identification et taux de vérification du rang-1 à 0.001 FAR pour les algorithmes 2D FR pour les expériences 1, 2 et 3.

Algorithme	Exp. 1 (I)	Exp. 2(I)	Exp. 3(I)	Exp.3 [2] (I)	Exp. 1 (V)	Exp. 2(V)	Exp. 3(V)
ACP	40.24%	30.02%	24.74%	23.1%	27.50%	21.18%	11.66%
LDA	64.74%	51.56%	28.94%	24.1%	50.69%	40.11%	15.30%
CLBP	92.90%	88.52%	88.52%	44.8%	81.51%	71.72%	71.72%

D'après ces résultats, la meilleure performance est obtenue en utilisant la méthode CLBP pour l'identification et la vérification avec une avance remarquable. Cela montre que, le CLBP, qui est une méthode basée sur la texture, est beaucoup plus approprié que les méthodes basées sur l'apparence, comme ACP et LDA.

6.3.2. Impact sur la Reconnaissance Faciale 3D

Pour l'évaluation des systèmes de reconnaissance faciale 3D, quatre algorithmes sont choisis où les surfaces faciales sont représentées comme cartes de disparité ou nuages de points. Les cartes de disparité peuvent être incluses dans la plupart des techniques 2D existantes, y compris les méthodes par sous-espace. Dans cette partie d'étude, pareillement au cas 2D, ACP et LDA sont sélectionnées afin d'être évaluées. En outre, deux approches différentes sont implémentées pour une reconnaissance faciale 3D qui utilise les nuages de points. Dans la première technique "nuage de points", la similarité (distance dans ce cas) entre deux visages est obtenue en faisant la moyenne des distances z entre les paires de sommets. Ainsi, une approximation de la différence de volume (VD) entre deux surfaces faciales est établie. Pour la seconde approche, nous implémentons la méthode présentée en section 5 qui emploie les paramètres de déformation TPS en tant que signatures biométriques. La Table 25 représente les résultats expérimentaux.

Table 25: Précisions d'identification et taux de vérification du rang 1 à 0.001 FAR pour les algorithmes 3D FR pour les expériences 1, 2 et 3.

Algorithme	Exp. 1(I)	Exp. 2(I)	Exp. 3(V)	Exp. 1(V)	Exp. 2(V)	Exp. 3(V)
ACP	64.11%	48.40%	33.96%	49.85%	35.22%	17.42%
LDA	68.47%	58.15%	42.03%	56.67%	42.18%	17.74%
VD	68.26%	51.95%	51.95%	56.97%	35.23%	35.23%
WP	94.46%	86.64%	86.64%	81.18%	60.79%	60.79%

Comme c'était le cas pour les expériences 2D, l'utilisation de bases de données externes a un effet négatif sur les précisions de reconnaissance. Les algorithmes ont de meilleures performances lorsqu'ils sont entrainés à la fois sur les images prises avant altération ainsi que sur les images prises après altération.

7. Conclusions

Dans cette thèse, nous avons examiné l'influence et les avantages des données 3D dans le domaine de la reconnaissance faciale 3D. Motivés par une panoplie d'applications performantes précédemment lancées, nous avons exploré les méthodes pour incorporer les modalités 3D dans le processus d'identification et de vérification des visages humains.

7.1. Résumé

Tout d'abord, nous avons examiné l'état de l'art dans l'utilisation des données 3D pour la reconnaissance faciale et présenté une critique minutieusement structurée. Ensuite, nous avons décrit les méthodes de prétraitement adoptées dans cette thèse et proposé un système de repérage automatique. À la section 4, un système a été proposé, dans lequel une enrolement combiné est réalisé par, à la fois, les informations 2D et 3D alors que les images cibles demeurent en 2D, et se présente comme une stratégie optimale pour fusionner les caractéristiques avantageuses des deux modalités. Ensuite, les propriétés discriminantes des paramètres de la déformation TPS ont été examinées au cours de la section 5 où il a été observé que les variations d'expression faciale ont un impact important sur les performances. Stimulés par ces résultats, nous proposons, à travers la section 6, d'utiliser un algorithme d'estimation du niveau de confiance pour une surface faciale en coopération avec des méthodes de reconnaissance faciale 3D basées sur régions. Enfin, dans la section 7, nous avons exploré un domaine de recherche relativement plus récent. L'impact des modifications partielles apportées à la surface faciale, à la région nasale en particulier, sur les performances de la reconnaissance a été analysé.

7.2. Perspectives:

Les études menées dans le contexte de cette thèse peuvent être développées de plus d'une façon. Pour débuter, les expressions simulées dans l'approche asymétrique proposée à la section 4, pourraient bien être analysées pour mesurer leurs similarités aux expressions réelles. À la lumière de ces évaluations, on pourrait essayer de comprendre les défauts du modèle animable ou du moteur d'animation et générer des images synthétiques plus réalistes, et par conséquent plus utiles pour obtenir un system plus robuste.

Pour l'estimation des niveaux de confiance régionaux, nous souhaiterons élargir les expériences en incluant la base de données Bosphorus. Ainsi, il serait possible d'utiliser deux bases de données en tant qu'ensemble d'entrainement et ensemble de test interchangeables, en nous offrant l'opportunité d'effectuer un entrainement avec un

nombre plus large d'échantillons et d'effectuer des comparaisons avec les autres approches testées avec le même protocole expérimental.

Works Cited

Abate Andrea F. [et al.] 2D and 3D face recognition: A survey [Journal] // Pattern Recognition Letters. - 2007. - 14 : Vol. 28. - pp. 1885-1906.

Achermann Bernard and Bunke Horst Classifying Range Images of Human Faces with Hausdorff Distance [Conference] // International Conference on Pattern Recognition. - 2000. - Vol. 2. - pp. 809-813.

Achermann Bernard, Jiang Xiaoyi and Bunke Horst Face Recognition Using Range Images [Conference] // International Conference on Virtual Systems and MultiMedia. - 1997. - pp. 129-136.

Akarun Lale, Gokberk Berk and Salah Albert Ali 3D Face Recognition For Biometric Applications [Conference] // European Signal Processing Conference. - 2005.

Al-Osaimi Faisal, Bennamoun Mohammed and Mian Ajmal Expression-invariant Non-rigid 3D Face Recognition: A Robust Approach to Expression-aware Morphing [Conference] // International Symposium on 3D Data Processing, Visualization and Transmission. - 2008. - pp. 19-26.

