NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
SCHOOL OF SCIENCES
Department of Informatics and Telecommunications
Program of Postgraduate Studies

PhD Thesis

Landmark Detection for Unconstrained Face Recognition

Panagiotis B. Perakis

European Union
European Social Fund

OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION & RELIGIOUS AFFAIRS, CULTURE & SPORTS
MANAGING AUTHORITY
Co-financed by Greece and the European Union

NSRF
2007-2013
programme for development
EUROPEAN SOCIAL FUND

ATHENS, July 2013

Version 5.0 EN P. Perakis (p.perakis@di.uoa.gr; takis@antinoos.gr)

PhD Thesis

Landmark Detection for Unconstrained Face Recognition

Panagiotis B. Perakis

SUPERVISOR: Theoharis Theoharis, Professor UoA, NTNU

THREE-MEMBER ADVISOR COMMITTEE:
Theoharis Theoharis, Professor UoA, NTNU
Ioannis Emiris, Professor UoA
Manolis Sangriotis, Associate Professor UoA

SEVEN-MEMBER EXAMINATION COMMITTEE:

Theoharis Theoharis
Professor UoA, NTNU

Ioannis Emiris
Professor UoA

Manolis Sangriotis
Associate Professor UoA

Sergios Theodoridis
Professor UoA

Georgios Papaioannou
Assistant Professor AUEB

Stavros Perantonis
Research Director NCSR Demokritos

Ioannis A. Kakadiaris
Professor UH

Examination Date: July 2, 2013

ABSTRACT

Facial landmark detection is a crucial first step in facial analysis for biometrics and numerous other applications. However, it has proved to be a very challenging task due to the numerous sources of variation in 2D and 3D facial data. The unconstrained acquisition of data from uncooperative subjects may result in facial scans with significant pose variations which can cause extensive occlusions that result in missing data. In this dissertation a novel method for 3D landmark detection and pose estimation, suitable for both frontal and side 3D facial scans, is presented. The proposed method exploits 3D and 2D information by using local shape descriptors to extract candidate interest points that are subsequently identified and labeled as anatomical landmarks. Additionally, a novel generalized framework for combining facial feature descriptors that can be used for landmark detection is introduced, and several feature fusion schemes are proposed and evaluated. However, feature detection methods which use general purpose shape descriptors cannot identify and label the detected candidate landmarks. Therefore, the topological properties of the human face need to be taken into consideration. To this end, a 3D *Facial Landmark Model* (FLM) of facial anatomical landmarks is introduced. Candidate landmarks, irrespectively of the way they are generated, can be identified and labeled by matching them with the FLM. Finally, a novel method for unconstrained face recognition is introduced. It employs the 3D landmark detector to provide an initial pose estimation and to indicate occluded areas with missing data for each facial scan. Subsequently, a 3D *Annotated Face Model* (AFM) is registered and fitted to the scan using facial symmetry to complete the occluded areas. Using a wavelet representation of the geometry and normal images produced from the fitted AFM, the proposed method can perform comparisons among interpose facial scans, unlike existing methods that require frontal only scans.

Subject Area Computer Graphics, Computer Vision, Image Processing, Pattern Recognition, Biometrics.

Keywords Biometrics, Face Recognition, Landmark Detection, Shape Models, Feature Descriptors, Feature Extraction, Feature Fusion, Pose Estimation, Partial Matching, Deformable Models.

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my advisor, Prof. Theoharis Theoharis, who first challenged me to pursue for the doctorate degree and has subsequently served as both teacher and advocate during the entire process, which ended founding a true friendship. The support, trust, and encouragement he has shown to me during the last years were frankly essential to the completion of this dissertation.

My thanks also go to my co-advisors, Prof. Ioannis Emiris and Associate Prof. Manolis Sangriotis for the scientific advice they have offered me, and also for reading the draft of this dissertation and providing many valuable comments that improved its presentation and contents.

I especially thank Prof. Ioannis A. Kakadiaris (Professor of Computer Science, Electrical & Computer Engineering, and Biomedical Engineering at the University of Houston) for his cooperation and support.

I am also grateful to all my colleagues, at the Computer Graphics Lab of the Informatics and Telecommunications Department of the University of Athens, for their collaboration, and especially Dr. Georgios Passalis for his contribution on the setup of the partial face recognition method. Giorgos provided me substantial help, through unnumbered discussions regarding both theoretical and practical issues related to my research work.

My parents, Byron and Paraskevi, receive my deepest gratitude for their love and the many years of support during my whole life.

I would finally like to dedicate this dissertation to my wife Maria, without whose understanding and encouragement, completion of the thesis would have been impossible, and to my little son Byron, for all of the joy and happiness that he brings.

My PhD research at the Department of Informatics and Telecommunications of the University of Athens has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

Contents

List of Figures

List of Tables

List of Algorithms

Preface

Biometrics is the science of establishing the identity of a person based on the physical (e.g., fingerprints, face, hand geometry, and iris) or behavioral (e.g., gait, signature, and keyboard dynamics) attributes associated with an individual [116].

Face recognition, as one of the primary biometric modalities, became more important owing to rapid advances in technologies such as digital cameras of visible or infrared spectrum, surveillance video cameras, 3D scanners, and increased demand on security. Face recognition has several advantages over other biometric technologies: it is non-intrusive, since the facial region is generally exposed, and potentially easy to use [75]. Thus, research and development in automatic face recognition followed naturally.

The performance of face recognition systems has improved significantly since the first automatic face recognition system was developed by Kanade [66]. Furthermore face recognition can now be performed in "realtime" for images captured under constrained situations. Although progress in face recognition has been encouraging, the task has also turned out to be a difficult endeavor, especially for unconstrained tasks where view point, illumination, inter-object occlusions, facial expressions and facial accessories vary considerably [75].

Face recognition is a task that humans perform routinely and effortlessly in their everyday life. Analysis and understanding of objects is one of the most fundamental tasks in our interaction with the surrounding world. For most of us, the most significant information about the surrounding world comes from our visual system. This information is actually a two-dimensional projection of the three dimensional world. When we see a picture, we recognize the depicted objects by relating them to concepts we have learned throughout our lives. In our every day experience, we are often unaware of how extremely complex the shape analysis performed by the brain is, because it is done mostly subconsciously, without involving our higher level of cognition. Faces are probably the most important class of objects in human perception. Infants learn about faces faster than other objects, suggesting that we may have special neural hardware for dealing with them [58].

In the era of computers, attempts to imitate the ability of the human visual system to analyze objects gave birth to the fields of computer vision and pattern recognition. Object recognition is one of the hardest problems in computer vision [120]. The most significant sources of difficulty in object recognition are the large changes in appearance of an object under different viewing, illumination and inter-object occlusion conditions. An object recognition system must be invariant to such changes, while being able to discriminate between different objects with similar appearance. This fundamental problem of IT systems is solved by a biological sensory system in the neocortex, offering recognition of objects relatively independent of size, contrast, spatial frequency, position in the retina or view angle [127].

In computer vision, as in neuroscience, object recognition is frequently divided into two schools of thought, which might be labeled *object-based* and *view-based*. In the object-based paradigm, the computational model of an object is inherently three-dimensional, and recognition is a matter of deciding which object is seen (classification), in which 3D orientation

(pose estimation). In the view-based paradigm, the many different appearances of an object are each modeled independently in 2D, and no explicit 3D computations are performed [127]. The view-based paradigm is supported by the evidence that not all views of an object are equally easy to be recognized by humans [94, 104]. This theory of object recognition proposes that we recognize objects by matching the visual information with internally stored view-point specific "prototypes" [38, 37]. These views are not simple snapshots; they allow recognition despite simple geometric distortions of the image, although, in the case of face recognition, this task is considerably impaired if the face images are shown upside down [110].

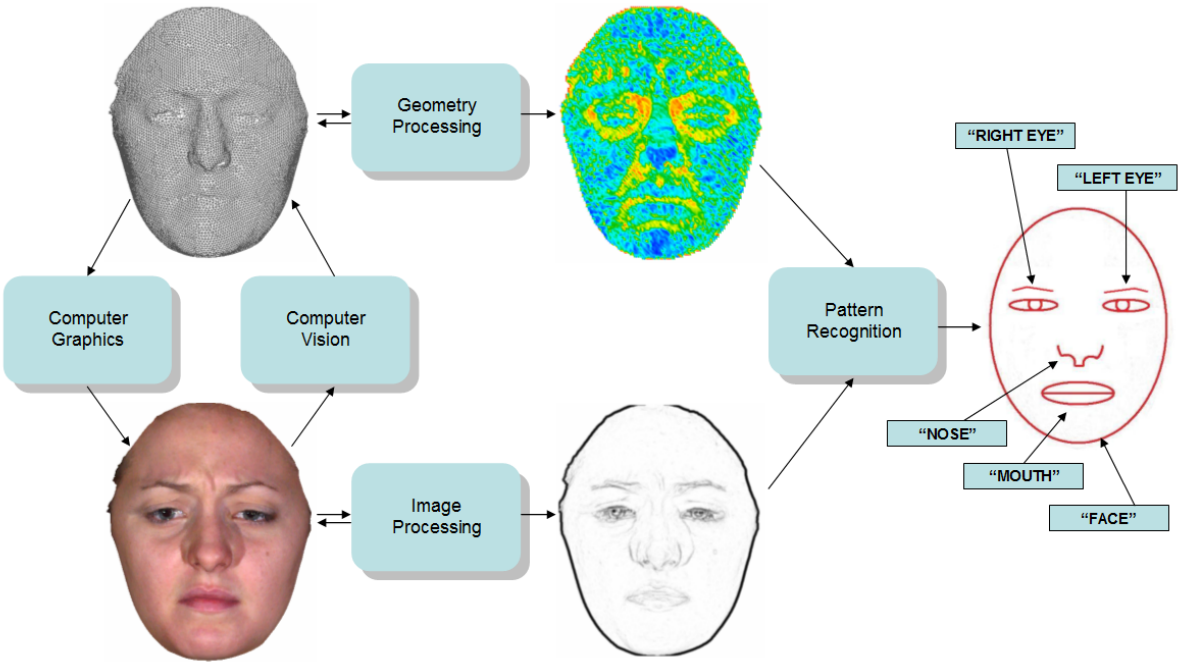


Figure 1: Depiction of the relation between the fields of computer graphics, computer vision, geometry and image processing and pattern recognition (adapted from [18]).

Similarity and *correspondence* are the two fundamental problems of shape analysis in computer vision. Because of the flexibility of facial tissues and our ability to express a wide range of emotions, the face is considered a non-rigid object. Therefore, face recognition falls into the category of non-rigid shape similarity problems [18]. Also, due to the anatomic constraints of the motion of the human neck, frontal to profile rotations of the face are very common. A natural way to estimate pose and facial similarity is based on landmark points' correspondence between the compared facial datasets.

Usually, some model that relates the shapes to the underlying objects is assumed. Since faces have a certain degree of flexibility and they are non-rigid, the models that represent this kind of objects have to be considered as deformable models. Being able to analyze the properties of such shapes and describe their behavior is a key concept in facial similarity and correspondence.

Modeling facial expressions as deformations and using deformation-invariant similarity criteria, we can distinguish between features resulting from expressions and those characterizing the person's identity, or in other words, make our face recognition system expression-

invariant. Also, by modeling faces as 3D shapes instead of 2D shapes we can avoid view-based modeling and make our face recognition system pose- and illumination-invariant.

Thus, in a broad sense, the face recognition and face modeling problem belongs to the realm of the following fields [18] (Fig. 1): *Computer vision* deals with extracting information about faces from their two-dimensional image representation, and the connection with their geometric representation, addressing the problem of 2D to 3D correspondence. *Computer graphics* on the other hand deals with the problem of how to realistically render an image of a face from its geometric (3D) and texture (2D) representation, addressing the problem of 3D to 2D correspondence and "morphing" between faces. *Geometry processing* deals with the geometric models of faces, trying to improve their quality, transform them or extract information from their 3D geometric representation. *Image processing* operates on facial images themselves, trying to improve their quality, transform them or extract information from their 2D representation. *Pattern recognition*, at the other end, deals with the assignment of labels to objects or to features extracted from an object or an image, addressing the problem of similarity and classification.

The above divisions are becoming less obvious in our days since methods from the above fields are interrelated. Considering images as geometric objects and operating on them using geometric tools created a revolution in image processing. Conversely, by representing geometric objects as images, many efficient and powerful methods can be borrowed from image processing and adapted to geometry processing. Projective and differential geometry, transformations in vector and affine spaces, illumination models, deformable models and morphing, all provide common tools to computer graphics and computer vision.

Thesis Overview

The uncontrolled conditions of real-world biometric applications pose a great challenge to any 3D face recognition approach. The unconstrained acquisition of data from uncooperative subjects may result in facial scans with significant pose and expression variations.

In this dissertation, an integrated novel method is proposed, in order to handle efficiently facial pose and expression variations, in the implementation of a pose- and expression-invariant face recognition system, that works under uncontrolled conditions and with uncooperative subjects.

The proposed landmark detection and face recognition system employs an automatic pose and expression invariant landmark detector, using local facial feature descriptors and a deformable Facial Landmark Model (FLM) to ensure global topological consistency of the detected landmarks. The landmark detector provides an initial pose estimation and indicates occluded areas with missing data for each facial scan resulting from pose variations. Facial landmark detection is a crucial first step for the registration of the facial datasets that have to be compared. Subsequently, an Annotated Face Model (AFM) is registered and fitted by deformation to each facial probe scan. During fitting, facial symmetry is used to complete the occluded areas of the face. Signature metadata are extracted using a wavelet transformation on the geometry and normal images of the fitted AFM. A similarity measure between signature metadata of probe and gallery facial datasets provide the face recognition results. This system is suitable for real-world applications as the only requirement is that

half of the face is visible to the sensor.

Structure of the Dissertation

This dissertation is organized as follows:

Introduction introduces the problem under consideration, the challenges, the motivation and the novelties that were introduced in addressing it.

Related Work describes related work in the fields of 3D and 2D facial landmark detection, facial features fusion and unconstrained partial face recognition.

Shapes and Landmarks presents the theoretical background of statistical shape analysis for describing shapes through landmarks, introduces the Facial Landmark Model (FLM) and defines its deformations. The FLM is constructed using Procrustes Analysis and Principal Component Analysis (PCA) over facial landmarks, pre-annotated on exemplar facial datasets.

Facial Data presents various facial dataset representations and the approaches used in this dissertation to process these 3D and 2D facial data. The algorithms presented in this chapter include: data cleaning and preprocessing, resolution and scale adjustments, curvature computations, 3D and 2D data registration, and 2D maps of 3D data. It also describes the Annotated Face Model (AFM) as a generic 3D geometric model of facial datasets.

Landmarks and Features presents various feature descriptors that are used in this dissertation to represent facial landmarks. These include the Shape Index, the Spin Image, the Extruded Points and the Edge Response descriptors. It also introduces various feature fusion schemes for the combination of these descriptors into a more descriptive resultant feature descriptor.

Landmark Detection presents the proposed landmark detection methods in detail, using the combination of the landmark feature models and the geometric landmark models (FLMs) for landmark consistency. Landmark detection is a key requirement for generic face recognition, calculating the coarse transformation for registration, and transforming a test scan into a canonical AFM.

Partial Face Recognition presents the proposed partial face recognition method in detail. It presents the two-step method used to register 3D facial surfaces, the deformation procedure involved in fitting the AFM to a facial dataset, the extraction of facial signature metadata using wavelet transformations of the geometry and normal images of the fitted AFM and finally their use in a similarity measure for face recognition.

Experimental Results presents the experimental results that evaluate the proposed methods. The proposed landmark detector achieves state-of-the-art accuracy, while the proposed partial face recognition method also achieves state-of-the-art performance, considerably outperforming existing methods, even when tested on the most challenging data, which contain scans with yaw variations up to 80° and strong expressions.

Conclusion discusses the conclusions of this dissertation and presents directions for future work.

List of Publications

Work from this dissertation has appeared in the following co-authored publications (Citations are according to "Google Scholar"):

Journals

- [100] **P. Perakis**, G. Passalis, T. Theoharis, and I. A. Kakadiaris, "3D facial landmark detection under large yaw and expression variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1552–1564, July 2013.

- [97] G. Passalis, **P. Perakis**, T. Theoharis, and I. A. Kakadiaris, "Using facial symmetry to handle pose variations in real-world 3D face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1938–1951, Oct. 2011. **Citations: 24**.

Conferences

- [101] **P. Perakis**, G. Passalis, T. Theoharis, G. Toderici, and I. A. Kakadiaris, "Partial matching of interpose 3D facial data for face recognition," in *Proc. 3^{rd} IEEE International Conference on Biometrics: Theory, Applications and Systems*, Arlington, VA, Sep. 28-30 2009, pp. 439–446. **Best-Reviewed Paper Award**. **Citations: 20**.

- [103] **P. Perakis**, T. Theoharis, G. Passalis, and I. A. Kakadiaris, "Automatic 3D facial region retrieval from multi-pose facial datasets," in *Proc. Eurographics Workshop on 3D Object Retrieval*, Munich, Germany, Mar. 30 - Apr. 3 2009, pp. 37–44. **Citations: 11**.

Book Chapters

- [63] I. A. Kakadiaris, G. Passalis, G. Toderici, E. Efraty, **P. Perakis**, D. Chu, S. K. Shah, and T. Theoharis, "Face recognition using 3D images," in *Handbook of Face Recognition*, 2^{nd} ed., S. Li and A. K. Jain, Eds. Springer-Verlag, July 2010, pp. 5–30.

- [65] I. A. Kakadiaris, G. Passalis, G. Toderici, **P. Perakis**, and T. Theoharis, "Face recognition, 3D-based," in *Encyclopedia of Biometrics*, S. Li, Ed. New York, NY: Springer, 2009, pp. 329–338.

Technical Reports

- [99] **P. Perakis**, G. Passalis, T. Theoharis, and I. A. Kakadiaris, "3D facial landmark detection & face registration: A 3D facial landmark model & 3D local shape descriptors approach," Computer Graphics Laboratory, University of Athens, Tech. Rep. TP-2010-01, Jan. 2010. **Citations: 3**.

- [102] **P. Perakis** and T. Theoharis, "Statistical Landmark Models: An ASM Approach," Computer Graphics Laboratory, University of Athens, Tech. Rep. TP-2008-03, Dec. 2008.

Panagiotis B. Perakis

– Athens, July 2013

1 Introduction

God ever geometrizes.

– Plato

1.1 Challenges & Motivation

In recent years, among many biometric modalities, the face has received the most interest. Not only is face recognition one of the most widely accepted modalities, but advances in processing power have allowed the development of more complex algorithms while still providing a rapid response to queries. Face recognition requires no contact with the subject, thus being more easily accepted by the public, compared to other biometrics such as fingerprints or iris detection.

Face recognition has been traditionally performed using 2D (visible spectrum) images, while hybrid approaches have used infrared images and 3D geometry. Infrared face recognition has not been widely adopted due to the high cost of the infrared cameras necessary to acquire the data. In contrast, as scanning methods have become more accessible due to lower cost and greater flexibility, 3D facial datasets are more easily available, and therefore the interest in developing algorithms that use 3D data has increased.

However, face recognition has proved to be a very challenging task due to the numerous sources of variation in 2D and 3D facial data. These variations can be environment-based (illumination conditions, occlusions by other objects or accessories), subject-based (pose and expression variations) and acquisition-based (image scale, distortion, noise, spikes and holes).

The main reason for using information from 3D data as a biometric is that the data acquired by 3D acquisition devices are invariant to pose and lighting conditions, these being the major challenges with which face recognition algorithms must cope. Moreover, image-based face recognition algorithms are more susceptible to impostors. Indeed, an impostor may use a printout of an image of a subject allowed to enter a facility in order to break in. To avoid this, a face recognition algorithm must be coupled with liveness test algorithms. Attempting such an attack on a system based on 3D data would be much more difficult, since the attackers would need to obtain an accurate 3D model (sculpture) of the person whom they would like to impersonate.

The challenges of a 3D face recognition system are the following [65]:

- *Robustness*: The system must perform robustly and reliably under a variety of conditions (e.g., lighting, pose variation, facial features variation).

- *Accuracy Gain*: A significant gain in accuracy with respect to 2D face recognition systems must justify the introduction of 3D recognition systems.
- *Efficiency*: 3D capture devices generate substantially more information than 2D cameras. Using this large volume of information is expensive in terms of computation time and storage requirements. Therefore, the algorithms developed need to be efficient both in time and space, by using appropriate metadata.
- *Automation*: The system must be completely automated. It is therefore not acceptable to assume user intervention, such as for the location of key landmarks in a 3D facial scan.
- *Capture Devices*: 3D capture devices were mostly developed for medical and other low-volume applications and suffer from a number of drawbacks, including artifacts, small depth of field, long acquisition time and multiple types of output. A deployable 3D face recognition system must be able to address these issues.

With the increase in the availability of 3D data, several 3D face recognition approaches have been proposed. These approaches aim to overcome the limitations of 2D face recognition by offering pose invariance. However, although they claim pose invariance, they mostly utilize frontal 3D scans assuming that the entire face is visible to the sensor. This assumption is not always valid in real-world applications, since unconstrained acquisition may lead to facial scans with extensive occlusions that result in missing data due to pose variations.

Thus, existing 3D face recognition methods, fail to address large pose variations and to confront the problem of missing facial areas in an automatic way (Chapter 2). The main assumption of these methods is that even though the head can be rotated with respect to the sensor, the *entire* face is always visible. However, this is true only for "almost frontal" scans or "reconstructed" complete face meshes, or "pre-aligned" scans to frontal pose. *Side scans usually have large missing areas, due to self-occlusion, that depend on pose variations.* These scans are very common in realistic scenarios such as uncooperative subjects or uncontrolled environments. Therefore, to take advantage of the full pose invariance potential of 3D face recognition, the problem of missing data must be addressed. Thus, in a face recognition system, an initial registration step, based on landmark points' correspondence, is necessary in order to make a system fully pose invariant [64, 97].

However, facial landmark detection also suffers from the same sources of variation in 2D and 3D facial data that face recognition does. Both 2D and 3D facial landmark detection suffer from occlusion, pose and expression variations. In addition, 2D facial landmark detection also suffers from illumination variations. Thus, a landmark detection algorithm must be pose-invariant to address the problem of missing facial areas and, at the same time, must be expression-invariant in order to allow the registration of the various instances of the face liable to expression variations.

1.2 Aim & Methodology

The main aim of the research presented in this dissertation is to *automatically detect landmarks on 3D facial scans that exhibit pose and expression variations, and hence consistently register and compare any pair of facial datasets subjected to missing data due to self-occlusion in a pose- and expression-invariant face recognition system.*

The proposed landmark detection and face recognition system employs an automatic pose- and expression-invariant landmark detector, using local facial feature descriptors and a deformable 3D Facial Landmark Model (FLM) to ensure global topological consistency of the detected landmarks.

At the training phase, a Facial Landmark Model (FLM) is created by first aligning the training landmark sets and calculating a mean landmark shape using Procrustes Analysis, and then applying Principal Component Analysis (PCA) to capture the shape variations. The FLM serves as a 3D geometric model of the landmark points. Also, templates for each shape descriptor that represents each landmark point are calculated from training facial datasets.

The shape templates serve as feature descriptors for each landmark point. The feature descriptors that have been used, depending on the case, include the *Shape Index*, a continuous map of principal curvature values of a 3D object's surface, the *Spin Image*, a local descriptor of the object's 3D point distribution, the *Extruded Points*, a local descriptor of a 3D object's points that extrude most and the *Edge Response* descriptor, a local descriptor of the 2D texture gradient of a 3D object.

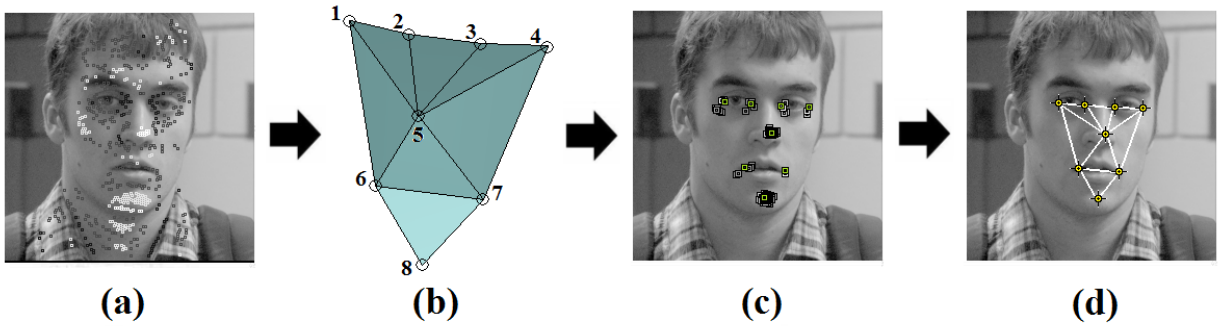


Figure 2: Process pipeline of landmark detection: (a) extracted candidate landmarks; (b) Facial Landmark Model (FLM); (c) landmark sets consistent with FLM; (d) resulting optimal landmark set.

At the detection phase, the algorithm first detects candidate landmarks on the queried facial datasets according to the similarity of the extracted facial features with the feature templates. The extracted candidate landmarks are then filtered out and labeled by matching them with the FLM (Fig. 2).

The landmark detector provides an initial pose estimation (frontal, right, left) and indicates occluded areas with missing data for each facial scan resulting from pose variations. Facial landmark detection is a crucial first step for the registration of the facial datasets that have to be compared.

Subsequently, a generic Annotated Face Model (AFM) is registered and fitted to each facial probe scan, using a subdivision-based deformable model framework. During fitting, facial symmetry is used to complete the occluded areas of the face. Signature metadata are extracted using a wavelet transformation on the geometry and normal images of the fitted AFM (Fig. 3). A similarity measure between signature metadata of probe and gallery facial datasets provide the face recognition results.

The presented method is extensively evaluated against a variety of 3D facial databases. The proposed 3D landmark detector achieves state-of-the-art accuracy (with $4.5 - 6.3$ mm

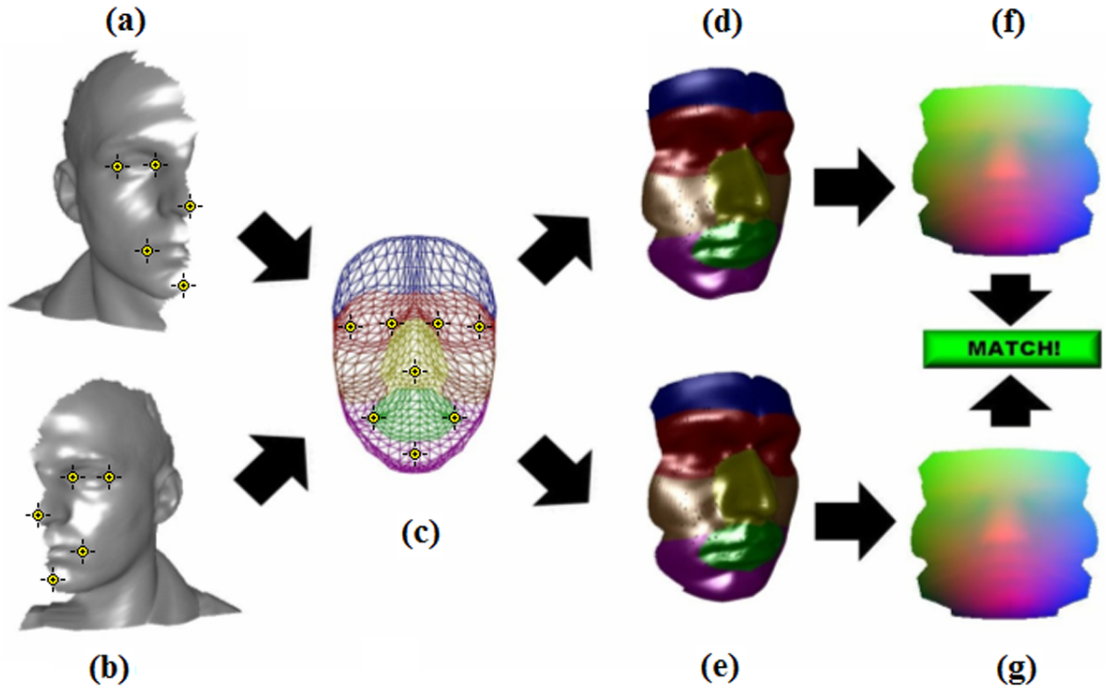


Figure 3: Interpose matching using the proposed method: (a) and (b) opposite side facial scans with extensive missing data and detected landmarks; (c) generic Annotated Face Model (AFM); (d) and (e) registered and deformed AFM for each scan (facial symmetry used); (f) and (g) extracted geometry images.

mean landmark localization error), and the proposed partial face recognition method state-of-the-art performance (with average rank-one recognition rate 83.7%), considerably outperforming existing methods, even when tested on the most challenging data, which contain scans with yaw variations up to 80° and strong expressions.

The proposed system is suitable for real-world applications as the only requirement is that half of the face is visible to the sensor, and its computational cost is low.

1.3 Contributions & Novelties

The contributions of this thesis fall under the fields of computer graphics, computer vision, geometry and image processing and pattern recognition (Fig. 1). Furthermore, it establishes new links between the synthesis methods of computer graphics and the analysis methods of computer vision, introducing novel techniques to the image-based and geometry-based approaches.

Specifically, the main contributions of the conducted research are the following:

- The design and optimization of a fast and fully automatic facial landmark detector. The proposed detector is useful for face alignment and is shown to be tolerant to changes in pose, lighting and expression, and fills a gap in existing research, which is dominated by methods that use pre-aligned or manually aligned data. By developing methods for the automatic registration of two facial scans in real time, a foundation for building fully automatic face processing algorithms is provided, without the need of being manually annotated with landmarks before analysis.

- A novel generalized framework of fusion methods and their application to landmark detection is investigated. The proposed fusion scheme transforms features to similarities and combines them to generate a resultant feature similarity score. The proposed approach of feature fusion offers dimensionality reduction, is easily extensible and works equally well for any feature extracted either from 3D or 2D facial data.

- Experimental analysis and study of a real-world face recognition system for use with uncooperative subjects. The proposed 3D face recognition method addresses the problem of missing facial areas due to large pose variations in an automatic way, completing missing facial data using symmetry. Experiments show that this 3D face recognition system is robust to the types of variations that would be expected in a real world application, such as yaw rotations and expression variations.

This dissertation also explores and improves upon the use of 3D/2D facial data, introducing several novelties to face processing research, including the following:

- Introduction of a deformable geometric 3D Facial Landmark Model (FLM) that incorporates size and expression variations of the human face.

- Introduction of three alternative FLMs (frontal, right and left) to address the problem of self-occluded data due to yaw rotations in order to detect landmarks on frontal and side facial scans.

- The three alternative FLMs (frontal, right and left) offer dimensionality reduction of the search space of consistent landmark sets.

- Efficient resampling of facial scans for creating a unified regular parametric surface. Such a surface mesh is able to represent the curved surface of a face as accurately as required.

- Efficiently performing differential operations on the u,v parametric surface of facial scans.

- Efficient representation of 3D differential information of facial surfaces with 2D image maps, using the u,v representation of facial data.

- Efficient registration of 3D geometry and 2D texture information, even from scans where the texture map may not be contiguous.

- Efficient fusing of feature descriptors from 2D and 3D data, using the u,v representation of facial data.

- Fusing feature descriptors at the similarity level offers significant dimensionality reduction of the resultant feature space.

- FLMs are used to classify "unseen" scans in categories such as frontal, right or left with almost no error.

- FLMs are used to classify areas of "unseen" face scans in categories such as nose, mouth, left eye etc.

- Efficient registration of "unseen" complete (frontal) and partial (semi-profile and profile) facial data to a generic Annotated Face Model (AFM).

- Alignment and registration can be done by using a small number of facial landmarks (5 or 8), instead of the thousands of points used in the AFM that is fitted afterwards.

- Registration using automatically detected landmark is done within an accuracy comparable to that of manual landmark placement.
- Efficient integration of occluded facial data by filling the model using symmetric facial data.
- The independent handling of left and right sides of frontal scans, fully allowing partial face matching.
- The applicability of the method to large pose variations (up to 80° of yaw rotation).
- The computational efficiency of the proposed method makes it suitable for real-world applications.

Thorough experimental analysis provides substantial evidence supporting the aforementioned assertions.

2 Related Work

The world is complex, dynamic, multidimensional;
the paper is static and flat.

– E. R. Tufte

Facial feature detectors can be distinguished into two main categories: those that detect feature points (landmarks) from the appearance characteristics of 2D intensity or color images and those that detect feature points from the geometric information of 3D objects or 2.5D scans. Facial feature detectors can also be classified into those that rely solely on geometric information and those that are supported by trained statistical feature models.

Landmark detectors use trained feature classifiers or 2D/3D appearance feature models/templates and 2D/3D geometry models for global topological consistency. There exist detectors that are based on 2D data, 3D data or fused combinations.

Even though existing 3D facial landmark detection methods claim pose invariance, they fail to address large pose variations. The main assumption of these methods is that even though the head can be rotated with respect to the sensor, the *entire* face is always visible. However, this is true only for "almost frontal" scans or "reconstructed" complete facial meshes. *Side scans usually have large missing areas, due to self-occlusion, and the size of the missing areas depends on the amount of pose variation.* These scans are very common in realistic scenarios such as in the case of imaging under uncontrolled conditions.

The same holds true for several 3D face recognition approaches that have been proposed. These approaches, although they aim to overcome the limitations of 2D face recognition by offering pose invariance, mostly utilize frontal 3D scans assuming that the entire face is visible to the sensor. This assumption is not always valid in real-world applications, since the unconstrained acquisition may lead to facial scans with extensive occlusions that result in missing data due to pose variations.

This Chapter describes related work in the fields of 3D and 2D facial landmark detection, facial features fusion for landmark detection and unconstrained partial face recognition.

2.1 3D Facial Landmark Detection

Development of 3D modeling and digitizing techniques has sparked research interest in 3D facial feature extraction for landmark detection and is reported in a number of publications.

Lu, Colbry, Stockman and Jain [83, 21, 84, 85, 20], in a series of publications, presented methods to locate the positions of eye and mouth corners, and nose and chin tips, based on a fusion scheme of shape index [34] on range maps and the "cornerness" response [56] on intensity maps. They also developed a heuristic method based on cross-profile analysis to locate the nose tip more robustly. Candidate landmark points were filtered out using a

static (non-deformable) statistical model of landmark positions, in contrast to the presented approach. The 3D feature extraction method presented in [21] addresses the problem of pose variations in a unified manner, and is tested against a composite database consisting of 953 scans from the FRGC database and 160 scan from a proprietary database with frontal scans extended with variations of pose, expressions, occlusions and noise. Their multimodal algorithm [83] uses 3D+2D information and is applicable to almost-frontal scans ($< 5°$ yaw rotation). It is tested against the FRGC database with 946 near frontal scans. The 3D feature extraction method presented in [84] also addresses the problem of pose variations, and is tested against the FRGC database with 953 near frontal scans along with their proprietary MSU database consisting of 300 multiview scans ($0°, \pm 45°$) from 100 subjects. Results of the methods [83, 84, 20] are presented in Table 15, and of the method [84] in Table 16, for comparison.

Conde *et al.* [22] introduced a global face registration method by combining clustering techniques over discrete curvature and spin images for the detection of eye inner corners and nose tip. The method was tested on a proprietary database of 51 subjects with 14 captures each (714 scans). Their database consists of scans with small pose variations ($< 15°$ yaw rotation). Although they presented a feature localization success rate of 99.66% on frontal scans and 96.08% on side scans, they do not define what a successful localization is.

Xu *et al.* [138] presented a feature extraction hierarchical scheme to detect the positions of nose tip and nose ridge. They introduced the "effective energy" notion to describe the local distribution of neighboring points and detect the candidate nose tips. Finally, an SVM classifier is used to select the correct nose tips. Although it was tested against various databases, no exact localization results were provided.

Lin *et al.* [78] introduced a coupled 2D and 3D feature extraction method to determine the positions of eye sockets by using curvature analysis. The nose tip is considered to be the extreme vertex along the normal direction of eye sockets. The method was used in an automatic 3D face authentication system, but was tested on only 27 human faces with various poses and expressions.

Segundo *et al.* [119] introduced a face and facial feature detection method by combining a method for 2D face segmentation on depth images with surface curvature information, in order to detect the eye corners, nose tip, nose base, and nose corners. The method was tested on the FRGC v2 database. Although they claim over 99.7% correct detections, they do not define a correct detection. Additionally, nose and eye corner detection presented problems when the face had a significant pose variation ($> 15°$ yaw and roll).

Wei *et al.* [136] introduced a nose tip and nose bridge localization method to determine facial pose. The method was based on a Surface Normal Difference algorithm and shape index estimation, and was used as a preprocessing step in pose-variant systems to determine the pose of the face. They reported an angular error of the nose tip - nose bridge segment less than $15°$ in 98% of the 2500 datasets of BU-3DFE facial database, which contains complete frontal facial datasets with capture range $\pm 45°$.

Mian *et al.* [88] introduced a heuristic method for nose tip detection. The method is based on a geometric analysis of the nose ridge contour projected on the $x - y$ plane. It is used as a preprocessing step to cut out and pose correct the facial data in a face recognition system. However, no clear localization error results were presented. Additionally, their nose tip detection algorithm has limited applicability to near frontal scans ($< 15°$ yaw and pitch).

Table 1: Comparison of 3D landmark detection methods

No	Ref	No of Lmks	Method	Test Datasets	Remarks
1	Colbry *et al.* [21] (2005), [20] (2006)	6	Shape Index map + 3D Statistical Landmark Contraints	FRGC v1 (953 scans) + propriatary DB (160 scans)	3D localization error in *mm* per landmark. No side scans. No expressions.
2	Conde *et al.* [22] (2005)	3	Discrete curvature + Spin Images + Clustering	Proprietary DB: 714 scans, almost-frontal	*NO localization results.*
3	Xu *et al.* [138] (2006)	1	Local distribution of points + SVM	Near-frontal	*NO localization results.*
4	Lin *et al.* [78] (2006)	3	Curvature and topological analysis	Proprietary DB: 27 faces, pose and expression variations	*NO localization results.*
5	Lu & Jain [84] (2006)	7	Shape Index map + Cornerness map + 3D Statistical Landmark Contraints	FRGC v1 (953 scans)	3D localization error in *mm* per landmark. No side scans. No expressions.
6	Segundo *et al.* [119] (2007)	8	Face segmentation + curvature analysis on depth images	FRGC v2: almost-frontal	*NO localization results.* Problems under significant pose variations.
7	Wei *et al.* [136] (2007)	1	Shape index + surface normal analysis	BU3DFE: 2,500 frontal scans	*NO localization results.*
8	Mian *et al.* [88] (2007)	1	Geometric analysis (heuristic)	Almost-Frontal	*NO localization results.* Problems under significant pose variations.
9	Faltemier *et al.* [41] (2008)	1	Curvature and Shape index analysis + Template matching	FRGC v2: 4,007 almost-frontal scans	*NO localization results.*
10	Faltemier *et al.* [42] (2008)	1	Rotated Profile Signatures	NDOff2007: 7,317 facial scans in various yaw and pitch angles.	*NO localization results.* 2D-assisted 3D method (it uses skin segmentation)
11	Dibeklioğlu *et al.* [32, 33] (2008)	4	Model of the local gradient of depth map	FRGC v1 + Bosphorus: Near-Frontal	*NO localization results.* Problems under significant pose variations.
12	Yu & Moon [142] (2008)	3	Trained model w. genetic algorithm	FRGC v1: Near-Frontal	3D errors per landmark (*mm*). Not applicable under significant pose variations.
13	Romero-Huertas & Pears [113] (2008)	3	convex and concave areas + graph model matching	FRGC v1 (509 scans) + FRGC v2 (3271 scans): Near-Frontal	*NO localization results.*
14	Nair & Cavallaro [92] (2009)	5	3D PDM + Shape Index + curvedness index	BU3DFE: 2,500 frontal scans	*NO localization results.* Not applicable under significant pose variations.
15	Perakis *et al.* [103] (2009), [100] (2013)	8	Shape Index + Spin Images + FLM	FRGC v2 + UND Ear: Frontal to profile scans	**State-of-the-art**. 3D errors per landmark (*mm*). Applicable under significant yaw and expression variations.

Faltemier *et al.* [41] introduced a heuristic method for nose tip detection. The method is a fusion of curvature and shape index analysis and a template matching algorithm using ICP. The nose tip detector had a localization error less than 10 *mm* in 98.2% of the 4007 facial datasets of FRGC v2 where it was tested. However, no exact localization distance error results were presented. They also introduced a method called "Rotated Profile Signatures" [42], based on profile analysis, to robustly locate the nose tip in the presence of pose, expression and occlusion variations. Their method was tested against NDOff2007 database which contains 7,317 facial scans, 406 frontal and 6,911 in various yaw and pitch angles. They reported a 96% to 100% success rate, with distance error threshold 10 *mm*, under significant yaw and pitch variations. Although their method achieved high success rate scores, it is a 2D-assisted 3D method since it uses skin segmentation to eliminate outliers, and is limited to the detection of the nose tip only. Finally, no exact localization distance error results were presented.

Dibeklioğlu, Salah and Akarun [32, 33] presented methods for detecting facial features

on 3D facial datasets to enable pose correction under significant pose variations. They introduced a statistical method to detect facial features, based on training a model of local features, from the gradient of the depth map. The method was tested against the FRGC v1 and the Bosphorus databases, but data with pose variations were not taken into consideration. They also introduced a nose tip localization and segmentation method using curvature-based heuristic analysis. However, the proposed system shows limited capabilities on facial datasets with yaw rotations greater than 45°. Additionally, even though the Bosphorus database used consists of 3,396 facial scans, they are obtained from 81 subjects. Finally, no exact localization distance error results were presented.

Yu and Moon [142] presented a nose tip and eye inner corners detection method on 3D range maps. The landmark detector is trained from example facial data using a genetic algorithm. The method was applied on 200 almost-frontal scans from FRGC v1 database. However, a limitation of the proposed system is that it is not applicable to facial datasets with large yaw rotations since it always uses the three aforementioned control points. Results of the method are presented in Table 15 for comparison reasons.

Romero-Huertas and Pears [113] presented a graph matching approach to locate the positions of nose tip and inner eye corners. They introduced the "distance to local plane" notion to describe the local distribution of neighboring points and detect convex and concave areas of the face. Finally, after the graph matching algorithm has eliminated false candidates, the best combination of landmark points is selected from the minimum Mahalanobis distance to the trained landmark graph model. The method was tested against FRGC v1 (509 scans) and FRGC v2 (3,271 scans) databases. They reported a success rate of 90% with thresholds for the nose tip at 15 mm, and for the inner eye corners at 12 mm.

Nair and Cavallaro [92] presented a method for detecting facial landmarks on 2.5D scans. Their method used the shape index and the curvedness index to extract candidate feature points (nose tip and inner and outer eye corners). A statistical shape model (PDM) of feature points is fitted to the facial dataset by using three control points (nose tip and left and right inner eye corners) for coarse registration, and the rest for fine registration. The localization accuracy of the landmark detector was tested against the BU-3DFE facial database, which only contains complete frontal facial datasets with capture range ±45°. Furthermore, their method is not applicable to missing data resulting from pose self-occlusion, since it always uses the aforementioned three control points for model fitting. Results of the method are presented in Table 15 for comparison purposes.

Finally, Perakis *et al.* [103, 100] presented methods for detecting facial landmarks (eye inner and outer corners, mouth corners, and nose and chin tips) on 2.5D scans. Local shape and curvature analysis utilizing shape index, extrusion maps and spin images were used to locate candidate landmark points. These are identified and labeled by matching them with a statistical facial landmark model. The method addresses the problem of extreme yaw rotations and missing facial areas, and it is tested against FRGC v2 and UND Ear databases.

2.2 2D Facial Landmark Detection

Two-dimensional landmark detection is based mostly on variations of the seminal work on Active Appearance Models of Cootes *et al.* [23, 27, 25, 28], and is reported in a number of publications.

Cootes *et al.* [25] presented an extension of the Active Appearance Model (AAM) algorithm, using multi-view 2D statistical landmark models (for −90°, −45°, 0°, +45°, +90° yaw

angles) to estimate head orientation and track faces through large angles. They learned a coupled model describing the relationship between the frontal appearance and the profile of a face to predict new views of a face and to constrain search algorithms which seek to locate a face in multiple views simultaneously.

Felzenszwalb and Huttenlocher [44] presented a computationally efficient framework for part-based modeling of objects based on the pictorial structure models. An object is represented by a collection of parts arranged in a deformable configuration. The appearance of each part is modeled separately, and the deformable configuration is represented by spring-like connections between pairs of parts. They demonstrated the technique by learning models that represent facial landmarks.

Vukadinovic and Pantic [134] presented a method for fully automatic detection of 20 facial feature points in images of expressionless faces using Gabor wavelet feature templates and GentleBoost based classifiers. The method tested on the Cohn-Kanade database, and achieved average detection rates of 93%.

Gu and Kanade [51] presented a method for aligning a 3D deformable model to a single face image. The model consists of a set of sparse 3D points and the view-based texture patches associated with every point. Assuming a weak perspective projection model, their algorithm iteratively deforms the model and adjusts the 3D pose to fit the image. Alignment experiments demonstrated that their approach can effectively handle unseen faces with a variety of pose and illumination variations.

Gu and Kanade [52] presented a face alignment system that is capable of dealing with exaggerated expressions, large occlusions, and a wide variety of image noises. Their system used a gradient-based landmark detector and a Gaussian shape distribution model. The inference algorithm iteratively examines the best candidate positions and updates face shape and pose, using an EM approach. Their system can effectively recover sufficient shape details from very noisy observations.

Romdhani and Vetter [112] introduced a probabilistic shape model based on the 3D Morphable Model (3DMM) that can be used to localize feature points in 2D images. Candidate feature points are detected using the SIFT descriptor. Using weak perspective projection of the 3D model to the 2D image and an optimization process based on maximum likelihood (ML) estimation, the relative position of the feature points, their appearance, scale, orientation and occlusion state are determined. Computational efficiency is obtained by using the Bellman principle and an early rejection rule based on 3D to 2D projection constraints. Evaluations of the detection algorithm on the CMU-PIE face images and on a large set of non-face images show high levels of accuracy (zero false alarms for more than 90% detection rate).

Cristinacce and Cootes [29] presented an extension of the Active Shape Model (ASM) framework. Two different types of non-linear boosted feature models represented by haar wavelets were trained using GentleBoost. The first type is a conventional feature detector classifier, which learns a discrimination function between the appearance of a feature and the local neighborhood. The second local model type is a boosted regression predictor which learns the relationship between the local neighborhood appearance and the displacement from the true feature location. They showed that the second regression model is much more efficient providing improved localization and increasing search speed.

Cristinacce and Cootes [28] introduced the Constrained Local Model (CLM) framework an extension of Active Appearance Model (AAM) framework. From a training set they constructed a joint model of the appearance of each feature together with their relative posi-

tions. The model is fitted to an unseen image in an iterative manner by generating templates using the joint model and the current parameter estimates, correlating the templates with the target image to generate response images and optimizing the shape parameters so as to maximize the sum of responses. They showed that the CLM algorithm is more robust and more accurate than the AAM search method, demonstrating improved localization accuracy on photographs of human faces.

Milborrow and Nicolls [89] presented extensions to the Active Shape Model, and use it to locate features in frontal views of upright faces. Their extensions were (i) fitting more landmarks than actually needed (ii) using 2D intensity gradients as landmark templates (iii) adding noise to the training set (iv) relaxing the constrains imposed by the shape model (v) trimming covariance matrices used for landmark similarity, and (vi) running two Active Shape Models in series.

Liang *et al.* [77] proposed a component-based discriminative approach for face alignment without requiring initialization. They first detect a number of candidate facial landmarks and construct an initial shape. Then, a discriminative search algorithm searches a new position for each facial landmark. The searching direction is determined by learned direction classifiers. Each landmark is represented by a Haar wavelet template, trained with Adaboost. Their approach gives excellent alignment results on the commonly used datasets created under controlled conditions, and also on a more challenging dataset containing facial images with a large range of variations in pose, lighting, expression and background.

Wu *et al.* [137] introduced the Boosted Ranking Model, a face model that is aligned by maximizing a score function, which was learned from training data. It is an extension of AAM and ASM approaches, with Haar wavelets used for landmark representation. GentleBoost was used for learning the detection scores.

Liu [79] introduced the Boosted Appearance Model (BAM), a discriminative framework for efficiently aligning images. During the modeling stage, a conventional Point Distribution Model (PDM) and a GentleBoost classifier, which acts as an appearance model, were trained. Haar wavelets were used for landmark representation. Using extensive experimentation, he showed that, compared to the AAM-based approach, this framework greatly improves the robustness, accuracy, and efficiency of face alignment by a large margin, especially for unseen data.