Alyuz Nese, Gokberk Berk and Akarun Lale A 3D Face Recognition System for Expression and Occlusion Invariance [Conference] // IEEE International Conference on Biometrics: Theory, Applications and Systems. - 2008. - pp. 1-7.

Amberg Brian, Knothe Reinhard and Vetter Thomas Expression invariant 3D face recognition with a Morphable Model [Conference] // International Conference on Automatic Face & Gesture Recognition. - 2008. - pp. 1-6.

Amberg Brian, Romdhani Sami and Vetter Thomas Optimal Step Nonrigid ICP Algorithms for Surface Registration [Conference] // IEEE Conference on Computer Vision and Pattern Recognition. - 2007. - pp. 1-8.

Amenta Nina, Choi Sunghee and Kolluri Ravi Krishna The Power Crust [Conference] // ACM Symposium on Solid Modeling and Applications (SMA). - 2001. - pp. 249-266.

ASAPS Cosmetic Surgery National Databank Statistics [Online]. - The American Society for Aesthetic Plastic Surgery, 2010. - December 03, 2011. - http://www.surgery.org/sites/default/files/Stats2010.pdf.

Bartlett Marian Stewart, Lades Martin H. and Sejnowski Terrence J. Independent Component Representations for Face Recognition [Journal] // Proceedings of SPIE. - 1998. - Vol. 3299. - pp. 528-539.

Belhumeur P. N., Hespanha J. P. and Kriegman D. J. Eigenfaces Vs. Fisherfaces: Recognition Using Class Specific Linear Projection [Journal] // EEE Transactions on Pattern Analysis and Machine Intelligence. - 1997. - 7 : Vol. 19. - pp. 711-720.

Ben Amor Boulbaba [et al.] 3D Face Recognition by ICP-Based Shape Matching [Conference] // International Conference on Machine Intelligence. - 2005.

Ben Amor Boulbaba, Ardabilian Mohsen and Liming Chen Toward a region-based 3D face recognition approach [Conference] // IEEE International Conference on Multimedia and Expo. - 2008. - pp. 101-104.

Berretti Stefano, Bimbo Alberto Del and Pala Pietro 3D Face Recognition Using Isogeodesic Stripes [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2010. - 12 : Vol. 32. - pp. 2162-2177.

Berretti Stefano, Bimbo Alberto Del and Pala Pietro Description and Retrieval of 3D Face Models Using Iso-geodesic Stripes [Conference] // ACM International Workshop on Multimedia Information Retrieval . - 2006. - pp. 13-22.

Berretti Stefano, Bimbo Alberto del and Pala Pietro Facial Curves Between Keypoints for Recognition of 3D Faces with Missing Parts [Conference] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. - 2011. - pp. 46-51.

Besl Paul J. and McKay Neil D. A Method for Registration of 3-D Shapes [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 1992. - 2 : Vol. 14. - pp. 239-256.

Beumier Charles and Acheroy Marc Automatic 3D Face Authentication [Journal] // Image and Vision Computing. - 2000. - 4 : Vol. 18. - pp. 315-321.

Bhatt H. S. [et al.] Evolutionary Granular Approach For Recognizing Faces Altered Due To Plastic Surgery [Conference] // IEEE International Conference on Automatic Face & Gesture Recognition and Workshops. - 2011. - pp. 720-725.

Bing Y., Ping C. and Lianfu J. Recognizing Faces With Expressions: Within-Class Space and Between-Class Space [Conference] // International Conference on Pattern Recognition. - 2005. - pp. 139-142.

Blanz Volker and Vetter Thomas A Morphable Model For The Synthesis Of 3D Faces [Conference] // Annual Conference on Computer Graphics and Interactive Techniques. - 1999. - pp. 187-194.

Blanz Volker and Vetter Thomas Face Recognition Based on Face Recognition Based on Fitting a 3D Morphable Model [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2003. - 9 : Vol. 25. - pp. 1063 - 1074 .

Blogger [Online]. - 2010. - January 2, 2012. - http://behind-the-knife-cosmetic-surgery.blogspot.com.

Boehnen C. and Russ T. A Fast Multi-Modal Approach to Facial Feature Detection [Conference] // IEEE Workshops on Application of Computer Vision. - 2005. - Vol. 1. - pp. 135-142.

Bolle Ruud M. [et al.] Guide to Biometrics [Book]. - [s.l.] : Springer, 2004.

Bookstein Fred L. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 1989. - 6 : Vol. 11. - pp. 567-585.

Bottino A., De Simone M. and Laurentini A. A Computer-Aided Technique For Planning Plastic Surgery Based On 3D Face Scans: Preliminary Results [Conference] // Proceedings of International Conference on 3D Body Scanning Technologies. - 2010.

Bowyer Kevin W., Chang Kyong and Flynn Patrick A Survey of Approaches and Challenges in 3D and Multi-modal 3D+2D Face Recognition [Journal] // Computer Vision and Understanding. - [s.l.] : Elsevier, 2006. - 15 : Vol. 1.

Bowyer Kevin W., Chang Kyong and Flynn Patrick A Survey Of Approaches To Three-Dimensional Face Recognition [Conference] // International Conference on Pattern Recognition. - 2004. - Vol. 1. - pp. 358-361.

Bronstein Alexander M., Bronstein Michael M. and Kimmel Ron Expression-Invariant 3D Face Recognition [Conference] // International Conference on Audio- and Video-Based Biometric Person Authentication. - 2003. - pp. 62-70.

Bronstein Alexander M., Bronstein Michael M. and Kimmel Ron Expression-Invariant Representations of Faces [Journal] // IEEE Transactions on Image Processing. - 2007. - 1 : Vol. 16. - pp. 188-197.

Bustard John D. and Nixon Mark S. 3D Morphable Model Construction for Robust Ear and Face Recognition [Conference] // IEEE Conference on Computer Vision and Pattern Recognition. - 2010. - pp. 2582-2589.

Cadavid S. and Abdel-Mottaleb M. Determining Discriminative Anatomical Point Pairings Using Adaboost For 3d Face Recognition [Conference] // International Conference on Image Processing. - 2009. - pp. 49-52.

Cartoux J. Y., Lapreste J. T. and Richetin M. Face Authentification or Recognition by Profile Extraction From Range Images [Conference] // Workshop on Interpretation of 3D Scenes. - 1989. - pp. 194-199.