Zhou *et al.* [144] introduced the Combined Active Shape Model. It exploits the Scale Invariant Feature Transform (SIFT) as landmark descriptors and the Active Shape Model (ASM) as the landmark model. In order to have a better representation of face images, the landmarks on the face region and the face contour are modeled and processed separately. The performance of the proposed Combined-ASM algorithm is tested on the BioID and FRGC v2 face image databases.

Zhou *et al.* [145] presented a 3D Active Shape Model (3DASM) algorithm to automatically locate facial landmarks from different views. The 3DASM is trained by setting different shape and texture parameters on a 3D Morphable Model (3DMM). Using 3DMM to synthesize training data, landmarks have one to one correspondence between the 2D points detected from the image and 3D points on 3DMM. The Scale Invariant Feature Transform (SIFT) was used as a landmark descriptor. During fitting, 3D rotation parameters are computed using PCA. The experimental results show that the method is robust to pose variations.

Table 2: Comparison of 2D landmark detection methods

No	Ref	No of Lmks	Method	Test Datasets	Remarks
1	Cootes *et al.* [25] (2002)		Multi-view models	Frontal to profile	*NO localization results.* Performs poorly on unseen faces and low-res images.
2	Felzenswalb & Huttenlocher [44] (2005)	5 - 7	Pictorial structure model	Near-frontal	*NO localization results.*
3	Vukadinovic & Pantic [134] (2005)	20	GentleBoost classifier using Gabor and texture templates	Cohn-Kanade DB: 300 images, frontal, no expressions	*NO localization results.* The lack of a global shape model can lead to non-plausible facial configurations.
4	Gu & Kanade [51] (2006)	83	Trained 3D model + 2D intensity image patches + weak perspective projection estimation by EM	CMU-PIE: Frontal to profile	*NO localization results.*
5	Gu & Kanade [52] (2008)	83	Trained 3D model + gradient landmark descriptors + weak perspective projection estimation by EM	CMU-PIE + AR + internet images: Frontal to profile, expressions, occlusions	*NO localization results.*
6	Romdhani & Vetter [112] (2007)	83	3D morphable model + SIFT + 2D/3D fitting process by weak perspective projection	CMU-PIE: Frontal to profile	*NO localization results.* Optimization process hampered by the curse of dimensionality
7	Cristinacce & Cootes [29] (2007)	22	2D ASM + Haar wavelets + Gentleboost	BioID + XM2VTS: Near-frontal	*NO localization results.*
8	Cristinacce & Cootes [28] (2008)	22	Constrained Local Model + Texture patches w. PCA	BioID + XM2VTS: Near-frontal	*NO localization results.*
9	Milborrow & Nicolls [89] (2008)	68	2D Extended ASM + similarity or affine transformation constraints + gradient descent optimization	BioID: Near-frontal	*NO localization results.*
10	Liang *et al.* [77] (2008)	33	Haar wavelets + Discriminative Classifier based on Adaboost	AR + FERET + CMU-PIE + LFW: 2,469 images frontal to side, expressions and lighting variations	*NO localization results.*
11	Wu *et al.* [137] (2008)	33	Boosted Ranking Model (BRM): 2D PDM + Haar-like local features + GentleBoost	ND1 + BioID: Frontal to side	*NO localization results.*
12	Liu [79] (2009)	33	Boosted Appearance Model (BAM): 2D PDM + Haar-like local features + GentleBoost	ND1 + FERET + IMM + BioID	*NO localization results.*
13	Zhou *et al.* [144] (2009)	58	2D ASM + SIFT	BioID + FRGC v2: 11,204 near-frontal images	*NO localization results.*
14	Zhou *et al.* [145] (2010)	45	3D ASM + SIFT + Probabilistic PCA	IMM: 80 images + CMU-PIE: 280 images	*NO localization results.*
15	Valstar *et al.* [133] (2010)	22	SVM regression for localization + Markov Random Fields for global consistency + Haar-like local features	MMI + FERET + BioID: 1,855 images w. expression and occlusions	**State-of-the-art**. 2D normalized errors per landmark (mean error / interoccular dist.). Not real time (50 *sec* per image).
16	Zeng *et al.* [143] (2010)	45	Personalized 3D landmark model + multiview DAISY descriptors + Data-Driven Sample Consensus (DDSAC)	CMU Multi-PIE + UHDB14	*NO localization results.* Cannot be used on "unseen" faces.
17	Efraty *et al.* [39, 40] (2011)	12	PDM of landmarks + IMRA local descriptors + Adaptive Bag-of-Words Representation	CMU Multi-PIE	*NO localization results.*
18	Belhumeur *et al.* [4] (2011)	29	Combined outputs of local descriptors using Bayesian inference with a consensus of non-parametric global model	LPFW: 300 images + BioID. Frontal to side, w. expressions and lighting variations.	*NO localization results.* **State-of-the-art**. Most accurate in uncontrolled imaging conditions. Not real time (1 *sec* per landmark)
19	Dantone *et al.* [30] (2012)	10	Image patches + Conditional Regression Forests	LFW: 2,469 images. Frontal to side w. expressions and lighting variations.	**State-of-the-art**. 2D normalized errors per landmark (mean error / interoccular dist.) Close to human accuracy + Real time.

Valstar *et al.* [133] presented a method based on a combination of Support Vector Regression and Markov Random Fields to search for a landmark point's location. Using Markov Random Fields they constrained the search space by exploiting the constellations that facial points can form. The regressors on the other hand learn a mapping between the appearance of the area surrounding a point and the positions of these points, which makes detection of the points robust to variations of appearance due to facial expression and moderate changes in head pose. They adopted Haar-like filters as the descriptors of the local appearance of the landmark points. The proposed point detection algorithm was tested on 1,855 images, and the results showed that it outperformed previous landmark detectors.

Zeng *et al.* [143] presented a personalized landmark localization method that leverages information available from 2D/3D gallery data. To realize a robust correspondence between gallery and probe key points, they used: (i) a hierarchical DAISY descriptor that encodes contextual information, (ii) a Data-Driven Sample Consensus (DDSAC) algorithm that leverages the image information to reduce the number of required iterations for robust transform estimation, and (iii) a 2D/3D gallery pre-processing step to build personalized landmark metadata (i.e., local descriptors and a 3D landmark model). They validated their approach on the Multi-PIE and UHDB14 databases.

Efraty *et al.* [40, 39] presented a fully-automated system for facial component-landmark detection based on multi-resolution isotropic analysis and adaptive bag-of-words descriptors incorporated into a cascade of boosted classifiers. Specifically, first each component-landmark detector is applied independently and then the information obtained is used to make inferences for the localization of multiple components. The advantage of their approach is that it has robustness to pose as well as illumination. They demonstrated that using their method for the initialization of a point landmark detector results in a performance comparable with that of state-of-the-art methods.

Belhumeur *et al.* [4] presented a novel approach which combines the output of local detectors with a non-parametric set of global models for localizing parts in images of human faces. Assuming that the global models generate the part locations as hidden variables, they derived a Bayesian objective function, which was optimized using a consensus of models for these hidden variables. They showed excellent performance on a database gathered from the internet and showed that their localizer achieved state-of-the-art performance on the less challenging BioID database.

Dantone *et al.* [30] presented a real-time method that estimates feature points even on low quality images, using the conditional regression forest approach on image patches for this task. In their experiments, they used the head pose as a global property and demonstrated that conditional regression forests outperform regression forests for facial feature detection. They evaluated the method on the challenging Labeled Faces in the Wild database where close-to-human accuracy was achieved while processing images in real-time.

2.3 Feature Fusion

Feature fusion in multi-modal biometrics: A number of studies showing the advantages of information fusion in pattern recognition and especially in multi-modal biometrics have appeared in the literature.

Xu *et al.* [140] (1992) grouped different combining methods into categories and proposed methods for classifier fusion at different levels (measurement, rank and abstract). These combining methods were applied to recognizing handwritten numerals. They reported a

significant improvement over the performance of individual classifiers.

Kittler *et al.* [70] (1998) have developed a theoretical framework for the combination approach to fusion at the matching score level of multimodal biometric applications. In their approach the matching scores of individual classifiers are interpreted as posterior probabilities and the resultant scores are the outcome of simple fixed fusion rules. They have experimented with several fusion rules (sum rule, product rule, max rule, min rule, median rule and majority voting) for face and voice biometrics and found that the sum rule outperformed the others. They also concluded that the sum rule is not significantly affected by the probability estimation errors and this explains its superiority.

Jain *et al.* [60] (2000) conducted experiments concerning the characteristics of combining twelve different classifiers using five different combination rules and six different feature sets generated from handwritten numerals (0-9). Reported results show that each case favors its own combining rule and that combining does not necessarily lead to improved peformance.

Duin [36] (2002) presents fusion techniques in a general abstract context. His presentation is an intuitive discussion on the use of trained combiners. There is no conducted experimentation.

Ross and Jain [115] (2003) addressed the problem of information fusion in biometric verification systems by combining face, fingerprint and hand geometry modalities using sum, decision-tree and LDA based methods. They reported that the sum rule outperforms the others.

Jain *et al.* [61] (2005) presented a thorough classification of information fusion approaches in biometric systems. They also experimented with different normalization techniques (min-max, z-score, median, sigmoid, tanh and Parzen) and fusion rules (sum rule, max rule and min rule and weighted-sum rule) to combine score from different matchers in a multimodal biometric recognition system. They concluded that the tanh normalization is the most robust and efficient for a recognition system, and that weighted summation of the matching scores resulted in a significant improvement in recognition rates.

Ross and Govindarajan [114] (2005) have experimented with fusion at the feature level in 3 different scenarios: (i) fusion of PCA and LDA coefficients of face; (ii) fusion of LDA coefficients corresponding to the R,B,G channels of a face image; and (iii) fusion of face and hand modalities. They concluded that it is difficult to predict the best fusion strategy for a given scenario.

Snelick *et al.* [124] (2005) examined the performance of multimodal biometric authentication systems using fusion techniques over fingerprint and face modalities on a population approaching 1,000 individuals. They also introduced adaptive normalization techniques and weighted fusion rules. They concluded that multimodal fingerprint and face biometric systems can achieve better performance than unimodal systems.

Gökberk and Akarun [47] (2006) have presented fusion techniques for 3D face recognition. Their fusion schemes combine four face classifiers which are used for the comparison of gallery and probe faces. Reported results show that their serial fusion technique offers the best solutions.

Theoharis *et al.* [130] (2008) presented a multimodal biometric recognition system using the fusion of face and ear modalities. They reported that the fused multimodal system achieved better performance (99.7% rank-one recognition rate) than the unimodal systems. The high reported accuracy was attributed to the low correlation of the two modalities.

Landmark feature fusion: In landmark detection literature on the other hand the combination of landmark descriptors is an under-studied issue.

Lu and Jain [83] (2005) used the combination of shape index response derived from the range map (3D) and the cornerness response from the intensity map (2D) to determine the positions of the corners of the eyes and the mouth. They used a fusion scheme of a pixel-wise summation of the normalized shape index and cornerness response values, for the "resultant" feature values of mouth and eye corners. A statistical 3D feature location model is applied after aligning the model with the nose tip for landmark topological consistency. This automatic feature extraction algorithm has been integrated in an automatic face recognition system.

Boehnen and Russ [12] (2005) used color images (2D) and range data (3D). A skin detection algorithm is applied using the YCbCr transformation of the initial RGB image. The face region that results from skin detection is refined by using z-erosion exploiting the range data. Thus, at first a face segmentation is applied; next, eye and mouth likelihood maps are calculated (using Cb and Cr values), to locate the corresponding landmarks. Thus this method is not a fusion method but merely a 2D/3D masking/filtering method. Their algorithm runs in approximately 4 *sec* on a 640×480 image with registered range data. On a database of 1,500 images their algorithm achieved a facial feature detection rate of 99.6%.

Table 3: Comparison of 2D/3D landmark detection methods

No	Ref	No of Lmks	Method	Test Datasets	Remarks
1	Lu & Jain [83] (2005)	7	Shape Index map + Cornerness map + 3D Statistical Landmark Contraints	FRGC v1	3D localization error in *mm* per landmark. No side scans. No expressions.
2	Boehnen & Russ [12] (2005)	4	Fusion scheme of 2D color + 3D depth images	UND: Frontal w. expressions	*NO localization results.* 4 *sec* per img. Frontal only.
3	Jahanbin *et al.* [59] (2011)	11	Trained Gabor jets on intensity and depth images. Fusion of intensity and depth images similarity maps of trained Gabor jets. + 2D statistical landmark constrain model w. fixed 2D search areas.	T3DFRD: 1,146 2D/3D co-registered frontal datasets	**State-of-the-art**. 3D errors per landmark (*mm*). 2D normalized errors per landmark (mean error / interoccular dist.) Most accurate in controlled imaging conditions. Applicable to cropped and aligned frontal faces only.

Jahanbin *et al.* [59] (2011) used Gabor jets to represent intensity (2D) and range (3D) data. Next, the jets of each pixel were compared (using the appropriate similarity measure) to a target bunch (describing the queried landmark) in order to create similarity maps for each modality and landmark class. Finally, intensity and and range similarity maps were combined into a "hybrid" resultant similarity map. For the calculation of the resultant similarity map different approaches of fusion were examined such as taking the pixel-wise sum, product or maximum of the similarity scores. They concluded that summation is the most appropriate for robust landmark detection. Their goal was the construction of a unified multimodal (2D + 3D) face recognition system with boosted performance. Although their landmark detection method is the most accurate in controlled imaging conditions, it is applicable to cropped and registered frontal faces only, since they use a 2D landmark constrain model which is not generalized.

Perakis *et al.* [103] (2009), [100] (2013) and Passalis *et al.* [97] (2011) presented a 3D facial landmark detection system using the fusion of shape index and spin image feature descriptors. Their fusion system operated in a cascade (sequential) fashion so that the candidate landmarks extracted from the shape index transformation were classified and filtered out according to their similarity with precalculated spin image templates. They also used a product rule fusion of landmarks' geometric distance to a landmark model and spin image similarities at the decision level. They reported high landmark detection accuracy under large facial yaw rotations.

2.4 Partial Face Recognition

Most face recognition methods focus on frontal scans only (see the surveys of Bowyer *et al.* [15] and Chang *et al.* [19]). As a result, the performance of these methods is not evaluated with data that exhibit significant pose variations. In previous work of our team [96, 64], a 3D face recognition method has been presented (ranked first in the shape-only section of NIST's Face Recognition Vendor Test 2006). However, only frontal scans were used as the method did not handle missing data. In subsequent work involving this thesis [101, 97], this method was extended to handle missing data by introducing a pose invariant landmark detection step. The methods that are evaluated using data with pose variations are mentioned below. Note that none of them handles the extreme pose variations and the extensive missing data that the proposed method does.

Face Recognition under extreme poses: Lu *et al.* [83, 84, 85], in a series of works, have presented methods to locate the positions of eye and mouth corners, and nose and chin tips, based on a fusion scheme of shape index on range maps and the "cornerness" response on intensity maps. They also developed a heuristic method based on cross-profile analysis to locate the nose tip more robustly. Candidate landmark points were filtered out using a static (non-deformable) statistical model of landmark positions, in contrast to the presented approach. Although they report a 90% rank-one matching accuracy in an identification experiment, no claims where made with respect to the effects of pose variations in Face Recognition. Note that their pure 3D approach [84] (evaluated using multiview scans with yaw rotations up to 45° from MSU) that can handle pose variations has worse 3D Landmark Detection accuracy than their multimodal approach [83] (evaluated using near frontal scans from FRGC v1).

Dibeklioglu *et al.* [32, 33] introduced a nose tip localization and segmentation method using curvature-based heuristic analysis to enable pose correction in a face recognition system that allows identification under significant pose variations. However, a limitation of the proposed system is that it is not applicable to facial scans with yaw rotations greater than 45°. Additionally, even though the Bosphorus database used consists of 3,396 facial scans, they are obtained from only 81 subjects.

Blanz *et al.* [11, 10, 111] presented works on 3D face reconstruction by fitting their 3D Morphable Model on 3D facial scans. Their method is a well established approach for producing 3D synthetic faces from scanned data. However, face recognition testing is performed on FRGC database with frontal facial scans, and on FERET database with faces under pose variations which do not exceed 40°.

Bronstein *et al.* [17] presented a face recognition method that can handle missing data. Their method is based on their previous work of "isometric embedding" for a canonical

representation of the face [16]. On a limited database of 30 subjects they reported high recognition rates. However, the database they use has no side scans. The scans with missing data that they use are derived synthetically by randomly removing certain areas from frontal scans.

In Nair and Cavallaro's [91] work on partial 3D face matching, the face is divided into areas and only certain areas are used for registration and matching. The assumption is that the areas of missing data can be excluded. Using a database of 61 subjects, they show that using parts of the face rather than the whole face yields higher recognition rates. As is the case with their subsequent work on 3D landmark detection [92], their method is not applicable to missing data resulting from pose self-occlusion, especially when holes exist around the nose region.

Lin *et al.* [78] introduced a coupled 2D and 3D feature extraction method to determine the positions of eye sockets using curvature analysis. The nose tip is considered as the extreme vertex along the normal direction of eye sockets. The method was used in an automatic 3D face authentication system but was tested on only 27 human faces with various poses and expressions.

Mian *et al.* [88] introduced a heuristic method for nose tip detection and used it in a face recognition system. The method is based on a geometric analysis of the nose ridge contour projected on the $x-y$ plane. It is utilized as a preprocessing step to cut out and pose correct the facial data. Even though it allows up to 90° roll variation, it requires yaw and pitch variation less than 15°, thus limiting the applicability to near frontal scans.

Facial Asymmetry: It is a well known fact that the human face is not perfectly symmetrical. The exact level of facial asymmetry was recently quantified in the work of Liu and Palmer [80]. It was shown that, given a reasonable range of sensor noise, facial asymmetry is statistically significant. Additionally, facial asymmetry has been used as a biometric in several works, such as the works of Kompanets [73], Liu *et al.* [81] and Mitra *et al.* [90]. In these works, facial asymmetry offered promising biometric results particularly in the presence of facial expressions. As pointed out by Liu and Palmer [80], facial asymmetry should not be ignored without a justification. The idea that partial facial data can be used for biometric purposes has also been investigated by Gutta *et al.* [55] in the 2D face recognition domain with promising results.

The proposed method, on the other hand, exploits facial symmetry. This method does not rely on the assumption that the human face is perfectly symmetrical. The main assumption is that the difference (caused by facial asymmetry) between the left and the right region of a subject's face is less than the difference between these regions and the regions of another subject's face. The experimental results presented in Chapter 8 justified this assumption for the databases that were used.

3 Shapes and Landmarks

Do you see that cloud,
that's almost in shape like a camel?

– W. Shakespeare

In a wide variety of disciplines it is of great practical importance to measure, describe and compare the shapes of objects. In biometric applications, computer vision and computer graphics, the class of objects is often the human face. In almost any application, requiring processing of 3D facial data, an initial registration step based on feature points (landmarks) correspondence is the most crucial step in order to make a system fully automatic. Facial landmark detection can be used for face registration, face recognition, facial expression recognition, facial shape analysis, segmentation and labeling of facial parts, facial region retrieval, partial face matching, facial mesh reconstruction, face relighting, face synthesis, face animation and motion capture.

Statistical shape analysis is concerned with methodology for analyzing shapes in order to estimate population average shapes and the structure of shape variability. The foundation of statistical shape analysis was the pioneering work of Kendall [68] (1984) and Bookstein [13] (1986). Main contributions in shape analysis were the "Snakes" paper by Kass, Witkin and Terzopoulos [67] (1988) and subsequent papers published as "Active Shape Models: Smart Snakes" [26](1992) and "Active Shape Models: their training and application" by Cootes, Taylor, Cooper and Graham [27] (1995). Snakes and Active Shape Models are both deformable models but, contrary to Snakes, Active Shape Models (ASMs) have global constraints w.r.t. shape. These constraints are learned through observation, giving the model flexibility, robustness and specificity, as the model only can synthesize plausible instances w.r.t. the observations.

This Chapter presents the theoretical background of statistical shape analysis for describing shapes through landmarks, introduces the Facial Landmark Model (FLM) and defines its deformations. The FLM is constructed using Procrustes Analysis and Principal Component Analysis (PCA) over facial landmarks which have been manually pre-annotated on exemplar facial datasets.

3.1 The Shape Space

The word "shape" is very commonly used in everyday language, but what do we actually understand by the concept of shape? Dryden and Mardia [35] adopt the definition by D.G. Kendall:

Shape *is all the geometrical information that remains when location, scale and rotational*

effects are filtered out from an object.

According to this, shape is, in other words, invariant to Euclidean similarity transformations. This is reflected in Fig. 4.

Two objects have the same *shape* if they can be translated, scaled and rotated to each other so that they match exactly. Scale is sometimes considered a distinguishing characteristic.

Rigid shape *is all the geometrical information that remains when location and rotational effects are filtered out from an object.*

So, two objects have the same *size-and-shape* if they can be translated and rotated to each other so that they match exactly, i.e rigid shapes are rigid-body transformations of each other.

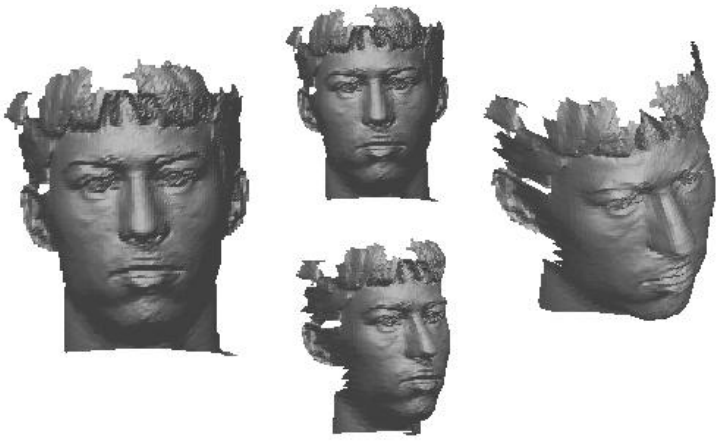


Figure 4: The same face shape under different Euclidean transformations.

The next question that naturally arises is: How should one describe a shape? One way to describe a shape is by locating a finite number of points on the outline or other specific points. Consequently, the concept of a *landmark* is adopted. According to Dryden & Mardia [35]:

Landmark *is a point of correspondence on each object that matches between and within populations.*

Dryden and Mardia [35] sort landmarks into the following categories:

Anatomical landmarks: *Points assigned by an expert that corresponds between organisms in some biologically meaningful way*, e.g. the corner of an eye.

Mathematical landmarks: *Points located on an object according to some mathematical or geometrical property*, e.g. a high curvature or an extremum point.

Pseudo-landmarks: *Constructed points on an object either on the outline or between anatomical or mathematical landmarks.*

Labeled landmarks: *Landmarks that are associated with a label (name or number), which is used to identify the corresponding landmark.*

Synonyms for landmarks include homologous points, interest points, nodes, anchor points, model points, key points, fiducial markers etc.

The mathematical representation of an n-point landmark shape in d dimensions can be defined by concatenating all landmark point coordinates into a $k = nd$ vector and establishing a *Shape Space* [35, 125, 23]. The *vector representation* for a 3D landmark shape $\mathbf{x} \in \mathbb{R}^k$ would then be:

$$\mathbf{x} = [[p_1]_x, \ldots, [p_n]_x, [p_1]_y, \ldots, [p_n]_y, [p_1]_z, \ldots, [p_n]_z]^T \ , \tag{1}$$

where $([p_i]_x, [p_i]_y, [p_i]_z)$ represent the coordinates of n landmark points $\mathbf{p}_i$ in the original Euclidean 3D space $\mathbb{R}^3$.

If a relationship between the distance in shape space and Euclidean distance in the original space can be established, then we have a *metric space*. This relationship is called a *shape metric*. A set of shapes actually forms a Riemannian manifold containing the shape object under consideration (*Kendall shape space*).

Often used shape metrics include the Hausdorff distance, the strain energy and the Procrustes distance. In the following the celebrated *Procrustes distance* will be used.

The squared Procrustes distance D_P between two shapes, $\mathbf{x}_1$ and $\mathbf{x}_2$, is simply a Euclidean metric in the $k = nd$ dimensional shape space $\mathbb{R}^k$:

$$D_P^2 = \|\mathbf{x}_1 - \mathbf{x}_2\|^2 = \sum_{j=1}^{k} ([x_1]_j - [x_2]_j)^2 \ . \tag{2}$$

The *centroid* $\mathbf{c}$ of a landmark shape is the center of mass (CM) of the physical system consisting of unit masses at each landmark $\mathbf{p}_i$. This is easily calculated in the original Euclidean 3D space $\mathbb{R}^3$ as:

$$\mathbf{c} = [c_x, c_y, c_z]^T = \left[\frac{1}{n}\sum_{i=1}^{n}[p_i]_x, \frac{1}{n}\sum_{i=1}^{n}[p_i]_y, \frac{1}{n}\sum_{i=1}^{n}[p_i]_z\right]^T \ . \tag{3}$$

The *centroid size* is used as a shape size metric. It is the square root of the sum of the squared Euclidean distances of each landmark $\mathbf{p}_i$ from the centroid $\mathbf{c}$:

$$S(\mathbf{x})^2 = \sum_{i=1}^{n} \|\mathbf{p}_i - \mathbf{c}\|^2 \tag{4}$$

in the original Euclidean 3D space. The centroid size has the property that $2nS(\mathbf{x})^2$ equals the sum of the inter-landmark distances.

3.1.1 Shape Alignment

Since shape has to be invariant to 3D Euclidean similarity transformations, translational, scale and rotational effects need to be filtered out by minimizing the *Procrustes distance* D_P:

$$D_P^2 = \|\mathbf{x}_i - \mathbf{x}_\mathbf{m}\|^2 = \sum_{j=1}^{k} ([x_i]_j - [x_m]_j)^2 \ , \tag{5}$$

between each example shape $\mathbf{x}_i$ and the mean shape $\mathbf{x_m}$.

The alignment procedure is commonly known as *Procrustes Analysis* [35, 125, 23] and is used to calculate the mean shape of the landmark shapes. Although there are analytic solutions, a typical iterative approach (adapted from [23]), is used. This approach is presented in Algorithm 1.

Algorithm 1 "Procrustes Analysis"

input: Example landmark shapes $\mathbf{x}_i$.
output: Landmarks' mean shape $\mathbf{x_m}$.

1: Compute the centroid $\mathbf{c}_i$ of each example shape $\mathbf{x}_i$.
2: Translate each example shape $\mathbf{x}_i$ so that its centroid $\mathbf{c}_i$ is at the origin (0,0,0).
3: Scale each example shape $\mathbf{x}_i$ so that its size is 1.
4: Assign the first example shape to the mean shape $\mathbf{x_m}$.
5: **repeat**
6: Assign the mean shape $\mathbf{x_m}$ to a reference shape $\mathbf{x}_0$.
7: Align all example shapes $\mathbf{x}_i$ to the reference shape $\mathbf{x}_0$ by an optimal rotation $\mathbf{R}$.
8: Recompute the mean shape $\mathbf{x_m}$.
9: Translate the mean shape $\mathbf{x_m}$ so that its centroid is at the origin (0,0,0).
10: Scale the mean shape $\mathbf{x_m}$ so that its size is 1.
11: Align the mean shape $\mathbf{x_m}$ to the reference shape $\mathbf{x}_0$ by an optimal rotation $\mathbf{R}$.
12: Compute the Procrustes distance $\|\mathbf{x}_0 - \mathbf{x_m}\|$ between the mean shape $\mathbf{x_m}$ and the reference shape $\mathbf{x}_0$.
13: **until** Convergence: $\|\mathbf{x}_0 - \mathbf{x_m}\| < \varepsilon$.
14: **return** Mean shape $\mathbf{x_m}$.

Thus, the mean shape of landmark shapes (Fig. 7) is created and all example shapes are aligned to it. The mean shape $\mathbf{x_m}$ is the *Procrustes mean* in $\mathbb{R}^k$

$$\mathbf{x_m} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \tag{6}$$

of all N example shapes $\mathbf{x}_i$.

Remarks: Note that in our case, where the size of the facial landmark shape is of great importance, scaling shapes to unit size is omitted in Algorithm 1. In such cases, the shapes are aligned by performing only the translational and rotational transformations. Thus, facial landmark distances are used as constraints that are incorporated during the training phase into the model. For this purpose, it is assumed that 3D facial data capture devices record the actual facial size. During the detection phase, landmark distances serve as constraints for rejecting outlier landmark shapes (Sections 3.2.2 and 6.2.1).

3.1.2 Alignment Transformations

As mentioned previously, to obtain a true representation of landmark shapes, location, scale and rotational effects need to be filtered out by bringing shapes to a common frame of reference. This is carried out by performing translational, scaling and rotational transformations. Notice that different approaches to alignment can produce different distributions of the aligned shapes.

Translation of an example shape $\mathbf{x}$ so that its centroid is at the origin is performed by applying to its n landmark points $\mathbf{p}_i$ the following transformation in 3D original space:

$$\mathbf{p}'_i = \mathbf{p}_i - \mathbf{c} \tag{7}$$

where $\mathbf{c}$ denotes the centroid of $\mathbf{x}$ and $i \in \{1, ..., n\}$.

Scaling of an example shape $\mathbf{x}$ to unit size is performed by applying to its n landmark points $\mathbf{p}_i$ the following transformation in 3D original space:

$$\mathbf{p}'_i = \alpha \mathbf{p}_i \tag{8}$$

where $\alpha = 1/S(\mathbf{x})$ is the scaling factor, $S(\mathbf{x})$ is the shape's size, and $i \in \{1, ..., n\}$.

Rotation in the original 3D space is slightly more complicated. A rotational transformation $R(\mathbf{x})$ has to be computed so as to minimize the Procrustes distance $\|R(\mathbf{x}) - \mathbf{x}_0\|$ between the transformed shape $R(\mathbf{x})$ and a reference shape $\mathbf{x}_0$. The rotational transformation $\mathbf{R}$ can be expressed as a product of three rotations around the three principal axes:

$$\mathbf{R} = \mathbf{R}_{x,\theta} \cdot \mathbf{R}_{y,\phi} \cdot \mathbf{R}_{z,\psi} \tag{9}$$

These can be expressed in a matrix form:

$$\mathbf{R}_{x,\theta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \tag{10}$$

$$\mathbf{R}_{y,\phi} = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \tag{11}$$

$$\mathbf{R}_{z,\psi} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

After setting partial derivatives of $\|R(\mathbf{x}) - \mathbf{x}_0\|^2$ w.r.t. each parameter to zero it is obtained that (Appendix A):

$$\theta = \tan^{-1}\left(\frac{S_{z0,y} - S_{y0,z}}{S_{y0,y} + S_{z0,z}}\right) \tag{13}$$

$$\phi = \tan^{-1}\left(\frac{S_{x0,z} - S_{z0,x}}{S_{z0,z} + S_{x0,x}}\right) \tag{14}$$

$$\psi = \tan^{-1}\left(\frac{S_{y0,x} - S_{x0,y}}{S_{x0,x} + S_{y0,y}}\right) \tag{15}$$

where:

$$S_{x0,x} = \sum_{j=1}^{n} x_{0j}x_j, \quad S_{x0,y} = \sum_{j=1}^{n} x_{0j}y_j, \quad S_{x0,z} = \sum_{j=1}^{n} x_{0j}z_j,$$

$$S_{y0,x} = \sum_{j=1}^{n} y_{0j}x_j, \quad S_{y0,y} = \sum_{j=1}^{n} y_{0j}y_j, \quad S_{y0,z} = \sum_{j=1}^{n} y_{0j}z_j,$$

$$S_{z0,x} = \sum_{j=1}^{n} z_{0j}x_j, \quad S_{z0,y} = \sum_{j=1}^{n} z_{0j}y_j, \quad S_{z0,z} = \sum_{j=1}^{n} z_{0j}z_j.$$

Hence, the rotational transformation of every landmark point $\mathbf{p}_i$ in the original 3D space gives:

$$\mathbf{p}'_i = \mathbf{R} \cdot \mathbf{p}_i \ , \tag{16}$$

with $i \in \{1, ..., n\}$.

Alignment of a shape $\mathbf{x}$ to a reference shape $\mathbf{x}_0$ is performed by minimizing the Procrustes distance in an iterative way, as described in Algorithm 2.

Algorithm 2 "Shape Alignment"

input: Landmark shapes $\mathbf{x}_0$ and $\mathbf{x}$.
output: Rotational transformation $\mathbf{R}$.

1: Translate $\mathbf{x}_0$ so that its centroid is at the origin (0,0,0).
2: Translate $\mathbf{x}$ so that its centroid is at the origin (0,0,0).
3: $\mathbf{R} := \mathbf{1}$.
4: **repeat**
5: Compute $\mathbf{R}_{x,\theta}$.
6: Apply $\mathbf{R}_{x,\theta}$ to the landmark points of shape $\mathbf{x}$.
7: $\mathbf{R} := \mathbf{R}_{x,\theta} \cdot \mathbf{R}$.
8: Compute $\mathbf{R}_{y,\phi}$.
9: Apply $\mathbf{R}_{y,\phi}$ to the landmark points of shape $\mathbf{x}$.
10: $\mathbf{R} := \mathbf{R}_{y,\phi} \cdot \mathbf{R}$.
11: Compute $\mathbf{R}_{z,\psi}$.
12: Apply $\mathbf{R}_{z,\psi}$ to the landmark points of shape $\mathbf{x}$.
13: $\mathbf{R} := \mathbf{R}_{z,\psi} \cdot \mathbf{R}$.
14: Compute the Procrustes distance $\|\mathbf{x}_0 - \mathbf{x}\|$ between the transformed shape $\mathbf{x}$ and the reference shape $\mathbf{x}_0$.
15: **until** Convergence: $\|\mathbf{x} - \mathbf{x}_0\| < \varepsilon$.
16: **return** $\mathbf{R}$.

Remarks:

a. Note that the proposed Algorithm 2 leaves us the discretion to permit certain rotations (e.g., only yaw rotations around the y-axis). For the case of a 2D shape only the $R_{z,\psi}(\mathbf{p}_i)$ transformation is applied.

b. Also, the implementation of Algorithm 2 avoids the *gimbal lock* problem, which lies in the fact that Euler angles (θ, ϕ, ψ) have to express a global rotation by a fixed order succession of only three global rotations around the main axes, $(\mathbf{R}_{x,\theta}, \mathbf{R}_{y,\phi}, \mathbf{R}_{z,\psi})$. In this implementation the gimbal lock problem is avoided since to express the overall rotation a single triplet of rotations is not used. Instead, as many rotations as are needed are used. These rotations are computed in an iterative way and are finally accumulated in the overall $\mathbf{R}$ transformation $(\mathbf{R}_{x,\theta}, \mathbf{R}_{y,\phi}, \mathbf{R}_{z,\psi}, \mathbf{R}_{x,\theta}, \mathbf{R}_{y,\phi}, \mathbf{R}_{z,\psi}, ...)$.

c. Algorithm 2 is a generalization in 3D of the method for aligning two shapes in 2D, presented by Cootes & Taylor in [23]. It is simple, transparent and converges in at most 8 iterations. It provides an efficient solution for registering 3D shapes, when a 1-1 correspondence between vertices is available, avoiding the use of more complex algorithms, such as the standard Iterative Closest Point (ICP) [8] (see also the Remarks in Sections 6.4 and 7.2).

3.1.3 Shape Variations

After bringing landmark shapes into a common frame of reference and computing the landmarks' mean shape, further analysis can be done to describe the shape variations. This shape decomposition is performed by applying Principal Component Analysis (PCA) to the aligned shapes.

Due to size normalization of Procrustes analysis, all shape vectors live in a hyper sphere manifold in shape space, which introduces non-linearities if large shape scalings occur. Since PCA is a linear procedure, all aligned shapes are at first projected to the tangent space of the mean shape (Fig. 5). This way, shape vectors lie in a hyper plane instead of a hyper sphere, and non-linearities are filtered out.

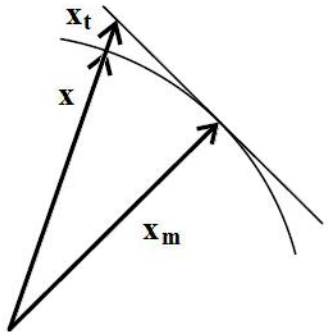


Figure 5: Tangent space projection $\mathbf{x_t}$ of a shape vector $\mathbf{x}$ to the mean shape $\mathbf{x_m}$.

The tangent space projection linearizes shapes by scaling them with a factor α:

$$\mathbf{x_t} = \alpha \mathbf{x} = \frac{\|\mathbf{x_m}\|^2}{\mathbf{x_m} \cdot \mathbf{x}} \mathbf{x} \tag{17}$$

where $\mathbf{x_t}$ is the tangent space projection of shape $\mathbf{x}$ and $\mathbf{x_m}$ is the mean shape. If no size normalization is applied, then tangent space projection is omitted.

Aligned shape vectors form a distribution in the $k = nd$ dimensional shape space. If landmark points were not representing a certain class of shapes, then they would be totally uncorrelated (i.e., purely random). On the other hand, if landmark points represent a certain class of shapes, as is in our case, then they will be correlated to some degree. This fact will be exploited by applying PCA to reduce dimensionality and obtain this correlation as shape deformations.

Since landmark points have a specific distribution, we can model this distribution by estimating a vector $\mathbf{b}$ of parameters that describes the landmark shape's deformations [27, 23, 24, 125].

After applying Procrustes analysis, the mean shape is calculated and example shapes are aligned and placed with their centroids at the origin and projected to the mean shape's tangent space. Typically, PCA is applied on variables with zero mean, but since this is not our case we replace each vector by $\mathbf{x} - \mathbf{x_m}$ [128]. The approach is described in Algorithm 3.

The covariance matrix $\mathbf{C_x}$ of the N aligned original example shape vectors $\mathbf{x}$ is computed according to

$$\mathbf{C_x} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i - \mathbf{x_m})(\mathbf{x}_i - \mathbf{x_m})^T \tag{18}$$

Algorithm 3 "Principal Component Analysis"

input: Example landmark shapes $\mathbf{x}_i$.

output: The mean shape $\mathbf{x_m}$, the eigenvectors $\mathbf{A}_i$ and the eigenvalues λ_i .

1: Apply Proctrustes Analysis to align example shapes $\mathbf{x}_i$ and compute their mean shape $\mathbf{x_m}$.
2: Transform the example shapes $\mathbf{x}_i$ to their projections $\mathbf{x}_t$ onto the tangent space.
3: Compute the covariance matrix $\mathbf{C_x}$ of the projected example shapes $\mathbf{x}_t$.
4: Compute the eigenvectors $\mathbf{A}_i$ and corresponding eigenvalues λ_i of $\mathbf{C_x}$, sorted in descending order.

If $\mathbf{A}$ contains (in columns) the $k = nd$ eigenvectors $\mathbf{A}_i$ of $\mathbf{C_x}$, by projecting aligned original example shapes to the eigenspace we uncorrelate them as

$$\mathbf{y} = \mathbf{A}^T \cdot (\mathbf{x} - \mathbf{x_m}) \tag{19}$$

and the covariance matrix $\mathbf{C_y}$ of projected example shapes $\mathbf{y}$

$$\mathbf{C_y} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{y_m})(\mathbf{y}_i - \mathbf{y_m})^T \tag{20}$$

becomes a diagonal matrix of the eigenvalues λ_i, so as to have

$$\mathbf{C_x} \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{C_y} \quad , \quad \mathbf{C_y} = \mathbf{A}^T \cdot \mathbf{C_x} \cdot \mathbf{A} = \mathbf{\Lambda} \tag{21}$$

and

$$\boldsymbol{\lambda} = diag(\mathbf{\Lambda}) \ . \tag{22}$$

The resulting transform is known as the *Karhunen-Loéve transform* (KLT), and achieves our original goal of creating mutually uncorrelated shapes [128].

To back-project uncorrelated shape vectors into the original shape space, we can use

$$\mathbf{x} = \mathbf{x_m} + \mathbf{A} \cdot \mathbf{y} \ . \tag{23}$$

Hence, if $\mathbf{A}$ contains (in columns) the p eigenvectors $\mathbf{A}_i$ corresponding to the p largest eigenvalues, we can approximate by $\mathbf{x}'$ any example shape $\mathbf{x}$ using

$$\mathbf{x} \approx \mathbf{x}' = \mathbf{x_m} + \mathbf{A} \cdot \mathbf{b} \tag{24}$$

where $\mathbf{b}$ is a p-dimensional vector given by

$$\mathbf{b} = \mathbf{A}^T \cdot (\mathbf{x} - \mathbf{x_m}) \tag{25}$$

The vector $\mathbf{b}$ is the projection of $\mathbf{x}$ onto the subspace spanned by the p most significant eigenvectors of the eigenspace (*principal components*). By selecting the p largest eigenvalues, the mean square error (MSE) between $\mathbf{x}$ and its approximation $\mathbf{x}'$ is minimized, since the last $k - p$ components are frozen to their respective mean values [128].

The vector $\mathbf{b}$ defines the deformation parameters of the model. By varying the components of $\mathbf{b}$ we can create shape deformations (Figs. 9, 10 and 11). By setting the following limits to each b_i:

$$b_i = \pm 3\sqrt{\lambda_i} \tag{26}$$

we can create marginal shape deformations, since each eigenvalue represents the data variance at the corresponding eigenspace axis [23, 128].

The number p of most significant eigenvectors and eigenvalues to retain (*modes of variation*) can be chosen so that the model represents a given proportion of the total variance of the data, that is the sum V_T of all the eigenvalues

$$\sum_{i=1}^{p} \lambda_i \geq f \cdot V_T \tag{27}$$

where factor f represents the percentage of the total shape variations of the training datasets.

Remarks: The computation of all eigenvalues and eigenvectors of a real symmetric matrix is done in a two step process [48, 109]: (i) application of the "Housholder reduction" algorithm for its reduction to tridiagonal form ($O(8n^3/3)$ operations), followed by (ii) the "QR iteration with implicit shifts" algorithm for computing its eigenvectors and eigenvalues ($O(n)$ operations per iteration). The above combination is the most efficient technique for finding all the eigenvectors and eigenvalues of a real symmetric matrix, according to [109].

3.2 The 3D Facial Landmark Models

The proposed method for 3D landmark detection and pose estimation uses 3D information to extract candidate interest points which are identified and labeled as anatomical landmarks by matching them with a Facial Landmark Model (FLM) [103, 101, 97, 100].

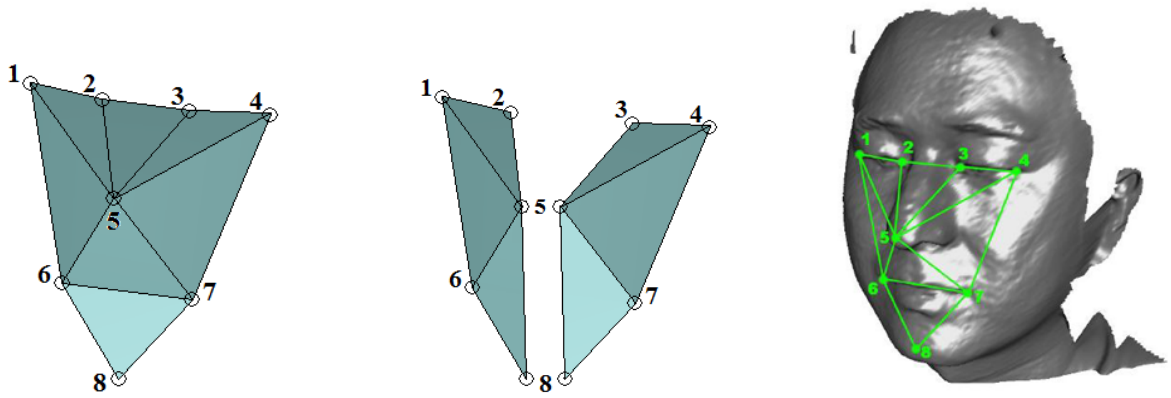


Figure 6: Depiction of: (a) FLM8 landmark model as a 3D object; (b) FLM5R and FLM5L landmark models; and (c) FLM8 landmark model overlaid on a 3D facial dataset.

For the creation of the FLM, a set of eight anatomical landmarks is used(Fig. 6):
(1) the Right Eye Outer Corner (REOC),
(2) the Right Eye Inner Corner (REIC),
(3) the Left Eye Inner Corner (LEIC),
(4) the Left Eye Outer Corner (LEOC),
(5) the Nose Tip (NT),
(6) the Mouth Right Corner (MRC),
(7) the Mouth Left Corner (MLC) and

(8) the Chin Tip (CT).

Note that five of these points are visible on profile and semi-profile face scans. Hence, the complete set of eight landmarks can be used for frontal and almost-frontal faces and two reduced sets of five landmarks (right and left) for semi-profile and profile faces. The right side landmark set and the left side landmark set contain the points (1, 2, 5, 6, 8), and (3, 4, 5, 7, 8), respectively.

Each of these sets of landmarks constitutes a corresponding Facial Landmark Model (FLM). Henceforth, the model of the complete set of eight landmarks will be referred to as FLM8 and the two reduced sets of five landmarks (right and left) as FLM5R and FLM5L, respectively.

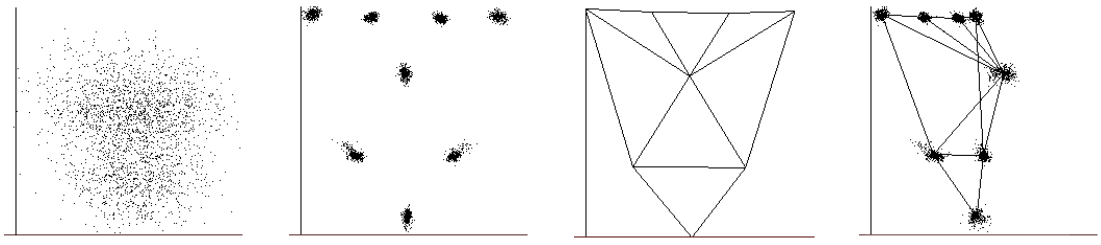
The main steps to create the FLMs are:

- The landmark models (FLM8, FLM5L and FLM5R) are computed from a manually annotated training set of 300 frontal facial scans of different subjects with varying expressions, which are chosen from the FRGC v2 database subset I (Fig. 51). The exact datasets that were used from the source databases for training (DB_TRAIN) can be found from the landmark annotation files available through the website [132]. Specifically, regarding faces there is a great variability in the visibility of landmarks according to pose changes. For this reason frontal face scans were used. It is important that the training set contains subjects that express the variations that the system is likely to face in practice (such as size, age and ethnicity). Training the FLMs with expressions allows the fitting procedure (Section 3.2.2) to capture candidate landmarks on faces exhibiting expressions.

- A statistical mean shape for each landmark set (FLM8, FLM5L and FLM5R) is computed from the manually annotated training set using Procrustes Analysis (Fig. 7).

- Variations of each Facial Landmark Model are computed using Principal Component Analysis (PCA) (Figs. 9, 10 and 11).

To retain the actual landmark shape model, Procrustes analysis and Principal Component Analysis have been carried out on the example shapes, according to Algorithm 3. Thus, the *Facial Landmark Model* (FLM) is created [103, 101, 97, 100], and is represented by the set $\{\mathbf{x}_m, \mathbf{A}_i, \lambda_i\}$, with $i \in \{1, ..., p\}$.



(a) (b) (c) (d)

Figure 7: Depiction of FLM8: (a) unaligned landmarks; (b) aligned landmarks; (c) landmarks' mean shape; and (d) landmark clouds and mean shape at 60°.

By applying PCA, we decompose landmark shape variations by projecting shapes to the eigenspace which has an ordered basis of eigenvectors. Thus each shape component

is ranked after the corresponding eigenvalue. This gives the components an order of significance (Fig. 8). Each eigenvalue represents the variance in eigenspace axes which are orthogonal. Note that the correlation matrix of shape vectors in the eigenspace has only diagonal elements: the eigenvalues (Fig. 12).

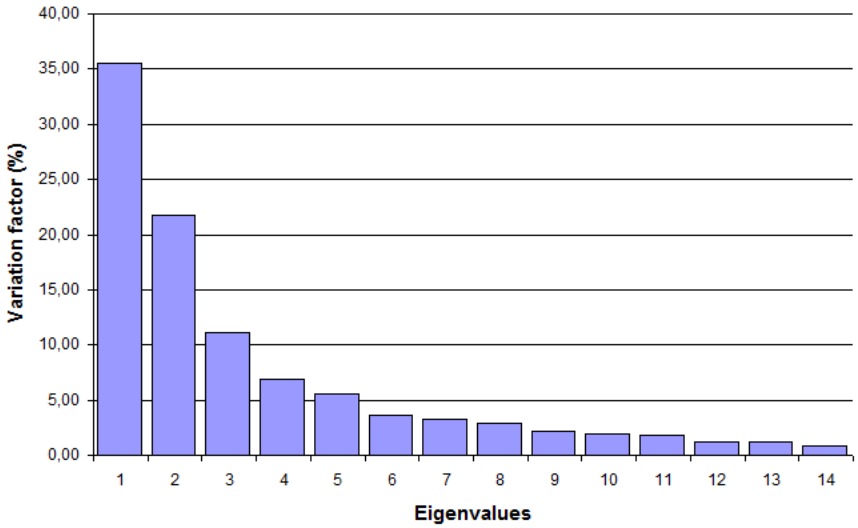


Figure 8: Landmark shape eigenvalues for FLM8 and percentage of total variations they capture.

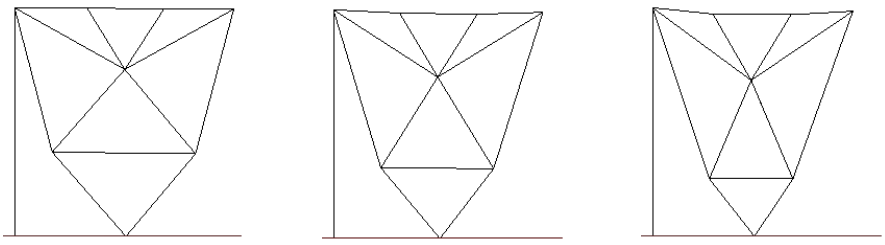
In FLM8 14 eigenvalues (out of the total 24) are incorporated, and in FLM5L and FLM5R seven eigenvalues (out of the total 15), which represent 99.0% of the total shape variations of each model. The least significant eigenvalues that are not incorporated into the FLMs are considered to represent noise [23, 128].

The incorporated eigenvalues represent the *principal modes of variation*. These variations can be rendered in the original 3D space as shape deformations $\mathbf{x}'$ of the mean landmark shape $\mathbf{x_m}$, using:

$$\mathbf{x}' = \mathbf{x_m} + \mathbf{A} \cdot \mathbf{b} \ . \tag{28}$$

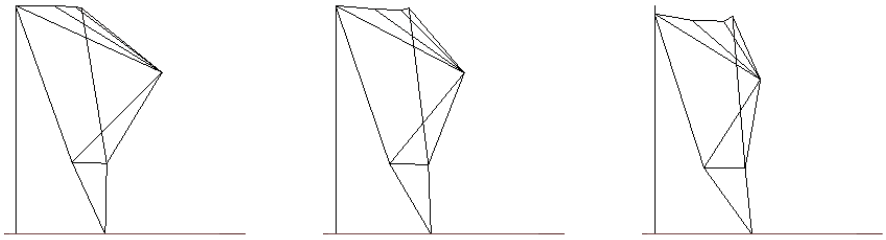
where $\mathbf{b}$ is a p-dimensional vector of deformation parameters.

By setting $b_i = \pm 3\sqrt{\lambda_i} = \pm 3\sigma_i$ and all the other $b_j = 0$ the marginal shape deformations for each mode of variation i are obtained, which represent $f_i = \frac{\lambda_i}{V_T}$ of the total shape variations incorporated into FLM [35, 23, 24].



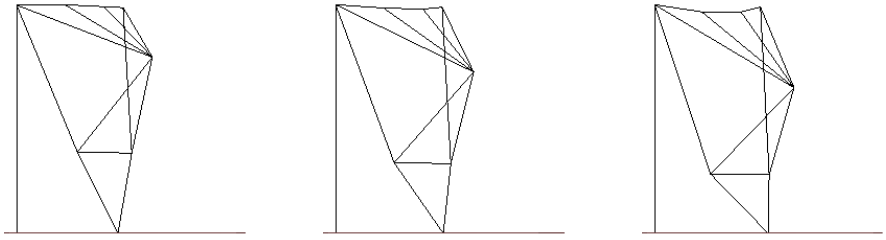
(a) $b_1 = -3\sqrt{\lambda_1}$ (b) $b_1 = 0$ (c) $b_1 = +3\sqrt{\lambda_1}$

Figure 9: First mode of FLM8 deformations at 0°.



(a) $b_2 = -3\sqrt{\lambda_2}$ (b) $b_2 = 0$ (c) $b_2 = +3\sqrt{\lambda_2}$

Figure 10: Second mode of FLM8 deformations at 70°.



(a) $b_3 = -3\sqrt{\lambda_3}$ (b) $b_3 = 0$ (c) $b_3 = +3\sqrt{\lambda_3}$

Figure 11: Third mode of FLM8 deformations at 60°.

The first mode captures the face size and shape (circular vs. oval) and represents 35.6% of the total shape variations of FLM8 (Fig. 9). The second mode captures the nose shape (peaked vs. flat) and represents 21.8% of the total shape variations of FLM8 (Fig. 10). The third mode captures the chin tip movement (down vs. up), due to open mouth and close mouth expressions, and represents 11.2% of the total shape variations of FLM8 (Fig. 11). The first three principal modes of FLM8 capture 59% of the total shape variations.

Remarks: The principal modes represent the marginal deformations of the landmark model (FLM), which are described by the deformation parameters b_i. These are used to establish whether a detected landmark shape is plausible or not (Section 3.2.2) and for computing the distance constraints of every pair of landmarks (Section 6.2.1). Note that, facial size is incorporated into the FLM by the first deformation parameter b_1. If scale normalization was applied, then size would not be incorporated into the FLM. Thus, at the detection phase, candidate landmark shapes consisting of outlier points (located on the hair or shirt), which are of "small sizes", would eventually be considered as plausible, resulting in more false detections.

3.2.1 Statistical Analysis of Landmarks

Unaligned landmark points (Fig. 7 a) are aligned by applying Procrustes analysis (Fig. 7 b). Point clouds of aligned landmarks seem to have a multivariate Gaussian distribution in 3D space. The axes of the Gaussians are analogous to the three standard deviations. The centroid of each landmark cloud coincides with the corresponding landmark point of the

mean shape (Fig. 7 c and d).

Point clouds of aligned landmarks represent landmark "movements" in 3D space. Looking at the correlation matrix it is obsereved that these "movements" are highly correlated (Fig. 12 a). Black squares denote negative correlation values, white, positive correlation values and mean gray, zero correlation.

Note that the shape vectors of the example shapes of FLM8 live in $\mathbb{R}^{24}$ shape space (8 landmarks $\times$ 3 coordinates each) and are presented by the vector

$$\mathbf{x} = [[p_1]_x, \ldots, [p_8]_x, [p_1]_y, \ldots, [p_8]_y, [p_1]_z, \ldots, [p_8]_z]^T \ , \tag{29}$$

where $([p_i]_x, [p_i]_y, [p_i]_z)$ represent the coordinates of the 8 landmark points $\mathbf{p}_i$ in the original Euclidean 3D space $\mathbb{R}^3$.

The main diagonal of the covariance matrix contains the variances of each shape vector component x_i

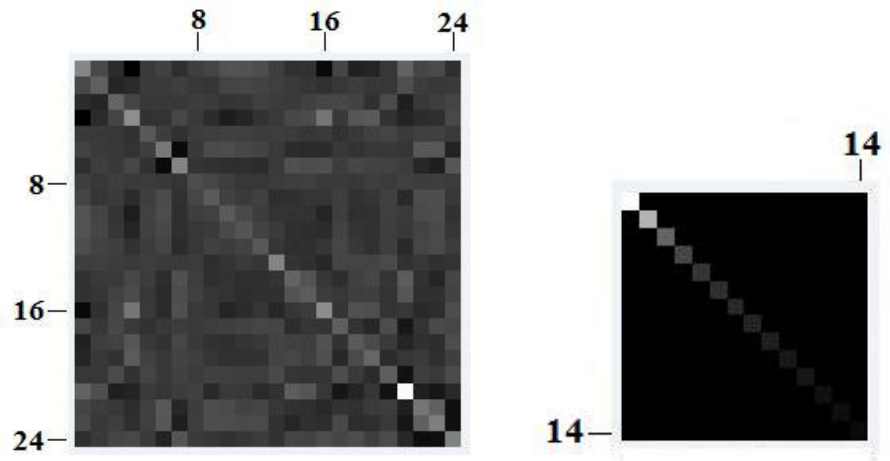
$$\text{Var}[x_i] = \frac{1}{N-1} \sum_{k=1}^{N} ([x_k]_i - [x_m]_i)^2 \tag{30}$$

and non-diagonal values the covariances between any two components x_i and x_j

$$\text{Covar}[x_i, x_j] = \frac{1}{N-1} \sum_{k=1}^{N} ([x_k]_i - [x_m]_i)([x_k]_j - [x_m]_j) \tag{31}$$

where $\mathbf{x_m}$ is the mean shape, $\mathbf{x_k}$ any example shape and N the examples number. The covariance matrix is symmetrical about the main diagonal, since $\text{Covar}[x_i, x_j] = \text{Covar}[x_j, x_i]$.

The values of the covariance indicate the strength of each relationship, and the sign whether the relationship is positive or negative. If the value is positive, the two components increase (or decrease) together. If it is negative, then if one component increases the other decreases.



(a) (b)

Figure 12: Statistical analysis of FLM8: (a) Correlation Matrix of its 24 components; (b) Selected 14 strongest eigenvalues.

Consider the black square (1,4) in Fig. 12(a); it represents the correlation between $([p_1]_x, [p_4]_x)$, which are the x coordinates of left and right eye outer corners. They are

negatively correlated; when the right eye moves right, the left eye moves left and vice versa. This is also indicated in the first mode of variations - circular vs. oval face (Fig. 9). Consider the black square (6,7); it represents the correlation between $([p_6]_x, [p_7]_x)$, which are the x coordinates of mouth left and right corners. They are also negatively correlated; when the mouth right corner moves right, the left corner moves left and vice versa. Black squares (24,22) and (24,23) represent the correlation between $([p_8]_z, [p_6]_z)$ and $([p_8]_z, [p_7]_z)$, which are the z coordinates of chin tip versus mouth left and right corners. There is a negative correlation, which means that chin tip and mouth corners move in opposite directions on the z-axis. This is also indicated in the third mode of variations - extruded vs. intruded chin (Fig. 11). Consider the gray squares of line (5); they represent the correlation of $([p_5]_x, [p_5]_y, [p_5]_z)$ coordinates of the nose tip with the other landmarks. We can conclude that the nose is mostly not correlated with any other landmark, because of the same gray color of the corresponding squares. It is the most robust facial landmark point. White square (21,21) represents the variance of $([p_5]_z)$, which is the z coordinate of the nose tip. We can observe that this has the maximum variance, which is also indicated in the second mode of variations - flat vs. peaked nose (Fig. 10).

3.2.2 Fitting Landmarks to the Model

General-purpose feature detection methods are not capable of identifying and labeling the detected candidate landmarks; some topological properties of faces must be taken into consideration. To address the problem of labeling the detected landmarks, the FLMs are used. Candidate landmarks, irrespective of the way they are produced, must be consistent with the corresponding FLM. This is accomplished by fitting a candidate landmark set to the FLM, and checking if the deformation parameters $\mathbf{b}$ fall within certain margins [23, 24].

Algorithm 4 "Landmark Fitting"

input: FLM $\{\mathbf{x_m}, \mathbf{A}_i, \lambda_i\}$ and probe landmark shape $\mathbf{y}$.
output: Acceptance of $\mathbf{y}$ (true/false).

1: Translate $\mathbf{y}$ so that its centroid is at the origin (0,0,0).
2: Scale $\mathbf{y}$ so that its size is 1, if $\mathbf{x_m}$ is also scaled.
3: **repeat**
4: Align $\mathbf{y}$ to the mean shape $\mathbf{x_m}$ by an optimal rotation $\mathbf{R}$.
5: Compute the Procrustes distance $\|\mathbf{y} - \mathbf{x_m}\|$ between $\mathbf{y}$ and the mean shape $\mathbf{x_m}$.
6: **until** Convergence: $\|\mathbf{y} - \mathbf{x_m}\| < \varepsilon$.
7: Compute the deformation parameters $\mathbf{b}$ of $\mathbf{y}$ from: $\mathbf{b} = \mathbf{A}^T \cdot (\mathbf{y} - \mathbf{x_m})$.
8: Accept $\mathbf{y}$ as a member of the shape's class if $\mathbf{b}$ satisfies certain constraints (Eqs. 33 and 34).

Fitting a set of landmark points $\mathbf{y}$ to the FLM $\{\mathbf{x}_m, \mathbf{A}_i, \lambda_i\}$ is done by minimizing the Procrustes distance $\|\mathbf{y} - \mathbf{x}_m\|$ in a simple iterative approach (adapted from [23]), as described in Algorithm 4. Then, by projecting $\mathbf{y}$ onto the shape eigenspace, its deformation parameters $\mathbf{b}$ are determined as:

$$\mathbf{b} = \mathbf{A}^T \cdot (\mathbf{y} - \mathbf{x}_m) \ . \tag{32}$$

A landmark shape $\mathbf{y}$ is considered as plausible if it is consistent with the marginal FLM deformations. Considering that certain b_i of $\mathbf{y}$ satisfy the deformation constraint

$$|b_i| \leq 3\sqrt{\lambda_i} \ , \tag{33}$$

then the candidate landmark shape $\mathbf{y}$ belongs to the shape class with probability

$$\Pr[\mathbf{y}] = \frac{\sum \lambda_i}{V_P} \ , \tag{34}$$

where λ_i are the eigenvalues that satisfy the deformation constraints and V_P is the sum of the eigenvalues that correspond to the selected p principal components, which represents the incorporated data variance in FLM. If $\Pr[\mathbf{y}]$ exceeds a certain threshold limit, the landmark shape is considered plausible, otherwise it is rejected as a member of the class. The threshold value is set to 0.99 so that only the weakest eigenvalue deformations may not be satisfied, since they can be considered as noise. Other criteria of declaring a landmark shape as plausible can also be used [23, 24].

4 Facial Data

Since we can't change reality,
let's change the eyes which see reality.

– N. Kazantzakis

Surfaces in the physical world have the property of varying continuously. Unfortunately, computers can only deal with discrete data. Thus, in order to perform any computation on a surface, we have first to approximate it by some discrete representation. The most basic problem in discrete surface representation is *sampling*. When we say that a surface is sampled, we imply a finite discrete set of points, called a *point cloud* (Fig. 13 a and d).

Obviously there are many ways to produce a point cloud out of a surface, and the natural question is how to decide whether one sampling is better than another. Intuitively, we wish the sampling to be as dense as possible, in order to better represent the underlying surface, and sparse enough so that the discrete representation does not increase storage and computational complexity costs.

If the neighborhood of each vertex can be mapped onto a disk (or to half-disk in case the surface has boundary vertices), we say that the mesh is a *manifold mesh* (Fig. 13 e and f). Equivalently, any edge in a manifold mesh belongs to at most two triangles. However, not every connectivity pattern results in a manifold mesh. For example, eight-neighbor connectivity produces a mesh where some edges are shared by four triangles, whereas six-neighbor connectivity produces a valid manifold mesh (Fig. 15).

This Chapter presents various facial data representations and the approaches used in this dissertation to process these 3D and 2D facial data. The algorithms presented in this chapter include: curvature calculations, issues with resolution and scale, data cleaning and preprocessing algorithms, and 2D maps of 3D data. It also describes the Annotated Face Model (AFM) as a generic 3D geometric model of facial datasets.

4.1 Facial Data Representations

4.1.1 Mesh Representation

In computational geometry and in computer graphics a surface can be approximated by a finite set of triangles. Such an approximation is called a *triangular mesh* (Fig. 13 b and e).

A mesh is usually defined as a structure of the form (V, F), consisting of a set of *vertices* V, and a set of triangular *facets* F. The facets can be represented as an $N_F \times 3$ matrix of indices, where the kth row is the set of vertices constituting the kth triangle,

$$\mathbf{f}_k = [[f_k]_1, [f_k]_2, [f_k]_3]^T \ , \quad [f_k]_i \in \{1, \ldots, N_V\} \ . \tag{35}$$

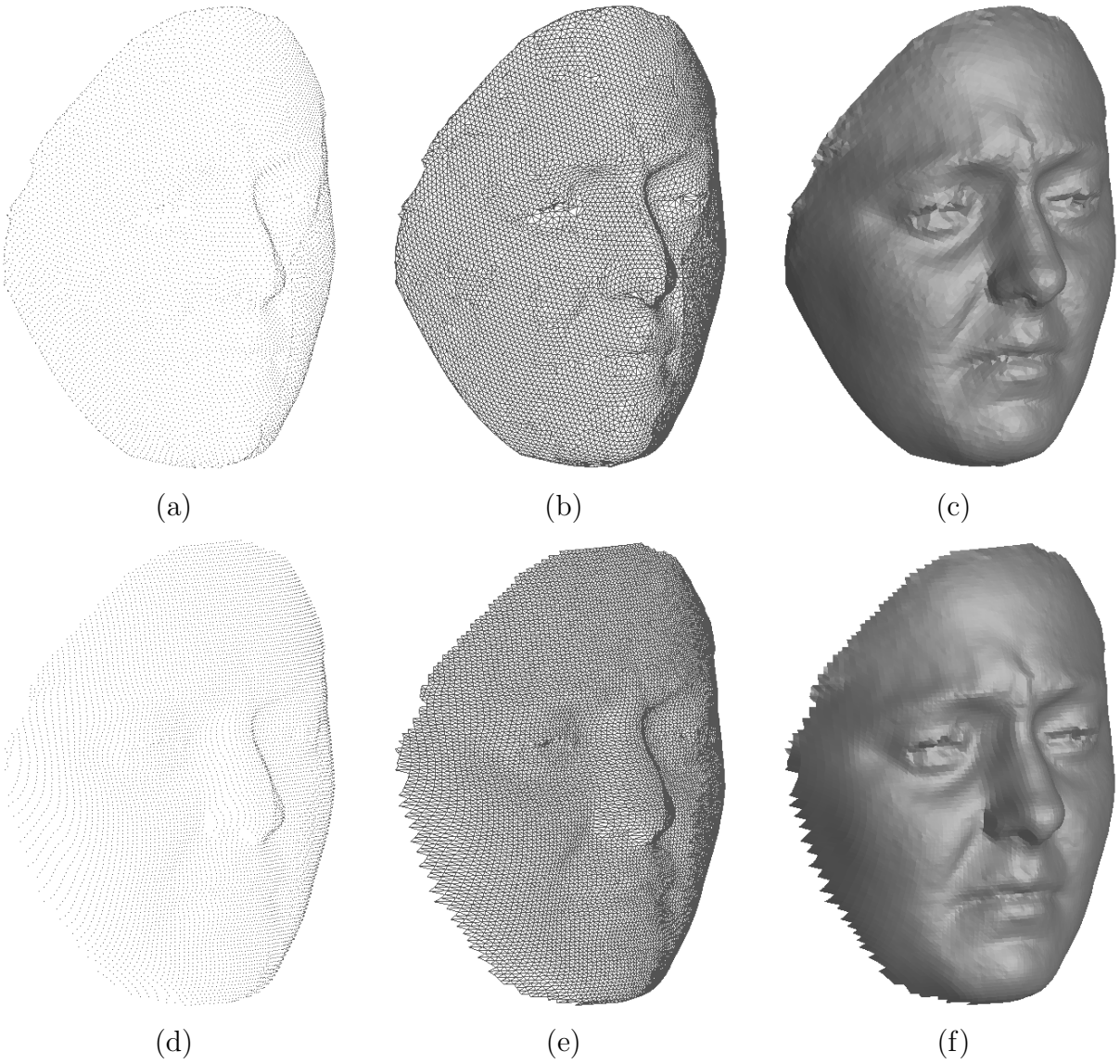
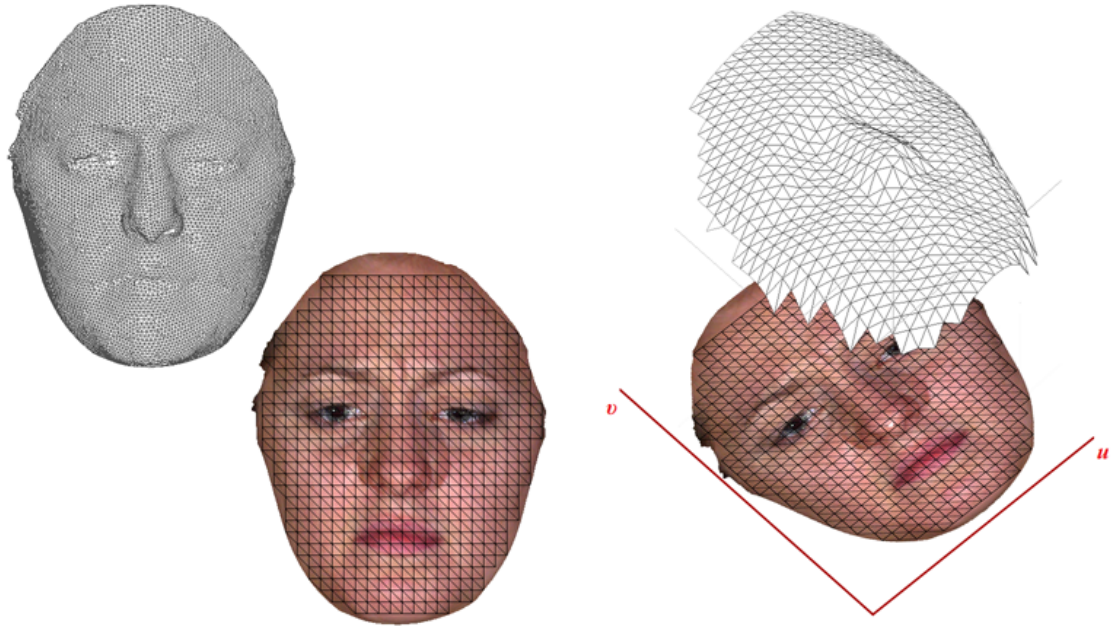


Figure 13: Facial data: (a) original point cloud; (b) original triangular mesh; (c) original face surface; (d) regularly sampled point cloud; (e) regular triangular mesh; and (f) face surface manifold.

N_F is the number of the triangular facets and N_V is the number of vertices. The vertices can be represented as an $N_V \times 3$ matrix of coordinates in $\mathbb{R}^3$, where the kth row is given by

$$\mathbf{p}_k = [[p_k]_x, [p_k]_y, [p_k]_z]^T \ . \tag{36}$$

Together the matrix of facets F and the matrix of coordinates V give a complete description of the triangular mesh. This piecewise planar approximation of an underlying smooth surface, defined as the union of triangular facets, is the most common representation of a discrete surface used in computer graphics.



(a) (b)

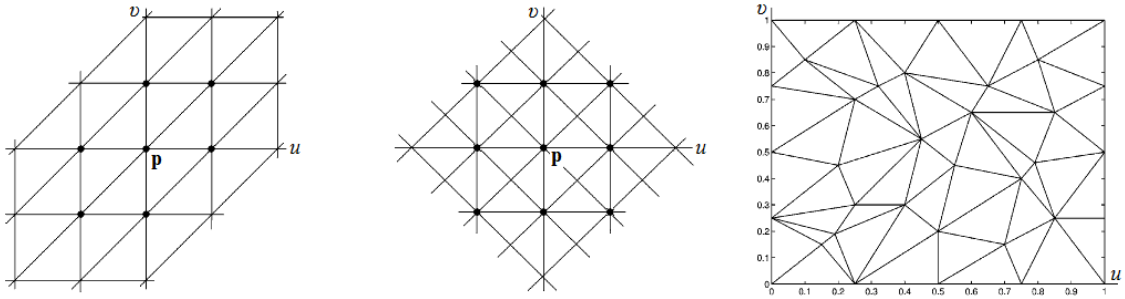
Figure 14: Facial mesh sampling: (a) irregular mesh and regular sampling; and (b) (u, v) parameterized regular mesh and registered texture image.

4.1.2 Parameterized Representations

Surfaces can be represented by a global bijective parameterization of the form

$$\mathbf{p}(u, v) = [x(u, v), y(u, v), z(u, v)]^T \ , \tag{37}$$

where (u, v) are coordinates on the parameterization domain, usually on the unit square $(u, v) \in [0, 1] \times [0, 1]$ (Figs. 14 and 15).



(a) (b) (c)

Figure 15: Parameterization domain: (a) regular mesh with 6-neighbor connectivity; (b) regular mesh with 8-neighbor connectivity; and (c) irregular mesh.

Sampling of surface properties in their preimage (bijective parametric domain or *texture atlas*) is convenient because the neighborhood of a point under consideration is known. We can sample a surface in a uniform way on the parameterization domain and create a Cartesian grid of values, which can be stored in matrix form, and displayed as a bitmap image. This way 2D maps of the 3D information of a surface can be stored and subsequently processed.

Algorithm 5 "Regular Orthographic Mesh Sampling"

input: Surface mesh (V, F) and texture image $I(u, v)$, $u_{samples}$, $v_{samples}$.
output: Geometry image I_G and texture image I_T.

Compute surface bounding box $(x_{min}, x_{max}, y_{min}, y_{max}, z_{min}, z_{max})$.
$u_{step} := (x_{max} - x_{min})/u_{samples}$.
$v_{step} := (y_{max} - y_{min})/v_{samples}$.
for $i := 1 \ldots u_{samples}$ **do**
  $x := x_{min} + (i - 1) * u_{step}$.
  for $j := 1 \ldots v_{samples}$ **do**
    $y := y_{min} + (j - 1) * v_{step}$.
    Consider line $(\mathbf{p}_0, \mathbf{p}_1)$ with $\mathbf{p}_0(x, y, z_{min})$ and $\mathbf{p}_1(x, y, z_{max})$.
    for $f := 1 \ldots \#facets$ **do**
      Get facet vertices $(\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2)$.
      Compute line-facet intersection point $\mathbf{q}(x, y, z)$ (Appendix B).
      if (intersection) **then**
        Compute texture coordinates (u, v) that correspond to $\mathbf{q}(x, y, z)$.
        Get texture value $I(u, v)$ using bilinear interpolation.
        BREAK
      end if
    end for
    if (intersection) **then**
      $I_G(i, j) := \mathbf{q}(x, y, z)$.
      $I_T(i, j) := I(u, v)$.
    end if
  end for
end for
return I_G and I_T.

Geometry Image Representation: A *geometry image* (Fig. 16 a) is the result of mapping all vertices of a 3D object (x, y and z coordinates) to a uniform 2D Cartesian grid representation (u, v coordinates) [53]. Thus, a geometry image is a regular contiguous sampling of a 3D model represented as a 2D image, with each u, v pixel corresponding to the original x, y, z coordinates (Eq. 37) [18, 1]. 2D geometry images have at least three channels assigned to each pair of u, v coordinates, encoding geometric information (x, y, z coordinates) and eventually a flag denoting data availability (i.e., holes). A simple orthographic mapping is presented in Algorithm 5, with each u, v pixel corresponding to the nearest x, y, z coordinate. For various techniques on surface parameterizations and remeshing see the survey of Alliez *et al.* [1].

Normal Image Representation: A *normal image* (Fig. 16 b) is an extension of the geometry image, mapping the normal components at the sampled points of a 3D object (n_x, n_y and n_z) to a 2D grid representation (u, v coordinates). 2D normal images have three channels assigned to each pair of u, v coordinates, encoding the normal information (n_x, n_y and n_z components), or two for normalized normals ($\hat{n}_x$, $\hat{n}_y$, since $\hat{n}_z = \sqrt{1 - \hat{n}_x^2 - \hat{n}_y^2}$), and eventually a flag denoting data availability.

Depth Image Representation: A particular case of parameterization is the so-called *Monge form*. This parameterization is given by $(u, v, z(u, v))$. For a Monge surface, the representation is even simpler if we store in a matrix the values of the function $z(u, v)$. This can be realized by a single-channel image, that is called *depth image* or *height field* (Fig. 16 c).

(a) (b) (c)

Figure 16: Facial data: (a) geometry image; (b) normal image; and (c) depth image.

4.2 Preprocessing

4.2.1 Range data preprocessing

A facial scan belongs to a subclass of 3D objects which is a surface S expressed in parametric form with native global (u, v) parameterization:

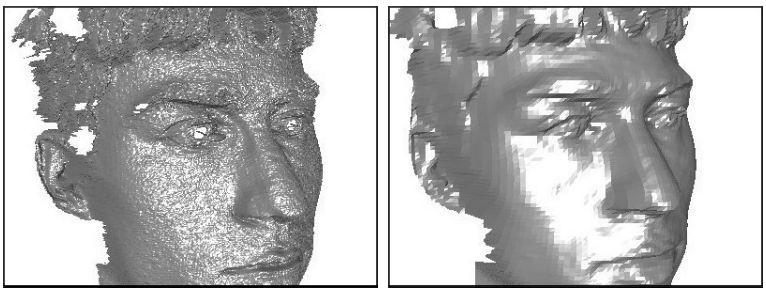
$$S(u, v) = \{\mathbf{p} \in \mathbb{R}^3 : \mathbf{p} = [x(u, v), y(u, v), z(u, v)]^T, (u, v) \in \mathbb{R}^2\} \tag{38}$$

Facial scans can also incorporate texture data acquired as a registered 2D image:

$$I(u, v), (u, v) \in \mathbb{R}^2 \tag{39}$$

This parameterization allows to map 3D information onto 2D space and vice-versa, thus the 3D and 2D information can be cross-referenced [103, 101, 97, 100].

Since differential geometry is used for describing local behavior of surfaces in a small neighborhood, such as surface curvature and surface normals, it is assumed that the surface S can be adequately modeled as being at least piecewise smooth, that is, at least be of class $\mathcal{C}^2$ (twice differentiable).



(a) (b)

Figure 17: Example of a facial scan (a) before and (b) after preprocessing.

The proposed method of landmark detection and face recognition can use both polygonal and range data obtained from optical or laser scanners. Scanners usually produce spikes and holes, even for surface regions visible to them, especially in areas like the eyebrows and the eyes, due to their inability to properly pick up the reflection from these surfaces. To eliminate such sensor-specific problems and convert data into a unified representation, certain preprocessing algorithms are applied directly on the range data before the conversion to polygonal data [64, 97]:

- *Median Cut*: To remove spikes a median cut filter with a 3×3 window was applied.
- *Hole Filling*: To remove holes, a hole filling procedure that uses bilinear interpolation was applied.
- *Smoothing*: A smoothing Gaussian filter with a 3×3 window was applied to remove white noise.
- *Subsampling*: The range data were subsampled at an $1:4$ ratio to reduce the computational cost.

An example of range data suffering from noise and holes is given in Fig. 17, before (a) and after (b) preprocessing.

4.2.2 Texture data registration

Preprocessing methods used in this dissertation also include algorithms for creating a regular (u, v) parametric surface. This is accomplished by an orthographic regular resampling of the 3D irregular surface mesh (see Algorithm 5). The resulting surface is able to represent the curved surface of a face as accurately as required for our purposes (mean edge length: 0.5 mm). By concurrently resampling the texture image, a unified representation of 3D and 2D data is accomplished by a (u, v) parametric map, even for facial scans where the texture map may not be contiguous (Fig. 18). Thus the 3D and 2D information can be cross-referenced (Eqs. 37 and 39).

(a) (b) (c)

Figure 18: Facial data: (a) original triangular mesh; (b) original texture image; (c) (u, v) registered face mesh and texture image after resampling.

4.2.3 RGB image preprocessing

Since the texture images are in RGB space we need to convert them into B/W intensity images, appropriate for the application of intensity differential operators. For this purpose, the L component of the CIE Lab color model is used, since it fixes to some extent the shadings due to illumination conditions, and gives more perceptually equalized intensity histograms.

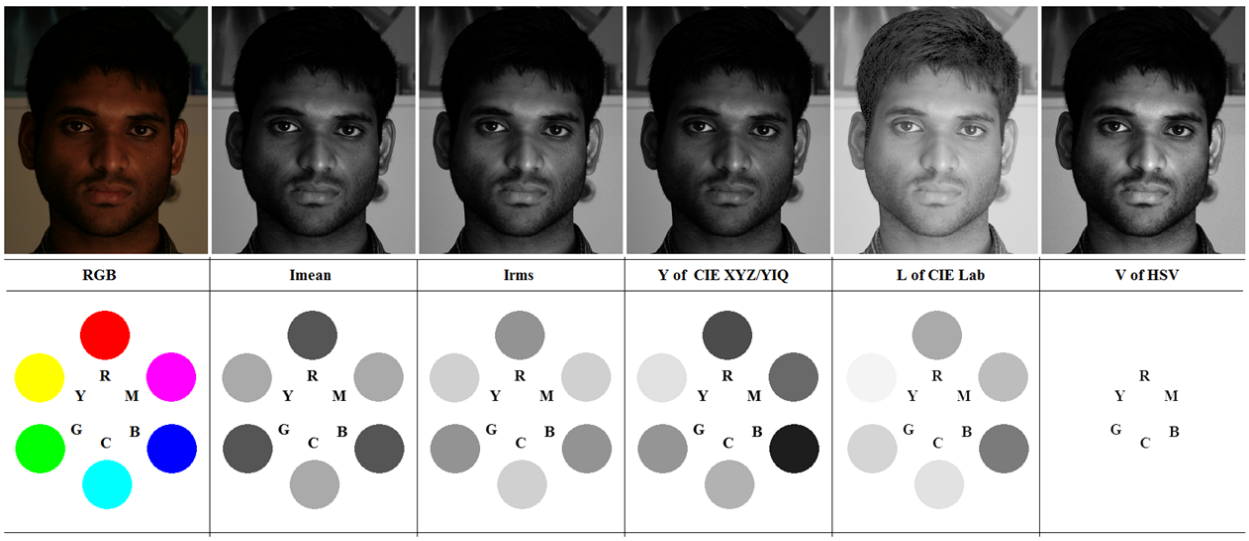


Figure 19: Depiction of various RGB to B/W transformations.

In Fig. 19, several methods for converting RGB color images to B/W intensity images are depicted:

a. The RGB mean value I_{mean} given by:

$$I_{mean} = (R + G + B)/3 \ . \tag{40}$$

b. The RGB rms value I_{rms} given by:

$$I_{rms} = \sqrt{(R^2 + G^2 + B^2)/3} \ . \tag{41}$$

c. The Value component V of the HSV color model given by:

$$V = \max(R, G, B) \ . \tag{42}$$

d. The Luminance (Luma) component Y of the XYZ/YIQ color model given by:

$$Y = 0.299 \ R + 0.587 \ G + 0.114 \ B \ . \tag{43}$$

e. The Luminance component L^* of CIE Lab color model given by:

$$L^* = (116.0 \ L - 16.0)/100.0 \ , \tag{44}$$

where

$$L = \begin{cases} Y^{\frac{1}{3}} & \text{if } \ Y \geq 0.00885 \\ 7.78703 \ Y + 0.13793 & \text{otherwise} \end{cases} \tag{45}$$

.

R, G, B are the red, green and blue components of the RGB color image.

In CIE Lab color model equal relative differences in color and brightness are equally quantified, according to a 1/3 power law which simulates *Weber's Law* of perception. Several high-end products, including Adobe Photoshop, use the CIE Lab model [76].

4.3 Differential Maps

Differential maps are used for describing the local behavior of surfaces in a small neighborhood, such as surface curvature and surface normals. The geometric parameters of a 2D surface (2-manifold) embedded in $\mathbb{R}^3$ are:
$S(\mathbf{p})$: Surface represented by the vertices $\mathbf{p}(u, v)$ of a parameterized triangular mesh
$\mathbf{T}(\mathbf{p})$: Tangent (1st derivative of $S(\mathbf{p})$ - represents velocity)
$\mathbf{N}(\mathbf{p})$: Normal on $S(\mathbf{p})$ (represents tangent plane)
$\mathbf{K}_N(\mathbf{p})$: Normal Curvature (represents radial acceleration of $S(\mathbf{p})$)
$\mathbf{K}_T(\mathbf{p})$: Tangential or Geodesic Curvature (represents transverse acceleration of $S(\mathbf{p})$)
$\mathbf{K}(\mathbf{p})$: Total Curvature (2nd derivative of $S(\mathbf{p})$ - represents net acceleration)
$\mathbf{K}_L(\mathbf{p})$: Laplace-Beltrami Operator (LBO)
dA : Infinitesimal Area of a point neighborhood dV on $S(\mathbf{p})$

Curvature vectors for the simple case of a planar line are depicted in Fig. 20.

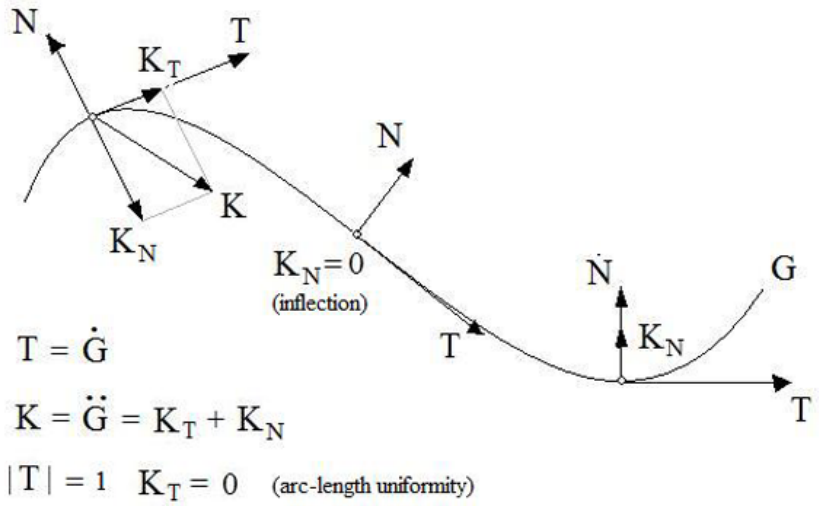


Figure 20: Depiction of curvature vectors on a planar line $\mathbf{G}$.

We have the following definitions:
a. Normal Curvature:

$$\mathbf{K}_N(\mathbf{p}) = \mathbf{K}_L(\mathbf{p}) = \nabla^2 S(\mathbf{p}) \ . \tag{46}$$

$\mathbf{K}_N$ is computed by applying the discrete Laplace-Beltrami $\mathbf{K}_L$ operator on the surface $S(\mathbf{p})$. Since every point $\mathbf{p}$ of $S(\mathbf{p})$ has 3 components, each component is considered as an independent scalar field for the application of the LBO. Thus, $\mathbf{K}_N$ is a vector field on the surface $S(\mathbf{p})$:

$$\mathbf{K}_N(\mathbf{p}) = \mathbf{K}_L(\mathbf{p}) = \frac{1}{dA} \sum_{dV} w \ d\mathbf{p} \ , \tag{47}$$

where w represents properly selected weights (see also Eq. 60).
b. Mean Curvature:

$$K_H(\mathbf{p}) = \frac{1}{2} \left[K_{min}(\mathbf{p}) + K_{max}(\mathbf{p}) \right] \ . \tag{48}$$

This is a scalar field on the surface $S(\mathbf{p})$, and is derived from the normal curvature $\mathbf{K}_N$:

$$K_H(\mathbf{p}) = \frac{1}{2} \left\| \mathbf{K_N}(\mathbf{p}) \right\| \ . \tag{49}$$

c. Gauss Curvature:

$$K_G(\mathbf{p}) = K_{min}(\mathbf{p}) \cdot K_{max}(\mathbf{p}) \ . \tag{50}$$

This is a scalar field on the surface $S(\mathbf{p})$, and is derived from the "*Gauss-Bonnet theorem*" (see also Eq. 63):

$$K_G(\mathbf{p}) = \frac{1}{dA} \left(2\pi - \sum_{dV} \theta \right) \ . \tag{51}$$

d. Min and Max Curvatures:

$$K_{min}(\mathbf{p}) = K_H(\mathbf{p}) - \sqrt{K_D(\mathbf{p})} \ , \tag{52}$$

$$K_{max}(\mathbf{p}) = K_H(\mathbf{p}) + \sqrt{K_D(\mathbf{p})} \ . \tag{53}$$

These are scalar fields on the surface $S(\mathbf{p})$, and are derived from K_H and K_D:

$$K_D(\mathbf{p}) = K_H^2(\mathbf{p}) - K_G(\mathbf{p}) \ . \tag{54}$$

Remarks:

a. The unit normal vector $\hat{\mathbf{n}}$ can be computed by averaging (with properly selected weights w_i) the unit normals $\hat{\mathbf{n}}_i$ on the facets around $\mathbf{p}$ [129], according to:

$$\mathbf{N}(\mathbf{p}) = \sum_{dV} w_i \hat{\mathbf{n}}_i \ , \tag{55}$$

and subsequently normalized such that

$$\hat{\mathbf{n}}(\mathbf{p}) = \frac{\mathbf{N}(\mathbf{p})}{\|\mathbf{N}(\mathbf{p})\|} \ . \tag{56}$$

It is a vector field on the surface $S(\mathbf{p})$, and always has an outward direction according to a CCW or CW convention. It also represents the tangent plane on the surface $S(\mathbf{p})$ since $\mathbf{N}(\mathbf{p}) \perp \mathbf{T}(\mathbf{p})$ (Fig. 21 a).

b. The normal curvature $\mathbf{K}_N$ is related to the mean curvature K_H, according to:

$$\mathbf{K}_N(\mathbf{p}) = 2 \cdot K_H(\mathbf{p}) \cdot \hat{\mathbf{K}}_N(\mathbf{p}) \ , \tag{57}$$

where $\hat{\mathbf{K}}_N$ is the normal curvature unit vector:

$$\hat{\mathbf{K}}_N(\mathbf{p}) = \frac{\mathbf{K}_N(\mathbf{p})}{\|\mathbf{K}_N(\mathbf{p})\|} \ . \tag{58}$$

c. The normal curvature unit vector $\hat{\mathbf{K}}_N$ is the same as the unit normal vector $\hat{\mathbf{n}}$, but contains the inflection points and curvature direction. The normal curvature $\mathbf{K}_N$ is the same as $\hat{\mathbf{K}}_N$, but also contains the magnitude of curvature K_H.

d. The principal curvatures $K_1 = K_{max}$ and $K_2 = K_{min}$ are the maximum and minimum of all normal curvatures for any direction of $\mathbf{T}$ on the tangent plane. $\hat{\mathbf{k}}_1$ and $\hat{\mathbf{k}}_2$ are the principal tangential directions which have the max and min curvatures (Fig. 21 a).

e. The mean curvature K_H is half the magnitude of normal curvature $\mathbf{K}_N$. It is the arithmetic mean of the principal curvatures K_{min} and K_{max}. It is not invariant to isometric deformations.

f. The Gauss curvature K_G is the product (geometric mean) of the principal curvatures K_{min} and K_{max}. K_G is an intrinsic property of $S(\mathbf{p})$ according to the "*Theorema Egregium*" [93, 7, 3]. It is thus invariant to isometric deformations (bendings but not stretches) [18].

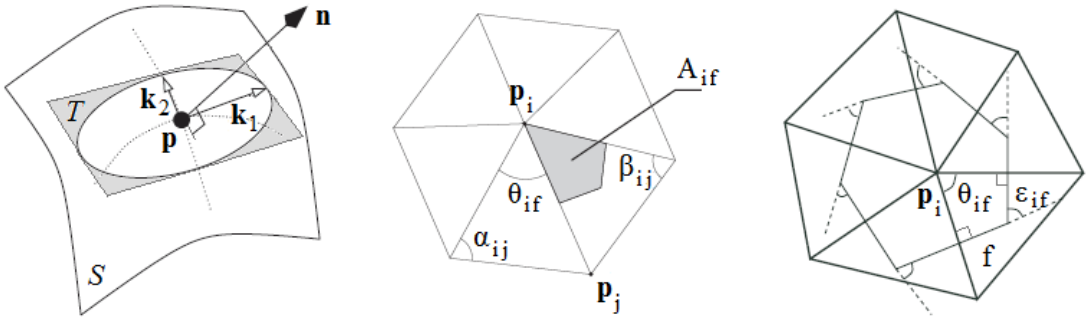
g. The Shape Index [34] can be computed from the min and max curvatures K_{min} and K_{max}, according to:

$$SI(\mathbf{p}) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{K_{max}(\mathbf{p}) + K_{min}(\mathbf{p})}{K_{max}(\mathbf{p}) - K_{min}(\mathbf{p})} \ , \tag{59}$$

and is a scalar field on the surface $S(\mathbf{p})$.

4.3.1 Curvature Computation

For the computation of curvature on manifold meshes several approaches have appeared in the literature. Desbrun *et al.* [31] introduced the Discrete Laplace-Beltrami Operator with cotagent weights and barycentric infinitesimal area. Meyer *et al.* [87] introduced the Discrete Laplace-Beltrami Operator with cotagent weights and Voronoi infinitesimal area. Belkin *et al.* [5, 6] introduced the Discrete Laplace-Beltrami Operator with Gaussian probability weights. Xu [139] gave a comparative study of the convergence of different discretizations of the Laplace-Beltrami Operator and Wardetzky [135] gave a comparative study of the properties of different discretizations of the Laplace-Beltrami Operator.



(a) (b) (c)

Figure 21: Surface local regions: (a) infinitesimal neighborhood on a surface patch; (b) 1-ring neighborhood of a mesh vertex; (c) Voronoi region of a mesh vertex.

For any vertex $\mathbf{p}_i$ on a manifold mesh we can compute the normal curvature vector $\mathbf{K}_N(\mathbf{p}_i)$ according to

$$\mathbf{K}_N(\mathbf{p}_i) = \frac{1}{A_i} \sum_{j=1}^{N_1(i)} \frac{1}{2} \left(\cot \alpha_{ij} + \cot \beta_{ij}\right) (\mathbf{p}_i - \mathbf{p}_j) \ , \tag{60}$$

where $\mathbf{p}_j$ are the vertices of the 1-ring around $\mathbf{p}_i$, $N_1(i)$ the number of 1-ring vertices, and α_{ij}, β_{ij} the two opposite angles to the edge $\mathbf{p}_i\mathbf{p}_j$ as depicted in Fig. 21 b. This discretization is according to Desbrun *et al.* [31] and is called *discrete Laplace-Beltrami operator with cotagent weights*.

The infinitesimal area A_i around $\mathbf{p}_i$ is computed according to

$$A_i = \sum_{f=1}^{N_1(i)} A_{if} \ , \tag{61}$$

where A_{if} denotes the barycentric subarea or the Voronoi subarea of a triangle f, and $N_1(i)$ the number of 1-ring triangles (facets) (Fig. 21 b). For the computation of A_{if} the barycentric area is used, hence

$$A_i = \frac{1}{3} A_{1i} \ , \tag{62}$$

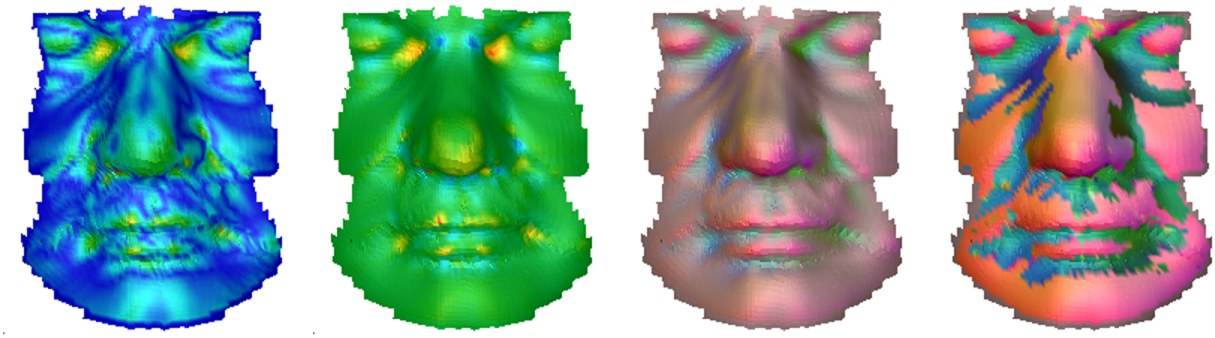
where A_{1i} is the whole 1-ring area around vertex $\mathbf{p}_i$. This discretization is according to Desbrun *et al.* [31].

The Gauss curvature $K_G(\mathbf{p}_i)$ at a vertex $\mathbf{p}_i$ is computed according to

$$K_G(\mathbf{p}_i) = \frac{1}{A_i} \left(2\pi - \sum_{f=1}^{N_1(i)} \varepsilon_{if} \right) \ , \tag{63}$$

where ε_{if} denotes the external angles of the Voronoi polygon, and $N_1(i)$ the number of 1-ring triangles (facets) (Fig. 21 c). For the computation of $K_G(\mathbf{p}_i)$ the Voronoi region around $\mathbf{p}_i$ is used, hence θ_{if} can substitute ε_{if}. This discretization is according to Meyer *et al.* [87], and is an application of the *Gauss-Bonnet theorem* [93, 7, 3].

4.3.2 Curvature Maps Representation



(a) (b) (c) (d)

Figure 22: Different types of curvature rendered as textures on the facial mesh: (a) mean curvature K_H; (b) Gauss curvature K_G; (c) normal curvature $\mathbf{K}_N$; and (d) unit normal curvature $\hat{\mathbf{K}}_N$.