Celiktutan Oya, Akakin Hatice Cinar and Sankur Bulent Multi-Attribute Robust Facial Feature Localization [Conference] // IEEE International Conference on Automatic Face & Gesture Recognition. - 2008. - pp. 1-6.

Chan T. and Vese L. Active Contours Without Edges [Journal] // IEEE Transactions on Image Processing. - 2001. - 2 : Vol. 10. - pp. 266-277.

Chang Kyong I., Bowyer Kevin W. and Flynn Patrick J. Adaptive Rigid Multi-region Selection for Handling Expression Variation in 3D Face Recognition [Conference] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. - 2005. - pp. 157-164.

Chang Kyong I., Bowyer Kevin W. and Flynn Patrick J. Face Recognition Using 2D and 3D Facial Data [Conference] // ACM Workshop on Multimodal User Authentication. - 2003. - pp. 25-32.

Chang Shoude, Rioux Marc and Domey Jacques Face Recognition with Range Images and Intensity Images [Journal] // Optical Engineering. - 1997. - 4 : Vol. 36. - pp. 1106-1112.

Chen Q. R., Cham W. K. and Tsui H. T. A Method for Estimating and Accurately Extracting The Eyebrow In Human Face Image [Conference] // International Conference on Image Processing. - 2002. - Vol. 3. - pp. 793-796.

Cheng Z. [et al.] Robust 3d Face Recognition In Uncontrolled Environments [Conference] // IEEE Conference on Computer Vision and Pattern Recognition. - 2008. - pp. 1-8.

Chua Chin-Seng, Han Feng and Ho Yeong-Khing 3D Human Face Recognition Using Point Signature [Conference] // IEEE International Conference on Automatic Face and Gesture Recognition. - 2000. - pp. 233-238.

CinemaSecrets [Online]. - January 6, 2012. - http://www.cinemasecrets.com.

Colbry Dirk [et al.] 3D Face Feature Extraction for Recognition [Report] / Department of Computer Science ; Michigan State University. - East Lansing, Michigan : [s.n.], 2004. - p. 33. - MSU-CSE-04-39.

Colbry Dirk and Stockman George Canonical Face Depth Map: A Robust 3D Representation for Face Verification [Conference] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). - 2007. - pp. 1-7.

Colbry Dirk, Stockman G. and Jain Anil K. Integrating Range and Texture Information for 3D Face Recognition [Conference] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. - 2005. - pp. 118-118.

Cook Jamie [et al.] Combined 2D / 3D Face Recognition using Log-Gabor Templates [Conference] // IEEE International Conference on Video and Signal Based Surveillance. - 2006. - pp. 83-88.

Cook Jamie [et al.] Face Recognition From 3D Data Using Iterative Closest Point Algorithm and Gaussian Mixture Models [Conference] // International Symposium on 3D Data Processing, Visualization and Transmission. - 2004. - pp. 502-509.

DB01 3D_RMA : 3D database [Online] // Signal and Image Centre - Royal Military Academy. - 1998. - February 13, 2012. - http://www.sic.rma.ac.be/~beumier/DB/3d_rma.html.

DB02 The 3D Face Database [Online]. - The University of York - Department of Computer Science, 2003. - February 13, 2012. - http://www-users.cs.york.ac.uk/~nep/research/3Dface/tomh/3DFaceDatabase.html.

DB03 GAVAB [Online]. - Universidad Rey Juan Carlos - Department of Computing, 2004. - February 13, 2012. - http://gavab.escet.urjc.es/recursos_en.html.

DB04 Face Recognition Grand Challenge [Online]. - 2004. - February 15, 2012. - http://www.nist.gov/itl/iad/ig/frgc.cfm.

DB05 BJUT-3D Face Database [Online]. - Multimedia and Intelligent Software Lab - Beijing University Of Technology, 2005. - February 13, 2012. - http://www.bjut.edu.cn/sci/multimedia/mul-lab/3dface/Release.htm.

DB06 icip2006.pdf [Online]. - Face Recognition and Artificial Vision Group - Universidad Rey Juan Carlos, 2006. - February 13, 2012. - http://www.frav.es/pdf/2006/icip2006.pdf.

DB07 CASIA 3D Face Database [Online]. - Center for Biometrics and Security Research, 2007. - February 13, 2012. - http://www.cbsr.ia.ac.cn/english/3DFace%20Databases.asp.

DB08 Bosphorus 3D Face Database [Online]. - Electrical and Electronics Engineering Department - Bogazici University, 2008. - February 13, 2012. - http://bosphorus.ee.boun.edu.tr/default.aspx.

DB09 Texas 3D Face Recognition Database (Texas 3DFRD) [Online]. - Laboratory for Image & Video Engineering - University of Texas, 2010. - February 13, 2012. - http://live.ece.utexas.edu/research/texas3dfr/.

DB10 NTU-CSP 3D Face Database [Online]. - Centre for Signal Processing - Nanyan Technological University, 2010. - February 13, 2012. - http://www3.ntu.edu.sg/home5/N030081/.

DB11 University of Milano Bicocca 3D Face Database [Online]. - The University of Milano Bicocca, 2011. - February 13, 2012. - http://www.ivl.disco.unimib.it/umbdb/.

DB12 Biometrics Data Sets [Online]. - Computer Vision Research Laboratory - Dept. of Computer Science and Engineering, University of Notre Dame, 2011. - February 13, 2012. - http://www.nd.edu/~cvrl/CVRL/Data_Sets.html.

De Marsico M. [et al.] Robust Face Recognition after Plastic Surgery Using Local Region Analysis [Journal] // Lecture Notes in Computer Science - Image Analysis and Recognition. - [s.l.] : Springer Berlin / Heidelberg, 2011. - pp. 191-200.

Dibeklioglu Hamdi, Salah Albert Ali and Akarun Lale 3D Facial Landmarking under Expression, Pose, and Occlusion Variations [Conference] // IEEE International Conference on Biometrics: Theory, Applications and Systems. - 2008. - pp. 1-6.

Due Trier Øivind, Taxt Torfinn and Jain Anil K. Data capture from maps based on gray scale topographic analysis [Conference] // International Conference on Document Analysis and Recognition. - 1995. - Vol. 2. - pp. 923-926.