The surface $S(\mathbf{p})$ can be represented as a *geometry image* map encoding in the three channels assigned to each pair of (u, v) coordinates, the (x, y, z) coordinates of the sampled points $\mathbf{p}$. Any vector field on the surface $S(\mathbf{p})$ can be represented (in analogy to the geometry image map) as an image map encoding in the three channels assigned to each pair of (u, v)

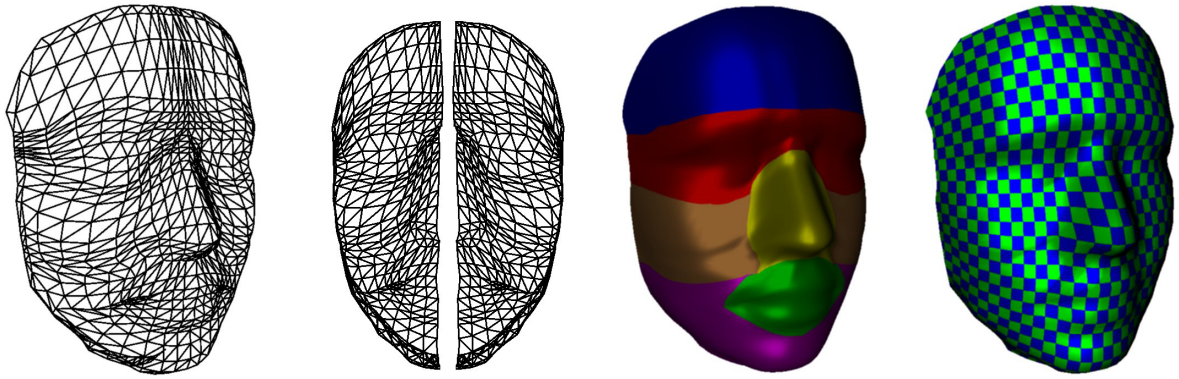
coordinates, the (x, y, z) components of the field. Any scalar field on the surface $S(\mathbf{p})$ can be represented (in analogy to the depth image map) as an image map encoding in the single channel assigned to each pair of (u, v) coordinates, the value of the field.

In Fig. 22 the different types of curvature are rendered as textures on the facial mesh. The mean curvature K_H and Gauss curvature K_G are rendered as single-channel image maps and the normal curvature $\mathbf{K}_N$ and unit normal curvature $\hat{\mathbf{K}}_N$ as three-channel image maps.

Note the differences between images in Figs. 22 d and 16 b. Although they both represent the normal vector, the normal image map has no discontinuities since the unit normal vector has always an outward direction. On the contrary the unit normal curvature map has discontinuities at the inflection points of the curvature vector.

4.4 The Annotated Face Model

In all steps of the proposed "Partial Face Recognition" method (registration, fitting and wavelet analysis), the *Annotated Face Model* (AFM) [64] is used. It is an anthropometrically correct 3D model of the human face [43]. The AFM needs to be constructed only once and consists of a triangular representation, a facial area annotation, annotated landmarks, and a (u, v) parameterization (see Fig. 23).



(a) (b) (c) (d)

Figure 23: Annotated Face Model (AFM): (a) full triangular mesh; (b) left & right triangular mesh; (b) annotated areas; (c) u, v parameterization.

The (u, v) parameterization allows the conversion of the AFM into the equivalent representation of *geometry image* [53]. Note that the (u, v) parameterization of the AFM offers an injective mapping from a roughly spherical surface in $\mathbb{R}^3$ to a plane in $\mathbb{R}^2$. This property is not violated even if its vertices are deformed, thus allowing the creation of a geometry image from a deformed AFM. Since the AFM is a topologically open model a simple cylindrical mapping technique was used to create the (u, v) parameterization. For a topologically closed and genus zero model (suitable for the full human head) Praun and Hoppe's octahedron-based parameterization [108] is more appropriate.

5 Landmarks and Features

We cannot all hope to combine the pleasing qualities
of good looks, brains, and eloquence.

– Homer

2D and 3D facial landmark detection is based on local descriptors of the 2D (intensity/color) or 3D (mesh/range) appearance of the face or of integral or differential transformations of it. Since a landmark detector has to possess the properties of repeatability and distinctiveness, local facial feature descriptors must be:
i) *robust*, to variations of facial data.
ii) *discriminative*, to distinguish between different anatomical landmarks.
iii) *descriptive*, to avoid similarity with outliers.
iv) *general*, to represent each landmark equally well on all "seen" faces.
v) *predictive*, to represent landmarks equally well on "unseen" faces.

To fulfill the above properties and constrain the detection process, landmark detectors use trained landmark classifiers or 2D/3D appearance landmark models/templates and 2D/3D geometry models for global topological consistency. 2D landmark detectors use view-based 2D geometry and appearance models or 3D geometry models. 3D landmark detectors use solely 3D geometry and 3D appearance models. Fused 2D/3D landmark detection methods use 3D geometry and 2D+3D appearance models. 2D and 3D landmark detection is based mostly on variations of the seminal work on Active Appearance Models of Cootes *et al.* [23, 27, 25, 28]. Fused 2D/3D landmark detection is presented in Boehnen & Russ [12], Jahanbin *et al.* [59], Lu & Jain [83], Passalis *et al.* [97] and Perakis *et al.* [103, 100].

A landmark detector, has four important levels (Fig. 24). At the *acquisition level* a sensor acquires the facial data. At the *feature extraction level* the data are transformed into features that represent the landmark classes. At the *matching score level* the extracted features are compared with feature templates that represent each landmark class in order to detect candidate landmarks with an associated matching score. Finally, at the *decision level* the matching scores (or ranks) are used to select a candidate landmark as the optimal solution for the queried landmark class, and assign to it the label of the class. Landmark detection can thus be considered as a two-fold problem: (i) a search problem for candidates, and (ii) an identification problem for the labeling of candidates.

This Chapter presents various feature descriptors that are used in this dissertation to represent facial landmarks. These include the Shape Index, the Spin Image, the Extruded Points and the Edge Response descriptors. It also introduces various feature fusion schemes for the combination of these descriptors into a more descriptive resultant feature descriptor.

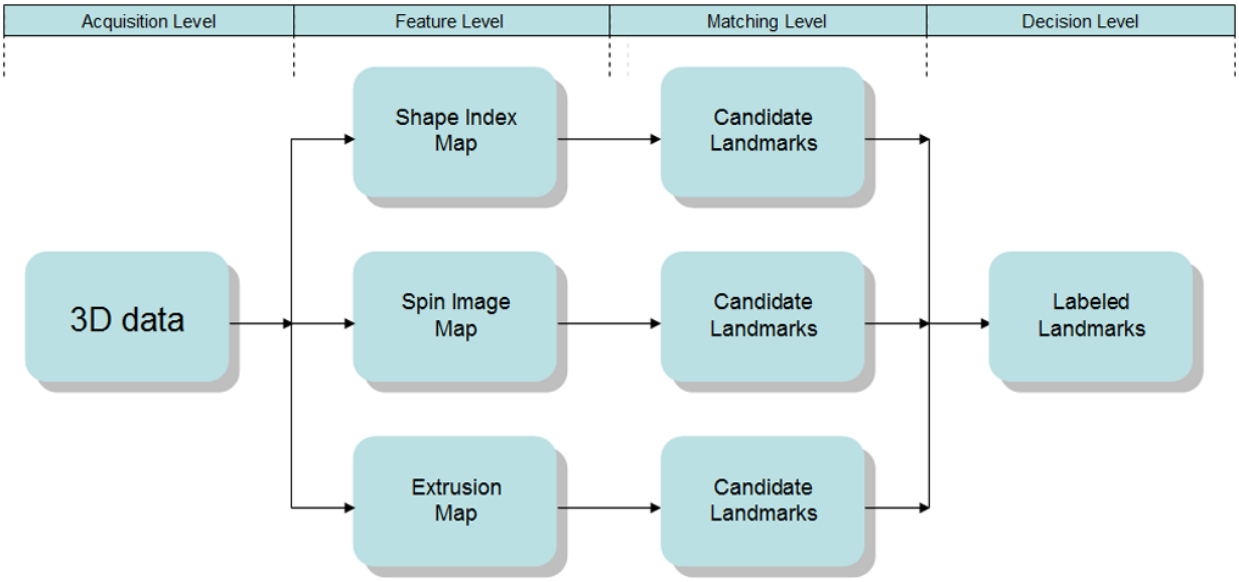


Figure 24: Pipeline of feature extraction for landmark detection.

5.1 Landmark Descriptors

To detect landmark points, three 3D local shape descriptors that exploit the 3D geometry-based information of the facial datasets and one 2D local appearance descriptor that exploits the 2D intensity-based information were used, depending on the case. The descriptors that are used are:
1) the *Shape Index* 3D descriptor (SI),
2) the *Spin Image* 3D descriptor (SS),
3) the *Extruded Points* 3D descriptor (EX) and
4) the *Edge Response* 2D descriptor (ER).

5.1.1 The Shape Index Descriptor

The *Shape Index* is extensively used for 3D landmark detection [20, 84, 85, 21, 83]. It is a continuous mapping of principal curvature values (K_{max}, K_{min}) of a 3D object point $\mathbf{p}$ into the interval [0,1], and is computed as:

$$SI(\mathbf{p}) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{K_{max}(\mathbf{p}) + K_{min}(\mathbf{p})}{K_{max}(\mathbf{p}) - K_{min}(\mathbf{p})} \ . \tag{64}$$

The Dorai and Jain definition is used here [34], an extension of Koenderink and van Doorn's original definition [72]. The shape index captures the intuitive notion of "local" shape of a surface. Every distinct surface shape corresponds to a unique value of shape index, except the planar shape. Points on a planar surface have an indeterminate shape index, since $K_{max} = K_{min} = 0$. Five well-known shape types and their locations on the shape index scale are as follows: Cup = 0.0, Rut = 0.25, Saddle = 0.5, Ridge = 0.75, and Cap = 1.0 (Fig. 25).

Shape index is computed from the principal curvature values of the surface spanned by the nearest neighbors of each vertex, a region of 5.5 mm radius on average.

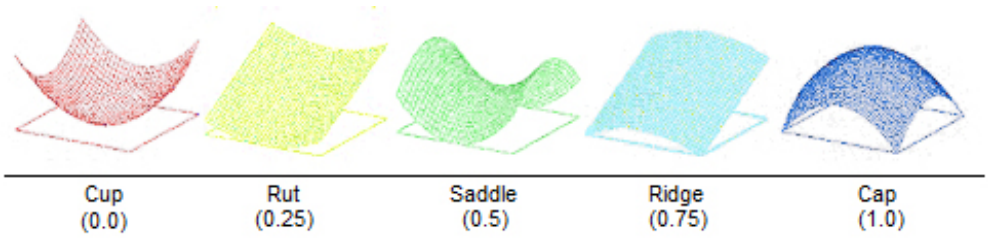
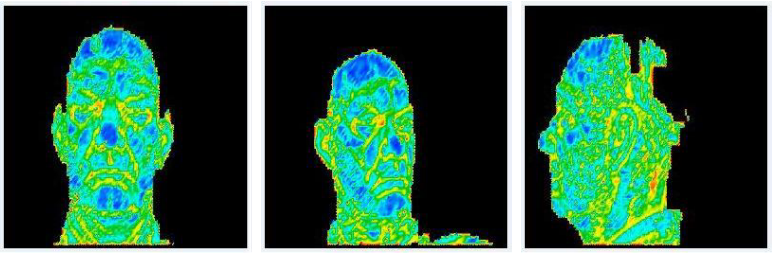


Figure 25: Depiction of Shape Index scale and corresponding "local" shape of a surface.

After computing the shape index values on a 3D facial dataset, a u, v mapping is performed, using the global u, v parameterization of the facial scan, in order to create a *shape index map* SI_{map}:

$$SI_{map}(u, v) \leftarrow SI(x, y, z) \ . \tag{65}$$



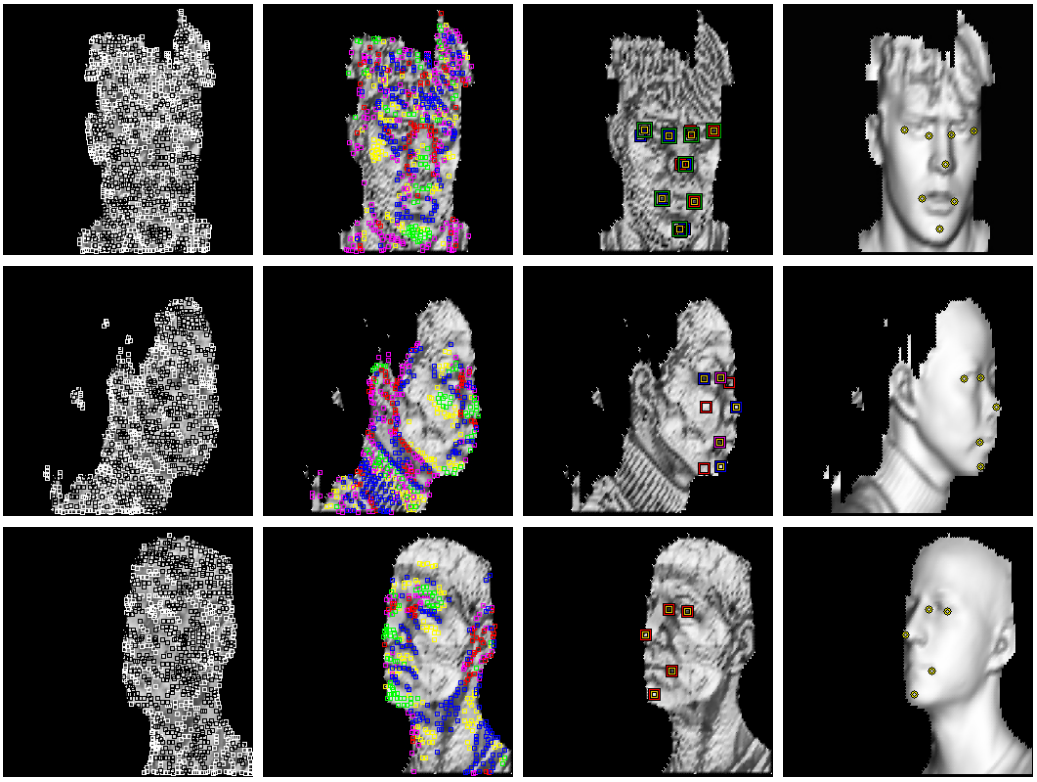
(a) (b) (c)

Figure 26: Depiction of shape index maps: (a) frontal face dataset; (b) 45° side face dataset; and (c) 60° side face dataset. (Blue denotes Caps, green Saddle, and red Cups.)

In a first approach (as published in [97, 101, 103]), local maxima ($SI_{map}(u, v) \rightarrow 1.0$) are candidate landmarks for nose tips and chin tips and local minima ($SI_{map}(u, v) \rightarrow 0.0$) for eye corners and mouth corners. This approach can be used in general when there are not trained target shape index values for each landmark class. Thus, local maxima and minima are detected on the shape index map (Fig. 26). The shape index's maxima and minima are sorted in descending order of significance according to their corresponding shape index values. The most significant subset of points for each group (Caps for nose and chin tips and Cups for eye and mouth corners) is retained (a maximum of 512 Caps and 512 Cups). In Fig. 29(a) and Fig. 27(a), black boxes represent Caps, and white boxes Cups.

In a second approach (as published in [100]), to locate interest points on the shape index map, we compute shape index target values that represent the landmarks used. Due to the symmetric nature of the face, shape index target values can represent only five landmark classes (without the distinction of left/right): the eye outer corner, eye inner corner, nose tip, mouth corner and chin tip landmarks. Shape index target values are statistically generated from 300 manually annotated frontal face scans of different subjects, from the FRGC v2 database, subset I (Fig. 51) with varying expressions. The shape index target values for each landmark class are obtained from the mode of the distribution of the shape index values of the associated landmark (Fig. 41f). These values are: 1.00 for nose tips, 0.90 for chin tips, 0.32 for mouth corners, 0.32 for eye outer corners and 0.16 for eye inner corners. The shape index candidate landmarks that are located for each class are kept in five lists sorted in descending order of significance according to their absolute difference from the

corresponding shape index target values. The most significant subset of points from each list is retained (a maximum of 1,024 points for each landmark class) (Fig. 32).



(a) (b) (c) (d)

Figure 27: Results of landmark detection and selection process using Shape Index + Spin Images [97, 101, 103]: (a) shape index's maxima and minima; (b) spin image classification; (c) extracted best landmark sets; and (d) resulting landmarks.

However, our experiments indicated that the shape index alone is not sufficiently robust for detecting landmarks on facial datasets in a variety of poses and expressions (the candidate landmarks are too many, having a large number of outliers that lead to false detections). Thus, candidate landmarks located from the shape index values serve as a basis, but are further classified and filtered out (Fig. 32).

5.1.2 The Extruded Points Descriptor

Experimentation indicated that the shape index is not sufficiently robust. For locating the nose and chin tips, the *extruded points descriptor* is proposed, a novel descriptor which is based on two common attributes of these two landmarks.

The first attribute is that they have large distances from the centroid of the face. To encode this feature the *radial map* (Fig. 28(a)) is introduced. The radial map is a 2D map that represents, at each u, v pixel, the distance of the corresponding (x, y, z) point from the centroid of the object, normalized to $[0, 1]$:

$$R_{map}(u, v) \leftarrow \|\mathbf{r}(x, y, z)\| \tag{66}$$

where $\mathbf{r}(x, y, z)$ is the radial vector.

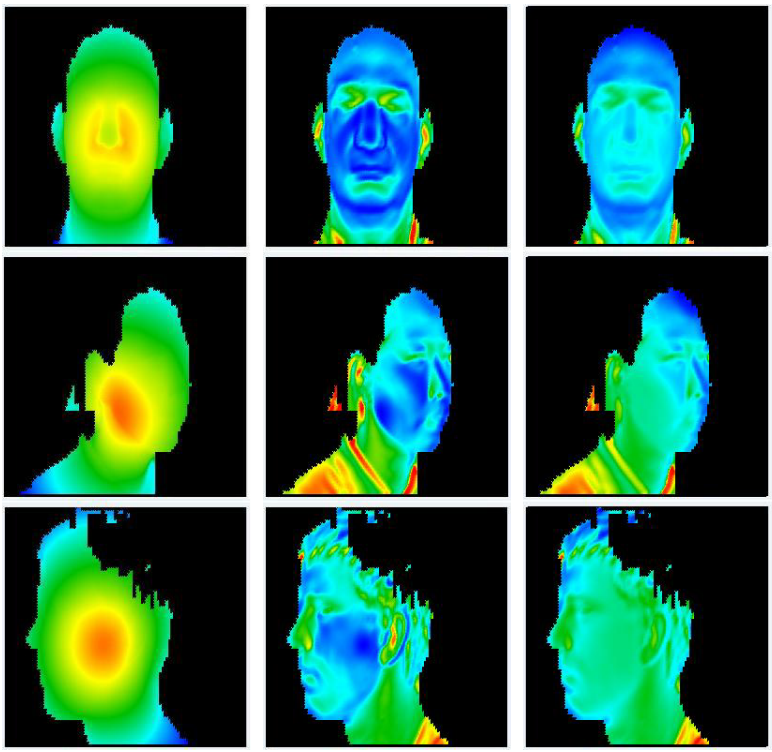
The second attribute is that most of the normals at nose and chin regions have an outward direction (with respect to the centroid). The *tangent map* (Fig. 28(b)) encodes this feature. It is a 2D map that represents, at each u, v pixel, the cosine value of the angle between the normal vector at the corresponding (x, y, z) point and the radial vector from the centroid of the object:

$$T_{map}(u, v) \leftarrow cos(\mathbf{r}(x, y, z), \mathbf{n}(x, y, z)) \tag{67}$$

Their product constitutes the *extrusion map* that represents the conjunction of the above two attributes, which is subsequently normalized to $[0, 1]$ (Fig. 28(c)):

$$E_{map}(u, v) = R_{map}(u, v) \odot T_{map}(u, v) \tag{68}$$

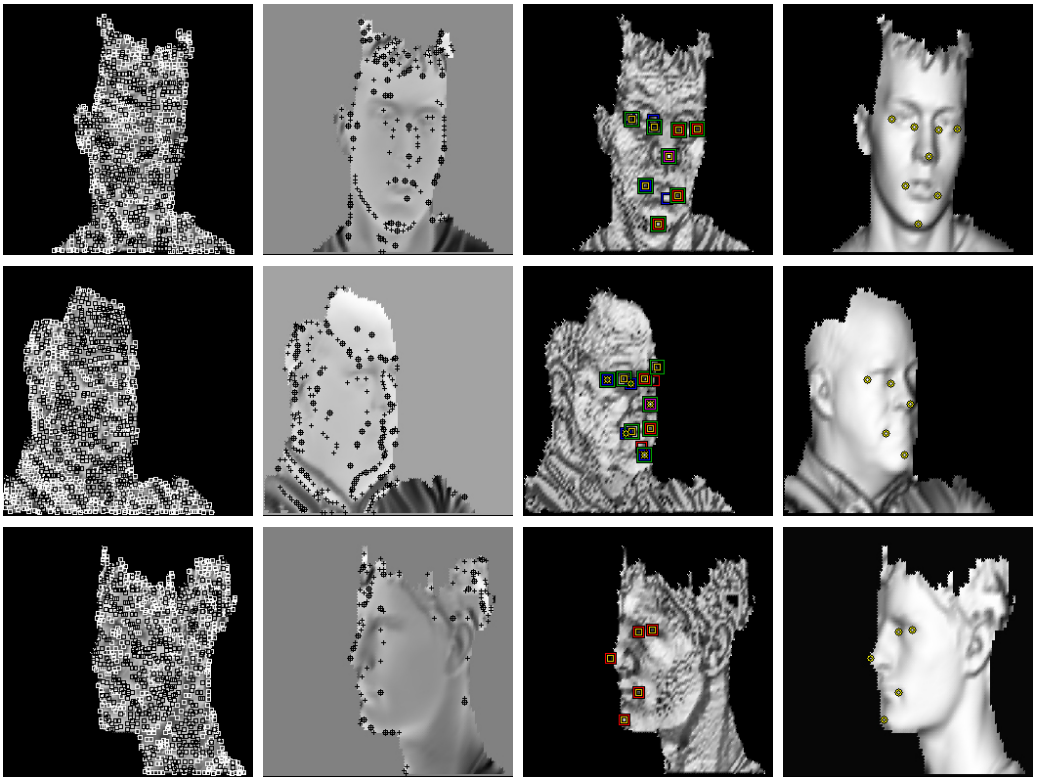
Since the extrusion map depends only on the position of the centroid, it can be considered pose invariant.



(a) (b) (c)

Figure 28: Depiction of extruded points: (a) radial map; (b) tangent map; and (c) extrusion map. (Blue denotes high values, and red low values.)

In this approach (as published in [101]), local maxima of the extrusion map ($E_{map}(u, v) \rightarrow 1.0$) that are also shape index maxima ($SI_{map}(u, v) \rightarrow 1.0$) are candidate landmarks for nose tips and chin tips. Located candidate nose and chin tips are sorted in descending order of significance according to their corresponding extrusion map values. The most significant subset of extruded points is retained (a maximum of 64 extruded points for nose and chin tips).



(a) (b) (c) (d)

Figure 29: Results of landmark detection and selection process using Shape Index + Extrusion Map [101]: (a) shape index's maxima and minima; (b) candidate nose and chin tips; (c) extracted best landmark sets; and (d) resulting landmarks.

By using the extrusion map, the number of candidate landmarks for nose and chin tips resulting from shape index's values alone are significantly decreased, and are more robustly localized. The shape index's minima are retained as candidate landmarks for eye and mouth corners (Fig. 29(a)) and extrusion map maxima are retained as candidate landmarks for the nose and chin tips (Fig. 29(b)). In Fig. 29(b), simple crosses represent extrusion map maxima and circled crosses represent extrusion map maxima that are also shape index's maxima: candidate nose and chin tips.

5.1.3 The Spin Image Descriptor

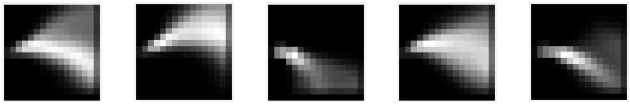
A *Spin Image* encodes the coordinates of points on the surface of a 3D object with respect to a local basis, a so-called *oriented point* [62]. An oriented point is the pair $(\mathbf{p}, \mathbf{n})$, where $\mathbf{n}$ is the normal vector at a point $\mathbf{p}$ of a 3D object. A spin image is a local descriptor of the global or local shape of the object, invariant under rigid transformations.

The spin image generation process can be visualized as a grid of bins spinning around the oriented point basis, accumulating points at each bin as it sweeps space. Therefore, a spin image at an oriented point $(\mathbf{p}, \mathbf{n})$ is a 2D grid accumulator of 3D points, as the grid is rotated around $\mathbf{n}$ by 360°.

Locality is expressed by the *Support Distance* parameter, which is:

$$\begin{aligned} (SupportDistance) &= (GridRows) \times (BinSize) \\ &= (GridColumns) \times (BinSize) \end{aligned}$$

A spin image at $(\mathbf{p}, \mathbf{n})$ is a signature of the shape of an object at the neighborhood of $\mathbf{p}$. For the purpose of representing facial features on 3D facial datasets, it was experimentally determined that a 16×16 spin image grid with 2 mm bin size should be used. This represents the local shape of the neighborhood of each landmark, spanned by a cylinder of 3.2 cm height and 3.2 cm radius.

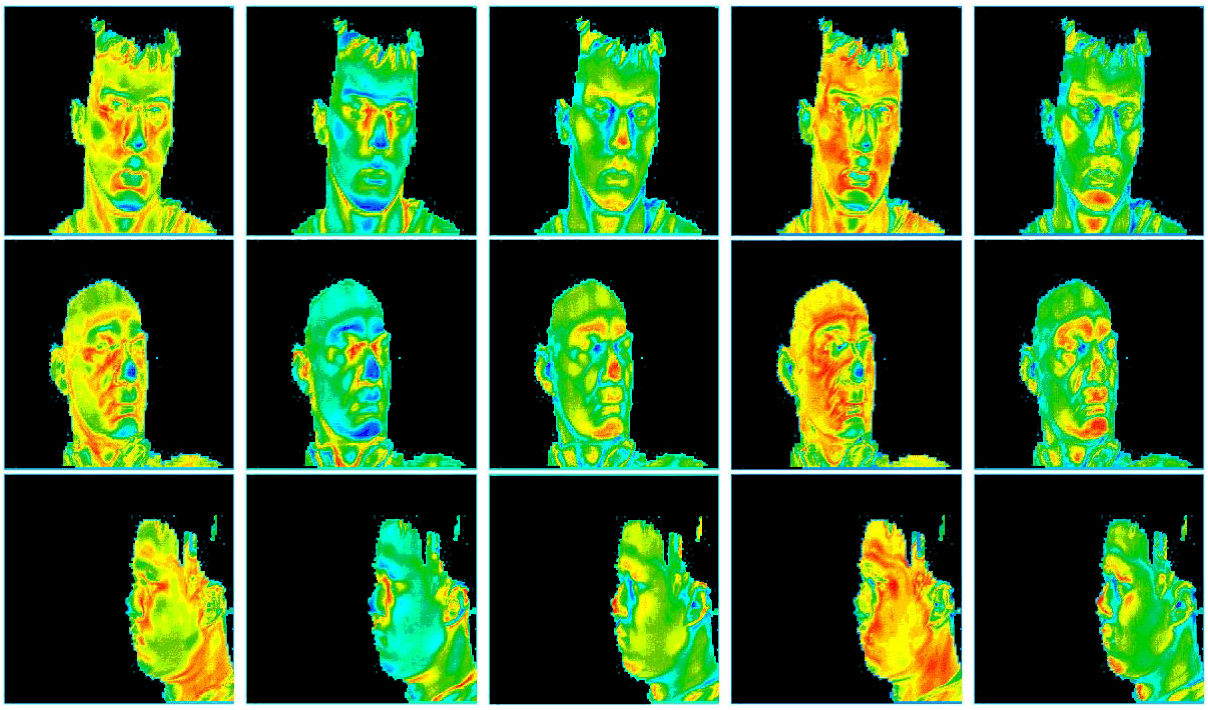


(a) (b) (c) (d) (e)

Figure 30: Depiction of spin image templates: (a) eye outer corner (EOC); (b) eye inner corner (EIC); (c) nose tip (NT); (d) mouth corner (MC); and (e) chin tip (CT).

In order to identify interest points on 3D facial datasets, spin image templates that represent the classes of the used landmarks are created. Due to the symmetric nature of the face, spin image templates can represent only five classes (without the distinction of left/right): the eye outer corner, eye inner corner, nose tip, mouth corner and chin tip landmarks.

Spin image templates are statistically generated from 300 manually annotated frontal face scans of different subjects, from the FRGC v2 database, subset I (Fig. 51) with varying expressions. They represent the mean spin images associated with the five classes of the landmarks (Fig. 30).



(a) (b) (c) (d) (e)

Figure 31: Depiction of spin image similarity maps: (a) eye outer corner; (b) eye inner corner; (c) nose tip; (d) mouth corner; and (e) chin tip. (Blue denotes low similarity values (-1), and red high similarity values $(+1)$.)

Landmark points can be identified according to a similarity measure of their spin images P with the five spin image templates Q that represent each landmark class. This *Spin Similarity* measure $SS(P,Q)$ is expressed by the normalized linear correlation coefficient:

$$SS(P,Q) = \frac{N\sum p_i q_i - \sum p_i \sum q_i}{\sqrt{[N\sum p_i^2 - (\sum p_i)^2]\,[N\sum q_i^2 - (\sum q_i)^2]}}\ , \tag{69}$$

where p_i, q_i denote each of the N elements of spin images P and Q, respectively [62].

Figure 31 depicts the *spin image similarity maps* of facial datasets for each spin image template (i.e., landmark class). It is a u,v mapping of the Spin Similarity measure $SS(P,Q)$ value between the spin image P of every facial dataset point and a spin image template Q:

$$SS_{map}(u,v) \leftarrow SS(P(x,y,z),Q)\ . \tag{70}$$

The *spin image similarity maps* SS_{map} (Fig. 31) provide an insight into the discriminating power of each spin image template. Spin image templates for the eye inner corner and the nose tip have the highest discriminating power, since high similarity areas are located at the expected facial regions, even though the nose tip template has some similarity with eyebrows and chin regions. The spin image template for the chin tip has a medium discriminating power, since it has similarity with eyebrows and nose regions. Finally, the spin image templates for the eye outer corner and the mouth corner have the lowest discriminating power, since there is high similarity between them, and also with other regions of the face, such as the cheeks and forehead. These error-prone regions can be filtered out by using the shape index values. This approach of using spin templates was used in the work published in [100, 97, 103].

(a) (b) (c) (d) (e)

Figure 32: Depiction of detected candidate landmarks on texture image (for viewing purposes only): (top) located landmarks according to similarity with shape index target values; and (bottom) filtered landmarks according to similarity with spin image templates: (a) eye outer corner; (b) eye inner corner; (c) nose tip; (d) mouth corner; and (e) chin tip.

Therefore, instead of searching all points of a facial dataset to determine the correspondence with the spin image templates, we use the shape index's candidate landmark points. Thus, the candidate landmark points of the five landmark classes (eye outer corner, eye inner corner, nose tip, mouth corner and chin tip) that are obtained from the shape index map are further filtered out according to the similarity $S(P,Q)$ of their spin images with the spin image templates representing each landmark class. These classified filtered landmarks are

sorted in descending order of significance according to their similarity measure with their corresponding spin image template and kept in five lists, one for each landmark class. The most significant subset from each list is retained (a maximum of 160 eye outer corners, 64 eye inner corners, 24 nose tips, 320 mouth corners and 128 chin tips). By using the spin images, the total number of candidate landmarks resulting from the shape index values are significantly decreased, and are more robustly localized (Fig. 32).

In Fig. 27(b), blue boxes represent the eye outer corner, red boxes the eye inner corner, green boxes the nose tip, purple boxes the mouth corner and yellow boxes the chin tip. Notice that some of the classified landmark boxes overlap due to similarity with different templates.

5.1.4 The Edge Response Descriptor

The *Edge Response* is based on the well known Harris corner and edge detector [56]. A response function $ER(u, v)$ encodes the intensity gradient of a point (u, v) on an image:

$$ER(u, v) = |I_x(u, v)| + |I_y(u, v)| \ , \tag{71}$$

where $I_x = \frac{\partial I}{\partial x}$ and $I_y = \frac{\partial I}{\partial y}$ denote the partial derivatives of the intensity image I in x and y respectively.

(a) (b) (c)

Figure 33: Depiction of edge response maps: (a) frontal face dataset; (b) 45° side face dataset; and (c) 60° side face dataset. (Blue denotes low edge response, and red denotes high edge response.)

$ER(u, v)$ is high in edge regions and close to zero in flat regions (Fig. 33). For computing $ER(u, v)$, the Sobel masks $\mathbf{F}_x$ and $\mathbf{F}_y$ are convolved with the intensity image for the calculation of I_x and I_y respectively [49], which are subsequently filtered by a Gaussian mask (7×7 pixels and $\sigma = 1.0$).

Sobel masks can be expressed in matrix form:

$$\mathbf{F}_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \tag{72}$$

$$\mathbf{F}_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \tag{73}$$

The Prewitt and Sobel masks are among the most used in practice for computing image gradients. However, Sobel masks have slightly superior noise-suppression characteristics, an important issue when dealing with derivatives [49].

The Edge Response (ER) descriptor is introduced for the evaluation of the feature fusion methods. To this end, the edge response descriptor is selected deliberately to represent a not so "good" descriptor, to exhibit the robustness of the fusion schemes. Edge Response is a "poor" descriptor for nose and chin tips (see Fig. 33).

5.2 Feature Fusion

Although many 2D/3D descriptors of facial features are used in the literature, a crucial issue has not been answered yet. How can these facial features be fused together in order to exploit their individual strengths and create a robust and accurate landmark detector?

Different feature descriptors can have complementary strengths and weaknesses, so combining them can increase system *accuracy*, *efficiency* and *robustness*, featuring *monotonicity*. Accuracy can be increased by exploiting data content from multiple sources (3D/2D) or the strengths of different data descriptors. In addition, using multiple descriptors can improve efficiency by limiting the landmarks' likelihood area. Finally, fusion can increase system robustness by limiting deficiencies inherent in using a single descriptor. For example a corner/edge detector is very sensitive in illumination variations, but the shape index is not. Thus, using multiple descriptors is a form of uncertainty reduction, since one descriptor may pick up what the other misses.

Fusion can be applied at the acquisition or feature extraction level (pre-classification fusion) and at the matching score or decision level (post-classification fusion) [61, 140]. Fusion at the matching score level can be viewed in two distinct ways. In the first, fusion is approached as a *classification* problem, while in the second, it is approached as a *combination* problem [61, 128]. In the classification approach, a composite feature vector (by weighted concatenation) is constructed using the values of the fused features, which is further classified by a composite classifier (e.g., Neural Network, k-NN, Decision Trees, SVM). In the combination approach, the matching scores of the fused features are combined to generate a single resultant feature score which is used for the final decision. The common characteristic of all combination techniques is that the individual feature classifiers are separately trained and the combination relies on simple fixed rules [128]. These rules are the *sum rule*, *product rule*, *max rule*, *min rule*, *median rule* and *majority voting* [70]. The various schemes for combining classifiers can be grouped into three main categories according to their architecture: (i) *parallel*, (ii) *cascading* (serial), and (iii) *hierarchical* (tree-like) [60].

An information fusion scheme should have the following fundamental properties, as described in [14]:

Neutrality: The result of a fusion scheme should not be biased by the order in which the input features are processed.

Consistency: The result of a fusion scheme with one input feature should be the same as the result of this single feature.

Monotonicity: The result of a fusion scheme of two input features should have better quality than the individual results of each feature.

Significance: The result of a fusion scheme should preserve the significance of the input feature measured values.

Conviviality: Expresses the complexity/simplicity of a fusion scheme.

Transparency: Expresses the ability to explain and replicate the result of a fusion scheme (black-box effect).

For landmark detection, although the construction of a composite feature classifier might be a potential solution, the combination method can be more easily applied to features whose values can be mapped to images, is more transparent (having also the strength of visualization), and possesses all the other fundamental properties required by a fusion scheme.

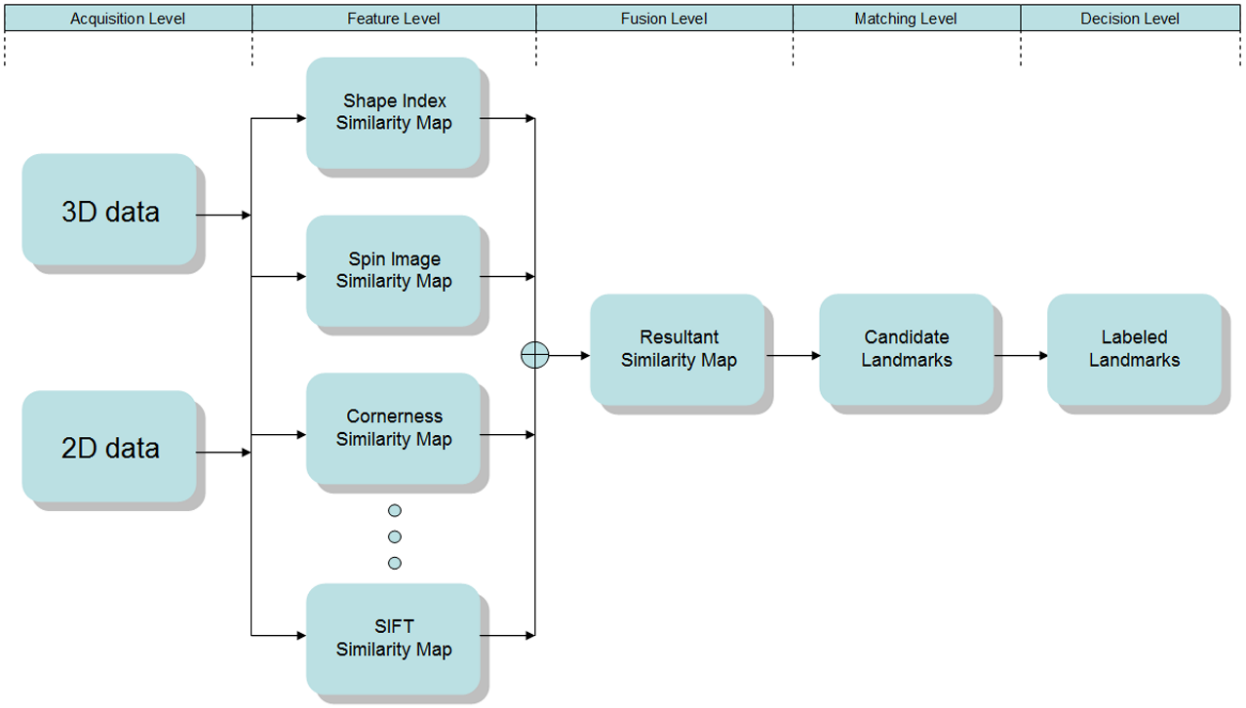


Figure 34: Pipeline of feature fusion procedure for landmark detection.

Feature fusion techniques have been proposed in the past (see Chapter 2), but in an entirely different context, that of multimodal biometrics or that of abstract feature fusion. The problem that is investigated in this dissertation is the behavior of fusion schemes under the strict context of landmark detection on facial datasets, which is an entirely different problem, since fusion techniques for landmark detection have to be also "locally consistent", which means that they have to boost results on a constrained area on facial surfaces; and this problem has not yet been investigated.

This dissertation provides a novel generalized framework of fusion methods and their application to landmark detection. This framework fills a gap in existing research, which is dominated by methods that use single landmark descriptors of 3D or 2D appearance of the face, without combining them (see Chapter 2).

The fusion scheme proposed acts after the "feature extraction level", transforms features to similarities and then combines them to generate a resultant feature similarity, which is considered as the matching score, and is used at the "matching level" for the detection of the queried landmarks (Fig. 34). The proposed approach of feature fusion is easily extensible by adding new feature-components in feature space and changing the resultant similarity appropriately. This approach works equally well for any feature extracted either from 3D or 2D facial data. The only prerequisite is the availability of a common (u,v) parameterization so that the 3D and 2D data can be combined at the "acquisition level".

The features used for facial landmark detection have very different characteristics, but in general can be distinguished in scalar features (such as the Shape Index and Cornerness/Edge Response), and vector features (1D/2D histogram features, such as the SIFT descriptor and Spin Images). For each scalar feature we can statistically compute a corresponding target value, while for each vector feature we can compute a corresponding vector target (template), which represent a landmark in feature space. A distance metric for a scalar feature could be the absolute difference of its value from the corresponding target value, and for a vector feature the absolute difference of its similarity with the corresponding template from the maximum similarity (1.00).

Thus, instead of fusing features by weighted concatenation, the features are first transformed to similarities with a target value or template, and then each feature similarity can act as a component in a normalized feature similarity space (Fig. 35), which can be fused together to form a resultant feature similarity, using simple combination rules (such as sum, product, max, min, AND, OR and threshold masking). In this manner a dramatic dimensionality reduction is achieved since, instead of using multiple components for a vector feature, only the similarity with its template is used.

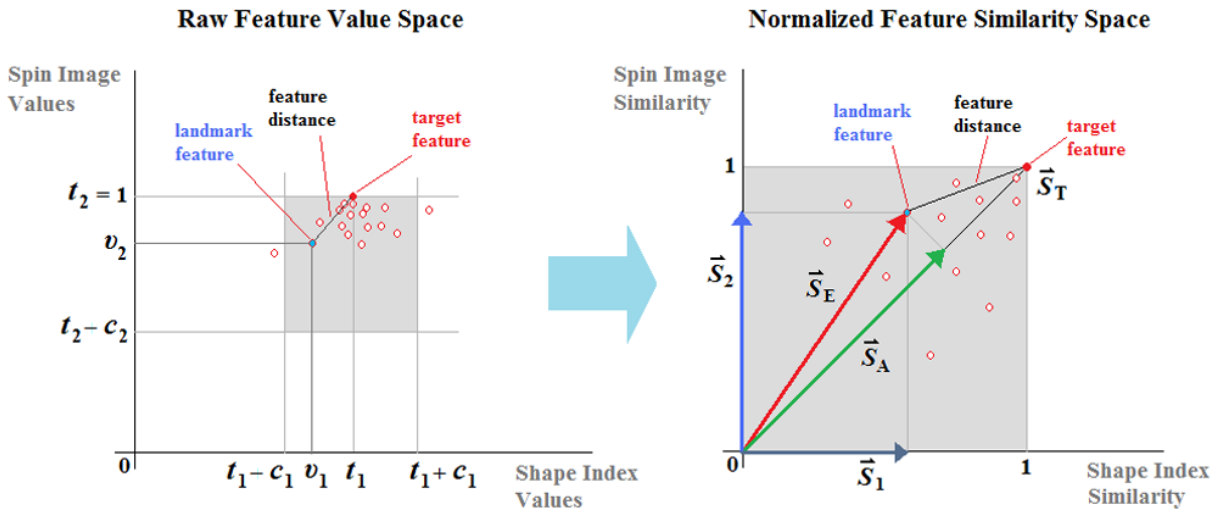


Figure 35: Example of the transformation from raw feature value space to normalized feature similarity space. Shape Index (v_1) and Spin Image (v_2) raw values are mapped onto Shape Index ($\mathbf{S_1}$) and Spin Image ($\mathbf{S_2}$) normalized similarity vectors. Note that the raw Spin Image values represent un-normalized similarity to the corresponding template.

Each feature for a landmark class has a target value or template (t_f) that describes the landmark in its feature space. Furthermore, we can consider a cut-off value (c_f) for each feature to incorporate the notion of an outlier. Feature values out of the range $[t_f - c_f, t_f + c_f]$ can be filtered out, so that threshold masking is implemented. The cut-off value can also be considered as a scaling factor for the normalization of each feature's range (Fig. 35).

The target and cut-off values can be estimated by examining the probability density function (pdf) of feature values or set to specific values based on a priori knowledge. A good choice for the target value could be the mean of the pdf of feature values and for the cut-off value could be a multiple of standard deviation (std) (e.g., $3 \times$ std as a first approximation), although the distribution of the values of every feature is not a Gaussian. Another choice

for the target value could be the mode or the median of the pdf and the cut-off value could be determined so that a certain proportion of feature values (e.g., 99%) are within the range $[t_f - c_f, t_f + c_f]$.

For a good normalization scheme, the estimates of target (location), cut-off (scale) parameters and of the normalization function must be robust and efficient and has to closely simulate the initial pdfs. In addition, a properly designed fusion method exploits information from each descriptor without degrading performance below that of the most accurate descriptor (monotonicity). This is the major challenge of adopting a fusion scheme.

5.2.1 Feature similarity mapping

Given a feature value v_f, a target value t_f and a cut-off value c_f for each feature descriptor f, a *normalized distance measure* to target D_f for each of the N feature descriptors of each landmark point is introduced:

$$D_f = \begin{cases} \frac{|v_f - t_f|}{c_f} & \text{if } |v_f - t_f| \leq c_f \\ 1 & \text{otherwise} \end{cases} \tag{74}$$

Note that the above definition is a generalization of the z-score normalization and median normalization [61].

A *normalized similarity measure* to target S_f can be derived from D_f as:

a. Linear mapping:

$$S_f = 1 - D_f \ . \tag{75}$$

This is the classic linear distance to similarity transformation [128].

b. Quadratic mapping:

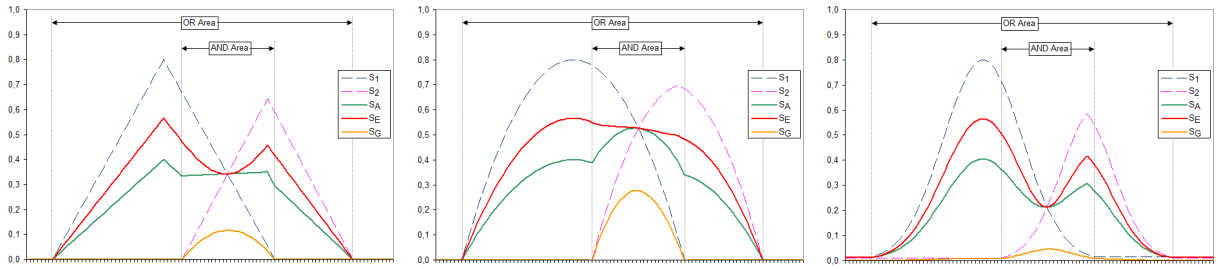
$$S_f = 1 - D_f^2 \ . \tag{76}$$

We introduce quadratic mapping, which favors close to target feature values. Note that D_f^2 behaves like the potential energy of elasticity.

c. Gaussian mapping:

$$S_f = exp(-\alpha D_f^2) \ , \tag{77}$$

where α is the drop-off parameter. We introduce Gaussian mapping, for smoothing out large distance measures. Note that the Gaussian tails can be cut at the cut-off values.



(a) (b) (c)

Figure 36: Depiction of fusion of similarities: (a) after linear mapping; (b) after quadratic mapping; and (c) after Gaussian mapping.

Remarks:

a. If the target value is the mean of the feature values and the cut-off is its standard deviation then Eq. 74 becomes

$$D_f = \frac{|v_f - \mu_f|}{\sigma_f} \ , \tag{78}$$

which is similar to the z-score normalization of feature values used in [61].

b. If the target value is the median of the feature values and the cut-off is its median absolute deviation (MAD) then Eq. 74 becomes

$$D_f = \frac{|v_f - median_f|}{median\left(|v_f - median_f|\right)} \ , \tag{79}$$

which is similar to the median normalization of feature values used in [61].

c. $D_f(c_f)$ is a decreasing function of c_f and $S_f(c_f)$ is an increasing function of c_f. As c_f increases, f-axis shrinks and similarity values approach maximum similarity (1.00 or w_f), on the contrary as c_f decreases f-axis dilates and similarity values deviate from maximum similarity (1.00 or w_f).

5.2.2 Feature similarity fusion

The resultant similarity measure to the target vector in the normalized similarity space describes the way by which the N feature descriptors can be fused together or combined into a resultant feature similarity for each queried landmark class:

a. Sum rule:

$$S_A = \frac{1}{N} \sum_{f=1}^{N} S_f \ , \tag{80}$$

which is the arithmetic mean or the Manhattan (L_1) metric (Fig. 35). Note that if the similarity measure is considered as the probability that the sample point is similar to the target, then this metric is equivalent to the *sum rule* for feature fusion [70, 128].

b. Root-mean-square rule:

$$S_E = \frac{1}{\sqrt{N}} \left(\sum_{f=1}^{N} S_f^2 \right)^{\frac{1}{2}} \ , \tag{81}$$

which is the root mean square (rms) of the similarities and actually a Euclidean (L_2) metric in the resultant similarity space. We introduce this novel *rms rule* so that feature similarities to targets can be considered as vectors and added according to vector addition (Fig. 35).

c. Product rule:

$$S_G = \left(\prod_{f=1}^{N} S_f \right)^{\frac{1}{N}} \ , \tag{82}$$

which is the geometric mean metric. Note that if the similarity measure is considered as the probability that the sample point is similar to the target, then this metric is equivalent to the *product rule* for feature fusion [70, 128].

d. Max rule:

$$S_{max} = \max_{f=1}^{N} \left(S_f \right) \ , \tag{83}$$

which is the L_∞ metric or *max rule* [70] and favors the feature with maximum similarity. Note that if the similarity measure is considered as a fuzzy variable, then this metric is equivalent to a fuzzy *OR rule* for feature fusion [128].

e. Min rule:

$$S_{min} = \min_{f=1}^{N} (S_f) \ , \tag{84}$$

which is the *min rule* [70] and favors the feature with minimum similarity. Note that if the similarity measure is considered as a fuzzy variable, then this metric is equivalent to a fuzzy *AND rule* for feature fusion [128].

Remarks:

a. If linear mapping and arithmetic mean is used, then the overall similarity measure is consistent with the overall distance measure.
$S_A = \frac{1}{N} \sum_{f=1}^{N} S_f$ and $S_f = 1 - D_f$, then
$S_A = \frac{1}{N} \sum_{f=1}^{N} (1 - D_f) \Rightarrow S_A = \frac{N}{N} - \frac{1}{N} \sum_{f=1}^{N} D_f \Rightarrow S_A = 1 - D_A$.

b. The S_A resultant similarity (L_1 metric) is equivalent to the normalized projection of the S_E similarity vector (L_2 metric) onto the target similarity vector S_T (Fig. 35) (i.e. it is a normalized inner product metric, or the *cosine similarity measure* [128].
$\frac{\overrightarrow{S_E}}{\sqrt{N}} \cdot \frac{\overrightarrow{S_T}}{\sqrt{N}} = \frac{1}{N} \sum_{f=1}^{N} S_f \cdot 1 = S_A$.

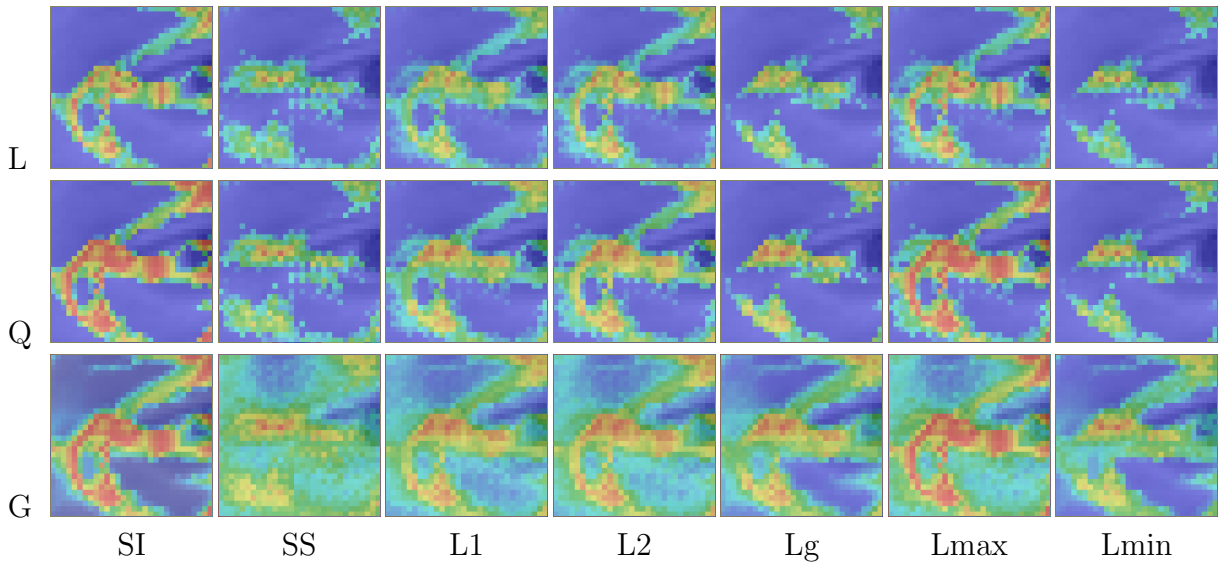


Figure 37: Depiction of the 2D similarity maps in the neighborhood of the Eye Outer Corner (EOC) for the various distance to similarity mappings and the various fusion methods: (blue) low similarity values (0.0); (green) medium similarity values (0.5); and (red) high similarity values (1.0). Rows depict: (top) L mapping; (middle) Q mapping; and (bottom) G mapping. Columns depict from left to right: SI similarity; SS similarity; L1 fusion; L2 fusion; Lg fusion; Lmax fusion; and Lmin fusion.

To illustrate the behavior of the proposed distance to similarity mappings and fusion schemes we depict the various combinations in Fig. 36. For simplicity the fusion of similarity

mapping functions is presented in a single dimension. We also depict in Fig. 37 the behavior of the proposed distance to similarity mappings and fusion schemes in the neighborhood of the Eye Outer Corner (EOC).

Remarks:

a. Linear mapping raises discontinuities in the superposed similarities. Due to the discontinuous behavior, linear mapping is expected to give unreliable results.

b. The "smoothest" results are given by the Gaussian and the Quadratic mapping. This behavior allows a "locally smoother" combination of the features that are fused.

c. S_G and S_{min} give results in the "AND Area" and S_A, S_E and S_{max} give results in the "OR Area". The "AND area" is more restricted and can be used as an "AND masking area" to restrict the search space of candidate landmarks. The "OR area" is wider, which has the implication of a larger number of candidate landmarks to be detected, raising the "curse of dimensionality" at the decision level.

d. S_G and S_{min} give almost the same peak, approximately in the middle of the initial peaks of the fused features, having a similar behavior to an "AND operator". This peak is "smoother" for S_G and "sharper" for S_{min}.

e. S_{max} gives the same peaks as the initial peaks of the fused features, having a behavior similar to an "OR operator".

f. S_{max} gives as a result the similarity of the most "intensive" feature. Selecting the most "intensive" feature is unreliable, because it could be the one that makes the largest errors.

g. S_{min} gives as a result the similarity of the least "intensive" feature (especially when feature areas overlap), and is not appropriate for landmark fusion, because it doesn't take into consideration the other features' similarities.

h. S_G and S_{min} may completely eliminate a feature's similarity peak which is not inside the "AND masking area", and thus are not appropriate for landmark fusion.

5.2.3 Weighted metrics

With the above metrics each feature contributes equally to the resultant similarity. Extended similarity metrics with weights per feature can also be considered:
a. Sum rule:

$$S_A = \frac{1}{W}\sum_{f=1}^{N} w_f S_f \ , \ W = \sum_{f=1}^{N} w_f \ . \tag{85}$$

b. Root-mean-square rule:

$$S_E = \frac{1}{\sqrt{W}}\left(\sum_{f=1}^{N} w_f S_f\right)^{\frac{1}{2}} \ , \ W = \sum_{f=1}^{N} w_f \ . \tag{86}$$

c. Product rule:

$$S_G = \left(\frac{1}{W}\prod_{f=1}^{N} w_f S_f\right)^{\frac{1}{N}} \ , \ W = \max_{f=1}^{N}(w_f) \ . \tag{87}$$

d. Max rule:

$$S_{max} = \frac{1}{W} \max_{f=1}^{N} (w_f S_f) \ , \ W = \max_{f=1}^{N} (w_f) \ . \tag{88}$$

e. Min rule:

$$S_{min} = \frac{1}{W} \min_{f=1}^{N} (w_f S_f) \ , \ W = \max_{f=1}^{N} (w_f) \ . \tag{89}$$

Remarks: The weights w_f act as scaling factors on the feature similarity components, and can take values $[0.0, 1.0]$. They actually correspond to the maximum similarity value a feature can take, which, as a first approximation, is proportional to the reliability of a feature with respect to other features.

5.2.4 Training of the descriptors

To train the landmark descriptors we used 300 frontal facial datasets of different subjects, manually annotated at the specific landmark positions. These datasets come from FRGC v2 database [107, 106] and contain subjects with varying expressions and illumination conditions. The available 3D scans were used to train the shape index and spin image descriptors and the corresponding 2D texture images to train the edge response descriptor. The exact datasets that were used from the source databases for training (DB_TRAIN) can be found from the landmark annotation files available through the website [132].

Table 4: Target (t) and cut-off (c) values of the landmark descriptors for each landmark class

	EOC		EIC		NT		MC		CT	
	t	c	t	c	t	c	t	c	t	c
SI	0.32	0.53	0.12	0.60	1.00	0.40	0.09	0.68	0.96	0.70
SS	1.00	0.48	1.00	0.80	1.00	0.75	1.00	0.72	1.00	0.56
ER	0.20	0.72	0.16	0.62	0.10	0.40	0.22	0.70	0.02	0.17

The pdf of the shape index values (SI) and edge response values (ER) for each landmark class were computed and used for the estimation of the shape index and edge response target and cut-off values. We computed spin image templates for each landmark class. Spin image templates represent the mean spin image associated with the five classes of landmarks (Fig. 30). The pdfs of the similarity values (SS) between the pre-computed spin image templates and the spin images of each landmark class, were computed for the estimation of the cut-off values. The spin image target values are set to the maximum similarity (1.00).

The estimated target and cut-off values for each descriptor (SI, SS, ER) and for each landmark class (EOC, EIC, NT, MC, CT) are presented in Table 4, and the correlation coefficients between the landmark descriptors for each landmark class are presented in Table 5. Note that the introduction of distance to similarity mappings improves the correlation coefficients in comparison to the raw values.

Table 5: Correlation coefficients between landmark descriptors for each landmark class

	EOC	EIC	NT	MC	CT
Raw values					
SI / SS	0.0358	−0.1242	0.3202	−0.1823	0.1925
SI / ER	0.1458	0.0024	−0.0895	0.0000	0.0001
SS / ER	−0.0377	−0.1358	−0.1794	−0.2481	−0.0075
Linear mapping similarity values (L)					
SI / SS	0.1781	0.1806	0.3202	0.2669	0.2290
SI / ER	0.1665	0.0360	0.0638	0.1354	−0.0265
SS / ER	0.1080	0.0813	0.1002	0.1991	−0.0013
Quadratic mapping similarity values (Q)					
SI / SS	0.2095	0.1965	0.3098	0.2366	0.5241
SI / ER	0.1968	−0.0101	0.0572	0.0543	−0.0222
SS / ER	0.1184	0.0907	0.0370	0.1849	−0.0093
Gaussian mapping similarity values (G)					
SI / SS	0.2084	0.1921	0.3170	0.2508	0.3459
SI / ER	0.2023	0.0003	0.0524	0.0882	−0.0241
SS / ER	0.1205	0.0989	0.0614	0.2052	−0.0018

5.2.5 Similarity mapping and fusion paradigms

To illustrate the characteristics of the proposed distance to similarity mappings and the fusion schemes we apply them for the detection of specific facial anatomical landmarks.
a. The landmark classes are:
1) the Eye Outer Corner (EOC)
2) the Eye Inner Corner (EIC)
3) the Nose Tip (NT)
4) the Mouth Corner (MC), and
5) the Chin Tip (CT).
b. The descriptors that are used are:
1) the Shape Index (SI)
2) the Spin Image (SS), and
3) the Edge Response (ER).
c. The distance to similarity mappings are:
1) the linear mapping (L)
2) the quadratic mapping (Q), and
3) the Gaussian mapping (G).
d. The fusion schemes are:
1) the sum rule using the arithmetic mean S_A (L1)
2) the rms rule using the Euclidean mean S_E (L2)
3) the product rule using the geometric mean S_G (Lg)
4) the max rule using S_{max} (Lmax)
5) the min rule using S_{min} (Lmin).

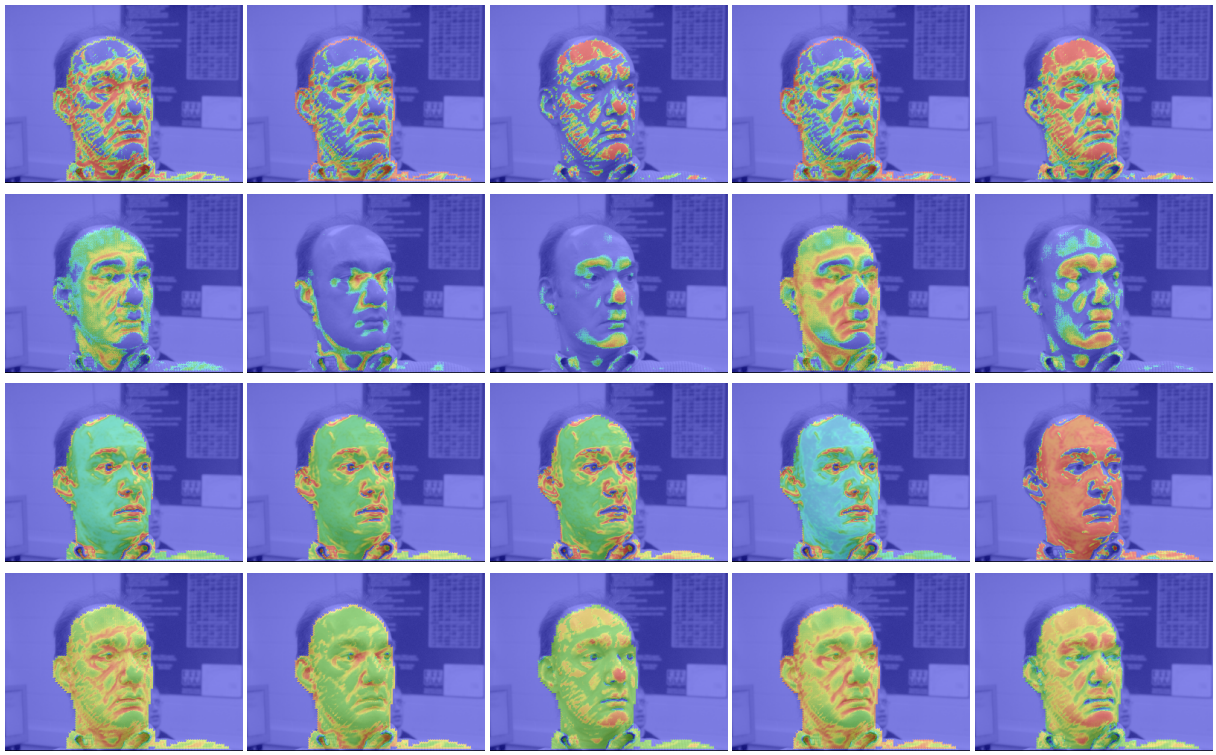


Figure 38: Depiction of feature similarity maps with Q−L2 fusion: (blue) low similarity values (0.0); (green) medium similarity values (0.5); and (red) high similarity values (1.0). (1^{st} row) SI similarity; (2^{nd} row) SS similarity; (3^{rd} row) ER similarity; and (4^{th} row) Q−L2 resultant similarity. (a) eye outer corner; (b) eye inner corner; (c) nose tip; (d) mouth corner; and (e) chin tip.

The resultant similarity of applying the rms rule for fusion after a quadratic mapping of the feature similarities (Q−L2 scheme) for the five landmark classes (eye outer corner, eye inner corner, nose tip, mouth corner, and chin tip) is depicted in Fig. 38.

The resultant similarity maps encode the likelihood of facial regions to represent the various landmark classes. Robustness and efficiency of a fusion scheme is evaluated according to the position and the size of the likelihood areas of each landmark class. The likelihood area of a landmark class is very important since its reduction means that fewer candidate landmarks have to be retained and fed to the "selection level" for labeling.

It is obvious that fusion reduces the likelihood areas of each individual feature descriptor and focus them on facial regions where they are expected to be found. Using a feature fusion scheme, error-prone regions that are resulted from a single feature descriptor are finally eliminated.

For instance (see Fig. 38), the error-prone regions resulted from shape index for eye outer corner mainly around the ear are finally eliminated. For eye inner corner, error-prone regions resulted from shape index around the eye outer corners are eliminated. Error prone regions for the nose tip resulted from shape index, mainly on the forehead, are finally eliminated. For the mouth corner, error-prone regions resulted from spin image, mainly around the eyes and the chicks, are finally reduced. Error-prone regions from the shape index around the nose tip are finally reduced to that of the chin tip.

Therefore, instead of searching landmark points on individual feature maps, the resultant

similarity map is used. Landmarks are detected as the the points with maximum likelihood on the resultant similarity map of each landmark class. Under this fashion the landmark detection problem is reduced to a search problem on a 2D discrete scalar field of values.

6 Landmark Detection

Law is order, and good law is good order.

– Aristotle

The proposed facial landmark detection method is based on local descriptors of the 2D (intensity/color) or 3D (mesh/range) appearance of the face. To apply a generalized framework for landmark detection, 2D, 3D or combined 2D/3D feature information is mapped onto 2D Cartesian grids, exploiting the (u, v) parameterization of the facial data.

Thus, the problem of detecting landmarks on facial datasets is converted to detecting landmarks on 2D grids or bitmaps. Landmark points are detected on a 2D map using general methods of locating extremum values. These detected mathematical landmarks do not necessarily represent the queried anatomical landmarks.

For this purpose, topological properties of faces are taken into consideration to ensure global topological consistency, and candidate mathematical landmarks, are filtered out and labeled by requiring consistency with the FLMs. Thus landmark detection is considered as a two-fold problem: (i) a search problem for candidates, and (ii) a classification problem for the labeling of candidates.

This Chapter presents the proposed landmark detection methods in detail, using the combination of the landmark feature models for landmark detection and the geometric landmark models (FLMs) for landmark consistency. Landmark detection is a key requirement for generic face recognition, calculating the coarse transformation for registration, and transforming a test scan into a canonical AFM.

6.1 Locating Landmarks on 2D Maps

To locate the most significant landmark points on a 2D map, general methods of locating extreme values are used. First, all 2D maps are normalized by linear stretching to [0,1] so that the problem of locating maximum or minimum is reduced to locating a single target value (i.e., 1 or 0). Then, if a 2D map is represented by its normalized values $I(u, v)$ and a target value V is searched within it, we can consider the function $|I(u, v) - V|$ as a transformation of the 2D map and search for its minimum values.

The localization of target values on a 2D map is implemented in Algorithm 6. This algorithm computes the value $|I(u, v) - V|$ and tests if it is within certain accepted variation limits t in order to reject unwanted values (outliers). Then it tests whether $|I(u, v) - V|$ is a local minimum within a window of neighbors by suppressing non minimum candidate points (hill climbing scheme). Finally, it tests wether the target value is a majority value (within some limits $|I(u, v) - V| \leq t$) in a window of neighbors (voting scheme). Thus, a list of

Algorithm 6 "Landmark Localization"

input: 2D map $I(u, v)$ and target value V.
output: List of landmark points.

```
for each point (u, v) do
    Compute |I(u, v) − V|.
    if |I(u, v) − V| ≤ t then
        if |I(u, v) − V| is a minimum in a window of neighbors then
            if |I(u, v) − V| is a majority value in a window of neighbors then
                add point in a descending ordered list of points according to: |I(u, v) − V|.
            end if
        end if
    end if
end for
return List of points.
```

candidate landmark points is returned, sorted in descending order of significance, according to the distance from target value $|I(u, v) - V|$.

6.2 Landmark Labeling & Selection

As mentioned in Section 3.2.2, detected geometric landmarks must be identified and labeled as anatomical landmarks. For this purpose, topological properties of faces must be taken into consideration. Thus, candidate geometric landmarks, irrespective of how they are generated, must be consistent with the FLMs. This is accomplished by applying the fitting procedure described in Section 3.2.2. The procedure for landmark detection, landmark labeling and registration for each facial dataset, is described in Algorithm 7.

In Fig. 29(c) and Fig. 27(c), blue boxes represent landmark sets consistent with the FLM5R, red boxes with the FLM5L, green boxes with the FLM8, and yellow boxes the best landmark set. Notice that some of the consistent landmarks overlap. Also note that the FLM8 consistent landmark set is not always the best solution; FLM5L and FLM5R are usually better solutions for side facial datasets (Fig. 29(d) and Fig. 27(d)).

The consistent landmark sets determine the pose of the face object under consideration from the alignment transformation with the corresponding FLM. Since the aim is to locate landmark sets on profile, semi-profile and profile faces, we retain the complete landmark solution only if estimated yaw-angle is within certain limits ($\pm 30^\circ$ around y-axis), otherwise the left or right landmark sets are preferred according to pose.

Finally, using the selected best solution, the registration transformation is calculated, the yaw-angle is estimated, and the facial dataset is classified as frontal, left side or right side.

Remarks:
Note that the use of candidate landmark sets with five landmarks has a dual purpose: (i) it is the potential solution for semi-profile and profile faces, and (ii) it reduces the combinatorial search space for creating the complete landmark sets in a divide-and-conquer manner. Instead of creating 8-tuples of landmarks out of N candidates, which generates N^8 combinations to be checked for consistency with FLM8, we create 5-tuples of landmarks, and check $N^5 + N^5 = 2N^5$ combinations for consistency with FLM5L and FLM5R. We retain 512

Algorithm 7 "Landmark Labeling & Selection"

1: Extract candidate landmarks from the geometric/appearance properties of the facial scans (Algorithm 6).
2: Create feasible combinations of 5 landmarks from the candidate landmark points, by using landmark constraints.
3: Compute the rigid transformation that best aligns the combinations of five candidate landmarks with the FLM5R and FLM5L (Algorithm 2).
4: Filter out those combinations that are not consistent with FLM5L or FLM5R, by applying the fitting procedure (Algorithm 4).
5: Sort consistent right (FLM5R) and left (FLM5L) landmark sets in descending order according to a distance metric from the corresponding FLM.
6: Fuse accepted combinations of 5 landmarks (left and right) in complete landmark sets of 8 landmarks.
7: Compute the rigid transformation that best aligns the combinations of eight landmarks with the FLM8 (Algorithm 2).
8: Discard combinations of landmarks that are not consistent with the FLM8, by applying the fitting procedure (Algorithm 4).
9: Sort consistent complete landmark sets in descending order according to a distance metric from the FLM8.
10: Select the best combination of landmarks (consistent with FLM5R, FLM5R or FLM8) based on the distance metric to the corresponding FLM.
11: Return the corresponding rigid transformation for registration (Algorithm 8).

landmark sets consistent with FLM5L and 512 landmark sets consistent with FLM5R. By fusing them and checking consistency with FLM8 we obtain an extra 512×512 combinations to be checked. Thus, by this approach $2N^5 + 512^2 \ll N^8$ combinations are checked, with $O(N^5) \ll O(N^8)$. For $N = 128$ we obtain approx. 69×10^9 instead of 72×10^{15} combinations to be checked.

6.2.1 Landmark Constraints

As previously mentioned, from the candidate landmark points we create combinations of five landmarks, one from each class. Since an exhaustive search of all possible combinations of the candidate landmarks is not feasible, two types of landmark position constraints are used to reduce the search space (pruning) by removing obvious outliers and thus speed up the search algorithm.

Absolute Distance constraint captures the fact that the distances between two landmark points must be within certain margins consistent with the absolute face dimensions. Distance constraints are created from the marginal shape variations of FLM8. For all modes of variation ($b_i = \pm 3\sqrt{\lambda_i}$), the minimum D_{min} and maximum D_{max} distance of every pair of landmarks $(\mathbf{r}_i, \mathbf{r}_j)$ are computed. We constrain candidate landmark distances $|\mathbf{r}_i - \mathbf{r}_j|$ within these margins plus a tolerance t, such that:

$$(1-t) \cdot D_{min}(\mathbf{r}_i, \mathbf{r}_j) \leq |\mathbf{r}_i - \mathbf{r}_j| \leq (1+t) \cdot D_{max}(\mathbf{r}_i, \mathbf{r}_j) \ , \tag{90}$$

where $\mathbf{r}_i$, $\mathbf{r}_j$ denote the positions of landmarks, with $i \neq j$.

Relative Position constraint captures the fact that the relative positions of landmark points must be consistent with the face shape. Considering the nose tip as a center, all other landmarks must lie in a counter-clockwise direction for FLM5L and in a clockwise direction for FLM5R. If we define a counter-clockwise direction $\mathbf{N}$, then the vectors from the nose tip to the other landmarks have also a counter-clockwise direction:

$$\mathbf{N} = (\mathbf{r}_m - \mathbf{r}_5) \times (\mathbf{r}_n - \mathbf{r}_5) \tag{91}$$

and

$$[(\mathbf{r}_i - \mathbf{r}_5) \times (\mathbf{r}_j - \mathbf{r}_5)] \cdot \mathbf{N} > 0 \tag{92}$$

where $\mathbf{r}_m$, $\mathbf{r}_n$, $\mathbf{r}_i$, $\mathbf{r}_j$ denote the positions of certain landmarks and $\mathbf{r}_5$ the position of the nose tip. For FLM5R: $(m, n) = (2, 1)$ and $(i, j) \in \{(1, 6), (6, 8)\}$, and for FLM5L: $(m, n) = (4, 3)$ and $(i, j) \in \{(7, 4), (8, 7)\}$.

The purpose of the above constraints is to speed up the search algorithm by removing only the outliers and not potential solutions. To avoid over-constraining we have used only the radial ordering of landmarks, expressed by Eq. 92 - instead of enforcing angles between landmarks - and a wide range for parameter t in Eq. 90.

6.2.2 Landmark Selection

To find the optimal solution, the three available consistent lists of landmark sets (left, right and complete) are sorted in descending order according to a distance measure from the corresponding model (FLM5L, FLM5R, FLM8). The landmark set (left, right or complete) that has the minimum distance measure is identified as the optimal solution (Figs. 41 and 27).

Since FLM5R, FLM5L, FLM8 have different dimensions in shape space, Procrustes distances cannot be used as a distance measure because they are not directly comparable:

$$D_P = \sqrt{\sum_{j=1}^{k} (x_j - y_j)^2} \tag{93}$$

where D_P is the Procrustes distance, $\mathbf{x}$ and $\mathbf{y}$ are the two shape vectors and k is the shape space dimension ($k = 24$ for FLM8 and $k = 15$ for FLM5R and FLM5L).

Thus, we must use alternative measures for the distance between two landmark shapes that can be comparable irrespectively of their dimensions.

An intuitive *normalized Procrustes distance* D_{NP}, that takes into consideration the shape space dimensions k, is:

$$D_{NP} = \frac{D_P}{k^2} \tag{94}$$

where D_P is the Procrustes distance, and k is the shape space dimension. The division by k^2 instead of k is preferred to give a bias to the complete solution.

A non-geometric measure of the quality of a landmark shape is its *mean spin image similarity* normalized to [0,1] (0 for high similarity and 1 for low similarity). Here, we take into consideration the spin image similarities between detected landmarks and spin image templates:

$$D_{SS} = \frac{1}{2}\left[1 - \frac{\sum_{i=1}^{n} SS(P_i, Q_i)}{n}\right] \tag{95}$$

where D_{SS} is the mean spin similarity distance, $S(P_i, Q_i)$ the similarity measure between the landmark spin image grid P_i and the corresponding template Q_i, and n the number of landmarks ($n = 8$ for FLM8 and $n = 5$ for FLM5R and FLM5L).

Thus, an intuitive *normalized Procrustes* $\times$ *mean spin similarity distance* D_{NPSS}, that takes into consideration the geometric distance and the spin image similarities can be defined as:

$$D_{NPSS} = D_{NP} \cdot D_{SS} \tag{96}$$

where D_{NP} is the normalized Procrustes distance and D_{SS} the mean spin image similarity.

We used the "normalized Procrustes" D_{NP} distance metric to select the best landmark set solution in "Method SIEM–NP" and "Method SISI–NP", and the "normalized Procrustes $\times$ mean spin similarity" D_{NPSS} distance metric in "Method SISI–NPSS" and "Method UR3D-S", where spin images are available.

6.3 Landmark Detection Methods

During the research conducted under this dissertation, several versions of the presented generalized framework for facial landmark detection were applied.

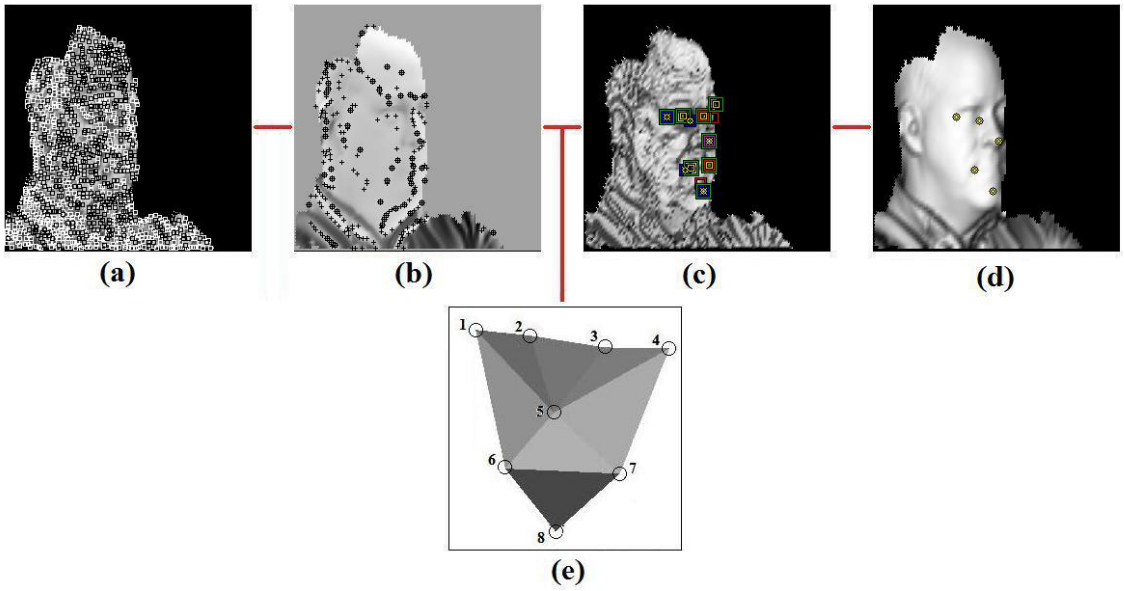


Figure 39: METHOD 1: SIEM–NP: Process pipeline for landmark detection: (a) shape index's maxima and minima; (b) extrusion map's candidate nose and chin tips; (c) extracted best landmark sets; (d) resulting landmarks; and (e) Facial Landmark Model (FLM) filtering.

These are summarized in the following:

METHOD 1: In this method, shape index's minima are the candidate landmarks for eye and mouth corners and shape index's maxima that are also Extrusion map's maxima are the candidate landmarks for the nose and chin tips (Fig. 39).

The FLMs were trained from 150 facial datasets from FRGC v2 database all having neutral expressions. To find the best solution, the *normalized Procrustes distance* D_{NP} (Eq. 94) was used.

This approach for landmark detection was used in a partial face recognition system based on symmetrical filling and published in [101].

This method is referred as **METHOD SIEM–NP**.

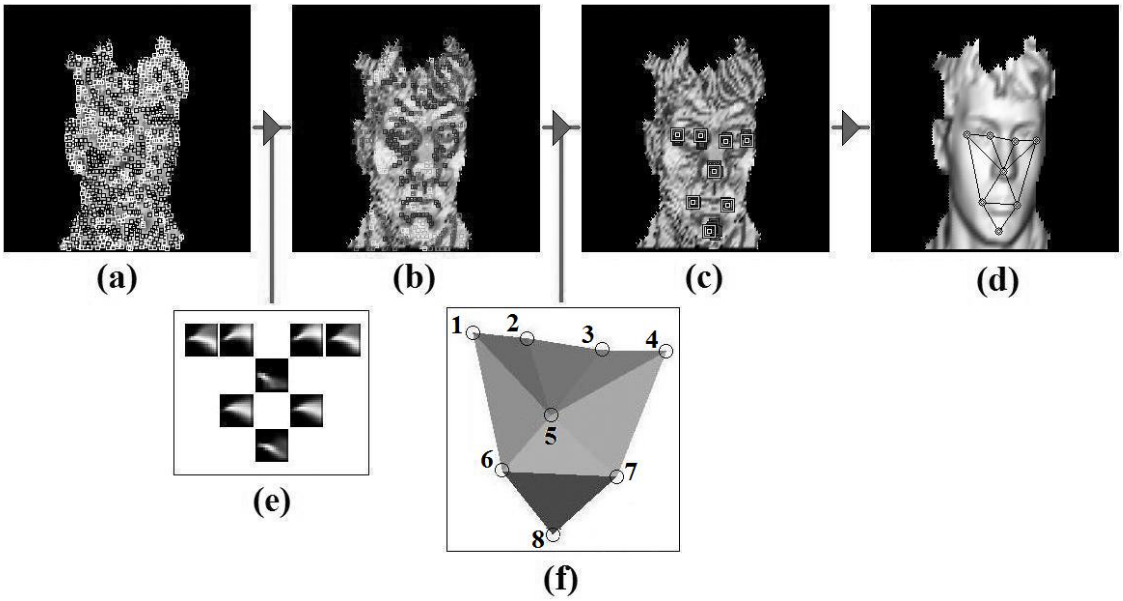


Figure 40: METHODS 2 and 3: SISI–NP and UR3D-S: Process pipeline for landmark detection: (a) shape index's maxima and minima; (b) spin image classification; (c) extracted best landmark sets; (d) resulting landmarks; (e) spin image templates filtering; and (f) Facial Landmark Model (FLM) filtering.

METHOD 2: In this method, shape index's maxima and minima are further classified into five classes by the spin image templates and are the candidate landmarks for eye outer corner, eye inner corner, nose tip, mouth corner and chin tip (Fig. 40).

The FLMs were trained from 150 facial datasets from FRGC v2 database all having neutral expressions. The spin image templates were not trained but were selected among representative exemplar facial datasets. To find the best solution, the *normalized Procrustes distance* D_{NP} (Eq. 94) was used.

This approach for landmark detection was used to locate landmarks in a manner that allows consistent retrieval of facial regions from 3D facial datasets and published in [103].

This method is referred as **METHOD SISI–NP**.

METHOD 3: In this method, shape index's maxima and minima are further classified into five classes by the spin image templates and are the candidate landmarks for eye outer corner, eye inner corner, nose tip, mouth corner and chin tip (Fig. 40).

The FLMs were trained from 150 facial datasets from FRGC v2 database all having neutral expressions. The spin image templates were trained from the 975 facial datasets of FRGC v2 that were also used for the detection experiments. To find the best solution, the *normalized Procrustes* $\times$ *mean spin similarity distance* D_{NPSS} (Eq. 96) was used.

In this work, compared to [101] and [103], a far more robust automatic 3D facial landmark detector was introduced.

This approach for landmark detection was used in a partial face recognition system based on symmetrical filling and published in [97].

This method is referred as **METHOD UR3D-S**.

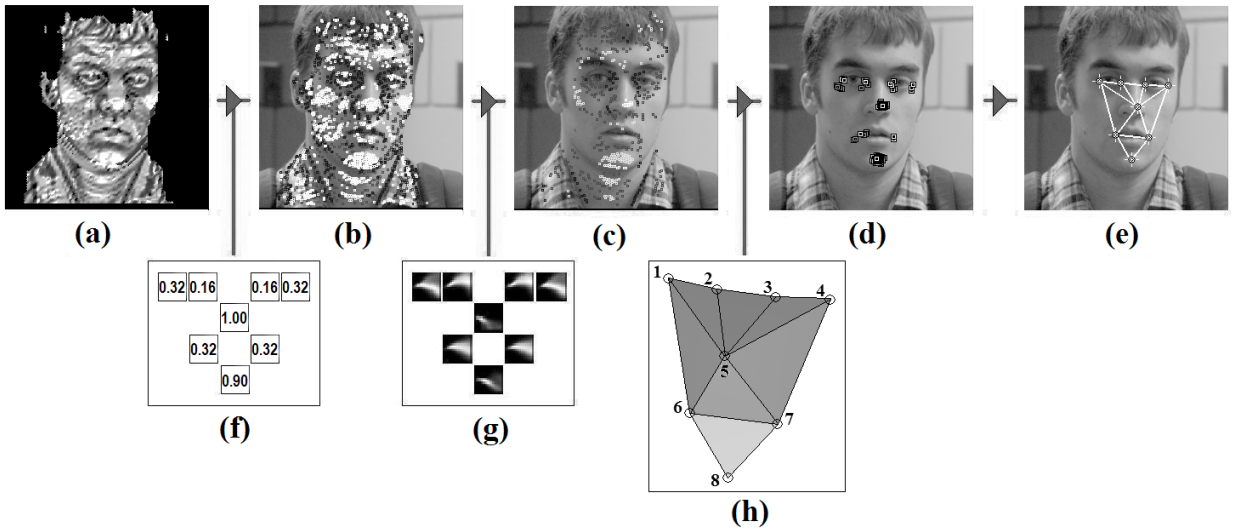


Figure 41: METHOD 4: SISI–NPSS: Process pipeline for landmark detection: (a) shape index map; (b) shape index's candidate landmarks; (c) spin image similarity filtering; (d) extracted landmark sets consistent with FLM; (e) resulting optimal landmark set; (f) shape index target values; (g) spin image templates; and (h) Facial Landmark Model (FLM).

METHOD 4: In this method, in addition to previously published work, the FLMs, the shape index target values and the spin image templates were trained from a specific subset of FRGC v2 database which contains 300 facial scans with varying expressions, that were not used in the evaluation experiments (Fig. 51).

To locate landmark points on the shape index map, shape index target values for each landmark class (eye outer corner, eye inner corner, nose tip, mouth corner and chin tip) were searched for on the shape index map. Subsequently, the candidate landmark points of the five landmark classes that are obtained from the shape index map are further filtered out according to the similarity $S(P, Q)$ of their spin images with the spin image templates representing each landmark class (Fig. 41). To find the best solution, the *normalized Procrustes* $\times$ *mean spin similarity distance* D_{NPSS} (Eq. 96) was used.

The inclusion of facial expressions into the FLMs and the use of separate shape index target values for each individual landmark resulted in an improved landmark detection accuracy (by up to 28%), and an improved landmark detection rate (by up to 16%), compared to the results that were obtained in previous work [97, 103].

This approach for landmark detection was published in [100], and the method is referred as **METHOD SISI–NPSS**.

Comparative results of the applied landmark detection methods are presented in Tables 15 and 16.

METHOD 5: In this method, the fusion schemes for combining landmark features (see Section 5.2) were incorporated into the landmark detection pipeline (Fig. 42). To locate

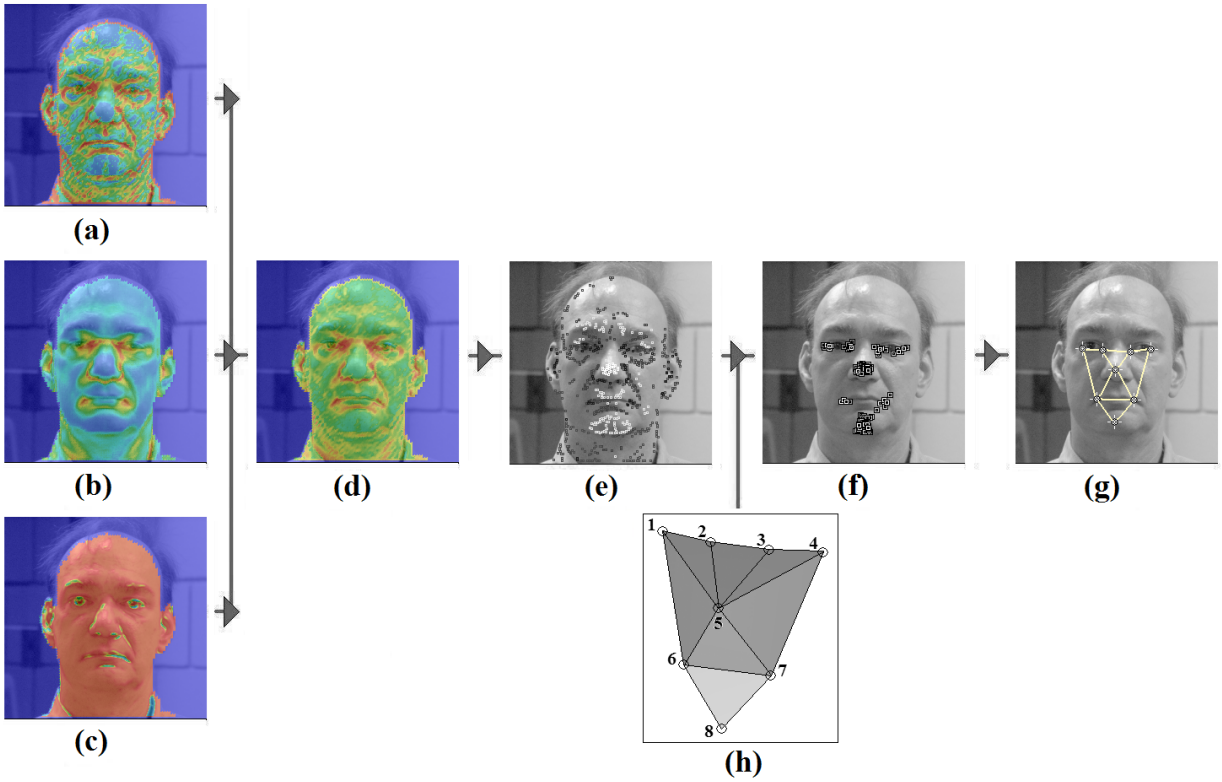


Figure 42: METHOD 5: Fusion scheme $\mathbf{Q-L2(SI+SS+ER)}$: Process pipeline for landmark detection: (a) shape index (SI) similarity map for EIC; (b) spin image (SS) similarity map for EIC; (c) edge response (ER) similarity map for EIC; (d) resultant similarity map for EIC; (e) candidate landmarks for all landmark classes; (f) extracted landmark sets consistent with FLM; (g) resulting optimal landmark set; and (h) Facial Landmark Model (FLM).

landmark points the shape index map, the spin image map and the edge response map were fused into a resultant similarity map, each for every landmark class. The candidate landmarks for each landmark class were searched on the corresponding resultant similarity map. Subsequently, the candidate landmark points of the five landmark classes were filtered out according to their consistency with the FLM. To find the best solution, the distance *normalized Procrustes distance* $\times$ *(1 – resultant similarity)* was used.

The adoption of the feature fusion method resulted in an improved landmark detection accuracy and an improved landmark detection rate, compared to the results that were obtained by METHOD SISI–NPSS. This is more clear for the fusion scheme $\mathbf{Q-L2(SI+SS)}$, which combines the shape index (SI) and spin image (SS) similarity maps. On the other hand the fusion scheme $\mathbf{Q-L2(SI+SS+ER)}$ did not improve the results as expected, because of the strong dependence of the edge response (ER) on the illumination conditions (half of the face in shadow) and the lack of correspondence between the acquired 2D and 3D images that is often present in FRGC v2 Database (see Section 8.1).

Comparative results of the fused landmark detection methods are presented in Table 19.

6.4 Face Registration & Pose Estimation

In a 3D face recognition system, alignment (registration) between the query and the stored datasets is necessary in order to make the probe and the gallery dataset comparable. Registration can be done against a common frame of reference, i.e. a *Reference Face Model* (RFM) of known coordinates (Fig. 46).

Registration of facial datasets to a reference face model can be accomplished, by minimizing the Procrustes distance between a set of landmark points on the facial dataset and the corresponding landmark points on the Reference Face Model. Landmark points $\mathbf{x}$ on the facial datasets have to be detected by applying one of the previously mentioned methods, and landmark points $\mathbf{x_0}$ on the Reference Face Model are manually annotated once at a preprocessing stage.

Alignment of a set of face landmark points $\mathbf{x}$ to the RFM landmark points $\mathbf{x_0}$ is done by minimizing the Procrustes distance in an iterative approach, as described in Algorithm 8.

Algorithm 8 "Face Registration"

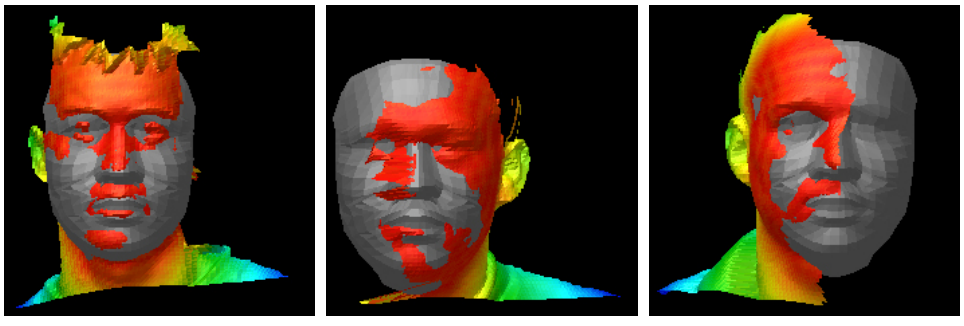
input: Reference landmark shape $\mathbf{x}_0$
and probe landmark shape $\mathbf{x}$.
output: Registration transformation $\mathbf{M}$.

1: Compute $\mathbf{T}$ to translate $\mathbf{x}$ so that its centroid is at the origin (0,0,0).
2: Compute $\mathbf{T}_0$ to translate $\mathbf{x}_0$ so that its centroid is at the origin (0,0,0).
3: **repeat**
4: Align $\mathbf{x}$ to the reference shape $\mathbf{x}_0$ by an optimal rotation $\mathbf{R}$.
5: Compute the Procrustes distance $\|\mathbf{x} - \mathbf{x}_0\|$ of $\mathbf{x}$ to the reference shape $\mathbf{x}_0$.
6: **until** Convergence: $\|\mathbf{x} - \mathbf{x}_0\| < \varepsilon$.
7: Apply $\mathbf{M} = \mathbf{T}_0^{-1} \cdot \mathbf{R} \cdot \mathbf{T}$ to register face data.

Thus, the final transformation to register a facial dataset with $\mathbf{v}_i$ vertices to the face model is given by:

$$\mathbf{v}_i' = \mathbf{T}_0^{-1} \cdot \mathbf{R} \cdot \mathbf{T} \cdot \mathbf{v}_i \tag{97}$$

and the pose is estimated from $\mathbf{R}$. Notice that scaling can be omitted when the probe and reference shapes are of the same size.



(a) (b) (c)

Figure 43: Reference face model (RFM) and probe face superposed after alignment: (a) frontal face dataset; (b) 45° left side face dataset; and (c) 60° right side face dataset. Gray colored mesh denotes the face model. Color on probe face denotes min distances of probe face vertices to model. (red: near to blue: far)

Note that the landmark set detected on the probe facial scan (complete, right or left) determines the set of the landmarks (FLM8, FLM5R or FLM5L) used for registration with the Reference Face Model. Fig. 43 depicts the registration of a profile facial dataset to a RFM. Note that a left side five landmark set detected on the facial scan has to be aligned with the left five landmark subset of the RFM, to have a correct global registration.

As a Reference Face Model the complete Facial Landmark Model (FLM8) (Fig. 7(c)) or the Annotated Face Model can be used (Fig. 23). The AFM is annotated into different areas (e.g., mouth, nose, eyes) and has predefined landmark points.

Remarks:

a. Note that a standard registration method such as the point-to-surface Iterative Closest Point (ICP) could not be used here, since without proper initialization it would be prone to false registration. Facial scans have many outlier data (such as shoulders), and also missing data due to self occlusion (profile data), that can mislead ICP registration with a complete facial model or a frontal facial dataset. Note that in FRGC v2 only frontal datasets are considered, and hence, the ICP was able to provide adequate registration results (see also the Remarks in Section 7.2).

b. An alternative approach to perform the registration would be to drive the procedure by an ICP algorithm, which uses our feature distance metrics instead of a (geometric) point-to-surface one. It is very common nowadays to run ICP-like methods on salient features instead of original point cloud data and many established ICP variants have adopted a similar or a hybrid methodology (such as [121, 118]).

6.4.1 Measurement of alignment quality

In order to evaluate the performance of the landmark detection algorithm and the quality of the alignment procedure two metrics were used.

Euclidian distance as a metric for face alignment quality: An overall measure which reflects the quality of the landmark detection process is the *mean Euclidian distance* between two landmark shapes in original 3D space:

$$D_{ME} = \frac{\sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{y}_i\|}{n} \tag{98}$$

where D_{ME} is the mean Euclidian distance, $\|\mathbf{x}_i - \mathbf{y}_i\|$ the Euclidian distance between the landmark points $\mathbf{x}_i$ and $\mathbf{y}_i$ of the two shapes and n the number of landmarks ($n = 8$ for frontal facial datasets and $n = 5$ for side facial datasets).

Mean Euclidian distance D_{ME} can be used to express the *mean localization error* of the detected landmarks. In such a case the $\mathbf{x}_i$ represent the detected landmarks and the $\mathbf{y}_i$ the manually annotated landmarks, which are considered as ground truth.