Edwards G. J., Cootes T. F. and Taylor C. J. Face Recognition Using Active Appearance Models [Conference] // European Conference on Computer Vision. - 1998. - Vol. 2. - pp. 581-595.

Ersi F. and Zelek J. S. Local Feature Matching for Face Recognition [Conference] // Canadian Conference on Computer and Robot Vision. - 2006. - p. 4.

Fabry Thomas, Smeets Dirk and Vandermeulen Dirk Surface Representations for 3D Face Recognition [Book Section] // Face Recognition / ed. Oravec Milos. - [s.l.] : InTech, 2010.

Fabry Thomas, Vandermeulen Dirk and Suetens Paul 3D Face Recognition using Point Cloud Kernel Correlation [Conference] // IEEE International Conference on Biometrics: Theory, Applications and Systems. - 2008. - pp. 1-6.

Faltemier Timothy C., Bowyer Kevin W. and Flynn Patrick J. A Region Ensemble for 3-D Face Recognition [Journal] // IEEE Transactions on Information Forensics and Security. - 2008. - 1 : Vol. 3.

Faltemier Timothy, Bowyer Kevin and Flynn Patrick 3D Face Recognition with Region Committee Voting [Conference] // International Symposium on 3D Data Processing, Visualization, and Transmission. - 2006. - pp. 318-325.

Feng S., Krim H. and Kogan I. A. 3D Face Recognition using Euclidean Integral Invariants Signature [Conference] // IEEE/SP 14th Workshop on Statistical Signal Processing. - 2007. - pp. 156-160.

Frauel Yann [et al.] Comparison of passive ranging integral imaging and active imaging digital holography for three-dimensional object recognition [Journal] // Applied Optics. - 2004. - 2 : Vol. 43. - pp. 452-462.

Georghiades Athinodoros S., Belhumeur Peter N. and Kriegman David J. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2001. - 6 : Vol. 23. - pp. 643-660.

Gokberk Berk and Akarun Lale Comparative Analysis of Decision-level Fusion Algorithms for 3D Face Recognition [Conference] // International Conference on Pattern Recognition. - 2006. - pp. 1018-1021.

Gokberk Berk, Irfanoglu M. O. and Akarun Lale 3d Shape-Based Face Representation And Feature Extraction For Face Recognition [Journal] // Image and Vision Computing. - 2006. - 8 : Vol. 24. - pp. 857-869.

Gordon Gaile G. Face Recognition Based on Depth Maps and Surface Curvature [Book Section] // Geometric Methods in Computer Vision. - [s.l.] : SPIE, 1991. - Vol. 1570.

Guan Yepeng Robust Eye Detection from Facial Image based on Multi-cue Facial Information [Conference] // IEEE International Conference on Control and Automation. - 2007. - pp. 1775-1778.

Gunlu Goksel and Bilge Hasan S. 3D Face Decomposition and Region Selection against Expression Variations [Conference] // 3D Face Decomposition and Region Selection against Expression Variations. - 2010. - pp. 1298-1301.

Gupta S. [et al.] Texas 3D Face Recognition Database [Conference] // IEEE Southwest Symposium on Image Analysis and Interpretation. - 2010. - pp. 97-100.

Gupta Shalini [et al.] 3D Face Recognition Founded on the Structural Diversity of Human Faces [Conference] // IEEE Conference on Computer Vision and Pattern Recognition. - 2007. - pp. 1-7.

Gupta Shalini [et al.] Three Dimensional Face Recognition Based On Geodesic and Euclidean Distances [Conference] // SPIE, Electronic Imaging: Vision Geometry XIV. - 2007. - Vol. 6499.

Hammal Z. and Caplier A. Eyes and Eyebrows Parametric Models for Automatic Segmentation [Conference] // IEEE Southwest Symposium on Image Analysis and Interpretation. - 2004. - pp. 138-141.

Hara F. [et al.] Automatic Feature Extraction of Facial Organs and Contour [Conference] // IEEE International Workshop on Robot and Human Communication. - 1997. - pp. 386-391.

Harmon Leon D. The Recognition of Faces [Journal]. - 1973. - Vol. 229.

Heo Jingu and Savvides Marios Face recognition across pose using view based active appearance models (VBAAMs) on CMU multi-PIE dataset [Conference] // International Conference on Computer Vision Systems. - 2008. - pp. 527-535.

Heseltine Thomas, Pears Nick and Austin Jim Three-Dimensional Face Recognition Using Combinations of Surface Feature Map Space Components [Journal] // Image and Vision Computing. - [s.l.] : Butterworth-Heinemann, 2008. - 3 : Vol. 26. - pp. 382-396.

Hesher Curt, Srivastava Anuj and Erlebacher Gordon A Novel Technique for Face Recognition Using Range Imaging [Conference] // International Symposium on Signal Processing and Its Applications. - 2003. - Vol. 2. - pp. 201-204.

Hoft Thomas [et al.] 2D+3D Face Imaging for Stand-off Biometric Identification [Conference] // Conference on Lasers and Electro-Optics (CLEO) . - 2011.

Horn Berthold K. P. Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View [Report]. - [s.l.] : MIT, 1970.

Hu Y., Zhou M. and Wu Z. An Automatic Dense Point Registration Method for 3D Face Animation [Conference] // International Congress on Image and Signal Processing. - 2009. - pp. 1-6.

Hu Y., Zhou M. and Wu Z. An Automatic Non-rigid Point Matching Method for Dense 3D Face Scans [Conference] // International Conference on Computational Science and Its Applications. - 2009. - pp. 215-221.

Hu Yuxiao [et al.] Automatic 3D Reconstruction for Face Recognition [Conference] // IEEE International Conference on Automatic Face and Gesture Recognition. - 2004. - pp. 843-848.

Huang Di [et al.] 3D Face Recognition using Distinctiveness Enhanced Facial Representations and Local Feature Hybrid Matching [Conference] // IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS). - 2010. - pp. 1-7.

Huang Di [et al.] Automatic Asymmetric 3D-2D Face Recognition [Conference] // Three-dimensional face recognition using geometric model. - 2010.

Huang Jennifer, Heisele Bernd and Blanz Volker Component-based Face Recognition with 3D Morphable Models [Conference] // International Conference on Audio- and Video-Based Biometric Person Authentication. - 2003. - Vol. 2688. - pp. 27-34.

Huang T. S., Blostein S. D. and Margerum E. A. Least-Squares Estimation Of Motion Parameters From 3-D Point Correspondences [Conference] // IEEE Conference on Computer Vision and Pattern Recognition. - 1986.

Husken Michael [et al.] Strategies and Benefits of Fusion of 2D and 3D Face Recognition [Conference] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. - 2005. - pp. 174-181.

Introna Lucas D. and Nissenbaum Helen Facial recognition technology: A survey of policy and implementation issues [Report]. - New York : Report of the Center for Catastrophe Preparedness and Response, NYU, 2009.

Irfanoglu Okan M., Gokberk Berk and Akarun Lale 3D Shape-Based Face Recognition Using Automatically Registered Facial Surfaces [Conference] // International Conference on Pattern Recognition. - 2004. - Vol. 4. - pp. 183-186.

Jahanbin Sina [et al.] Three Dimensional Face Recognition Using Iso-Geodesic and Iso-Depth Curves [Conference] // IEEE International Conference on Biometrics: Theory, Applications and Systems. - 2008. - pp. 1-6.

Jain Anil K., Nandakumar K. and Ross Arun Score Normalization in Multimodal Biometric Systems [Journal] // Pattern Recognition. - 2005. - 12 : Vol. 38. - pp. 2270-2285.

Jain Anil K., Ross Arun and Prabhakar Salil An Introduction to Biometric Recognition [Journal]. - 2004. - 1 : Vol. 14. - pp. 4 - 20.

Jones Thouis R., Durand Fredo and Desbrun Mathieu Non-Iterative, Feature-Preserving Mesh Smoothing [Journal] // ACM Transactions on Graphics. - 2003. - 3 : Vol. 22. - pp. 943-949.

Ju X. and Siebert J. P. Individualising Human Animation Models [Conference] // Eurographics. - 2001.

Kakadiaris Ioannis A. [et al.] Three-Dimensional Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2007. - 4 : Vol. 29. - pp. 640-649.

Kanade Takeo Computer Recognition of Photographs of Human Faces [Report]. - 1973.

Kapoor A. and Picard R. W. Real-Time, Fully Automatic Upper Facial Feature Tracking [Conference] // IEEE International Conference on Automatic Face and Gesture Recognition. - 2002. - pp. 8-13.

Kittler J. [et al.] 3D Assisted 2D Face Recognition: Methodology [Book Section] // Progress in Pattern Recognition, Image Analysis and Applications. - [s.l.] : Springer Berlin / Heidelberg, 2005. - Vol. 3773.

Kittler J. [et al.] 3D Assisted Face Recognition: A Survey of 3D Imaging, Modelling and Recognition Approaches [Conference] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops . - 2005. - pp. 114-120.

Kittler J. [et al.] On Combining Classifiers [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 1998. - 3 : Vol. 20. - pp. 226-239.

Koudelka Melissa L., Koch Mark W. and Russ Trina D. A Prescreener for 3D Face Recognition Using Radial Symmetry and the Hausdorff Fraction [Conference] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. - 2005. - pp. 168-175.

Kuo P. and Hannah J. An Improved Eye Feature Extraction Algorithm Based On Deformable Templates [Conference] // IEEE International Conference on Image Processing. - 2005. - Vol. 2. - pp. 1206-1209.

Kuriakin V. [et al.] Mpeg-4 Compliant 3d Face Animation [Conference] // International Conference on Computer Graphics. - 2001. - pp. 54-58.

Kyong I. C., Bowyer Kevin W. and Flynn Phillip J. Adaptive Rigid Multi-Region Selection For Handling Expression Variation In 3d Face Recognition [Conference] // EEE Conference on Computer Vision and Pattern Recognition - Workshops. - 2005.

Lavagetto F. and Pockaj R. The Facial Animation Engine: Toward A High-Level Interface For The Design Of Mpeg-4 Compliant Animated Faces [Journal] // IEEE Transactions on Circuits and Systems for Video Technology. - 1999. - 2 : Vol. 9. - pp. 277-289.

Lee John C. and Milios Evangelos Matching Range Images of Human Faces [Conference] // International Conference on Computer Vision. - 1990. - pp. 722-726.

Lee Mun Wai and Ranganath Surendra Pose-Invariant Face Recognition Using A 3D Deformable Model [Journal] // Journal of Pattern Recognition. - 2003. - 8 : Vol. 36. - pp. 1835-1846.

Lee T. Y. [et al.] Three-Dimensional Facial Model Reconstruction And Plastic Surgery Simulation [Journal] // IEEE Transactions on Information Technology in Biomedicine. - 1999. - 3 : Vol. 3. - pp. 214-220.

Lee Yeunghak and Yi Taihong 3D Face Recognition Using Multiple Features for Local Depth Information [Conference] // EURASIP Conference focused on Video/Image Processing and Multimedia Communications. - 2003. - Vol. 1. - pp. 429-434.

Li Huibin [et al.] Expression Robust 3D Face Recognition via Mesh-Based Histograms of Multiple Order Surface Differential Quantities [Conference] // IEEE International Conference on Image Processing. - 2011. - pp. 3053-3056.

Li Huibin [et al.] Learning Weighted Sparse Representation of Encoded Facial Normal Information for Expression-Robust 3D Face Recognition [Conference] // International Joint Conference on Biometrics (IJCB). - 2011. - pp. 1-7.

Li Xiaoli and Da Feipeng 3D Face Recognition by Deforming the Normal Face [Conference] // International Conference on Pattern Recognition (ICPR). - 2010. - pp. 3975-3978.

Li Xiaoxing and Zhang Hao Adapting Geometric Attributes for Expression-Invariant 3D Face Recognition [Conference] // EEE International Conference on Shape Modeling and Applications. - 2007. - pp. 21-32.

Liew A. W. C., Leung S. H. and Lau W. H. Lip Contour Extraction Using a Deformable Model [Conference] // International Conference onImage Processing, Proceedings. - 2000. - Vol. 2. - pp. 255-258.

Liu Dang-hui, Shen Lan-sun and Lam Kin-man Image Synthesis and Face Recognition Based on 3D Face Model and Illumination Model [Book Section] // dvances in Natural Computation. - [s.l.] : Springer Berlin / Heidelberg, 2005. - Vol. 3611.

Liu Xin [et al.] A Lip Contour Extraction Method Using Localized Active Contour Model with Automatic Parameter Selection [Conference] // International Conference on Pattern Recognition. - 2010. - pp. 4332-4335.