Mean Euclidian distance D_{ME} can also be used to express the *alignment quality* of the probe face with the RFM. In such a case $\mathbf{x}_i$ represent the detected landmarks and $\mathbf{y}_i$ the annotated landmarks on the RFM.

Hausdorff distance as a metric for face alignment quality: After the alignment of a test face scan with a reference face model (i.e., the AFM) the "modified directed Hausdorff distance" D_{MH} is used as a metric that reflects the quality of the landmark detection and alignment procedure. This metric can be used in situations where there are no annotated landmarks which could be considered as ground truth.

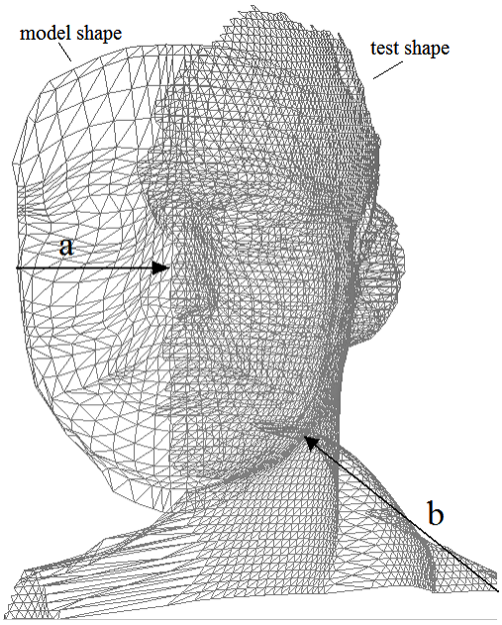


Figure 44: Face model (M) and test face (T) after alignment: (a) D_h is biased due to the lack of points of a left test face: $D_h(M,T) = \|\mathbf{a}\|$; and (b) D_h is biased due to the lack of points of model: $D_h(T,M) = \|\mathbf{b}\|$.

Consider two point sets:
$M = \{\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p\}$, that represents a face shape model, and
$T = \{\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_q\}$, that represents a test face shape, where $\mathbf{m}_i,\ \mathbf{t}_j \in \mathbb{R}^3$.

The *standard Hausdorff* distance is defined as:

$$D_H(M,T) = \max(D_h(M,T), D_h(T,M)) \ , \tag{99}$$

where

$$D_h(M,T) = \max_i(\min_j(\|\mathbf{m}_i - \mathbf{t}_j\|)) \ , \tag{100}$$

is the *directed Hausdorff* distance from M to T [98].

The directed Hausdorff distance expresses the Euclidean distance $\|\mathbf{m}_i - \mathbf{t}_j\|$ of the farthest point of M from any point of T, i.e., the maximum value of the minimum Euclidean distances of the points of M from any point of T.

The *modified directed Hausdorff distance* D_{MH}, of a face model M to a test face T, is defined, according to [46], as:

$$D_{MH}(M,T) = \frac{1}{p}\sum_{i=1}^{p} \min_j(\|\mathbf{m}_i - \mathbf{t}_j\| \ , \tag{101}$$

where $\|\mathbf{m}_i - \mathbf{t}_j\|$ is the Euclidian distance between the face model vertices $\mathbf{m}_i$ and the test face vertices $\mathbf{t}_j$, and p the number of the face model vertices.

The $D_{MH}(M,T)$ expresses the mean value of the minimum Euclidian distances $\|\mathbf{m}_i - \mathbf{t}_j\|$ of the vertices of the face model M, to which a test face scan T is registered. To get comparative results for D_{MH} we used as a model for frontal databases all the vertices of the complete AFM, for left-side databases the left side vertices of the AFM, and for right-side databases the right side vertices of the AFM.

Remarks:

a. The directed Hausdorff distance $D_h(T,M)$ of T from M is sensitive to points that belong to the T shape (i.e., shoulders, hair) and have no corresponding points in M (Fig. 44 b). Thus, it can't be used as a reliable measure of the quality of the alignment procedure. The same holds true for the undirected Hausdorff distance $D_H(M,T)$, since it includes it.

b. A more appropriate measure is the $D_h(M,T)$ of M from T, since it measures the distances of the points that belong to the model, avoiding outlier points that bias the maximum distance. This is not actually true for side scans, since there are no corresponding points in the missing half of test face (Fig. 44 a).

c. The $D_{MH}(M,T)$ is a more proper measure, since it expresses the mean of the least distances of the model shape to the test shape and not only the maximum of them as $D_h(M,T)$ does.

d. The $D_{MH}(M,T)$ is larger for semi-profile and profile scans due to the lack of points that correspond to the complete face model (Fig. 44 a). Thus it is not a good measure for comparing the results from different databases that contain different poses.

e. To avoid these over-determined Hausdorff distance values that result from comparing a profile test face with a complete face model, we split the model into two halves, one right M_R and one left M_L, and compute the distances: $D_{MH}(M_R,T)$ and $D_{MH}(M_L,T)$ instead.

f. The modified directed Hausdorff distance could also be used after the deformation fitting process of "Partial Face Recognition" to measure its quality.

Fig. 43 depicts the face model (AFM) and test face superposed after alignment. Gray color denotes the face model. The color on the test face (red for near to blue for far) denotes min distances of test face vertices to model, $\min_j(\|\mathbf{t}_i - \mathbf{m}_j\|)$. You can observe that $D_h(T,M) = \max_i(\min_j(\|\mathbf{t}_i - \mathbf{m}_j\|))$ is located on the shoulder edge.

7 Partial Face Recognition

... and everything under the sun is in tune,
but the sun is eclipsed by the moon.

– Pink Floyd

Face recognition is the procedure of recognizing an individual from their facial attributes or features and belongs to the class of biometrics recognition methods. *3D face recognition* is a method of face recognition that exploits the 3D geometric information of the human face. It employs data from 3D sensors that capture information about the shape of a face. Recognition is based on matching metadata extracted from the 3D shapes of faces. In an *identification* scenario the matching is one-to-many, in the sense that a probe is matched against all of the gallery data to find the best match above some threshold. In an *authentication* or *verification* scenario the matching is one-to-one, in the sense that the probe is matched against the gallery entry for a claimed identity, and the claimed identity is taken to be authenticated if the quality of match exceeds some threshold. 3D face recognition has the potential to achieve better accuracy than its 2D counterpart by utilizing features that are not sensitive in lighting conditions, head orientation, differing facial expressions and make-up.

With the increase in the availability of 3D data, several 3D face recognition approaches have been proposed. These approaches aim to overcome the limitations of 2D face recognition by offering pose invariance. However, they mostly use frontal 3D scans assuming that the entire face is visible to the sensor. This assumption is not always valid in real-world applications, since the unconstrained acquisition may lead to facial scans with extensive occlusions that result in missing data. Therefore, to take advantage of the full pose invariance potential of 3D face recognition, the problem of missing data must be addressed.

In this dissertation, previous work on face recognition [64, 96] is extended and integrated, and a method suitable for real-world applications, that combines pose invariance and high recognition rates is presented. The proposed method for partial face recognition allows matching among interpose facial scans, and solves the missing data problem by using facial symmetry on occluded areas (see Fig. 3).

The pose of each facial scan is determined by detecting facial landmarks, allowing an initial registration with an Annotated Face Model (AFM). The AFM is subsequently fitted to the facial scan using a subdivision-based deformable model framework that is extended to allow symmetric fitting. The symmetric fitting alleviates the missing data problem allowing the creation of geometry and normal images that are pose invariant.

The geometry and normal images are then transformed into a wavelet domain representation. These metadata representations constitute biometric signatures, which are directly comparable with each other using a L_1 distance metric, allowing efficient matching in both

identification and verification scenarios. The novelty of the proposed method is that the signature is independent from the initial pose and the missing data caused by occlusions (as long as half of the face with respect to the yaw axis is visible in the scan). Specifically, in order to perform interpose matching we require that the following landmarks are visible on the same side of the face: inner and outer eye corner, nose tip, mouth corner and chin tip. This allows seamless comparisons among frontal, left and right side scans, making the proposed method suitable for real-life biometric applications.

The processing pipeline of each facial scan consists of the following fully automated steps (see Fig. 45):

1. *Preprocessing:* Standard preprocessing techniques are used to filter the raw data.
2. *3D Landmark Detection:* The landmark detector is used for pose estimation (determining if it is a frontal, left or right scan).
3. *Registration:* The raw data are registered to the AFM using a two-stage approach.
4. *Symmetric Deformable Model Fitting:* The AFM is fitted to the data using facial symmetry. The fitted model is then converted to a geometry image and to a normal image.
5. *Wavelet Analysis:* A wavelet transform is applied on the geometry and normal image and the wavelet coefficients are stored as a signature metadata representation.

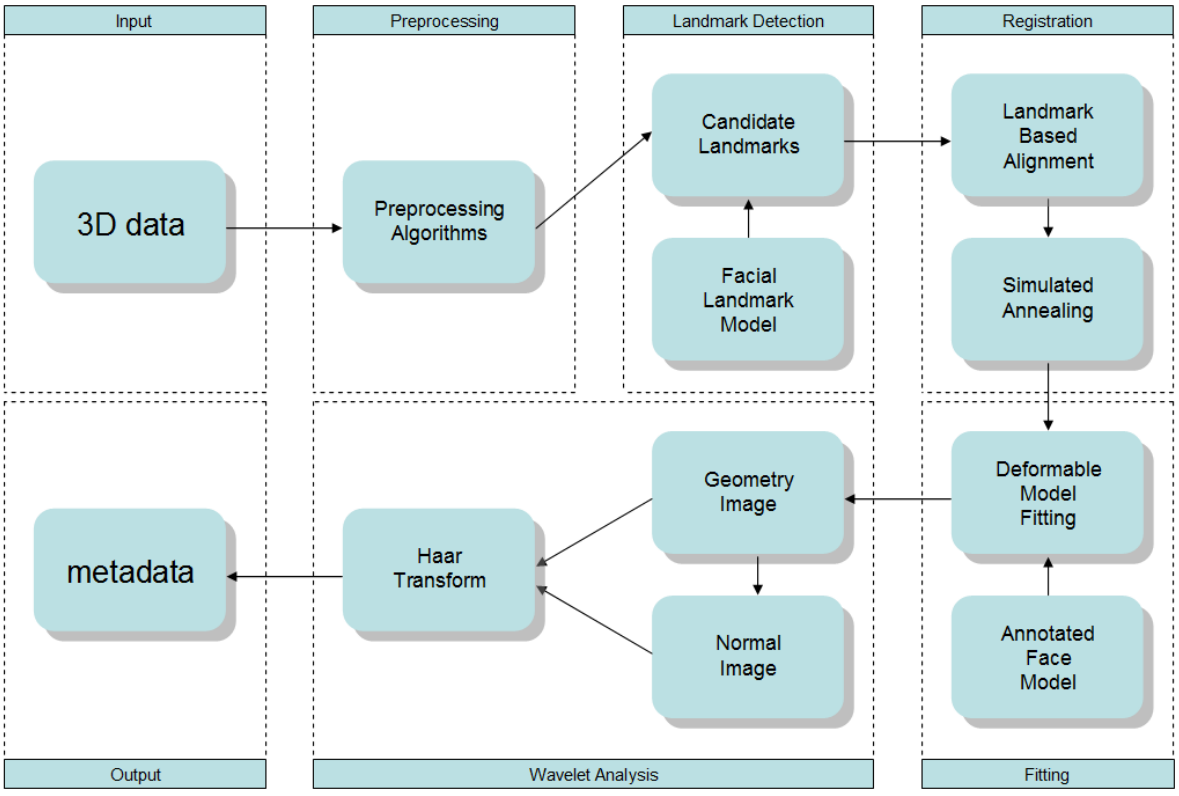


Figure 45: Pipeline of the Partial Face Recognition method.

7.1 3D Landmark Detection

The method for 3D landmark detection and pose estimation uses 3D information to extract candidate interest points which are identified and labeled as landmarks by matching them with the Facial Landmark Model (FLM) as already described. To detect landmark points, two 3D local shape descriptors that exploit the 3D geometry-based information of facial scans are used: the shape index and the spin images.

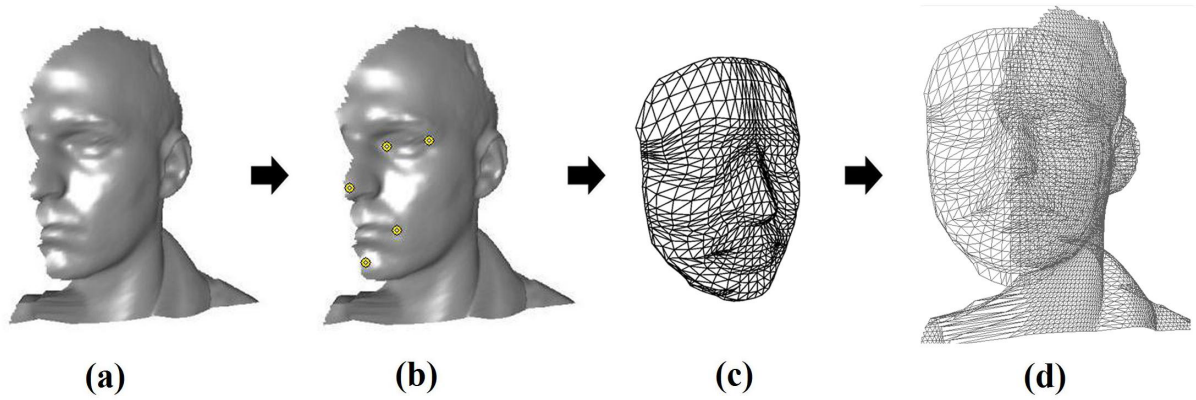


Figure 46: Face registration based on detected landmarks using the proposed method: (a) facial scan with extensive missing data; (b) extracted landmarks; (c) generic Annotated Face Model (AFM); and (d) registered facial scan with AFM.

The side of the face object under consideration (frontal, right or left) is determined from the resulting optimal landmark set (FLM8, FL5R or FLM5L) and the pose yaw-angle is calculated from the alignment transformation with the corresponding FLM. Once anatomical landmarks are localized and the face object is classified as frontal, left side or right side, the corresponding rigid transformation is computed in order to register the facial scans to the generic Annotated Face Model (AFM), as described in Section 6.4 (see Fig. 46).

7.2 AFM Registration

In order to fit the AFM to each facial scan, they both must be defined in the same coordinate system (Fig. 46). To this end, the facial scans are registered with the AFM using a two-stage approach. Firstly, the landmarks detected in the previous step provide an initial registration and secondly, an algorithm based on Simulated Annealing fine tunes the registration.

The landmark set detected on a facial scan (frontal, right or left) determines which of the FLM8, FLM5R and FLM5L will be used to aid registration with the AFM. However, in practice when a frontal scan is detected, the FLM8 is not utilized, but it is considered as a pair of side scans (therefore computing two independent registrations using FLM5R and FLM5L). In this case the remaining steps of the method are repeated twice, and two independent metadata representations are finally derived.

To improve the registration, the algorithm presented by Papaioannou *et al.* [95] is used; it uses a global optimization technique (Simulated Annealing [69, 122]) applied to depth

images. The Simulated Annealing process minimizes the following objective function:

$$D_Z = \sum_{i=1}^{r} \sum_{j=1}^{r} |Z_m(i,j) - Z_d(i,j)| \ , \tag{102}$$

where r is the spatial resolution of the buffers and $\mathbf{Z}_m$ and $\mathbf{Z}_d$ are the z-buffers of model and data respectively (normalized to $[0,1]$). For side scans, only one half of the model's z-buffer is used in the objective function. The other half is excluded as it would have been registered with areas that may have missing data.

Since it is assumed that the initial registration is roughly correct, the Simulated Annealing algorithm is only allowed to produce limited translations and rotations. Its purpose is only to fine-tune the registration; it cannot alleviate errors caused by erroneous landmark detection.

Remarks: Note that, in this implementation of partial face recognition method, the step that used the standard ICP algorithm [8] for fine-tuning the registration before the application of the z-buffer simulated annealing step, in the full-frontal face recognition method presented in [64], is omitted. The use of the standard ICP algorithm (without a 1-1 correspondence between vertices) in current method deteriorated the face recognition results, due to misregistration of the probe and gallery datasets, especially when a profile dataset had to be aligned with a frontal one.

7.3 Symmetric Deformable Model Fitting

The subdivision-based deformable model framework presented in [64] is utilized to fit the AFM to each facial scan (already registered by the previous step). During fitting, the AFM deforms in order to capture the shape of the facial scan. The forces that drive this deformation are called *external forces*. The forces that resist this deformation are called *internal forces* and correspond to the elastic properties of the model's surface (e.g., strain energy, material stiffness).

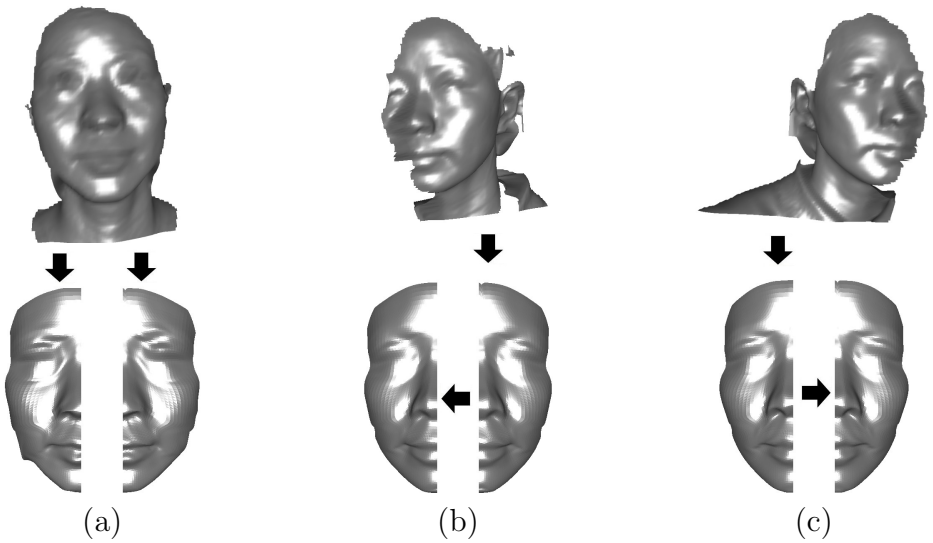


Figure 47: Symmetric fitting of the Annotated Face Model (AFM): (a) frontal face dataset; (b) left side face dataset; and (c) right side face dataset.

This framework is modified properly to incorporate the notion of *symmetric fitting* in order to handle missing data. The fitting step can now handle the left and right sides of the AFM independently (Fig. 47). The idea is that facial symmetry can be used to avoid the computation of the external forces on areas of possible missing data. The internal forces are not affected and remain unmodified in order to ensure the continuity of the fitted surface. As a result, when fitting the AFM to facial scans classified as left side, the external forces are computed on the left side of the AFM and mirrored to the right side (and vice versa for right side scans). This technique can also be applied to frontal scans, since they are handled as a pair of independent left and right side scans, and the above rule is used accordingly. Therefore, for each frontal scan, two fitted AFMs are computed: one that has the left side mirrored to the right and another that has the right side mirrored to the left.

The basic equation of the deformable model framework is given by Newton's second law:

$$\mathbf{M}_q \frac{d^2\mathbf{q}}{dt^2} + \mathbf{D}_q \frac{d\mathbf{q}}{dt} + \mathbf{K}_q\mathbf{q} = \mathbf{f}_q \ . \tag{103}$$

The term $\mathbf{q}$ is the control points vector that determines the degrees of freedom of the AFM (each point having three degrees of freedom). The term $\mathbf{M}_q$ is the mass matrix and is multiplied with the acceleration vector in order to control the kinetic energy. $\mathbf{D}_q$ is the damping matrix and is multiplied with the velocity vector in order to control the energy dissipation.

Note that for data fitting purposes we set:

$$\mathbf{M}_q \frac{d^2\mathbf{q}}{dt^2} = \mathbf{0} \ \text{ and } \ \mathbf{D}_q \frac{d\mathbf{q}}{dt} = \mathbf{0} \ , \tag{104}$$

since they represent the translational effects of the external forces.

The term $\mathbf{f}_q$ represents the external forces vector; during fitting, it consists of forces that pull the control points vector toward the surface of the facial scan. Finally, the term $\mathbf{K}_q$ is the stiffness matrix and it determines the elastic properties of the AFM that resist the external forces. It can be decomposed into three matrices $\mathbf{K}_q = \mathbf{K}_\alpha + \mathbf{K}_\beta + \mathbf{K}_\gamma$. $\mathbf{K}_\alpha$ is related to the first order strain energy, $\mathbf{K}_\beta$ to the second order strain energy and $\mathbf{K}_\gamma$ is related to the spring forces energy:

$$\begin{aligned} E_\alpha &= \frac{1}{2}\kappa_\alpha \mathbf{q}^T \mathbf{K}_\alpha \mathbf{q} \ , \\ E_\beta &= \frac{1}{2}\kappa_\beta \mathbf{q}^T \mathbf{K}_\beta \mathbf{q} \ , \\ E_\gamma &= \frac{1}{2}\kappa_\gamma \mathbf{q}^T \mathbf{K}_\gamma \mathbf{q} \ , \end{aligned} \tag{105}$$

where κ_α, κ_β, κ_γ are the individual weights.

The analytical equations are solved using an iterative Finite Element Method (FEM) approximation. In current implementation, the subdivision-based FEM approximation proposed by Mandal [86] is employed. This approximation solves the above equations in an iterative way. The AFM is used as the control mesh of a subdivision surface. At each step, the internal and external forces are computed on the limit surface and by using the inverse subdivision matrices they are transferred to the control mesh ($\mathbf{q}$).

Details of this implementation can be found in [64], but some key aspects are the following:

- The resolution of the control mesh determines the degrees of freedom but does not affect the accuracy of the approximation (which is determined by the resolution of the limit surface).

- The Loop subdivision scheme [82] is used because it produces a limit surface with $\mathcal{C}^2$ continuity, and only 1-neighborhood area information is needed for each vertex.

- For the computation of the external forces, multiple nearest neighbor searches between the AFM and the surface of the facial scan are needed. To decrease the computational cost a space partitioning technique (Octrees [71, 45]) is employed.

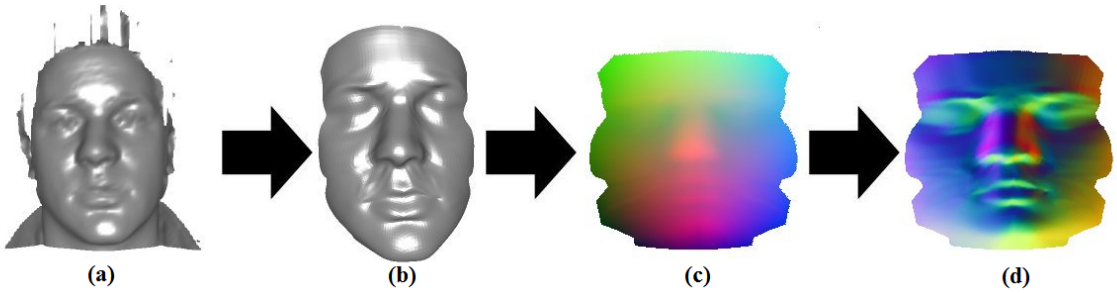


Figure 48: Pipeline of face fitting: (a) raw facial data; (b) deformed AFM to facial data; (c) Geometry image of deformed AFM; and (d) Normal image of deformed AFM.

When the deformation stops, the annotated model acquires the shape of the raw data. Since the deformation has not violated the injective nature of AFM (u, v) parameterization, the deformed AFM can be converted to a geometry image. The normal image is also computed, equivalent to the first spatial derivative of the geometry image (Fig. 48).

7.4 Wavelet Analysis

A wavelet transform is applied on the derived geometry and normal images in order to extract a descriptive and compact biometric signature [64]. As explained above, even if half of the face is missing from the facial scan, the derived geometry and normal images describe the full face. When facial symmetry is used (for side scans) there is redundant information, as half of the geometry and normal image is the mirror of the other half. However, both sides are kept in order to have a common representation that is independent of the initial pose.

Each channel of the geometry and normal image is treated as a separate image for the wavelet analysis (thus resulting in six channels, three for each type of image). The Walsh wavelet transform [126, 105] for images is a decimated wavelet decomposition using tensor products of the full Walsh wavelet packet system. The 1D Walsh wavelet packet system is constructed by repeated application of the Haar filterbank, a two-channel multirate filterbank based on the Haar conjugate mirror filter. The choice of Haar wavelets was based on their excellent localization properties.

The application of the Haar filterbank is conceptually simple and computationally efficient. The Haar wavelet transform is performed by applying a low-pass filter and a high-pass

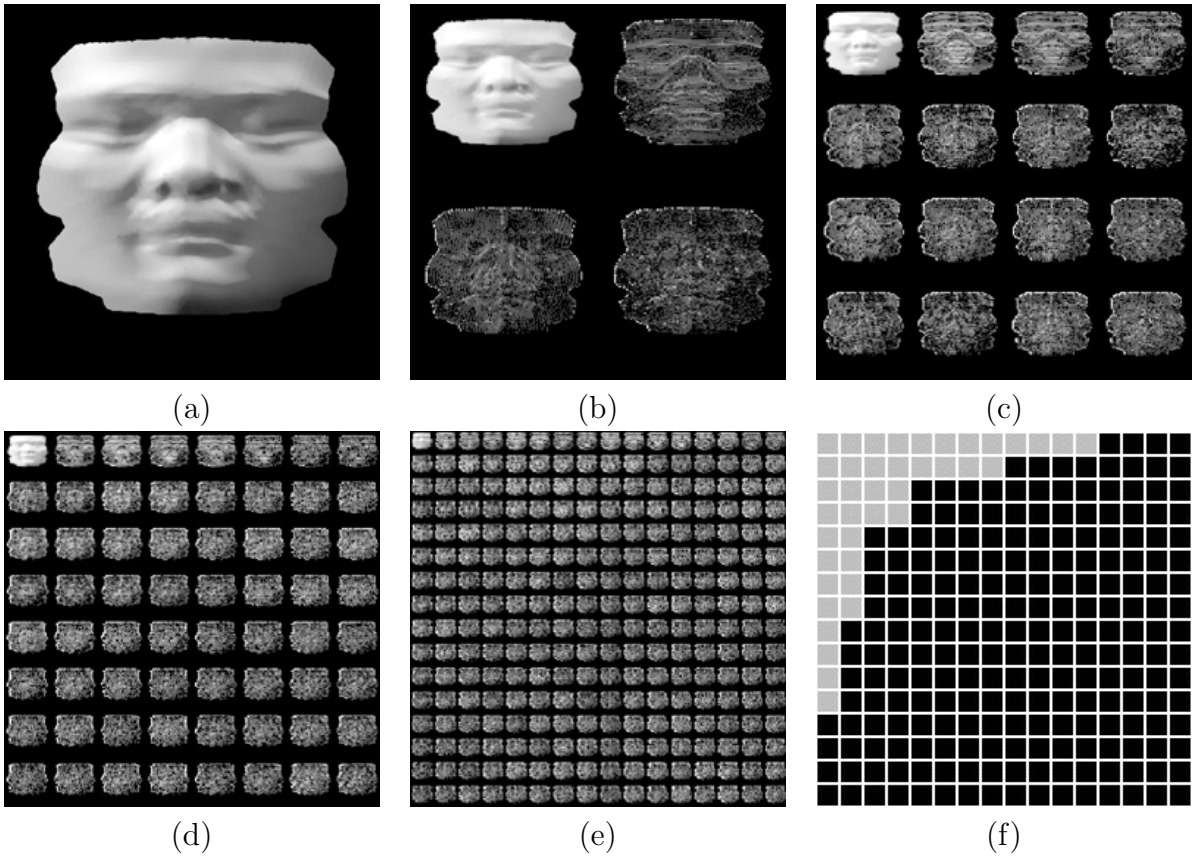


Figure 49: Wavelet analysis of a frontal facial normal image (the intensity of the coefficients was adjusted for visualization purposes): (a) original image, (b-e) 1^{st}, 2^{nd}, 3^{rd} and 4^{th} level Walsh transform, (f) mask that selects 15% of the wavelet packets.

filter on a one-dimensional input, then repeating the process on the two resulting outputs. The low–pass and high–pass Haar filters are g and h, respectively:

$$g = \frac{1}{\sqrt{2}}[1 \quad 1] \quad \text{and} \quad h = \frac{1}{\sqrt{2}}[1 \quad -1] \; . \tag{106}$$

Since we are working with images, there will be four outputs for each level of the Haar wavelet:

$$g^T g = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \; g^T h = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \; h^T g = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \text{ and } h^T h = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{107}$$

corresponding to low-low, low-high, high-high and high-low filters respectively, acting on the rows and columns of the image array. The low-low filter encodes the the local mean of pixel values, while the rest encode the horizontal, vertical and diagonal edges of the image.

A level 4 decomposition is computed, meaning that the filters are applied four times for creating the full Walsh transform, which yields $(2^2)^4$ wavelet packets. Since the geometry and normal images are of resolution 256×256 and at each decomposition a $1:2$ subsampling is applied, each wavelet packet has a resolution of 16×16.

The level 4 Walsh decomposition produces $256 = (2^2)^4$ wavelet packets, arranged in a 16×16 array of images (see Fig. 49 (e)) [105]. Since not all packets encode the same

amount of information, it is possible to ignore most of the packets without loosing significant information from the original image. To this end, 40 wavelet packets are kept (roughly 15%) to create an efficient and compact metadata representation (Fig. 49 (f)), that can be used as a face signature. The wavelet packets with the minimum variation among the scans of the same subject as well as the maximum variation among the scans of different subjects are favored. The selection of the 40 wavelet packets was optimized using a test database with frontal facial scans.

The coefficients contained within the metadata representation can be directly compared without having to reconstruct the original image using a weighted L_1 metric. The distance d_G between a probe and a gallery geometry image (for each x, y, z component) is measured as:

$$d_G = \sum_{i=1}^{n} w_i |[\mathbf{G}_P]_i - [\mathbf{G}_G]_i| \ , \tag{108}$$

where $\mathbf{G}_P$ and $\mathbf{G}_G$ are the wavelet coefficients derived from the geometry images of the probe and gallery scans respectively, n is the number of these coefficients, and w_i is a weight mapping function.

The total distance D_G is the sum of the distances computed on all x, y, z components of a geometry image:

$$D_G = [d_G]_x + [d_G]_y + [d_G]_z \ . \tag{109}$$

The localization properties of the Walsh transform allow per area mapping, therefore for k annotated facial areas (according to AFM), the w_i weights will have k distinct values. For the experiments, these values were selected empirically based on the biometric importance of each area [64].

The distance D_N between a probe and a gallery normal image is similarly computed.

The final distance between a probe and a gallery scan is given by:

$$D = D_G + w_N D_N \ , \tag{110}$$

where w_N is a normalization weight.

Since the ratio of the average L_1 difference between two geometry images over the average L_1 difference between two normal images is approximately $1:8$, w_N is set equal to 8. Note that the normal images, being the first spatial derivative of the geometry images, are less sensitive to positional (but not rotational) errors introduced during registration. As a result, the interpose matching favors the normal images over the geometry images.

8 Experimental Results

Tell me and I forget.
Teach me and I remember.
Involve me and I learn.

– B. Franklin

In this Chapter the experimental results that evaluate the proposed methods are presented. For this purpose the largest publicly available 3D face and ear databases were combined. In order to evaluate performance for the landmark detection methods and the landmark feature fusion methods, the facial datasets were manually annotated.

The proposed 3D landmark detector achieves state-of-the-art accuracy (with $4.5-6.3\ mm$ mean landmark localization error), and the proposed partial face recognition method state-of-the-art performance (with average rank-one recognition rate 83.7%), considerably outperforming existing methods, even when tested with the most challenging data, which contain scans with yaw variations up to 80° and strong expressions. Experimental results of the landmark feature schemes imply that the quadratic distance to similarity mapping in conjunction with the rms rule for fusion exhibits the best performance, improving the landmark detector accuracy and robustness (to $3.5-5.5\ mm$ mean landmark localization error).

8.1 Face Databases

A short description of the face databases widely available to the research community is given below. The evaluation of the proposed algorithms for landmark detection and partial face recognition was based on some of these databases.

3D-Databases:
The **FRGC** (Face Recognition Grand Challenge) [107, 106] database from the University of Notre Dame (UND) contains 4,950 facial scans and is divided into two completely disjoint subsets: FRGC v1 and FRGC v2. The hardware used to acquire these range data was a Minolta Vivid 900/910 laser range scanner, with a resolution of 640×480.

The **FRGC v1** database contains 943 range images of 275 individuals, acquired before Spring 2003 (*FRGC 3D Training Set*). Subjects have neutral expressions and almost frontal pose.

The **FRGC v2** database contains a total of 4,007 range images of 466 individuals, acquired between Fall 2003 and Spring 2004 (*FRGC 3D Validation Set*). Subjects have almost frontal poses, various facial expressions (e.g., happiness and surprise) and various illumination conditions (e.g. half of the face shaded). FRGC v2 is considered more challenging than FRGC v1.

The **Ear Database** from the University of Notre Dame (UND) [131], collections F and G. This database (which was created for ear recognition purposes) contains side scans with a vertical rotation of 45°, 60° and 90°. In the 90° side scans, both sides of the face are occluded from the sensor, therefore these were excluded since they contain no useful information. The UND database contains 119 side scans at ±45° (119 subjects, 119 left and 119 right) and 88 side scans at ±60° (88 subjects, 88 left and 88 right).

The **MSU** [84] database from the Michigan State University contains 300 multiview 3D facial scans from 100 individuals. For each subject, three scans where captured with yaw angles of less than −45°, 0° (frontal) and more than +45°.

The **BU-3DFE** [141] database from the University of New York at Binghamton contains 2,500 3D facial data of 100 individuals. The system used to acquire these data consists of six digital cameras and two light pattern projectors evenly positioned at 45° at each side of the subject. The system creates a single complete 3D triangular surface mesh of the face (20,000 - 35,000 triangles), by merging the cameras' viewpoints. Subjects perform seven universal expressions (i.e., *neutral*, *happiness*, *surprise*, *fear*, *sadness*, *disgust* and *anger*).

The **T3DFRD** (Texas 3D Face Recognition DB) [54] database contains 1,149 pairs of high resolution, pose normalized, preprocessed, and perfectly co-registered color and range images of 118 adult human subjects acquired using a stereo camera. The images are accompanied with information about the subjects' gender, ethnicity, facial expression, and the locations of 25 manually located anthropometric facial fiducial points.

The **NDOff2007** [42] database which contains 7,317 facial scans, 406 frontal and 6,911 in various yaw and pitch angles, acquired from 406 subjects. Pitch angles vary in the range of $[-45°, +45°]$ and yaw angles vary in the range of $[-90°, +90°]$. Cross rotations incorporating both yaw and pitch angles are also available.

The **Bosphorus** [117] database consists of 3,396 facial scans, which are obtained from 81 subjects in various poses, expressions and occlusion conditions. Many of the male subjects have facial hair like beard and moustache. There are three types of head poses which correspond to seven yaw angles, four pitch angles, and two cross rotations which incorporate both yaw and pitch.

2D-Databases:

The **AR** [2] contains over 4,000 color images corresponding to 126 people's faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants.

The **FERET** [50] database contains 14,126 color images from 1,199 people. It includes changes in appearance through time, controlled pose variations and facial expressions.

The **CMU-PIE** [123] database contains 41,368 color images from 64 people. Each person was imaged across 13 different poses, under 43 different illumination conditions, and with 4 different expressions.

The **Bio-ID** [9] database contains 1,521 gray level images with a resolution of 384 by 286 pixels. Each image shows a frontal view of the face of one out of 23 different test persons. Image are acquired under real world conditions, with varying background, lighting and expressions. The set contains the eye positions.

The **LFW** (Labeled Faces in the Wild) [57] database contains more than 13,000 color images of faces collected from the web. Each face has been labeled with the name of the 1,680 persons pictured. The images have not a specific resolution since they were cropped

by the Viola-Jones face detector. No landmark annotations are available.

The **LFPW** (Labeled Face Parts in the Wild) [74] database contains 1,432 Images from the internet with 29 points labeled on each image. Extremely challenging, real-world dataset, which contains large variation in pose, illumination, expression, image quality and severe occlusions.

Remarks:

a. The FRGC DBs include high resolution data, three-dimensional scans, and image sequences of each individual. Each of these data types were taken for the special purpose of face recognition by machine and are potentially more informative than simple and moderate resolution images, since one of the major goals of the FRGC was to study how higher fidelity data can help make face recognition more accurate.

b. While there are large numbers of images with uncontrolled lighting in the FRGC data sets, these images contain a great deal less natural variations. Although variation in clothing, pose, background, and other variables exists in the FRGC databases, one may sum up these differences as controlled variations.

c. Regarding the FRGC data set, during the acquisition of data there was a significant time-lapse between the operation of the laser range finder and the optical camera in the FRGC data acquisition, which caused the acquired 2D and 3D images to often be out of correspondence. This time-lapse also caused inconsistencies in facial expressions between the range and portrait images captured in single subject sessions. In addition, since laser scanning is not instantaneous, some of the faces in the FRGC are distorted due to head movements during acquisition.

d. By contrast, stereo imaging captures both the shape and the texture image of the face simultaneously, hence each range and texture pair are perfectly co-registered in the T3DFRD. Furthermore, accuracy assessment of fiducial detection requires access to publicly available ground-truth of manually pinpointed fiducial points (not provided by FRGC).

e. T3DFRD has been acquired using a stereo imaging system at a high resolution of $0.32mm$ along the x, y and z dimensions. By comparison, images in the FRGC database were acquired at a lower average resolution of 0.98 mm along the x and y dimensions and 0.5 mm along the z dimension.

f. The BU-3DFE database has been also acquired using a stereo imaging system with capture devices at $\pm 45^{\circ}$. It contains only cropped and reconstructed frontal facial datasets.

8.2 Landmark Detection

8.2.1 Test Databases

For the performance evaluation of the proposed landmark detector, the largest publicly available 3D face and ear databases were combined. To evaluate the performance of the method against yaw variations, frontal, semi-profile and profile facial datasets were used. To evaluate the tolerance of the method against expression variations, subjects with varying

degrees of expressions were included. To have a measure of the landmark detection error, the used facial datasets were manually annotated at the queried landmark points.

Thus the following collection of facial datasets were created:
(i) a database with 975 frontal facial datasets obtained from 149 different subjects, selected from the FRGC v2 database (Fig. 51), including subjects with varying degrees of expressions (45.44% "neutral", 36.41% "mild" and 18.15% "extreme"), acquired under varying illumination conditions (e.g. half of the face shaded). This database will henceforth be referred as DB00F (Fig. 50 (a)).
(ii) a composite frontal-to-profile database with the datasets of 39 common subjects found in the FRGC v2 database and in the UND Ear database. This database consists of 117 (3x39) facial scans having three poses, frontal (39 scans) and 45° left (39 scans) and right (39 scans), and will henceforth be referred as DB00F45RL.
(iii) two semi-profile databases with 118 left and 118 right 45° side datasets, which come from 118 different subjects, obtained from the UND Ear database. These databases will be referred as DB45L and DB45R respectively (Fig. 50 (b-c)).
(iv) two profile databases with 87 left and 87 right 60° side datasets, which come from 87 different subjects, obtained from the UND Ear database. These databases will be referred as DB60L and DB60R, respectively (Fig. 50 (d-e)).

Figure 50: Front view of scans from the used UND databases: (a) frontal (from FRGC v2); (b) 45° right (from Ear DB); (c) 45° left (from Ear DB); (d) 60° right (from Ear DB); (e) 60° left (from Ear DB). Note the extensive missing data in (b-e).

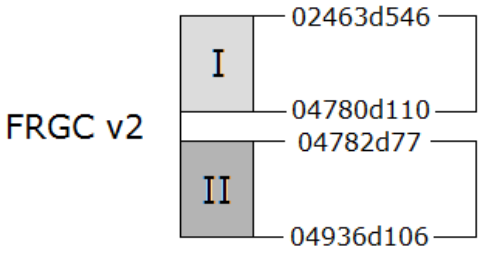


Figure 51: FRGC v2 partitioning: (I) 300 facial scans for training FLMs, shape index target values and spin image templates; and (II) 975 facial scans for testing.

Note that, even though the creators of the UND Ear database marked the side scans as 45° and 60°, the measured maximum angle of rotation is 70° and 80° respectively (Table 13). Also, note that the evaluation databases contain facial scans that are not cropped and reconstructed to contain facial only data.

In the evaluation databases, only facial datasets with all landmark points visible were included (eight for frontal scans and five for side scans). The visible landmark points were manually annotated and are considered as our ground truth. The exact datasets that were used from the source databases for testing can be found from the landmark annotation files available through the website [132].

8.2.2 Landmark Detection Evaluation

To evaluate the performance of the proposed landmark detection method, we conducted the following two experiments: In **Experiment 1** we evaluated the performance of *Method SISI–NPSS* against yaw variations, and in **Experiment 2** we evaluated the tolerance of *Method SISI–NPSS* against expression variations.

The performance evaluation of a landmark detector is generally presented by computing the following values, which represent the localization accuracy of the detected landmarks:

Absolute Distance Error: The Euclidean distance in physical units (e.g., *mm*) between the position of the detected landmark and the manually annotated landmark, which is considered ground truth.

Detection Success Rate: The percentage of successful detections of a landmark over a test database. Successful detection is considered as the detection of a landmark with Absolute Distance Error under a certain threshold (e.g., 10 *mm*).

In our experiments, the *localization error* is represented by the mean and standard deviation of the absolute distance error of the detected landmarks. Also, the overall mean distance error of the eight landmark points for the frontal datasets and of the five landmark points for the side datasets was computed.

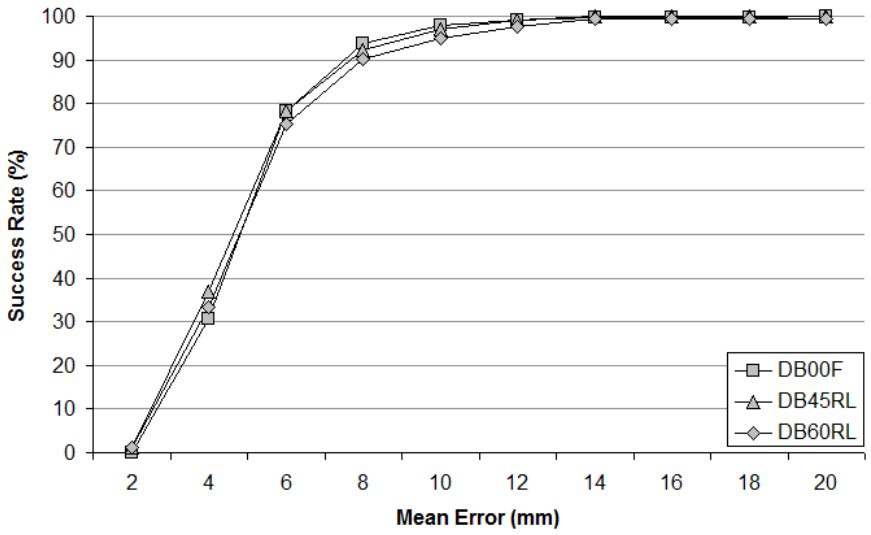


Figure 52: Mean Error Cumulative Distribution of METHOD SISI–NPSS on DB00F, DB45RL and DB60RL.

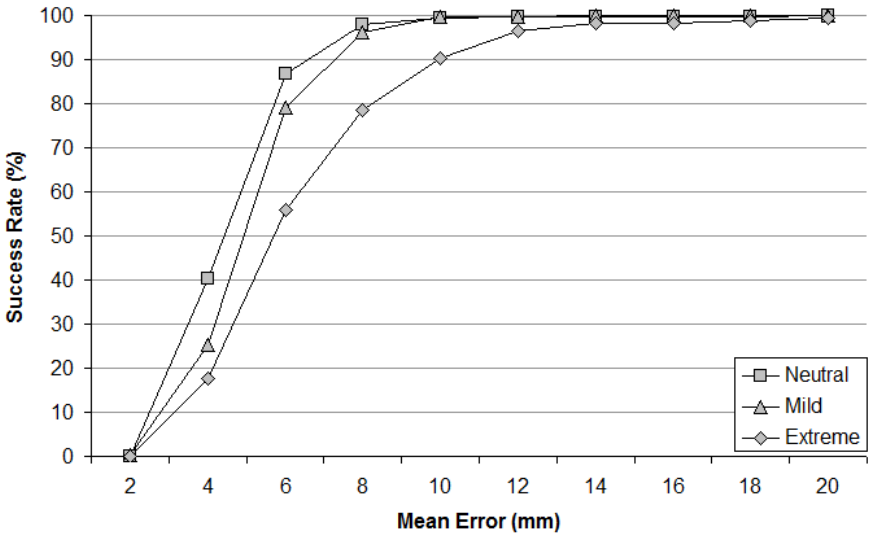


Figure 53: Mean Error Cumulative Distribution of METHOD SISI–NPSS on DB00F "neutral", "mild" and "extreme".

The *success rate of landmark localization* with an absolute distance error threshold of 10 mm is reported in the result tables. Note that, as pointed out in [97], our UR3D-S face recognition method can tolerate landmark localization errors up to 10 mm.

The yaw angle of probe faces is computed and its mean value, standard deviation, and minimum and maximum values are presented. The yaw angle results from the rotational transformation of the optimal solution that fits the probe face to the corresponding FLM and thus the probe face is classified as frontal, left side or right side. Side detection can be crucial in determining follow-up actions in a biometric system. The side detection rate, reported in the result tables, is the percentage of correct side estimations of the probe faces with respect to their ground truth side and whose detected landmarks also have an overall mean distance error under 30 mm.

We depict the Cumulative (Error) Distribution graph of the mean distance error in Fig. 53 to show the method's tolerance to expression variations and in Fig. 52 to show the method's robustness to yaw rotations. In these graphs the x-axis represents the mean distance error between the manually annotated landmarks and the automatically detected landmarks in intervals of 2 mm, and the y-axis represents the percentage of face datasets with a mean distance error up to a certain x-value, out of all gallery datasets.

Summary results for METHOD SISI–NPSS on all tested databases are presented in Table 6. The results clearly indicate that the proposed method exhibits high accuracy and robustness both to yaw and expression variations. The mean error is under 6.3 mm, with standard deviation under 2.6 mm on all tested facial scans. Also note that the mean error is under 10 mm for at least 90.4% of the tested facial scans and the facial side was correctly estimated on over 98.9% of the tested facial scans.

Specifically, the best results were obtained for the frontal facial scans category and the worst for the 60° facial scans. This is due to the fact that, as the yaw angle increases, landmark detection becomes more difficult, mainly due to distortions on their shape index and spin image values caused by the missing data around the nose and chin tip regions (Figs. 50b-e). The results that assess the robustness of METHOD SISI–NPSS against yaw

Table 6: Summary results for METHOD SISI–NPSS

Database	Mean Error			Side Detection Rate
	mean (mm)	stdev (mm)	≤ 10 (mm)	
DB00F	5.00	1.85	97.85%	99.90%
DB00F-neutral	4.52	1.51	99.32%	100.00%
DB00F-mild	4.95	1.46	99.72%	100.00%
DB00F-extreme	6.28	2.60	90.40%	99.44%
DB00F45RL	4.97	1.92	97.44%	100.00%
DB45R	5.03	1.92	96.61%	100.00%
DB45L	4.75	1.91	97.46%	100.00%
DB60R	4.95	1.80	96.55%	98.85%
DB60L	5.30	2.49	93.10%	100.00%

variations are presented in Tables 12 and 13 and Fig. 52.

The most robust facial features are the nose tip and eye inner corners, with a lower mean error and standard deviation across yaw rotations and expression variations. This is due to the fact that they have more distinct geometry which is more easily captured by the detectors, and there are no substantial changes in their shape index and spin image values due to the deformations resulting from facial expressions. The least robust facial feature appears to be the mouth corners mainly due to the fact that they do not have enough distinct geometry and are also prone to changes in their shape index and spin image values due to the deformations resulting from facial expressions. The results that assess the tolerance of METHOD SISI–NPSS against expression variations are presented in Table 14 and Fig. 53.

8.2.3 Comparative Results

For comparison of the performance of the presented landmark detection method against other state-of-the-art methods, landmark localization errors are presented in Tables 15 and 16. Note that each method uses a different facial database, making direct comparisons difficult. However, these results indicate that METHOD SISI–NPSS outperforms previous methods for the following reasons: (i) it is more accurate, since it gives smaller mean localization distance errors for almost all landmarks, and (ii) it is more robust, since it gives smaller standard deviations for the localization distance error.