Lu Xiaoguang [et al.] Face Recognition with 3D Model-Based Synthesis [Conference] // International Conference on Biometric Authentication. - 2004. - pp. 139-146.

Lu Xiaoguang and Jain Anil K. Automatic Feature Extraction for Multiview 3D Face Recognition [Conference] // International Conference on Automatic Face and Gesture Recognition. - 2006. - pp. 585-590.

Lu Xiaoguang and Jain Anil K. Deformation Analysis for 3D Face Matching [Conference] // IEEE Workshops on Application of Computer Vision. - 2005. - pp. 99-104.

Lu Xiaoguang and Jain Anil K. Deformation Modeling for Robust 3D Face Matching [Conference] // IEEE Conference on Computer Vision and Pattern Recognition. - 2006. - pp. 1377-1383.

Lu Xiaoguang and Jain Anil K. Integrating Range and Texture Information for 3D Face Recognition [Conference] // IEEE Workshops on Application of Computer Vision, 2005. - 2005. - pp. 156-163.

Lu Xiaoguang and Jain Anil K. Matching 2.5D Face Scans to 3D Models [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2006. - 1 : Vol. 28. - pp. 31-43.

Lu Xiaoguang and Jain Anil K. Multimodal Facial Feature Extraction for Automatic 3D Face Recognition [Report] / Department of Computer Science ; Michigan State University. - 2005. - MSU-CSE-05-22.

Lu Xiaoguang Image Analysis for Face Recognition - A Brief Survey [Report]. - 2003.

Lu Xiaoguang, Colbry Dirk and Jain Anil K. Three-Dimensional Model Based Face Recognition [Conference] // International Conference on Pattern Recognition. - 2004. - Vol. 1. - pp. 362-366.

Maes Chris [et al.] Feature Detection on 3D Face Surfaces for Pose Normalisation and Recognition [Conference] // IEEE International Conference on Biometrics: Theory Applications and Systems. - 2010. - pp. 1-6.

Malassiotis Sotiris and Strintzis Michael G. Pose and Illumination Compensation for 3D Face Recognition [Conference] // International Conference on Image Processing (ICIP). - 2004. - Vol. 1. - pp. 91-94.

Mao Z. [et al.] Constructing Dense Correspondences To Analyze 3d Facial Change [Conference] // International Conference on Pattern Recognition. - 2004. - Vol. 3. - pp. 144-148.

Martinez A. M. and Kak A. C. PCA versus LDA [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2001. - 2 : Vol. 23. - pp. 228-233.

Maurer Thomas [et al.] Performance of Geometrix ActiveID^TM 3D Face Recognition Engine on the FRGC Data [Conference] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. - 2005. - pp. 154-160.

McCool Chris [et al.] Normalisation of 3D Face Data [Conference] // International Conference on Signal Processing and Communication Systems. - 2007.

McCool Chris, Sanchez-Riera Jordi and Marcel Sebastien Feature Distribution Modelling Techniques for 3D Face Verification [Journal] // Pattern Recognition Letters. - 2010. - 11 : Vol. 31. - pp. 1324-1330.

McKeon Robert and Russ Trina Employing Region Ensembles in a Statistical Learning Framework for Robust 3D Facial Recognition [Conference] // IEEE International Conference on Biometrics: Theory Applications and Systems. - 2010. - pp. 1-7.

Medioni Gerard [et al.] Non-Cooperative Persons Identification at a Distance with 3D Face Modeling [Conference] // IEEE International Conference on Biometrics: Theory, Applications, and Systems, BTAS. - 2007. - pp. 1-6.

Medioni Gerard G. and Waupotitsch Roman Face Recognition and Modeling in 3D [Conference] // International Workshop on Analysis and Modeling of Faces and Gestures. - 2003. - pp. 232-233.

Mian A. S., Bennamoun M. and Owens R. A. Region-based Matching for Robust 3D Face Recognition [Conference] // British Machine Vision Conference. - 2005. - pp. 207-217.

Mian Ajmal S., Bennamoun Mohammed and Owens Robyn An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2007. - 11 : Vol. 29. - pp. 1927-1943.

Mian Ajmal, Bennamoun Mohammed and Owens Robyn Automatic 3D Face Detection, Normalization and Recognition [Conference] // International Symposium on 3D Data Processing, Visualization, and Transmission. - 2006. - pp. 735-742.

Mian Ajmal, Bennamoun Mohammed and Owens Robyn Keypoint Identification and Feature-Based Keypoint Identification and Feature-Based [Conference] // International Confrence on Biometrics. - 2007. - Vol. 4642. - pp. 163-171.

Moreno Ana Belen [et al.] Face Recognition Using 3d Surface-Extracted Descriptors [Conference] // Irish Machine Vision and Image Processing Conference. - 2003.

Most M. Battle of the Biometrics [Article] // Digital ID World Magazine. - November 2003. - pp. 16-18.

Mpiperis Iordanis, Malassiotis Sotiris and Strintzis Michael G. 3-D Face Recognition With the Geodesic Polar Representation [Journal] // IEEE Transactions on Information Forensics and Security. - 2007. - 3 : Vol. 2. - pp. 537-547.

Mpiperis Iordanis, Malassiotis Sotiris and Strintzis Michael G. Bilinear Models for 3-D Face and Facial Expression Recognition [Journal] // IEEE Transactions on Information Forensics and Security. - 2008. - 3 : Vol. 3. - pp. 498-511.

Nagamine Takashi, Uemura Tetsuya and Masuda Isao 3D Facial Image Analysis for Human Identification [Conference] // IAPR International Conference on Pattern Recognition. - 1992. - Vol. 1. - pp. 324-327.

Nayar Shree K., Ikeuchi Kasushi and Kanade Takeo Shape from Interreflections [Conference] // IEEE International Conference on Computer Vision (ICCV). - 1990. - pp. 2-11.

Nayar Shree K., Ikeuchi Katsushi and Kanade Takeo Determining Shape and Reflectance of Hybrid Surfaces by Photometric Sampling [Journal] // IEEE Transactions on Robotics and Automation. - 1990. - 4 : Vol. 6. - pp. 418-431.

Nikolaidis A., Kotropoulos C. and Pitas I. Facial Feature Extraction Using Adaptive Hough Transform, Template Matching and Active Contour Models [Conference] // International Conference on Digital Signal Processing. - 1997. - Vol. 2. - pp. 865-868.