Comparative results of landmark localization errors on almost-frontal facial datasets are presented in Table 15. Yu's method [142] exhibits the minimum mean localization error for the nose tip, but has a large standard deviation. Lu's method [83] exhibits the minimum mean localization error for the mouth corners, but is not a pure 3D method, since it is assisted by 2D intensity data. Finally, Colbry's method [20] seems to perform well for all landmarks, comparatively close to the proposed method, but has larger standard deviations. Note that the FRGC v1 database used in Yu *et al.* [142], Lu *et al.* [84], Lu *et al.* [83], and Colbry [20] is considered less challenging than the FRGC v2 used in our experiments, since FRGC v1 contains subjects with neutral expressions, while FRGC v2 contains subjects with various facial expressions. Furthermore, the database used by Colbry [20] contains a small

portion ($\approx 5\%$) of proprietary datasets with pose variations, occlusions and expressions. The BU-3DFE database [141] used in Nair *et al.* [92] contains frontal only 3D facial datasets, which were created by the fusion of facial data acquired at $\pm 45^{\circ}$ yaw, from 100 subjects that perform seven universal expressions.

Comparative results of landmark localization errors on mixed (frontal and profile) facial datasets are presented in Table 16. To the best of our knowledge, Lu's method [84] is the only method in which localization errors on both frontal and profile facial datasets were presented. These results indicate that METHOD SISI–NPSS outperforms Lu's method in both accuracy and robustness. The proprietary MSU database used in Lu *et al.* [84] contains 300 3D facial scans from 100 subjects, three scans for each subject captured at 0 and $\pm 45^{\circ}$ yaw angles. The DB00F45RL database used in our experiments, despite having fewer subjects, is considered more challenging, since yaw angles lie in the range $[-65^{\circ}, +67^{\circ}]$ (Tables 12 and 13).

The inclusion of facial expressions into the FLMs and the use of separate shape index target values for each individual landmark resulted in an improved accuracy of our landmark detector (by up to 28%), and an improved detection rate (by up to 16%), compared to our early results that appeared in [97].

8.2.4 Evaluation of Fusion Schemes

For the purposes of this evaluation, two dataset collections were used: (i) the DB00F, and (ii) the DB00F45RL.

The evaluation of the performance of the proposed distance to similarity mappings and fusion schemes for landmark detection is not a straight-forward task, since there are many factors that characterize performance. As already stated, fusion techniques are expected to improve system's *accuracy*, *efficiency* and *robustness*. An equally important characteristic of a fusion scheme is that of *monotonicity*, i.e., the addition of a new feature descriptor should improve prior results.

Thus, performance is evaluated according to these four characteristics. *Accuracy* is evaluated according to the distance between the selected optimal landmark and the manually annotated landmark, which is considered as ground-truth. The selected optimal landmark is the 1^{st} rank candidate landmark for each landmark class (i.e., the candidate landmark which has the maximum resultant similarity score). *Efficiency* is evaluated according to the reduction of the likelihood area of a landmark class (see the high similarity areas in Figs. 38 and 37). The likelihood area of a landmark class is very important since its reduction means that fewer candidate landmarks have to be retained and fed to the "selection level". *Robustness* is evaluated by the use of testing datasets which contain subjects acquired under large yaw rotations, varying expressions and different illumination conditions, and also by the use of five different landmark classes. *Monotonicity* is evaluated according to the accuracy improvement between the use of individual descriptors, the fusion of the two richest descriptors, the shape index (SI) and the spin image (SS), and the fusion with the addition of a third poorer descriptor, the edge response (ER).

A qualitative performance evaluation of the proposed fusion schemes according to the aforementioned characteristics is presented in Table 7. Detailed landmark localization error analysis is presented in Tables 17 and 18.

Table 7: Qualitative evaluation of proposed fusion schemes

	Accuracy	Efficiency	Robustness	Monotonicity
L−L1	Fair	**High**	Fair	Fair
L−L2	Fair	Low	Fair	Fair
L−Lg	**High**	Fair	Fair	Fair
Q−L1	**High**	**High**	Fair	Fair
Q−L2	**High**	**High**	**High**	**High**
Q−Lg	**High**	Fair	Fair	Fair
G−L1	**High**	**High**	**High**	**High**
G−L2	**High**	**High**	Fair	Fair
G−Lg	**High**	Fair	Fair	Fair
L−Lmax	Low	Low	Low	Low
Q−Lmax	Low	Low	Low	Low
G−Lmax	Low	Low	Low	Low
L−Lmin	Unreliable	Fair	Fair	Low
Q−Lmin	Unreliable	Fair	Fair	Low
G−Lmin	Unreliable	Fair	Fair	Low

Current experimental findings are similar to those of [60], which are summarized in the following:
i) There is no single combination rule that scores best for all cases.
ii) Combining does not necessarily lead to improved performance.
iii) There are cases where none of the combining rules does better than the best individual detector.

Despite these general findings a more detailed examination of the results shows that there are some fusion schemes that perform better in most cases and can be adopted, and others that perform quite poorly and should be avoided (see also the Remarks of Section 5.2.2).

Current results show that, in general, the Quadratic (Q) and Gaussian (G) mappings behave better than the Linear (L) mapping. For the Linear mapping the product rule (Lg) behaves better than other rules. For the Quadratic mapping the rms rule (L2) behaves better than other rules. For the Gaussian mapping the sum rule (L1) behaves better than other rules. Quadratic and Gaussian mappings have almost the same performance.

The introduction of the Edge Response (ER) descriptor improves the results for the EOC, EIC and MC landmarks, but degrades the results for NT and CT. Note that, although ER is a poor descriptor, the improvement in accuracy is more dramatic in MC and EOC where the ER descriptor is more correlated with the SI and SS descriptors. Also note that the decline in accuracy is more dramatic in NT and CT where the ER descriptor is uncorrelated with the SI and SS descriptors (Table 5).

Accuracy improvement is more dramatic when the information fused is correlated. In correlated features the performance of one descriptor predicts to some extent the performance of the other and strengthens the results. On the other hand highly uncorrelated features have similarity peaks that do not coincide and degrade the results. Efficiency improvement is achieved by excluding obvious non-matches, reducing the number of candidate landmarks, for each landmark class. Fusion, also, reduces system sensitivity to sample-specific, poor-quality or erroneous descriptors.

We can thus deduce that the best performance in terms of accuracy is exhibited by the Q-L2 and G-L1 fusion schemes, with the Q-L2 exhibiting a slight better performance than the G-L1 in landmarks' likelihood area reduction. Q-L2 and G-L1 also exhibit high robustness in yaw, expression and illumination variations, and strong monotonicity.

Also landmark localization using the Q-L2 fusion scheme (see METHOD 5 in Section 6.3) improved the accuracy and robustness of the landmark detector (with $3.5 - 5.5$ mm mean landmark localization error), indicating the superiority of the fusion approach. Comparative results are presented in Table 19.

8.3 Partial Face Recognition

8.3.1 Test Databases

Combined UND Databases:
To evaluate the performance of the proposed partial face recognition method, a combination of the largest publicly available 3D face and ear databases was used. For frontal facial scans, the FRGC v2 database [107, 106] was used. It contains a total of 4,007 near frontal range images, obtained from 466 subjects having various facial expressions (e.g., happiness, surprise). For side facial scans, the Ear Database from the University of Notre Dame (UND) [131] was used. This database (which was created for ear recognition purposes) contains side scans with yaw rotations of 45°, 60° and 90°. In the 90° side scans, both sides of the face are occluded from the sensor; therefore they contain no useful information for face recognition purposes. Thus, only the 45° side scans (118 subjects, 118 left and 118 right) and the 60° side scans (87 subjects, 87 left and 87 right) were used. Even though the creators of the database marked these side scans as 45° and 60°, the measured maximum angle of rotation is 70° and 80° respectively (Table 13). However, when referring to these scans the database notation (45° and 60°) will be used. Unfortunately, not all subjects exist in both databases. The number of common subjects between the frontal scans and the 45° side scans is 39 and between the frontal scans and the 60° side scans is 32.

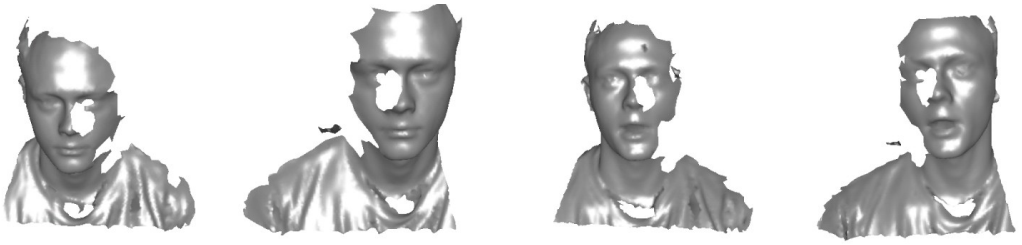
For the conducted experiments the following collections were defined:

- UND45LR: Contains 45° side scans from 118 subjects. For each subject, the left scan is considered gallery and the right is considered probe. *Total: 236 scans.*
- UND60LR: Contains 60° side scans from 87 subjects. For each subject, the left scan is considered gallery and the right is considered probe. *Total: 174 scans.*
- UND00LR: Gallery set has one frontal scan for each of the 466 subjects. Probe set has two 45° side scans (left and right) for each of the 39 subjects and two 60° side scans (left and right) for each of the 32 subjects. *Total: 608 scans.*

In all cases there is only one gallery scan per subject. Also, all subjects present in a probe set are also present in the gallery set (the opposite is not always true).

UH Databases:
In addition to the UND databases a database with data collected at the University of Houston was used. The database contains 1,075 left and 1,075 right scans of 281 subjects. The novelty of this database is that each pair of left and right side scans was acquired simultaneously (see

Fig. 54). They are acquired using a 3dMD system [64]. This system consists of one left and one right optical scanner that acquire data simultaneously but are independent of each other. These side scans are considered comparable to the UND's 45° scans. During the acquisition of the data, each subject was asked to remove any accessories (e.g., glasses). An initial scan was acquired while the subject assumed a neutral facial expression. Subsequently, several scans were acquired while the subject was reading loudly a predefined text (thus assuming arbitrary facial expressions). All scans of each subject were acquired on the same day.



(a) (b) (c) (d)

Figure 54: Scans from the UH database from a single subject: (a,b) Right and left scans with neutral expression were acquired simultaneously, (c,d) Right and left scans with open mouth were acquired simultaneously.

For the conducted experiments the following collections were defined:

- UHDB7L: Contains left side scans from 281 subjects. For each subject, one scan is considered gallery and the rest are considered probe. The minimum and maximum left scans per subject are 1 and 6 respectively. *Total: 1,075 scans.*
- UHDB7R: Contains right side scans from 281 subjects. For each subject, one scan is considered gallery and the rest are considered probe. The minimum and maximum right scans per subject are 1 and 6 respectively. *Total: 1,075 scans.*
- UHDB7LR-M: Contains multiple left and right side scan pairs from 281 subjects. For each subject, one left and one right scan (acquired simultaneously) are considered gallery and the rest are considered probes. The minimum and maximum pairs of scans (left and right) per subject are 1 and 6 respectively. *Total: 2,150 scans.*
- UHDB7LR-S: Contains a single left and right side scan pair from 281 subjects. For each subject, the left scan is considered gallery and the right scan is considered probe. The left and right scans were acquired simultaneously. *Total: 562 scans.*

In all cases there is only one gallery scan per subject (with the exception of UHDB7LR-M where there are two). Also, all subjects present in a probe set are also present in the gallery set (the opposite is not always true).

8.3.2 Landmark Detection Evaluation

In order to evaluate the performance of the landmark detection algorithm on the databases described in Section 8.3.1 we manually annotated landmarks on several facial scans. Great care was given to minimize the subjectiveness of the manual process so that it can be considered ground truth. Landmarks were manually annotated on the 466 frontal scans of

UND00LR (Fig. 50 (a)), on the 118 left and 118 right side scans of UND45LR (Fig. 50 (b-c)) and on the 87 left and 87 right side scans of UND60LR (Fig. 50 (d-e)).

In all cases, the overall mean distance error and its standard deviation between the manually annotated landmarks and the automatically detected landmarks was calculated to represent the landmarks' *localization error*. This error is expressed with the *mean Euclidian distance* D_{ME}, defined in Eq. 98, between the detected landmark points and the annotated landmark points.

Another metric which reflects the quality of the landmark detection algorithm for registration purposes is the *modified directed Hausdorff distance* D_{MH}, between the face model M and the test face T, which is defined in Eq. 101.

The $D_{MH}(M,T)$ expresses the mean value of the minimum Euclidian distances $|\mathbf{m}_i - \mathbf{t}_j|$ of the vertices of the face model M, to which a test facial scan T is registered. In order to compute this metric, only the automatically detected landmarks were used for registration (without the Simulated Annealing step). To get comparative results for D_{MH} we used as a model for frontal databases all the vertices of the complete AFM, for left-side databases the left side vertices of the AFM, and for right-side databases the right side vertices of the AFM.

Table 8: Summary results for landmark detection and face registration

Database	D_{MH}		D_{ME}		
	mean (mm)	stdev (mm)	mean (mm)	stdev (mm)	> 10 (mm)
UND00LR - front	4.61	1.04	5.77	1.81	3.4%
UND45LR - right	3.90	0.95	5.83	2.49	4.2%
UND45LR - left	4.03	1.22	6.02	2.45	6.8%
UND60LR - right	4.37	3.11	5.87	2.47	6.9%
UND60LR - left	4.32	2.41	6.08	2.53	11.5%
UHDB7R	4.72	2.46	-	-	-
UHDB7L	4.76	2.86	-	-	-

The results are summarized in Table 8. The last column reports the percentage of landmarks with D_{ME} more than 10 mm. Note that there were no manually annotated landmarks for the UHDB7R and UHDB7L databases, so only the D_{MH} metric is reported.

8.3.3 Face Recognition Performance Evaluation

Using the databases described in Section 8.3.1 several identification experiments were performed. The proposed method tackles the problem of matching arbitrary facial scans (left, right or frontal). This is considerably harder than matching only frontal scans, since a lot of the facial information is missing and it is not known a priori whether each scan is left, right or frontal. In all experiments the Cumulative Match Characteristic (CMC) graphs and the rank-one recognition rates are reported. The automatic landmark detector was used in all cases unless stated otherwise.

Matching facial scans of the same side:
In this experiment, the performance of the proposed method was evaluated using scans of

the same side for both gallery and probe sets. This is not a realistic scenario, but allows for the evaluation of the proposed method without the need to use facial symmetry. The only database suitable for this purpose is the UH Database as it has multiple left and multiple right side scans of each subject.

Table 9: Rank-one Recognition Rate between facial scans of the same side

	Rank-one Rate
UHDB7L	85.8%
UHDB7R	86.8%

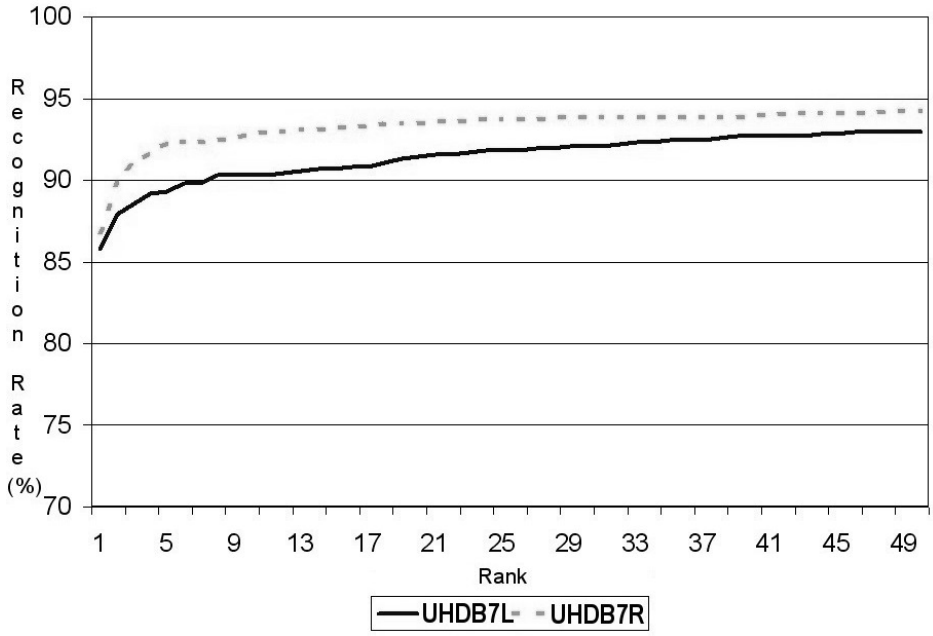


Figure 55: CMC graphs for matching left (gallery) with left (probe) side scans (for UHDB7L) and right (gallery) with right (probe) side scans (for UHDB7R).

The performance for matching left (gallery) with left (probe) side scans (UHDB7L) and right (gallery) with right (probe) side scans (UHDB7R) was measured. The CMC graphs are presented in Fig. 55 and the rank-one rates are given in Table 9.

Matching facial scans of arbitrary side:
In this experiment, the performance of the proposed method using scans of arbitrary sides for gallery and probe sets was evaluated. This is a realistic scenario, as the side scans (with extensive occlusions that lead to missing data) are very common in real world applications with unconstrained acquisition. The proposed method can match any combination of left, right or frontal facial scans with the use of facial symmetry. Moreover, the proposed method automatically detects the side of the scan. For this experiment we utilized the UND45LR, UND60LR, UND00LR, UHDB7LR-M and UHDB7LR-S databases and the rank-one rates are given in Table 10.

Table 10: Rank-one Recognition Rate between facial scans of arbitrary side

	Rank-one Rate
UND45LR	86.4%
UND60LR	81.6%
UND00LR	76.8%
UHDB7LR-M	89.1%
UHDB7LR-S	79.4%

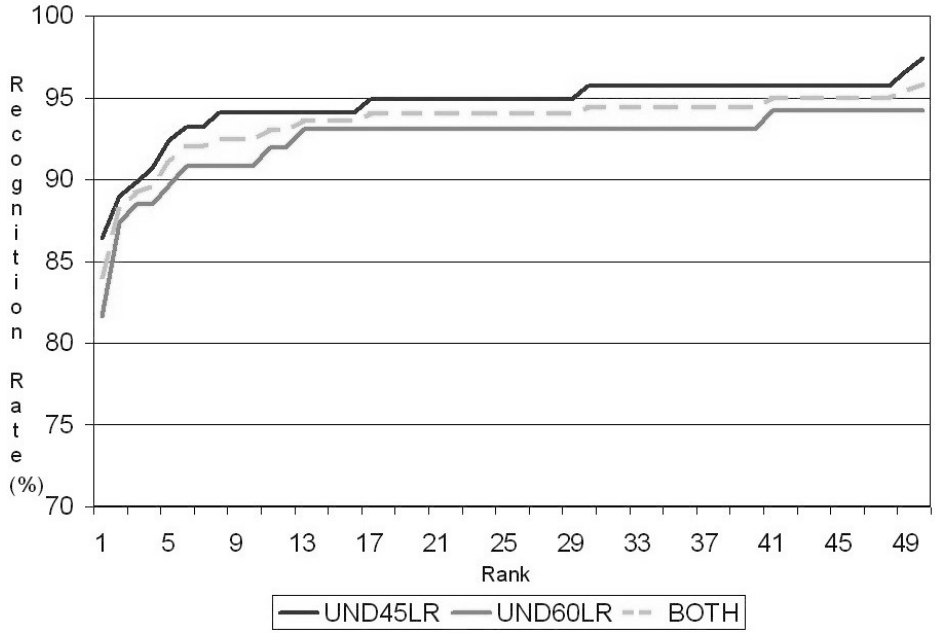


Figure 56: CMC graphs for matching left (gallery) with right (probe) side scans using UND45LR, UND60LR and the combination of the two.

In the cases of UND45LR and UND60LR, for each subject, the gallery set contains a single left side scan while the probe set contains a single right side scan. Therefore, facial symmetry is always used in order to perform identification. As expected, the 60° side scans yield lower results as they are considered more challenging compared to the 45° side scans (see Fig. 56).

In the case of UND00LR, the gallery set contains a frontal scan for each subject, while the probe set contains left and right side scans. This scenario is very common when the enrollment of subjects is controlled but the identification is uncontrolled. In Fig. 57 the CMC graph is given (UND00LR's probe set is also split in left-only and right-only subsets). Compared to UND45LR and UND60LR, there is a decrease in the performance of the proposed method in UND00LR. One could argue that since the gallery set consists of frontal scans (without missing data), there should be an increase in performance. However, UND00LR has the largest gallery set (it includes all of the 466 subjects found in the FRGC v2 database) making it the most challenging database in current experiments.

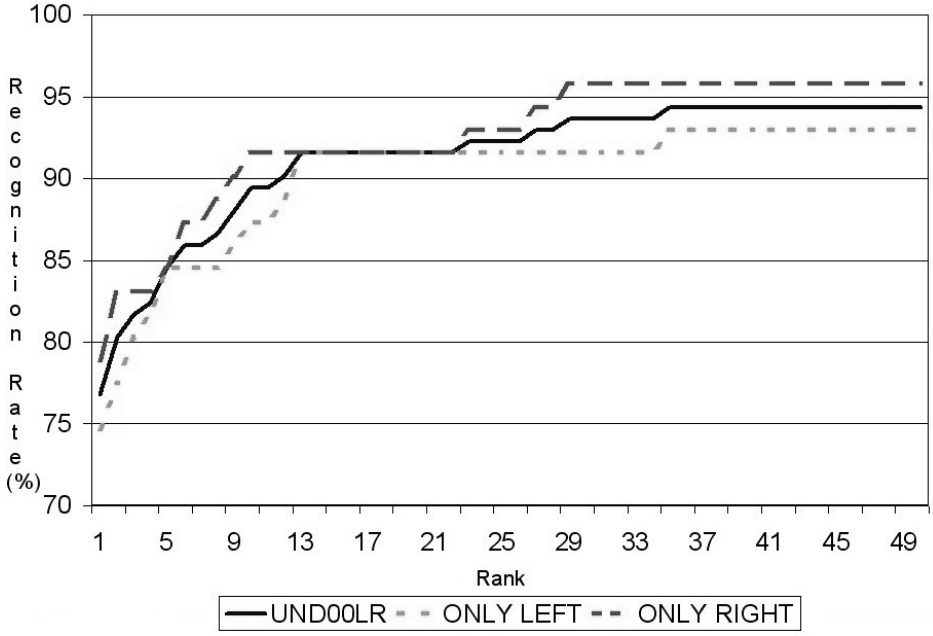


Figure 57: CMC graphs for matching frontal (gallery) with left, right and both (probe) side scans using UND00LR.

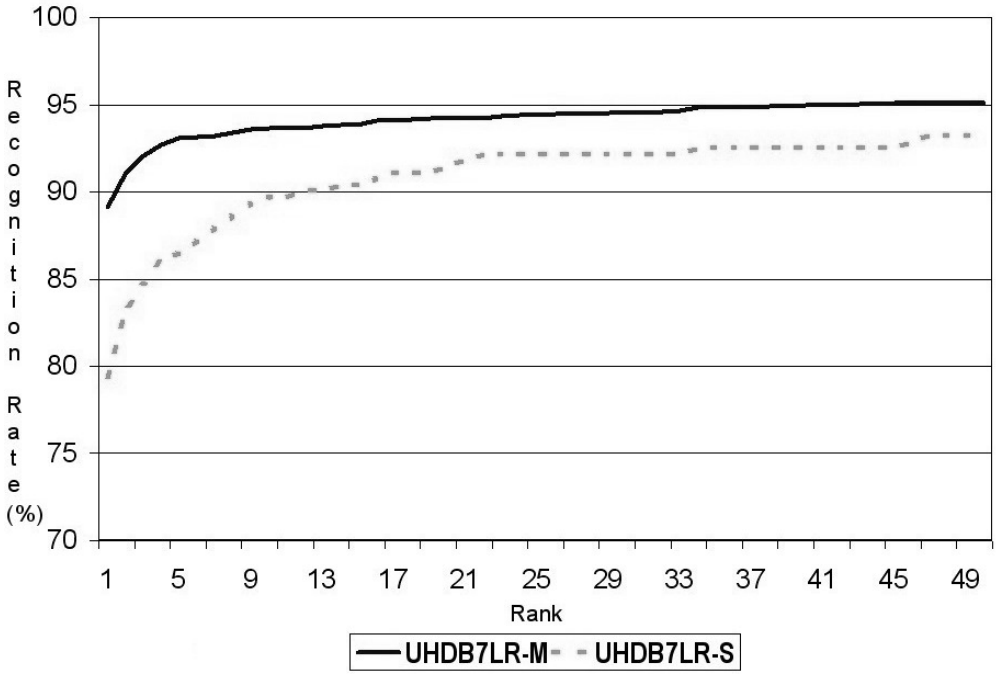


Figure 58: UHDB7LR-M: matching left and right (gallery) with left and right (probe) side scans. UHDB7LR-S: matching left (gallery) with right (probe) side scans.

In the case of UHDB7LR-M, for each subject, the gallery set contains a left and right side scan pair, while the probe set contains multiple left and right side scan pairs. As expected, since the gallery set has two scans per subject, the performance on this database is the highest among all databases. The performance difference is substantial compared to

UND00LR (89.1% versus 76.8% rank-one). This indicates that one pair of left and right side scans is more descriptive than one frontal scan.

In the case of UHDB7LR-S, even though it is a subset of UHDB7LR-M, is considered more challenging (see Fig. 58). This is because the gallery set contains a single left side scan while the probe set contains a single right side scan. Compared to UND45LR and UND60LR (which have a similar probe/gallery setup), the performance on UHDB7LR-S is lower (79.4% versus 86.4% and 81.6% rank-one). However UHDB7LR-S is considerably larger, it has 281 subjects versus 118 and 87 subjects for UND45LR and UND60LR respectively.

Automatic versus manual landmarks:
In the last experiment, the performance of the proposed method with an ideal landmark detector was evaluated. To this end, the manually annotated landmarks for the UND45LR and UND60LR databases described in Section 8.3.2 were used. The CMC graphs are depicted in Figs. 59 and 60 and the rank-one rates are given in Table 11.

Table 11: Rank-one Recognition Rate for automatic and manual landmarks

	Rank-one Rate Automatic Landmarks	*Rank-one Rate Manual Landmarks*
UND45LR	86.4%	91.5%
UND60LR	81.6%	90.8%

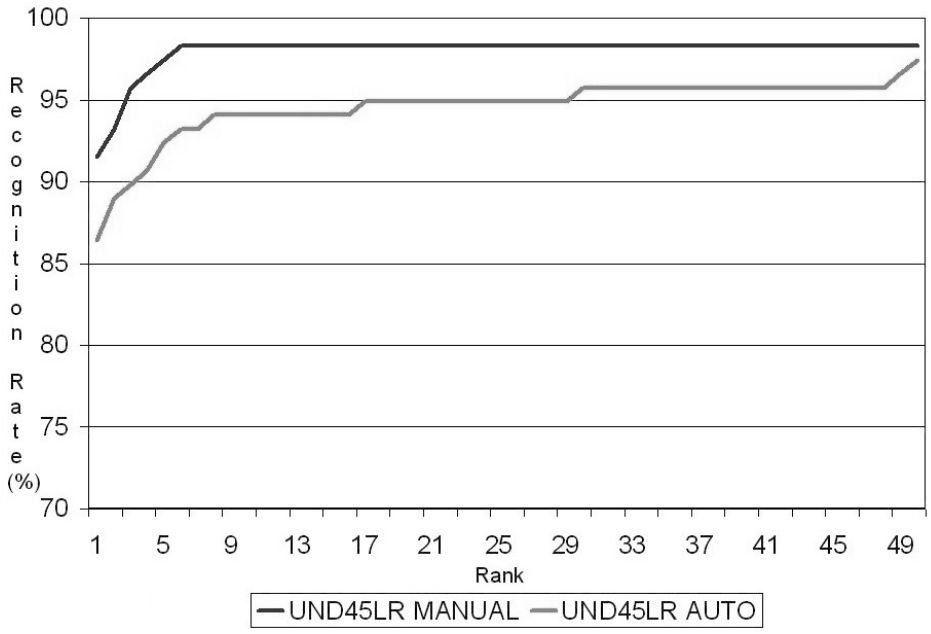


Figure 59: CMC graphs for matching left (gallery) and right (probe) side scans using automatic and manual landmarks on UND45LR

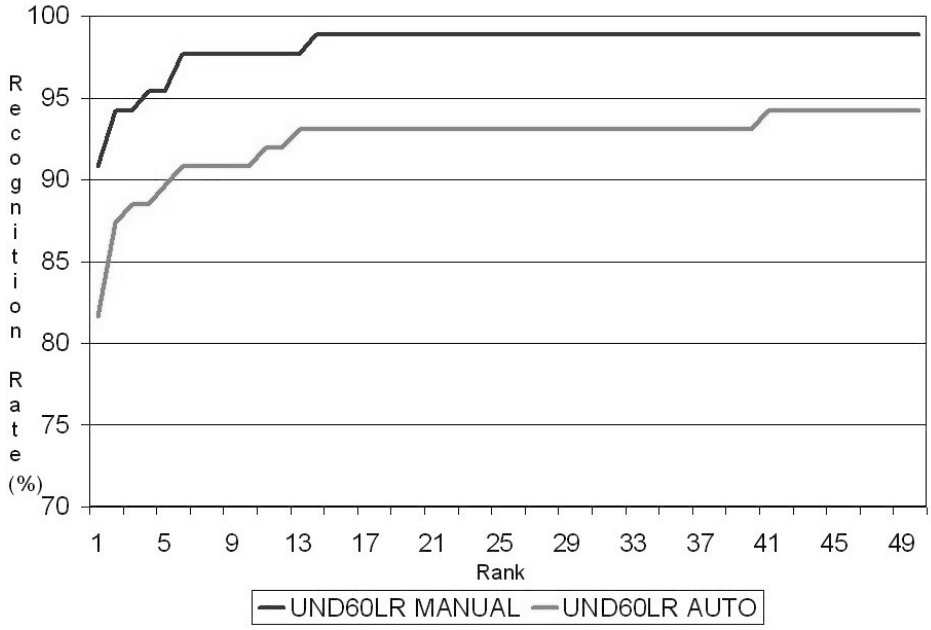


Figure 60: CMC graphs for matching left (gallery) and right (probe) side scans using automatic and manual landmarks on UND60LR

As expected, there is an increase in rank-one recognition rate for both databases (5.1% and 9.2% increase for UND45LR and UND60LR respectively).

Interestingly, the percentage of automatic landmarks with mean localization error more than 10 mm (compared to manually annotated landmarks) is on average 5.5% and 9.2% on UND45LR and UND60LR respectively (see the last column of Table 8). This indicates that the proposed method for partial face recognition can tolerate landmark mean localization errors up to 10 mm. If this error is significantly above this threshold, then the pose estimation is probably invalid and the subsequent fitting fails to extract meaningful geometrical information.

8.3.4 Discussion

As mentioned in Chapter 2, most of the face recognition methods that have been proposed do not handle data with significant pose variations along the yaw axis. On the contrary, the proposed method can handle extensive occlusions (that result in missing data) caused by such pose variations. The only limitation is that no more than half of the face is missing (so that facial symmetry can be used). The flexibility to seamlessly handle left, right and frontal scans is important in an unconstrained acquisition scenario. Therefore, it is considered that focusing on data with arbitrary pose variations is a necessity for real-world applications.

For evaluation purposes the most challenging databases in terms of pose variations and missing data were used. It is demonstrated that the proposed method can match frontal and left or right side scans by using facial symmetry. Unavoidably, the use of facial symmetry has an impact on recognition rates, as human faces are not completely symmetric. In the above experiments, the average rank-one recognition rate was 83.7%. It is anticipated that the average recognition rate can be increased by improving certain steps of the proposed method (e.g., landmark detection) before the limit imposed by facial asymmetry is reached.

Compared to the published work in [101], the currently proposed method, as published in [97], offers significantly better results. When using automatically detected landmarks, the rank-one recognition rates increased by 19% and 17%, for UND45LR and UND60LR respectively. This indicates that the landmark detector proposed in [97] is far more robust and accurate compared to the one proposed in [101]. When using manually selected landmarks, the rank-one recognition rates increased by 9% and 22%, for UND45LR and UND60LR respectively (with respect to [101]). Thus, the improvements on the other steps of the method also offer increased accuracy.

Another important difference of published work in [97] compared to the work published in [101] is that each frontal scan is currently handled as a pair of left and right side scans (producing two signatures that are matched independently). This is why the largest performance increase was on a database with frontal scans in the gallery set and side scans in the probe set (20% increased recognition rate on UND60LR compared to the average recognition rate of DB60F and DB45F in [101]).

8.4 Computational Cost

8.4.1 Landmark Detection

For the evaluation of the proposed landmark detection method's computational efficiency, a PC with the following specifications was used: Intel Core i5 2.5 GHz with 4 GB RAM. Using this PC, 6.68 *sec* on average was required to locate the landmarks for each facial scan. The average time taken for each step of the method is: Data loading 0.04 *sec*, shape index computation and landmark localization 0.26 *sec*, spin image computation and landmark filtering 0.31 *sec*, FLM5L-FLM5R matching and landmark labeling 5.05 *sec*, and FLM8 matching and optimal landmark set selection 1.02 *sec*. The procedures for determining the optimal rotation for the alignment of the landmark shapes to the FLMs require at most 8 iterations to converge.

Speedups through parallelization are also possible and thus the computational efficiency of the presented landmark detector makes it applicable to real-world applications.

8.4.2 Face Recognition

For the evaluation of the proposed partial face recognition method's computational efficiency, a PC with the following specifications was used: Intel Core 2 Duo 2.2 GHz with 2 GB RAM and NVIDIA GeForce 8600GTS graphics card. Using this PC, 18 *sec* on average are required to process a facial scan: 9 *sec* to localize the facial landmarks plus 9 *sec* to extract the biometric signature (geometry and normal images). The procedures of determining the optimal rotation for the alignment of the landmark shapes to the AFM require at most 8 iterations to converge. The Simulated Annealing step requires 2 *sec*. It may take up to 2,000 iterations to converge, but the computation is very efficient (requires 2 *sec*) as the z-buffers are created using the GPU. The fitting step takes 64 iterations to converge and requires 7 *sec*. The creation of the signature from the deformed AFM requires just a few milliseconds. Finally, the signatures can be matched at a rate of 15,000 $matches/sec$.

There are a number of independent tasks that can be speed up through parallel processing techniques. Nevertheless, the combination of a signature creation step within reasonable time (18 *sec*) and a signature matching procedure with an extremely low computational cost

(15,000 $matches/sec$) makes the proposed method for partial face recognition suitable for real-world scenarios.

Remark: Note that the presented times are only indicative and cannot be attributed solely to the methods' algorithmic procedures, since they were recorded in a typical Windows PC, where a multitasking OS is running.

9 Conclusion

There will come a time
when you believe everything is finished.
That will be the beginning.

– L. L' Amour

The automatic detection of facial landmarks is a key area in most facial processing applications as it often constitutes their first step. Examples of such applications (apart from facial recognition which has been widely explored in this thesis) include facial shape analysis, facial expression understanding, facial expression transfer for animation, facial motion capture etc.

In this thesis a novel automatic facial landmark detection method has been proposed. It offers pose invariance and robustness to large missing (self-occluded) facial areas with respect to large yaw variations and high tolerance to large expression variations. The proposed approach consists of methods for landmark localization that exploit the 3D facial geometry and the modeling ability of trained landmark models. It has been evaluated using the most challenging 3D facial databases available, which contain scans with yaw variations of up to 80° and strong expressions. In these databases it achieved state-of-the-art accuracy (with $4.5 - 6.3$ mm mean landmark localization error), significantly outperforming existing methods.

Although it is possible to consider extensions for improving accuracy (e.g., by including the nostrils' base or another anatomical landmark into the FLMs, or by applying heuristic methods of post-processing for fine-tuning the positions of landmarks), such improvements are likely to be marginal and at the expense of the method's simplicity and speed.

Also, a novel generalized framework of fusion methods and their application to landmark detection has been presented. The proposed fusion scheme acts after the "feature extraction level", transforms features to similarities and then combines them to generate a resultant feature similarity, which is considered as the matching score used at the "matching level" for the detection of the queried landmarks. The proposed feature fusion scheme is easily extensible to new feature-components in feature space, offers significant dimensionality reduction and works equally well for features extracted from 3D or 2D facial data.

For the proposed fusion scheme different distance to similarity mappings (linear, quadratic and Gaussian) and different fusion rules (sum rule, rms rule, product rule, max rule and min rule) have been evaluated according to *accuracy*, *efficiency*, *robustness* and *monotonicity*. The results indicate that the quadratic distance to similarity mapping in conjunction with the rms rule for fusion (Q-L2) exhibits the best performance. Landmark localization using this fusion scheme achieved state-of-the-art accuracy (with $3.5 - 5.5$ mm mean landmark localization error), indicating the superiority of the fusion approach.

Finally, a novel 3D face recognition method suitable for real-world biometric applica-

tions was proposed. Unlike most previous methods that require frontal scans, the proposed method can perform partial matching among interpose facial scans, even when extensive data are missing. It exploits the 3D landmark detector to provide an initial pose estimation and to indicate occluded areas with missing data for each facial scan. By using facial symmetry to complete missing facial data, it can handle seamlessly frontal and side facial scans. Competitive results were presented on databases with the most challenging pose variations. The proposed partial face recognition method exhibits state-of-the-art performance (with average rank-one recognition rate 83.7%), considerably outperforming existing methods.

Appendix

A Rotational Alignment of 3D Shapes

Given two 3D shapes a rotational transformation $R(\mathbf{x})$ has to be computed so as to minimize the Procrustes distance between the transformed shape $R(\mathbf{x})$ and a reference shape $\mathbf{x}_0$

$$E = \|R(\mathbf{x}) - \mathbf{x}_0\| \ , \tag{111}$$

which is considered as the alignment error.

The rotational transformation $\mathbf{R}$ can be expressed as a product of three rotations around the three principal axes:

$$\mathbf{R} = \mathbf{R}_{x,\theta} \cdot \mathbf{R}_{y,\phi} \cdot \mathbf{R}_{z,\psi} \tag{112}$$

These can be expressed in a matrix form:

$$\mathbf{R}_{x,\theta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \tag{113}$$

$$\mathbf{R}_{y,\phi} = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \tag{114}$$

$$\mathbf{R}_{z,\psi} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{115}$$

To simplify the notation, we define the following sums:

$$S_{x0,x} = \sum_{j=1}^{n} x_{0j}x_j, \quad S_{x0,y} = \sum_{j=1}^{n} x_{0j}y_j, \quad S_{x0,z} = \sum_{j=1}^{n} x_{0j}z_j,$$

$$S_{y0,x} = \sum_{j=1}^{n} y_{0j}x_j, \quad S_{y0,y} = \sum_{j=1}^{n} y_{0j}y_j, \quad S_{y0,z} = \sum_{j=1}^{n} y_{0j}z_j,$$

$$S_{z0,x} = \sum_{j=1}^{n} z_{0j}x_j, \quad S_{z0,y} = \sum_{j=1}^{n} z_{0j}y_j, \quad S_{z0,z} = \sum_{j=1}^{n} z_{0j}z_j,$$

$$S_{x,y} = \sum_{j=1}^{n} x_j y_j, \qquad S_{y,z} = \sum_{j=1}^{n} y_j z_j, \qquad S_{z,x} = \sum_{j=1}^{n} z_j x_j,$$

$$S_{x,x} = \sum_{j=1}^{n} x_j x_j, \qquad S_{y,y} = \sum_{j=1}^{n} y_j y_j, \qquad S_{z,z} = \sum_{j=1}^{n} z_j z_j.$$

where n is the number of landmarks of each shape.

We also consider the similarity transformations:

$$\mathbf{M}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & -b \\ 0 & b & a \end{bmatrix} \tag{116}$$

which represents a rotation around x-axis by $\theta = \tan^{-1}\left(\frac{b}{a}\right)$ and a scaling by $s^2 = a^2 + b^2$,

$$\mathbf{M}_y = \begin{bmatrix} a & 0 & b \\ 0 & 1 & 0 \\ -b & 0 & a \end{bmatrix} \tag{117}$$

which represents a rotation around y-axis by $\phi = \tan^{-1}\left(\frac{b}{a}\right)$ and a scaling by $s^2 = a^2 + b^2$, and

$$\mathbf{M}_z = \begin{bmatrix} a & -b & 0 \\ b & a & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{118}$$

which represents a rotation around z-axis by $\psi = \tan^{-1}\left(\frac{b}{a}\right)$ and a scaling by $s^2 = a^2 + b^2$.

After applying the $\mathbf{M}_x$ transformation we have:

$$E = \sum_{j=1}^{n} [(M_x(x_j) - x_{0j})^2 + (M_x(y_j) - y_{0j})^2 + (M_x(z_j) - z_{0j})^2] \Rightarrow$$

$$E = \sum_{j=1}^{n} [(x_j - x_{0j})^2 + (ay_j - bz_j - y_{0j})^2 + (by_j + az_j - z_{0j})^2] \ . \tag{119}$$

Setting partial derivatives of E w.r.t. each parameter to zero we obtain that:

$$\frac{\partial E}{\partial a} = \sum_{j=1}^{n} [(ay_j - bz_j - y_{0j})y_j + (by_j + az_j - z_{0j})z_j] = 0 \Rightarrow$$

$$aS_{y,y} - bS_{y,z} + bS_{y,z} + aS_{z,z} = S_{y0,y} + S_{z0,z} \Rightarrow$$

$$a = \frac{S_{y0,y} + S_{z0,z}}{S_{y,y} + S_{z,z}} \ . \tag{120}$$

and

$$\frac{\partial E}{\partial b} = \sum_{j=1}^{n} [-(ay_j - bz_j - y_{0j})z_j + (by_j + az_j - z_{0j})y_j] = 0 \Rightarrow$$

$$-aS_{y,z} - bS_{z,z} + bS_{y,y} + aS_{y,z} = -S_{y0,z} + S_{z0,y} \Rightarrow$$

$$b = \frac{S_{z0,y} - S_{y0,z}}{S_{y,y} + S_{z,z}} \ . \tag{121}$$

finally resulting in:

$$\theta = \tan^{-1}\left(\frac{S_{z0,y} - S_{y0,z}}{S_{y0,y} + S_{z0,z}}\right) \tag{122}$$

After applying the $\mathbf{M}_y$ transformation we have:

$$E = \sum_{j=1}^{n} [(M_y(x_j) - x_{0j})^2 + (M_y(y_j) - y_{0j})^2 + (M_y(z_j) - z_{0j})^2] \Rightarrow$$

$$E = \sum_{j=1}^{n} [(ax_j - bz_j - x_{0j})^2 + (y_j - y_{0j})^2 + (-bx_j + az_j - z_{0j})^2] \ . \tag{123}$$

Setting partial derivatives of E w.r.t. each parameter to zero we obtain that:

$$\frac{\partial E}{\partial a} = \sum_{j=1}^{n} [(ax_j - bz_j - x_{0j})y_j + (-bx_j + az_j - z_{0j})z_j] = 0 \Rightarrow$$

$$aS_{x,x} + bS_{z,x} - bS_{z,x} + aS_{z,z} = S_{x0,x} + S_{z0,z} \Rightarrow$$

$$a = \frac{S_{x0,x} + S_{z0,z}}{S_{x,x} + S_{z,z}} \ . \tag{124}$$

and

$$\frac{\partial E}{\partial b} = \sum_{j=1}^{n} [(ax_j - bz_j - x_{0j})z_j + (-bx_j + az_j - z_{0j})x_j] = 0 \Rightarrow$$

$$aS_{z,x} + bS_{z,z} + bS_{x,x} - aS_{z,x} = S_{x0,z} - S_{z0,x} \Rightarrow$$

$$b = \frac{S_{x0,z} - S_{z0,x}}{S_{x,x} + S_{z,z}} \ . \tag{125}$$

finally resulting in:

$$\phi = \tan^{-1}\left(\frac{S_{x0,z} - S_{z0,x}}{S_{z0,z} + S_{x0,x}}\right) \tag{126}$$

After applying the $\mathbf{M}_z$ transformation we have:

$$E = \sum_{j=1}^{n} [(M_z(x_j) - x_{0j})^2 + (M_z(y_j) - y_{0j})^2 + (M_z(z_j) - z_{0j})^2] \Rightarrow$$

$$E = \sum_{j=1}^{n} [(ax_j - by_j - x_{0j})^2 + (bx_j + ay_j - y_{0j})^2 + (z_j - z_{0j})^2] \ . \tag{127}$$

Setting partial derivatives of E w.r.t. each parameter to zero we obtain that:

$$\frac{\partial E}{\partial a} = \sum_{j=1}^{n} [(ax_j - by_j - x_{0j})x_j + (bx_j + ay_j - y_{0j})y_j] = 0 \Rightarrow$$

$$aS_{x,x} - bS_{x,y} + bS_{x,y} + aS_{y,y} = S_{x0,x} + S_{y0,y} \Rightarrow$$

$$a = \frac{S_{x0,x} + S_{y0,y}}{S_{x,x} + S_{y,y}} \ . \tag{128}$$

and

$$\frac{\partial E}{\partial b} = \sum_{j=1}^{n} [-(ax_j - by_j - x_{0j})y_j + (bx_j + ay_j - y_{0j})x_j] = 0 \Rightarrow$$

$$-aS_{x,y} + bS_{y,y} + bS_{x,x} + aS_{x,y} = -S_{x0,y} + S_{y0,x} \Rightarrow$$

$$b = \frac{S_{y0,x} - S_{x0,y}}{S_{y,y} + S_{z,z}} \ . \tag{129}$$

finally resulting in:

$$\psi = \tan^{-1}\left(\frac{S_{y0,x} - S_{x0,y}}{S_{x0,x} + S_{y0,y}}\right) \tag{130}$$

Setting the values of θ, ϕ, and ψ into Eqs. 113, 114, and 115 we approximate the similarity transformations by its rotational component, ignoring the scaling factors. Thus, repeatedly applying the extracted rotational transformations, the alignment error is reduced below a predefined threshold.

B Line–Triangle Intersection

When resampling a surface mesh, it is necessary to compute the intersection between a line segment and a triangle [129].

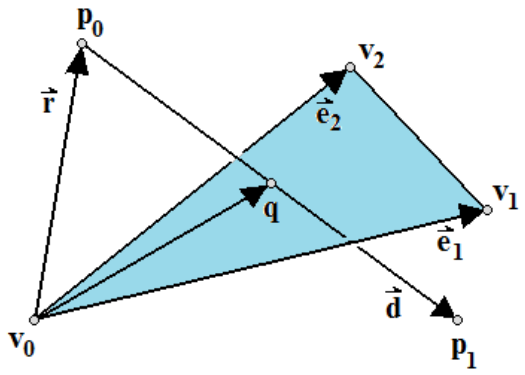


Figure 61: Intersection of a line segment and a triangle in 3D space.

Consider a triangle $(\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2)$ and a line segment $(\mathbf{p}_0, \mathbf{p}_1)$. The line segment can be written in parametric form as

$$\mathbf{L}(t) = \mathbf{p}_0 + t\mathbf{d} \ , \tag{131}$$

with $0 \leq t \leq 1$ and

$$\mathbf{d} = \mathbf{p}_1 - \mathbf{p}_0 \ . \tag{132}$$

Their intersection point $\mathbf{q}(x, y, z)$ can be written in parametric form as

$$\mathbf{q}(u, v) = \mathbf{v}_0 + u(\mathbf{v}_1 - \mathbf{v}_0) + v(\mathbf{v}_2 - \mathbf{v}_0) \ , \tag{133}$$

where $u, v \geq 0$ and $u + v \leq 1$ for the point to be inside the triangle.

Since $\mathbf{q}(x, y, z)$ satisfies Eq. 131, we have

$$\mathbf{p}_0 + t\mathbf{d} = \mathbf{v}_0 + u(\mathbf{v}_1 - \mathbf{v}_0) + v(\mathbf{v}_2 - \mathbf{v}_0) \ \Rightarrow$$

$$\mathbf{p}_0 - \mathbf{v}_0 = -t\mathbf{d} + u(\mathbf{v}_1 - \mathbf{v}_0) + v(\mathbf{v}_2 - \mathbf{v}_0) \ .$$

Setting $\mathbf{e}_1 = \mathbf{v}_1 - \mathbf{v}_0$, $\mathbf{e}_2 = \mathbf{v}_2 - \mathbf{v}_0$, and $\mathbf{r} = \mathbf{p}_0 - \mathbf{v}_0$, we have

$$\mathbf{r} = \begin{bmatrix} -\mathbf{d} & \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} \cdot \begin{bmatrix} t \\ u \\ v \end{bmatrix} \ \Rightarrow$$

$$\begin{bmatrix} t \\ u \\ v \end{bmatrix} = \begin{bmatrix} -\mathbf{d} & \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix}^{-1} \cdot \mathbf{r} \ \Rightarrow$$

$$\begin{bmatrix} t \\ u \\ v \end{bmatrix} = \frac{1}{\begin{vmatrix} -\mathbf{d} & \mathbf{e}_1 & \mathbf{e}_2 \end{vmatrix}} \cdot \begin{bmatrix} \begin{vmatrix} \mathbf{r} & \mathbf{e}_1 & \mathbf{e}_2 \end{vmatrix} \\ \begin{vmatrix} -\mathbf{d} & \mathbf{r} & \mathbf{e}_2 \end{vmatrix} \\ \begin{vmatrix} -\mathbf{d} & \mathbf{e}_1 & \mathbf{r} \end{vmatrix} \end{bmatrix} \cdot \mathbf{r} \ , \tag{134}$$

where $|\ \mathbf{a}\ \mathbf{b}\ \mathbf{c}\ | = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$.