Ocegueda Omar [et al.] UR3D-C: Linear Dimensionality Reduction for Efficient 3D Face Recognition [Conference] // International Joint Conference on Biometrics (IJCB). - 2011. - pp. 1-6.

Ojala T., Pietikainen M. and Maenpaa T. Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - [s.l.] : 24, 2002. - Vol. 7.

Pan Gang [et al.] Removal of 3D Facial Expressions: A Learning-Based Approach [Conference] // IEEE Conference on Computer Vision and Pattern Recognition. - 2010. - pp. 2614-2621.

Pan Gang and Wu Zhaohui 3D Face Recognition From Range Data [Journal] // International Journal of Image and Graphics. - 2005. - 3 : Vol. 5.

Pan Gang, Wu Zhaohui and Pan Yunhe Automatic 3D Face Verification From Range Data [Conference] // International Conference on Multimedia and Expo, 2003. - 2003. - Vol. 3. - pp. 133-136.

Pandzic I. S. and Forchheimer R. Mpeg-4 Facial Animation: The Standard, Implementation And Applications [Book]. - New York : John Wiley & Sons, 2003.

Papatheodorou Theodoros and Rueckert Daniel Evaluation of Automatic 4D Face Recognition Using Surface and Texture Registration [Conference] // IEEE International Conference on Automatic Face and Gesture Recognition. - 2004. - pp. 321-326.

Park Unsang, Tong Yiying and Jain Anil K. Age-Invariant Face Recognition [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2010. - 5 : Vol. 32. - pp. 947-954.

Passalis G. [et al.] Evaluation of 3D Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach [Conference] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. - 2005. - pp. 171-179.

Passalis Georgios [et al.] Using Facial Symmetry to Handle Pose Variations in Real-World 3D Face Recognition [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2011. - 10 : Vol. 33. - pp. 1938-1951.

Phillips Jonathan P. [et al.] Distinguishing Identical Twins by Face Recognition [Conference] // IEEE International Conference on Automatic Face & Gesture Recognition and Workshops. - 2011. - pp. 185-192.

Phillips Jonathon P. [et al.] FRVT 2002 Evaluation Report [Report]. - 2003.

Phillips Jonathon P. [et al.] FRVT 2006 and ICE 2006 Large-Scale Experimental Results [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2009. - 5 : Vol. 32. - pp. 831-846.

Phillips Jonathon P. [et al.] Overview of the Face Recognition Grand Challenge [Conference] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. - 2005. - Vol. 1.

Prabhu Utsav, Heo Jingu and Savvides Marios Unconstrained Pose-Invariant Face Recognition Using 3D Generic Elastic Models [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2011. - 10 : Vol. 33. - pp. 1952-1961.

Queirolo Chaua C. [et al.] 3D Face Recognition Using Simulated Annealing and the Surface Interpenetration Measure [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2010. - 2 : Vol. 32. - pp. 206-219.

Rabi S. A. and Aarabi P. Face Fusion: An Automatic Method for Virtual Plastic Surgery [Conference] // International Conference on Information Fusion. - 2006. - pp. 1-7.

Rara Ham M. [et al.] Distant Face Recognition Based On Sparse-Stereo Reconstruction [Conference] // IEEE International Conference on Image Processing (ICIP). - 2009. - pp. 4141-4144.

Ross Arun A., Nandakumar A. and Jain Anil K. Handbook of Multibiometrics [Book]. - [s.l.] : Springer, 2006.

Russ Trina, Boehnen Chris and Peters Tanya 3D Face Recognition Using 3D Alignment for PCA [Conference] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. - 2006. - pp. 1391-1398.

Samir Chafik, Srivastava Anuj and Daoudi Mohamed Three-Dimensional Face Recognition Using Shapes of Facial Curves [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2006. - 11 : Vol. 28. - pp. 1858-1863.

Savran Arman [et al.] Bosphorus Database For 3D Face Analysis [Conference] // COST Workshop on Biometrics and Identity Management. - 2008.

Scheenstra Alize, Ruifrok Arnout and Veltkamp Remco C. A Survey of 3D Face Recognition Methods [Book Section] // Lecture Notes in Computer Science. - [s.l.] : SpringerVerlag, 2005.

Schneider D. and Eisert P. Automatic and Robust Semantic Registration of 3D Head Scans [Conference] // European Conference on Visual Media Production. - 2008.

Singh N. Face Recognition by Capturing Eye Illumination Spot [Conference] // Applied Imagery Pattern Recognition Workshop. - 2004. - pp. 153-158.

Singh R. [et al.] Plastic Surgery: A New Dimension to Face Recognition [Journal] // IEEE Transactions on Information Forensics and Security. - 2010. - 3 : Vol. 5. - pp. 441-448.

Singh R. and Vatsa M. Effect of Plastic Surgery on Face Recognition: A Preliminary Study [Conference] // EEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. - 2009. - pp. 72-77.

Smeets Dirk [et al.] Fusion of an Isometric Deformation Modeling Approach using Spectral Decomposition and a Region-based Approach using ICP for Expression-invariant 3D Face Recognition [Conference] // International Conference on Pattern Recognition (ICPR). - 2010. - pp. 1172-1175.

Smeets Dirk [et al.] Objective 3D face recognition: Evolution, approaches and challenges [Journal] // Forensic Science International. - 2010. - 1-3 : Vol. 201. - pp. 125-132.

Smeets Dirk [et al.] Sometric Deformation Modeling Using Singular Value Decomposition For 3d Expression-Invariant Face Recognition [Conference] // IEEE International Conference on Biometrics: Theory, Applications, and Systems. - 2009. - pp. 1-6.

Smeets Dirk [et al.] Symmetric Surface-Feature based 3D Face Recognition for Partial Data [Conference] // International Joint Conference on Biometrics (IJCB). - 2011. - pp. 1-6.

Snelick R. D. [et al.] Multimodal Biometrics: Issues in Design and Testing [Conference] // ACM International Conference on Multimodal Interface. - 2003.

Spreeuwers Luuk Fast and Accurate 3D Face Recognition Using Registration to an Intrinsic Coordinate System and Fusion of Multiple Region Classifiers [Journal] // Inteurnal of computer Vision. - 2011. - 3 : Vol. 93. - pp. 389-414.

Srivastava Anuj, Liu Xiuwen and Hesher Curt Face Recognition Using Optimal Linear Components of Range Images [Journal] // Image and Vision Computing. - 2006. - 3 : Vol. 24. - pp. 291-299.

Swets Daniel L. and Weng Juyang Using Discriminant Eigenfeatures for Image Retrieval [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 1996. - 8 : Vol. 18. - pp. 831-836.

Szeptycki P., Ardabilian Mohsen and Chen Liming A Coarse-To-Fine Curvature Analysis-Based Rotation Invariant 3d Face Landmarking [Conference] // International Conference on Biometrics: Theory Applications and Systems. - 2009.

Tanaka Hiromi T., Ikeda Masaki and Chiaki Hisako Curvature-based Face Surface Recognition Using Spherical Correlation - Principal Directions for Curved Object Recognition [Conference] // IEEE International Conference on Automatic Face and Gesture Recognition. - 1998. - pp. 372-377.

Tarres F., Rama A. and Torres L. A Novel Method for Face Recognition Under Partial Occlusion or Facial Expression Variations [Conference] // International Symposium on Multimedia Systems and Applications (ELMAR). - 2005.

Tena J. R. [et al.] A Validated Method for Dense Non-rigid 3D Face Registration [Conference] // IEEE International Conference on Video and Signal Based Surveillance. - 2006. - pp. 81-81.

Ter Haar Frank B. and Veltkamp Remco C. A 3D Face Matching Framework [Conference] // IEEE International Conference on Shape Modeling and Applications. - 2008. - pp. 103-110.

Trier O. D., Taxt T. and Jain Anil K. Data Capture From Maps Based On Gray Scale Topographic Analysis [Conference] // International Conference on Document Analysis and Recognition. - 1995. - pp. 923-926.

Tsalakanidou Filareti, Malassiotis Sotiris and Strintzis Michael G. Integration of 2D and 3D Images for Enhanced Face Authentication [Conference] // IEEE International Conference on Automatic Face and Gesture Recognition. - 2004. - pp. 266-271.

Turk Matthew and Pentland Alex Eigenfaces for Recognition [Journal] // Journal of Cognitive Neuroscience. - 1991. - 1 : Vol. 3. - pp. 71-86.

Verlinde P. [et al.] Applying Bayes Based Classifiers For Decision Fusion In A Multi-Modal Identity Verification System [Conference] // International Symposium on Pattern Recognition. - 1999.

Visage visage|life [Online] // Visage Technologies. - Visage Technologies – The Character Animation Company. - February 14, 2012. - http://www.visagetechnologies.com/products_life.html.

Wang J. [et al.] 3D Facial Expression Recognition Based On Primitive Surface Feature Distribution [Conference] // IEEE Conference on Computer Vision and Pattern Recognition. - 2006. - pp. 1399-1406.

Wang Yingjie, Chua Chin-Seng and Ho Yeong-Khing Facial Feature Detection and Face Recognition From 2D and 3D Images [Journal] // Pattern Recognition Letters. - [s.l.] : Elsevier, 2001. - 10 : Vol. 23.

Wang Yueming [et al.] Exploring Facial Expression Effects in 3D Face Recognition Using Partial ICP [Conference] // Asian Conference on Computer Vision. - 2006. - pp. 581-590.

Wang Yueming, Liu Jianzhuang and Tang Xiaoou Robust 3D Face Recognition by Local Shape Difference Boosting [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2010. - 10 : Vol. 32. - pp. 1858-1870.

Wang Yueming, Pan Gang and Wu Zhaohui 3D Face Recognition in the Presence of Expression: A Guidance-based Constraint Deformation Approach [Conference] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). - 2007. - pp. 1-7.

webshots [Online]. - January 07, 2012. - http://good-times.webshots.com.

Wei-Yang L. [et al.] 3d Face Recognition Under Expression Variations Using Similarity Metrics Fusion [Conference] // IEEE International Conference on Multimedia and Expo. - 2007. - pp. 727-730.

Wiskott L. [et al.] Face Recognition by Elastic Bunch Graph Matching [Conference] // International Conference on Computer Analysis of Images and Patterns. - 1997. - Vol. 1296. - pp. 456-463.

Wong Hau-San, Cheung Kent K. T. and Ip Horace H. S. 3D Head Model Classification by Evolutionary Optimization of the Extended Gaussian Image Representation [Journal] // Pattern Recognition. - 2004. - 12 : Vol. 37. - pp. 2307-2322.

Woodham Robert J. Photometric Stereo: A Reflectance Map Technique for Determining Surface Orientation From A Single View [Conference] // SPIE Image Understanding Systems and Industrial Applications. - 1978. - Vol. 155. - pp. 136-143.

Xiaoxing L., Tao J. and Hao Z. Expression-Insensitive 3d Face Recognition Using Sparse Representation [Conference] // IEEE Conference on Computer Vision and Pattern Recognition. - 2009. - pp. 2572-2582.

Xu Chenghua [et al.] A New Attempt to Face Recognition Using 3D Eigenfaces [Conference] // Asian Conference on Computer Vision. - 2004. - Vol. 2. - pp. 884-889.

Yacoob Yaser and Davis Larry Labeling of Human Face Components from Range Data [Conference] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. - 1993. - pp. 592-593.

Yang Yi-Jun [et al.] Optimal Parameterizations of Bézier Surfaces [Book Section] // Advances in Visual Computing. - [s.l.] : Springer Berlin / Heidelberg, 2006. - Vol. 4291.

Zhang Xiaozheng and Gao Yongsheng Face recognition across pose: A review [Journal] // Pattern Recognition. - 2009. - 11 : Vol. 42. - pp. 2876-2896.

Zhang Zhengyou [et al.] Robust and Rapid Generation of Animated Faces From Video Images: A Model-Based Modeling Approach [Journal] // International Journal of Computer Vision - Special Issue on Research at Microsoft Corporation. - 2004. - 1 : Vol. 58. - pp. 93-119.

Zhao Wen Yi and Chelappa Rama SFS Based View Synthesis for Robust Face Recognition [Conference] // IEEE International Conference on Automatic Face and Gesture Recognition. - 2000. - pp. 285-292.

Zhou Dianle, Petrovska-Delacretaz Dijana and Dorizzi Bernadette 3D Active Shape Model for Automatic Facial Landmark Location Trained with Automatically Generated Landmark Points [Conference] // International Conference on Pattern Recognition. - 2010. - pp. 3801-3805.