Thus, finally

$$t = \frac{(\mathbf{r} \times \mathbf{e}_1) \cdot \mathbf{e}_2}{(-\mathbf{d} \times \mathbf{e}_1) \cdot \mathbf{e}_2} \ , \tag{135}$$

$$u = \frac{(-\mathbf{d} \times \mathbf{r}) \cdot \mathbf{e}_2}{(-\mathbf{d} \times \mathbf{e}_1) \cdot \mathbf{e}_2} \ , \tag{136}$$

$$v = \frac{(-\mathbf{d} \times \mathbf{e}_1) \cdot \mathbf{r}}{(-\mathbf{d} \times \mathbf{e}_1) \cdot \mathbf{e}_2} \ , \tag{137}$$

and

$$\mathbf{q}(x, y, z) = \mathbf{p}_0 + t\mathbf{d} \ . \tag{138}$$

C 3D Landmark Detection Results

Table 12: Experiment 1: Performance of METHOD SISI–NPSS against yaw variations

Database	**DB00F**			**DB00F45RL**		
Side Detection Rate	974 / 975		99.90%	117 / 117		100.00%
Yaw mean ± stdev	$+0.93° \pm 4.03°$			$+1.15° \pm 41.35°$		
Yaw [min ... max]	$[-17.98° \ \ldots \ +16.98°]$			$[-65.23° \ \ldots \ +66.82°]$		
Localization Error (mm)	mean	stdev	≤ 10 mm	mean	stdev	≤ 10 mm
REOC	5.58	3.33	88.82%	5.32	3.71	88.46%
REIC	4.15	2.35	96.92%	4.65	2.45	96.15%
LEIC	4.41	2.49	97.33%	4.90	2.96	96.15%
LEOC	5.83	3.42	88.10%	6.06	4.13	80.77%
NT	4.09	2.41	98.56%	4.41	2.68	98.29%
MRC	5.56	3.93	88.62%	5.01	2.97	92.31%
MLC	5.42	3.84	88.82%	4.91	2.88	96.15%
CT	4.92	3.74	94.77%	4.80	3.52	93.16%
Mean Error	5.00	1.85	97.85%	4.97	1.92	97.44%

Table 13: Experiment 1: Performance of METHOD SISI–NPSS against yaw variations

Database	**DB45R**			**DB45L**			**DB60R**			**DB60L**		
Side Detection Rate	118 / 118		100.00%	118 / 118		100.00%	86 / 87		98.85%	87 / 87		100.00%
Yaw mean ± stdev	$+44.20° \pm 8.20°$			$-45.57° \pm 8.95°$			$+57.47° \pm 7.22°$			$-58.51° \pm 8.06°$		
Yaw [min ... max]	$[+16.81° \ \ldots \ +68.04°]$			$[-69.22° \ \ldots \ -16.19°]$			$[+30.24° \ \ldots \ +80.81°]$			$[-82.52° \ \ldots \ -30.79°]$		
Localization Error (mm)	mean	stdev	≤ 10 mm	mean	stdev	≤ 10 mm	mean	stdev	≤ 10 mm	mean	stdev	≤ 10 mm
REOC	5.63	3.76	85.59%	-	-	-	6.01	3.52	80.46%	-	-	-
REIC	4.71	2.69	95.76%	-	-	-	4.89	3.11	93.10%	-	-	-
LEIC	-	-	-	5.05	2.79	94.92%	-	-	-	5.01	3.02	95.40%
LEOC	-	-	-	5.42	3.42	88.98%	-	-	-	5.14	3.19	91.95%
NT	4.87	2.90	95.76%	4.64	2.81	95.76%	3.94	2.35	96.55%	4.13	1.85	100.00%
MRC	4.84	3.50	88.98%	-	-	-	4.68	3.36	91.95%	-	-	-
MLC	-	-	-	4.21	2.93	96.61%	-	-	-	5.37	5.12	85.06%
CT	5.12	4.95	91.53%	4.45	3.57	96.61%	5.23	4.72	89.66%	6.86	6.03	85.06%
Mean Error	5.03	1.92	96.61%	4.75	1.91	97.46%	4.95	1.80	96.55%	5.30	2.49	93.10%

Table 14: Experiment 2: METHOD SISI–NPSS tolerance to expression variations on DB00F

Expression	**Neutral**			**Mild**			**Extreme**			**All**		
Side Detection Rate	443 / 443		100.00%	355 / 355		100.00%	176 / 177		99.44%	974 / 975		99.90%
Localization Error (mm)	mean	stdev	≤ 10 mm	mean	stdev	≤ 10 mm	mean	stdev	≤ 10 mm	mean	stdev	≤ 10 mm
REOC	5.38	3.14	90.97%	5.76	3.42	88.17%	5.71	3.57	84.75%	5.58	3.33	88.82%
REIC	3.95	2.19	97.52%	4.28	2.35	97.46%	4.38	2.68	94.35%	4.15	2.35	96.92%
LEIC	4.37	2.51	98.19%	4.48	2.33	96.90%	4.40	2.74	96.05%	4.41	2.49	97.33%
LEOC	5.66	3.37	89.16%	5.95	3.38	87.61%	6.02	3.59	86.44%	5.83	3.42	88.10%
NT	3.99	2.24	99.10%	3.92	2.06	98.59%	4.67	3.25	97.18%	4.09	2.41	98.56%
MRC	4.25	2.30	99.10%	5.36	3.10	90.14%	9.26	5.88	59.32%	5.56	3.93	88.62%
MLC	4.35	2.40	97.52%	5.21	3.14	91.27%	8.55	5.87	62.15%	5.42	3.84	88.82%
CT	4.21	2.36	98.42%	4.66	2.70	96.34%	7.27	6.45	82.49%	4.92	3.74	94.77%
Mean Error	4.52	1.51	99.32%	4.95	1.46	99.72%	6.28	2.60	90.40%	5.00	1.85	97.85%

Table 15: Comparison of METHOD SISI–NPSS against state-of-the-art on almost-frontal complete facial datasets

Mean Localization Error (mm)										
Method		Test DB (scans)	REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
Yu *et al.* [142]	(GA model)	FRGC v1 (200)	4.74	5.59	-	-	2.18	-	-	-
Nair *et al.* [92]	(w/o PDM)	BU-3DFE (2350)	25.01	26.68	31.84	34.39	14.59	-	-	-
	(w PDM)		12.11	11.89	20.46	19.38	8.83	-	-	-
Lu *et al.* [84]	(3D)	FRGC v1 (953)	8.30	8.20	9.50	10.30	8.30	-	6.00	6.20
Lu *et al.* [83]	(3D+2D)	FRGC v1 (946)	6.00	5.70	7.10	7.90	5.00	-	3.60	3.60
Colbry [20]	(w/o CFDM)	FRGC v1 (953)	5.50	6.30	-	-	4.10	11.00	6.90	6.70
	(w CFDM)	+ propr. (160)	5.60	6.00	-	-	4.00	11.70	5.40	5.40
Perakis *et al.* [103]	(SISI–NP)	FRGC v2 (975)	7.02	7.46	8.13	9.21	5.23	6.71	8.30	9.83
Passalis *et al.* [97]	(UR3D-S)	FRGC v2 (975)	5.03	5.48	5.79	5.62	4.91	6.31	5.65	6.47
Perakis *et al.* [100]	(SISI–NPSS)	FRGC v2 (975)	4.15	4.41	5.58	5.83	4.09	4.92	5.56	5.42
Stdev of Localization Error (mm)										
Method		Test DB (scans)	REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
Yu *et al.* [142]	(GA model)	FRGC v1 (200)	9.76	16.08	-	-	6.83	-	-	-
Nair *et al.* [92]	(w/o PDM)	BU-3DFE (2350)	-	-	-	-	-	-	-	-
	(w PDM)		-	-	-	-	-	-	-	-
Lu *et al.* [84]	(3D)	FRGC v1 (953)	17.20	17.20	17.10	18.10	19.40	-	16.90	17.90
Lu *et al.* [83]	(3D+2D)	FRGC v1 (946)	3.30	3.00	5.90	5.10	2.40	-	3.30	2.90
Colbry [20]	(w/o CFDM)	FRGC v1 (953)	4.90	5.00	-	-	5.10	7.60	8.60	9.30
	(w CFDM)	+ propr. (160)	4.80	4.70	-	-	5.40	7.30	6.80	6.70
Perakis *et al.* [103]	(SISI–NP)	FRGC v2 (975)	3.18	3.07	3.79	4.25	3.28	4.32	4.53	4.47
Passalis *et al.* [97]	(UR3D-S)	FRGC v2 (975)	2.47	2.59	3.45	3.47	2.49	4.43	4.34	4.26
Perakis *et al.* [100]	(SISI–NPSS)	FRGC v2 (975)	2.35	2.49	3.33	3.42	2.41	3.74	3.93	3.84

Table 16: Comparison of METHOD SISI–NPSS against state-of-the-art on mixed (frontal and profile) facial datasets

Mean Localization Error (mm)										
Method		Test DB (scans)	REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
Lu *et al.* [84]	(3D)	MSU (300)	9.00	7.10	13.60	13.30	6.40	-	6.70	5.20
Passalis *et al.* [97]	(UR3D-S)	FRGC v2 + Ear (117)	5.97	6.87	6.51	6.71	4.60	6.59	5.52	6.10
Perakis *et al.* [100]	(SISI–NPSS)	FRGC v2 + Ear (117)	4.65	4.90	5.32	6.06	4.41	4.80	5.01	4.91
Stdev of Localization Error (mm)										
Method		Test DB (scans)	REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
Lu *et al.* [84]	(3D)	MSU (300)	13.10	9.20	11.90	10.10	13.40	-	12.90	9.00
Passalis *et al.* [97]	(UR3D-S)	FRGC v2 + Ear (117)	3.13	2.92	3.68	3.76	3.01	4.16	3.58	4.17
Perakis *et al.* [100]	(SISI–NPSS)	FRGC v2 + Ear (117)	2.45	2.96	3.71	4.13	2.68	3.52	2.97	2.88

D Feature Fusion Results

Table 17: Landmark localization error (mm) results of Shape Index (SI), Spin Image (SS) and Edge Response (ER) fusion, in DB00F and DB00F45RL

DB00F – Landmark localization error (mm)						
	EOC	EIC	NT	MC	CT	Mean
SI	11.72	7.71	14.66	**5.98**	10.81	10.18
SS	**7.31**	**4.42**	**3.84**	8.47	**7.56**	**6.32**
ER	12.26	13.05	10.54	9.27	11.74	11.37
L−L1	6.40	4.60	4.12	**4.82**	**7.16**	5.42
L−L2	6.72	4.74	4.19	4.78	7.24	5.53
L−Lg	**6.31**	**4.52**	**4.08**	4.85	7.23	**5.40**
Q−L1	6.21	4.15	3.97	4.90	7.31	5.31
Q−L2	**6.19**	**4.14**	3.97	**4.87**	**7.28**	**5.29**
Q−Lg	6.20	4.15	**3.95**	4.92	7.29	5.30
G−L1	6.19	**4.14**	**3.97**	**4.86**	**7.28**	**5.29**
G−L2	**6.16**	4.15	3.98	4.89	**7.28**	**5.29**
G−Lg	6.21	4.15	**3.97**	4.90	7.31	5.31
L−Lmax	11.93	11.57	14.66	8.45	11.63	11.65
Q−Lmax	12.17	11.50	14.69	8.49	12.05	11.78
G−Lmax	12.17	11.50	14.69	8.49	12.05	11.78
L−Lmin	7.21	3.97	3.88	5.23	8.41	5.74
Q−Lmin	7.21	3.97	3.88	5.23	8.41	5.74
G−Lmin	7.21	3.97	3.88	5.23	8.41	5.47

DB00F45RL – Landmark localization error (mm)						
	EOC	EIC	NT	MC	CT	Mean
SI	10.99	7.20	12.51	**4.68**	11.26	9.33
SS	**9.16**	**4.83**	**3.68**	7.03	**7.24**	**6.39**
ER	11.31	12.10	11.79	9.16	12.29	11.33
L−L1	**6.97**	**4.94**	4.40	**4.09**	**7.56**	**5.59**
L−L2	7.22	5.11	4.88	**4.09**	7.57	5.77
L−Lg	6.98	4.95	**4.20**	4.14	7.69	**5.59**
Q−L1	6.89	**4.59**	3.82	**3.83**	7.80	5.39
Q−L2	6.80	**4.59**	3.82	**3.83**	**7.73**	**5.35**
Q−Lg	**6.77**	**4.59**	**3.80**	**3.83**	7.79	5.36
G−L1	**6.80**	**4.59**	**3.82**	**3.83**	**7.73**	**5.35**
G−L2	6.85	4.64	3.84	**3.83**	**7.73**	5.38
G−Lg	6.89	**4.59**	**3.82**	**3.83**	7.80	5.39
L−Lmax	11.89	10.86	12.51	7.91	11.96	11.03
Q−Lmax	12.01	10.79	12.51	7.91	12.44	11.13
G−Lmax	12.01	10.79	12.51	7.91	12.44	11.13
L−Lmin	8.53	4.64	3.53	4.42	7.88	5.80
Q−Lmin	8.53	4.64	3.53	4.42	7.88	5.80
G−Lmin	8.53	4.64	3.53	4.42	7.88	5.80

Table 18: Landmark localization error (mm) results of Shape Index (SI) and Spin Image (SS) fusion, in DB00F and DB00F45RL

DB00F – Landmark localization error (mm)						
	EOC	EIC	NT	MC	CT	Mean
SI	11.72	7.71	14.66	**5.98**	10.81	10.18
SS	**7.31**	**4.42**	**3.84**	8.47	**7.56**	**6.32**
L–L1	7.58	4.81	**3.85**	5.85	7.30	5.88
L–L2	7.70	4.84	**3.85**	5.81	**7.16**	**5.87**
L–Lg	**7.54**	**4.80**	**3.85**	**5.80**	7.38	**5.87**
Q–L1	7.54	4.73	**3.84**	**5.84**	**7.28**	5.85
Q–L2	**7.52**	**4.72**	3.85	**5.84**	**7.28**	**5.84**
Q–Lg	7.53	4.73	3.85	5.87	7.29	5.85
G–L1	**7.52**	**4.72**	3.85	**5.84**	**7.28**	**5.84**
G–L2	7.53	**4.72**	**3.84**	**5.84**	**7.28**	**5.84**
G–Lg	7.54	4.73	**3.84**	**5.84**	**7.28**	5.85
L–Lmax	11.72	7.71	14.66	6.06	10.81	10.19
Q–Lmax	11.72	7.72	14.66	6.06	10.81	10.19
G–Lmax	11.72	7.72	14.66	6.06	10.81	11.78
L–Lmin	7.34	4.61	3.84	5.91	7.39	5.82
Q–Lmin	7.34	4.61	3.84	5.91	7.39	5.82
G–Lmin	7.34	4.61	3.84	5.91	7.39	5.82

DB00F45RL – Landmark localization error (mm)						
	EOC	EIC	NT	MC	CT	Mean
SI	10.99	7.20	12.51	**4.68**	11.26	9.33
SS	**9.16**	**4.83**	**3.68**	7.03	**7.24**	**6.39**
L–L1	8.82	5.11	**3.67**	5.04	7.38	6.00
L–L2	8.80	5.06	**3.67**	5.03	7.53	6.02
L–Lg	**8.53**	**5.05**	**3.67**	**4.99**	**7.35**	**5.92**
Q–L1	8.39	4.98	**3.62**	**4.72**	**7.53**	5.85
Q–L2	**8.33**	**4.97**	**3.62**	**4.72**	**7.53**	**5.83**
Q–Lg	8.39	**4.97**	**3.62**	**4.72**	7.54	5.85
G–L1	**8.33**	**4.97**	**3.62**	**4.72**	**7.53**	**5.83**
G–L2	8.34	**4.97**	3.67	**4.72**	**7.53**	5.85
G–Lg	8.39	4.98	**3.62**	**4.72**	**7.53**	5.85
L–Lmax	11.00	7.23	12.51	4.68	11.26	9.34
Q–Lmax	10.99	7.20	12.51	4.68	11.26	9.33
G–Lmax	10.99	7.20	12.51	4.68	11.26	9.33
L–Lmin	8.53	4.64	3.53	4.42	7.88	5.80
Q–Lmin	9.20	4.88	3.51	5.03	7.27	5.98
G–Lmin	9.20	4.88	3.51	5.03	7.27	5.98

Table 19: Comparison of performance of Q−L2 Fusion Method against SISI–NPSS

Database	**DB00F**								
Method	SISI–NPSS			Q−L2(SI+SS)			Q−L2(SI+SS+ER)		
Side Detection Rate	974 / 975		99.90%	974 / 975		99.90%	975 / 975		**100.00**%
Yaw mean ± stdev	+0.93° ± 4.03°			+1.05° ± 3.81°			+1.40° ± 3.79°		
Yaw [min ... max]	[−17.98° ... + 16.98°]			[−17.37° ... + 17.68°]			[−17.30° ... + 17.82°]		
Localization Error (mm)	mean	stdev	≤ 10 mm	mean	stdev	≤ 10 mm	mean	stdev	≤ 10 mm
REOC	5.58	3.33	88.82%	4.78	3.06	93.54%	**4.48**	2.94	**95.08**%
REIC	4.15	2.35	96.92%	4.00	2.26	97.95%	**3.53**	2.23	**98.26**%
LEIC	4.41	2.49	**97.33**%	**4.13**	2.46	**97.33**%	4.26	2.54	96.62%
LEOC	5.83	3.42	88.10%	**5.39**	3.24	90.05%	5.53	3.16	**91.28**%
NT	4.09	2.41	98.56%	**3.65**	2.33	**98.77**%	3.78	2.39	98.67%
MRC	5.56	3.93	88.62%	4.21	4.01	92.10%	**3.91**	3.77	**92.41**%
MLC	5.42	3.84	88.82%	**4.48**	3.89	**91.08**%	4.51	3.98	90.97%
CT	4.92	3.74	94.77%	4.13	3.65	95.79%	**4.09**	3.43	**96.21**%
Mean Error	5.00	1.85	97.85%	4.34	1.91	98.05%	**4.26**	1.86	**98.46**%

Database	**DB00F45RL**								
Method	SISI–NPSS			Q−L2(SI+SS)			Q−L2(SI+SS+ER)		
Side Detection Rate	117 / 117		**100.00**%	117 / 117		**100.00**%	117 / 117		**100.00**%
Yaw mean ± stdev	+1.15° ± 41.35°			+0.93° ± 41.25°			+1.21° ± 41.27°		
Yaw [min ... max]	[−65.23° ... + 66.82°]			[−68.19° ... + 68.95°]			[−67.44° ... + 70.22°]		
Localization Error (mm)	mean	stdev	≤ 10 mm	mean	stdev	≤ 10 mm	mean	stdev	≤ 10 mm
REOC	5.32	3.71	88.46%	4.74	3.17	**93.59**%	**4.60**	3.40	**93.59**%
REIC	4.65	2.45	96.15%	4.35	2.01	**98.72**%	**4.10**	2.37	**98.72**%
LEIC	4.90	2.96	**96.15**%	**4.86**	3.23	93.59%	5.23	3.53	91.03%
LEOC	6.06	4.13	80.77%	**5.39**	3.16	87.18%	5.48	3.32	**91.03**%
NT	4.41	2.68	**98.29**%	**4.14**	2.77	97.44%	4.51	2.77	97.44%
MRC	5.01	2.97	92.31%	4.02	1.95	**100.00**%	**3.54**	1.91	98.72%
MLC	4.91	2.88	96.15%	**3.73**	2.64	**97.44**%	4.21	4.01	94.87%
CT	**4.80**	3.52	93.16%	4.86	3.44	**94.87**%	5.29	4.06	90.60%
Mean Error	4.97	1.92	**97.44**%	**4.60**	1.77	**97.44**%	4.79	1.80	**97.44**%

Notation

$\mathbb{Z}$	Integers
$\mathbb{Z}^{m \times n}$	$m \times n$ integer grid
$\mathbb{R}$	Real numbers
$\mathbb{R}^m$	m-dimensional vector space
$\mathbb{R}^{m \times n}$	$m \times n$ vector space
$\mathcal{C}^m$	Class of m-times continuously differentiable functions
$\mathcal{C}^\infty$	Class of smooth functions
$\lvert \cdot \rvert$	Absolute value
$\lVert \cdot \rVert$	Norm
$\lVert \cdot \rVert_p$	p-norm
$\mathbf{a} = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}^T$	Vector $\mathbf{a} \in \mathbb{R}^m$
$a_i = [a]_i = a(i)$	i^{th} component of a vector
$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$	Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$
$a_{ij} = [A]_{ij} = A(i,j)$	(i,j) element of a matrix
$\mathbf{A}_j = \mathbf{A}(:,j)$	j^{th} column-vector of a matrix
$\mathbf{A}^T$, $\mathbf{a}^T$	Transpose of a matrix or vector
$\mathbf{A}^{-1}$	Inverse of a matrix
$det(\mathbf{A})$	Determinant of a matrix
$trace(\mathbf{A})$	Trace of a matrix
$rank(\mathbf{A})$	Rank of a matrix
$diag(\mathbf{A})$	Diagonal elements of a matrix in vector form
$\mathbf{1}$	Identity matrix
$\odot$	Element-wise (Hadamard) product
f , $\mathbf{f}$	Scalar/vector function
df	Differential of f
∇f	Gradient of f
$\nabla^2 f$	Laplacian of f
$\Pr[\cdot]$	Probability
$\mathrm{Var}[\cdot]$	Variance
$\mathrm{Covar}[\cdot,\cdot]$	Covariance
D	Distance measure
S	Similarity measure
g	Low-pass Haar filter
h	High-pass Haar filter
$\leftarrow$, $\rightarrow$	Mapping
$:=$	Assignment
$\ldots$	Up to
$\#$	Number (of)
$O(\cdot)$	Algorithm complexity

Acronyms

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
AAM	Active Appearance Model
AFM	Annotated Face Model
AR	Ohio State University Face DB
ASM	Active Shape Model
AUEB	Athens University of Economics and Business, Greece
B/W	Black and White image - gray scale color space
BAM	Boosted Appearance Model
BMP	BitMaP image
BRM	Boosted Ranking Model
BU3DFE	3D Facial Expressions DB of University of New York at Binghampton
CCW	Counter-Clock-Wise
CIE	Commission Internationale de L' Eclairage
CLM	Constrained Local Model
CM	Center of Mass or centroid
CMC	Cumulative Match Characteristic
CMU-PIE	Carnegie Mellon University, Pose, Expression, and Illumination Face DB
CPU	Central Processing Unit
CT	Chin Tip
CW	Clock-Wise
DB	Data Base
DDSAC	Data-Driven SAmple Consensus
EER	Equal Error Rate
EOC	Eye Outer Corner
EIC	Eye Inner Corner
EM	Expectation Maximization
FAR	False Acceptance Rate
FEM	Finite Element Method
FERET	Facial Recognition Technology DB
FLM	Facial Landmark Model
FRGC	Face Recognition Grand Challenge
FRR	False Rejection Rate
FRVT	Face Recognition Vendor Test
GPU	Graphics Processing Unit
HSV	Hue, Saturation, Value color space
ICP	Iterative Closest Point
IMRA	Isotropic Multi-Resolution Analysis
IT	Information Technology
k-NN	k-Nearest Neighbor
KLT	Karhunen-Loéve Transform
$L^*a^*b^*$	CIE perceptually equalized color space
LBO	Laplace-Beltrami Operator

LEIC	Left Eye Inner Corner
LEOC	Left Eye Outer Corner
LFW	Labeled Faces in the Wild database
LFPW	Labeled Face Parts in the Wild database
MC	Mouth Corner
ML	Maximum Likelihood
MLC	Mouth Left Corner
MRC	Mouth Right Corner
MSE	Mean Square Error
MSU	Michigan State University, USA
NCSR	National Centre for Scientific Research of Greece
NDOff	University of Notre Dame Off Pose Facial DB
NIST	National Institute of Standards and Technology, USA
NT	Nose Tip
NTNU	Norwegian University of Science and Technology
NTSC	National Television Standards Committee
OS	Operating System
PC	Personal Computer
PCA	Principal Component Analysis
pdf	probability density function
PDM	Point Distribution Model
RAM	Random Access Memory
RANSAC	RANdom SAmple Consensus
RFM	Reference Face Model
RGB	Red, Green, Blue color space
rms	root-mean-square
REIC	Right Eye Inner Corner
REOC	Right Eye Outer Corner
ROC	Receiver Operating Characteristic
SIFT	Scale Invariant Feature Transform
stdev	standard deviation
SVM	Support Vector Machine
T3DFRD	Texas 3D Face Recognition Database
UH	University of Houston, Texas, USA
UND	University of Notre Dame, USA
UoA	National & Kapodistrian University of Athens, Greece
XYZ	Tristimulus color space
YCbCr	Component digital video color space
YIQ	NTSC tristimulus TV color space

References

[1] P. Alliez, G. Ucelli, C. Gotsman, and M. Attene, "Recent advances in remeshing of surfaces," in *Shape Analysis and Structuring, Mathematics and Visualization*. Springer, 2008.

[2] AR, "Ohio State University Face Database," http:// www2.ece.ohio-state.edu/ ∼aleix/ ARdatabase.html, 2012.

[3] C. Bär, *Elementary Differential Geometry*. Cambridge University Press, 2010.

[4] P. Belhumeur, D. Jacobs, D. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," in *Proc.* 24^{th} *IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, USA, Jun. 21-25 2011, pp. 545–552.

[5] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 13, pp. 1373–1396, 2003.

[6] M. Belkin, J. Sun, and Y. Wang, "Discrete Laplace operator on meshed surfaces," in *Proc.* 24^{th} *Annual Symp. on Computational Geometry*, College Park, MD, USA, 2008, pp. 278–287.

[7] M. Berger, *A Panoramic View of Riemannian Geometry*. Springer, 2003.

[8] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[9] BIO-ID, "BioID Face Database," http:// www.bioid.com/ index.php? q=downloads/ software/ bioid-face-database.html, 2012.

[10] V. Blanz, K. Scherbaum, and H.-P. Seidel, "Fitting a morphable model to 3D scans of faces," in *Proc.* 11^{th} *IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, Oct. 14-20 2007, pp. 1–8.

[11] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[12] C. Boehnen and T. Russ, "A fast multi-modal approach to facial feature detection," in *Proc.* 7^{th} *IEEE Workshop on Applications in Computer Vision*, vol. 1, Jan. 5-7 2005, pp. 135–142.

[13] F. L. Bookstein, "Size and shape spaces for landmark data in two dimensions," *Statistical Science*, vol. 1, pp. 181–242, 1986.

[14] E. Bossè, A. Guitouni, and P. Valin, "An essay to characterise information fusion systems," in *Proc.* 9^{th} *International Conference on Information Fusion*, Florence, Italy, Jul. 10-13 2006, pp. 1–7.

[15] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, Jan. 2006.

[16] A. Bronstein, M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *Int'l J. Computer Vision*, vol. 64, no. 1, pp. 5–30, 2005.

[17] ——, "Robust expression-invariant face recognition from partially missing data," in *Proc. European Conference on Computer Vision*, Graz, Austria, 2006, pp. 396–408.

[18] ——, *Numerical Geometry of Non-Rigid Shapes*. Springer, 2008.

[19] K. Chang, K. Bowyer, and P. J. Flynn, "An evaluation of multi-modal 2D+3D face biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 619–624, Apr. 2005.

[20] D. Colbry, "Human face verification by robust 3D surface alignment," Ph.D. dissertation, Michigan State University, 2006.

[21] D. Colbry, G. Stockman, and A. Jain, "Detection of anchor points for 3D face verification," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, Jun. 20-25 2005, p. 118.

[22] C. Conde, R. Cipolla, L. J. Rodrígez-Aragón, A. Serrano, and E. Cabello, "3D facial feature location with spin images," in *Proc. IAPR Conference on Machine Vision Applications*, Tsukuba Science City, Japan, May 16 – 18 2005, pp. 418–421.

[23] T. Cootes and C. Taylor, "Statistical models of appearance for computer vision," University of Manchester, Tech. Rep., Oct. 2001.

[24] T. Cootes, C. Taylor, H. Kang, and V. Petrovic, *Handbook of Face Recognition*. Springer, 2005, ch. Modeling Facial Shape and Appearance, pp. 39–63.

[25] T. Cootes, K. Walker, and C. Taylor, "View-based active appearance models," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, Mar. 26-30 2002, pp. 227–232.

[26] T. Cootes and C. Taylor, "Active shape models: Smart snakes," in *Proc. British Machine Vision Conference*, Leeds, UK, Sep. 22-24 1992.

[27] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[28] D. Cristinacce and T. Cootes, "Automatic feature localization with constrained local models," *Pattern Recognition*, vol. 41, no. 10, pp. 3054–3067, Oct. 2008.

[29] ——, "Boosted regression active shape models," in *Proc. British Machine Vision Conference*, University of Warwick, United Kingdom, Sep. 10-13 2007, pp. 880–889.

[30] M. Dantone, J. Gall, G. Fanelli, and L. van Gool, "Real-time facial feature detection using conditional regression forests," in *Proc.* 25^{th} *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 16-21 2011.

[31] M. Desbrun, M. Meyer, P. Schröder, and A. H. Barr, "Implicit fairing of irregular meshes using diffusion and curvature flow," in *Proc. SIGGRAPH*, 1999, pp. 317–324.

[32] H. Dibeklioğlu, "Part-based 3D face recognition under pose and expression variations," Master's thesis, Boğaziçi University, 2008.

[33] H. Dibeklioğlu, A. Salah, and L. Akarun, "3D facial landmarking under expression, pose, and occlusion variations," in *Proc. 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*, Arlington, VA, Sep. 20 - Oct. 1 2008, pp. 1–6.

[34] C. Dorai and A. K. Jain, "COSMOS - a representation scheme for 3D free-form objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 10, pp. 1115–1130, Oct. 1997.

[35] I. Dryden and K. Mardia, *Statistical Shape Analysis*. Wiley, 1998.

[36] R. Duin, "The Combining Classifier: To train or not to train?" in *Proc. 16th International Conference on Pattern Recognition*, vol. 2, Quebec City, Canada, Aug. 11-15 2002, pp. 765–770.

[37] S. Edelman, "Representation of similarity in 3D object discrimination," *Neural Computation*, vol. 7, pp. 407–422, 1995.

[38] S. Edelman and H. H. Buelthoff, "Orientation depedence in the recognition of familiar and novel views of 3D objects," *Vision Research*, vol. 32, pp. 2385–2400, 1992.

[39] B. Efraty, M. Papadakis, A. Profitt, S. Shah, and I. Kakadiaris, "Facial component-landmark detection," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, Mar. 21-25 2011, pp. 278–285.

[40] ——, "Pose invariant facial component-landmark detection," in *Proc. 18th IEEE International Conference on Image Processing*, Sept. 11-14 2011, pp. 569–572.

[41] T. Faltemier, K. Bowyer, and P. Flynn, "A region ensemble for 3-D face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 62–73, Mar. 2008.

[42] ——, "Rotated profile signatures for robust 3D feature detection," in *Proc. 8th IEEE International Conference on Automatic Face and Gesture Recognition*, Amsterdam, The Netherlands, Sep. 17-19 2008, pp. 1–7.

[43] L. Farkas, *Anthropometry of the head and face*, 2nd ed., L. G. Farkas, Ed. Raven Press, 1994.

[44] P. Felzenszwalb and D. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, Jan. 2005.

[45] A. Finkel and J. Bentley, "Quad trees: A data structure for retrieval of composite keys," *Acta Informatica*, vol. 4, no. 1, pp. 1–9, 1974.

[46] Y. Gao, "Efficiently comparing face images using a modified Hausdorff distance," in *Proc. IEEE Conference on Vision, Image and Signal Processing*, Dec. 2003, pp. 346–350.

[47] B. Gökberk and L. Akarun, "Comparative analysis of decision-level fusion algorithms for 3D face recognition," in *Proc.* 18^{th} *International Conference on Pattern Recognition*, vol. 3, Hong Kong, China, Aug. 20-24 2006, pp. 1018 – 1021.

[48] G. H. Golub and C. F. van Loan, *Matrix Computations*. Johns Hopkins University Press, 1996.

[49] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2^{nd} ed. Prentice-Hall, 2002.

[50] R. Gross, *Handbook of Face Recognition*. Springer, 2005, ch. Face Databases, pp. 301–327.

[51] L. Gu and T. Kanade, "3D alignment of face in a single image," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, NY, Jun. 17-22 2006, pp. 1305–1312.

[52] ——, "A generative shape regularization model for robust face alignment," in *Proc.* 10^{th} *European Conference on Computer Vision*, Marseille, France, Oct. 12-18 2008, pp. 413–426.

[53] X. Gu, S. Gortler, and H. Hoppe, "Geometry images," in *Proc. SIGGRAPH*, San Antonio, TX, Jul. 2002, pp. 355–361.

[54] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, "Texas 3D face recognition database," in *Proc. IEEE Southwest Symposium on Image Analysis & Interpretation*, May 23-25 2010, pp. 97–100.

[55] S. Gutta, V. Philomin, and M. Trajkovic, "An investigation into the use of partial-faces for face recognition," in *Proc.* 5^{th} *IEEE international Conference on Automatic Face and Gesture Recognition*, Washington DC, May 20-21 2002, pp. 28–33.

[56] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc.* 4^{th} *Alvey Vision Conference*, 1988, pp. 147–151.

[57] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, Oct. 2007.

[58] R. J. K. Jacob, "The face as a data display," *Human Factors*, vol. 18, pp. 189–200, 1976.

[59] S. Jahanbin, H. Choi, and A. Bovik, "Passive multimodal 2-D+3-D face recognition using gabor features and landmark distances," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 4, pp. 1287–1304, Dec. 2011.

[60] A. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Inteligence*, vol. 22, no. 1, pp. 4–37, Jan. 2000.

[61] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, Dec. 2005.

[62] A. E. Johnson, "Spin Images: A Representation for 3-D Surface Matching," Ph.D. dissertation, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Aug. 1997.

[63] I. A. Kakadiaris, G. Passalis, G. Toderici, E. Efraty, P. Perakis, D. Chu, S. K. Shah, and T. Theoharis, "Face recognition using 3D images," in *Handbook of Face Recognition*, 2^{nd} ed., S. Li and A. K. Jain, Eds. Springer-Verlag, July 2010, pp. 5–30.

[64] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 640–649, Apr. 2007.

[65] I. Kakadiaris, G. Passalis, G. Toderici, P. Perakis, and T. Theoharis, "3D-based face recognition," in *Encyclopedia of Biometrics*, S. Li, Ed. New York, NY: Springer, 2009, pp. 329–338.

[66] T. Kanade, "Picture processing by computer complex and recognition of human faces," Ph.D. dissertation, Kyoto University, 1973.

[67] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contours models," *International Journal of Computer Vision*, 1988.

[68] D. G. Kendall, "Shape manifolds, procrustean metrics and complex projective spaces," *Bulletin of the London Mathematical Society*, vol. 16, pp. 81–121, 1984.

[69] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 22, no. 4598, pp. 671–680, 1983.

[70] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis an Machine Inteligence*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[71] A. Klinger, "Pattern and search statistics," *Optimizing Methods in Statistics*, pp. 303–337, 1971.

[72] J. Koenderink and A. van Doorn, "Surface shape and curvature scales," *Image and Vision Computing*, vol. 10, pp. 557–565, Oct. 1992.

[73] L. Kompanets, "Biometrics of asymmetrical face," in *Proc.* 1^{st} *International Conference on Biometric Authentication (ICBA)*, 2004, pp. 67–73.

[74] LFPW, "Labeled Face Parts in the Wild Database," http:// www.kbvt.com/ LFPW/, 2012.

[75] S. Li and A. K. Jain, in *Handbook of Face Recognition*. Springer, 2005, ch. Introduction, pp. 1–11.

[76] Z.-N. Li and M. S. Drew, *Fundamentals of Multimedia*. Pearson Education, 2004.

[77] L. Liang, R. Xiao, F. Wen, and J. Sun, "Face alignment via component-based discriminative search," in *Proc. European Conference on Computer Vision*, Marseille, France, Oct. 12-18 2008, pp. 72–85.

[78] T. Lin, W. Shih, W. Chen, and W. Ho, "3D face authentication by mutual coupled 3D and 2D feature extraction," in *Proc.* 44^{th} *ACM Southeast Regional Conference*, Melbourne, FL, Mar. 10 – 12 2006, pp. 423–427.

[79] X. Liu, "Discriminative face alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1941–1954, Nov. 2009.

[80] Y. Liu and J. Palmer, "A quantified study of facial asymmetry in 3D faces," in *Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, Oct. 17 2003, pp. 222–229.

[81] Y. Liu, R. L. Weaver, K. Schmidt, N. Serban, and J. Cohn, "Facial asymmetry: A new biometric," Carnegie Mellon University - Robotics Institute, Tech. Rep. CMU-RI-TR-01-23, August 2001.

[82] C. Loop, "Smooth subdivision surfaces based on triangles," Master's thesis, Department of Mathematics, University of Utah, 1987.

[83] X. Lu and A. Jain, "Multimodal facial feature extraction for automatic 3D face recognition," Michigan State University, Tech. Rep. MSU-CSE-05-22, Oct. 2005.

[84] ——, "Automatic feature extraction for multiview 3D face recognition," in *Proc.* 7^{th} *International Conference on Automatic Face and Gesture Recognition*, Southampton, UK, Apr. 10-12 2006, pp. 585–590.

[85] X. Lu, A. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 31–43, 2006.

[86] C. Mandal, "A dynamic framework for subdivision surfaces," Ph.D. dissertation, University of Florida, 1998.

[87] M. Meyer, M. Desbrun, P. Schröder, and A. H. Barr, "Discrete differential geometry operators for triangulated 2-manifolds," *Visualization and Mathematics III*, pp. 35–57, 2003.

[88] A. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1927–1943, Nov. 2007.

[89] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Proc.* 10^{th} *European Conference on Computer Vision*, Marseille, France, Oct. 12-18 2008, pp. 504–513.

[90] S. Mitra, M. Savvides, and B. V. Kumar, "Face identification using novel frequency-domain representation of facial asymmetry," in *Proc. IEEE Transactions on Information Forensics and Security*, Sep. 2006, pp. 350–359.

[91] P. Nair and A. Cavallaro, "Matching 3D faces with partial data," in *Proc. British Machine Vision Conference*, 2008.

[92] ——, "3-D face detection, landmark localization, and registration using a point distribution model," *IEEE Transactions on Multimedia*, vol. 11, no. 4, pp. 611–623, June 2009.

[93] B. O'Neil, *Elementary Differential Geometry*. Academic Press, 1997.

[94] S. E. Palmer, E. Rosh, and P. Case, in *Attention and Performance*, J. Long and A. Baddeley, Eds. Lawrence Erlbaum Associates, Hillsdale, N.J., 1981, ch. Canonical perspective and perception of objects, pp. 135–151.

[95] G. Papaioannou, E. Karabassi, and T. Theoharis, "Reconstruction of three-dimensional objects through matching of their parts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 114–124, Jan. 2002.

[96] G. Passalis, I. Kakadiaris, and T. Theoharis, "Intra-class retrieval of non-rigid 3D objects: Application to face recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 218–229, 2007.

[97] G. Passalis, P. Perakis, T. Theoharis, and I. Kakadiaris, "Using facial symmetry to handle pose variations in real-world 3D face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1938–1951, Oct. 2011.

[98] H.-O. Peitgen, H. Jürgens, and D. Saupe, *Chaos & Fractals: New Frontiers of Science*, 2^{nd} ed. Springer, 2004.

[99] P. Perakis, G. Passalis, T. Theoharis, and I. Kakadiaris, "3D facial landmark detection & face registration: A 3D facial landmark model & 3D local shape descriptors approach," Computer Graphics Laboratory, University of Athens, Tech. Rep. TP-2010-01, Jan. 2010.

[100] ——, "3D facial landmark detection under large yaw and expression variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1552–1564, July 2013.

[101] P. Perakis, G. Passalis, T. Theoharis, G. Toderici, and I. Kakadiaris, "Partial matching of interpose 3D facial data for face recognition," in *Proc. 3^{rd} IEEE International Conference on Biometrics: Theory, Applications and Systems*, Arlington, VA, Sep. 28-30 2009, pp. 439–446.

[102] P. Perakis and T. Theoharis, "Statistical Landmark Models: An ASM approach," Computer Graphics Laboratory, University of Athens, Tech. Rep. TP-2008-03, Dec. 2008.

[103] P. Perakis, T. Theoharis, G. Passalis, and I. Kakadiaris, "Automatic 3D facial region retrieval from multi-pose facial datasets," in *Proc. Eurographics Workshop on 3D Object Retrieval*, Munich, Germany, Mar. 30 - Apr. 3 2009, pp. 37–44.

[104] D. I. Perrett, M. W. Oram, M. H. Harries, R. Bevan, J. K. Hietanen, P. J. Benson, and S. Thomas, "Viewer-centered and object-centered coding of heads in the Macaque temporal cortex," *Experimental Brain Research*, vol. 86, pp. 159–173, 1991.

[105] M. Petrou and C. Petrou, *Image Processing: The Fundamentals*, 2^{nd} ed. Whiley and Sons, Ltd, 2010.

[106] P. Phillips, T. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 large-scale experimental results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 831–846, 2010.

[107] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, 2005, pp. 947–954.

[108] E. Praun and H. Hoppe, "Spherical parametrization and remeshing," in *Proc. SIGGRAPH*, San Diego, CA, July 2003, pp. 340–349.

[109] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes - The Art of Scientific Computing*, 3^{rd} ed. Cambridge University Press.

[110] G. Rhodes, in *Cognitive and Computational Aspects of Face Recognition*, T. Valentine, Ed. Routledge, N.Y., 1995, ch. Face recognition and configurational coding.

[111] S. Romdhani, V. Blanz, C. Basso, and T. Vetter, *Handbook of Face Recognition*. Springer, 2005, ch. Morphable Models of Faces, pp. 217–245.

[112] S. Romdhani and T. Vetter, "3D probabilistic feature point model for object detection and recognition," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, Jun. 17-22 2007, pp. 1–8.

[113] M. Romero-Huertas and N. Pears, "3D facial landmark localization by matching simple descriptors," in *Proc.* 2^{nd} *IEEE International Conference on Biometrics: Theory, Applications and Systems*, Arlington, VA, Sep. 20 - Oct. 1 2008.

[114] A. Ross and R. Govindarajan, "Feature level fusion using hand and face biometrics," in *Proc. SPIE Conference on Biometric Technology for Human Identification II*, Orlando, USA, Mar. 2005, pp. 196–204.

[115] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, 2003.

[116] A. Ross and A. K. Jain, "Biometrics, overview," in *Encyclopedia of Biometrics*, S. Li, Ed. New York, NY: Springer, 2009, pp. 168–172.

[117] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, *Biometrics and Identity Management: First European Workshop, BIOID 2008*. Springer Berlin Heidelberg, 2008, vol. 5372, ch. Bosphorus Database for 3D Face Analysis, pp. 47–56.

[118] A. V. Segal, D. Haehnel, and S. Thrun, "Generalized-ICP," in *Robotics: Science and Systems*, 2009.

[119] M. Segundo, C. Queirolo, O. Bellon, and L. Silva, "Automatic 3D facial segmentation and landmark detection," in *Proc.* 14^{th} *International Conference on Image Analysis and Processing*, Modena, Italy, Sep. 10-14 2007, pp. 431–436.

[120] L. G. Shapiro and G. C. Stockman, *Computer Vision*. Prentice-Hall, 2001.

[121] G. Sharp, S. Lee, and D. Wehe, "ICP registration using invariant features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 90–102, Jan. 2002.

[122] P. Siarry, G. Berthiau, F. Durbin, and J. Haussy, "Enhanced simulated annealing for globally minimizing functions of many-continuous variables," *ACM Transactions on Mathematical Software*, vol. 23, no. 2, pp. 209–228, 1997.

[123] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces," Robotics Institute, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-01-02, Jan. 2001.

[124] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain, "Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems," *IEEE Transactions on Pattern Analysis an Machine Inteligence*, vol. 27, no. 3, pp. 450–455, Mar. 2005.

[125] M. Stegman and D. Gomez, "A brief introduction to statistical shape analysis," Technical University of Denmark, Tech. Rep., Mar. 2002.

[126] E. Stollnitz, T. DeRose, and D. Salesin, *Wavelets for computer graphics: Theory and applications*. Morgan Kaufmann Publishers, Inc, 1996.

[127] L. Tarassenko and M. Denham, in *Cognitive Systems: Information Processing Meets Brain Science*. Elsevier, 2006, ch. Sensory Processing, pp. 85–104.

[128] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 3^{rd} ed. Academic Press, 2006.

[129] T. Theoharis, G. Papaioannou, N. Platis, and N. Partikalakis, *Graphics & Visualization: Principles and Algorithms*. A K Peters, 2008.

[130] T. Theoharis, G. Passalis, G. Toderici, and I. Kakadiaris, "Unified 3D face and ear recognition using wavelets on geometry images," *Pattern Recognition*, vol. 41, no. 3, pp. 796–804, Mar. 2008.

[131] UND, "University of Notre Dame Biometrics Data Sets," http:// www.nd.edu/ ∼cvrl/ CVRL/ Data_Sets.html, 2012.

[132] UoA-CGL, "Facial Landmarks Annotation Files," http:// graphics.di.uoa.gr/ Research/ Publications/ facial-landmarks.zip, 2012, ver. 3.

[133] M. Valstar, B. Martinez, X. Binefa, and M. Pantic, "Facial point detection using boosted regression and graph models," in *Proc.* 23^{rd} *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, USA, Jun. 13-18 2010, pp. 2729–2736.

[134] D. Vukadinovic and M. Pantic, "Fully automatic facial feature point detection using gabor feature based boosted classifiers," in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, Waikoloa, Hawaii, USA, Oct. 10-12 2005, pp. 1692–1698.

[135] M. Wardetzky, S. Mathur, F. Kalberer, and E. Grinspun, "Discrete Laplace operators: No free lunch," in *Proc. Symp. Geometry Processing*, 2007, pp. 33–37.

[136] X. Wei, P. Longo, and L. Yin, *LNCS, Advances in Biometrics*. Springer, 2007, ch. Automatic Facial Pose Determination of 3D Range Data for Face Model and Expression Identification, pp. 144–153.

[137] H. Wu, X. Liu, and G. Doretto, "Face alignment via boosted ranking model," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, Jun. 23-28 2008, pp. 1–8.

[138] C. Xu, T. Tan, Y. Wang, and L. Quan, "Combining local features for robust nose location in 3D facial data," *Pattern Recognition Letters*, vol. 27, no. 13, pp. 62–73, 2006.

[139] G. Xu, "Convergence of discrete Laplace-Beltrami operators over surfaces," *Computers and Mathematics with Applications*, vol. 48, no. 3-4, pp. 347–360, 2004.

[140] L. Xu, A. Krzyzak, and C. Suen, "Methods for combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on System, Man, and Cybernetics*, vol. 22, no. 3, pp. 418–435, May 1992.

[141] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato, "A 3D facial expression database for facial behavior research," in *Proc.* 7^{th} *International Conference on Automatic Face and Gesture Recognition*, Southampton, UK, Apr. 10-12 2006, pp. 211–216.

[142] T. Yu and Y. Moon, "A novel genetic algorithm for 3D facial landmark localization," in *Proc.* 2^{nd} *IEEE International Conference on Biometrics: Theory, Applications and Systems*, Arlington, VA, Sep. 20 - Oct. 1 2008.

[143] Z. Zeng, T. Fang, S. Shah, and I. Kakadiaris, "Personalized 3D-aided 2D facial landmark localization," in *Proc.* 10^{th} *Asian Conference on Computer Vision*, vol. 6493, Queenstown, New Zealand, Nov. 8-12 2010, pp. 633–646.

[144] D. Zhou, D. Petrovska-Delacretaz, and B. Dorizzi, "Automatic landmark location with a combined active shape model," in *Proc. IEEE* 3^{rd} *International Conference on Biometrics: Theory, Applications, and Systems*, Sep. 2009, pp. 49–55.

[145] ——, "3D Active Shape Model for automatic facial landmark location trained with automatically generated landmark points," in *Proc.* 20^{th} *International Conference on Pattern Recognition (ICPR)*, Telecom SudParis, Paris, France, Aug. 23-26 2010, pp. 3801–3805.