ADAPTIVE PCA-BASED MODELS TO RECONSTRUCT 3D FACES FROM SINGLE 2D IMAGES

ASHRAF Y. A. MAGHARI

UNIVERSITI SAINS MALAYSIA

2014

ADAPTIVE PCA-BASED MODELS TO RECONSTRUCT 3D FACES FROM SINGLE 2D IMAGES

by

ASHRAF Y. A. MAGHARI

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosphy

March 2014

ACKNOWLEDGEMENTS

IN THE NAME OF ALLAH

First and topmost, praise be to Allah for providing me with the capacity to accomplish this work.

I am deeply indebted to my supervisors, Dr. Ibrahim Venkat, Assoc. Prof. Dr. Bahari Belaton, and Assoc. Prof. Dr. Iman Yi Liao for their unfailing interest, guidance and wisdom to complete this thesis. They helped me understand and appreciate what research is and made my work meaningful. Their supervision, support, confidence, and valuable comments and suggestions often guided me throughout my research.

Gratitude is also extended to the School of Computer Sciences, Universiti Sains Malaysia, which not only provided all the facilities needed to accomplish my study but also partly supported my work. My thanks are also extended to all academic and non-academic members of the School of Computer Sciences, specially the members in our group meetings for creating a friendly and resourceful academic environment.

I wish to acknowledge the Malaysian Ministry of Higher Education for supporting me with the MIS scholarship.

I would like to thank all friends at USM, especially Dr. Hussam F. Abushawish who helped me with editing my thesis.

Finally, and most important, I would like to thank my parents, my sisters and brothers, and my own family - my wife and my children for their unconditional love and constant support over the years.

TABLE OF CONTENTS

CHAPTER 5 – A NOVEL ADAPTIVE PCA-BASED MODEL FOR 3D FACE SHAPE RECONSTRUCTION

CHAPTER 6 – A NOVEL DISTANCE-BASED MODEL FOR OPTIMAL SELECTION OF REGULARIZATION PARAMETER

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3DMM 3D Morphable Model

AAM Active Appearance Model

ASM Active Shape Model

ETR Eigenvalue Tikhonov Regularization

HMM Hidden Markov Model

ICP Iterative Closest Points

MFF Multi-feature Fitting

MRF Markov Random Field

PCA Principle Component Analysis

RP Representational Power

SFM Shape From Motion

SFS Shape-from-Shading

SNO Stochastic Newton Optimization

STR Standard Tikhonov Regularization

SVD Singular Value Decomposition

TPS Thin-Plate Spline

MODEL ADAPTIF BERSANDARKAN PCA UNTUK PEMBINAAN SEMULA WAJAH 3D DARIPADA IMEJ TUNGGAL 2D

ABSTRAK

Model wajah statistik berasaskan contoh menggunakan Analisis Komponen Utama (PCA) telah digunakan secara meluas bagi pembinaan semula wajah dalam bentuk 3D dan pengecamannya. Tumpuan utama tesis ini adalah untuk meningkatkan ketepatan dan kecekapan kaedah berasaskan PCA untuk membina semula wajah dalam bentuk 3D. Lebih tepat lagi, tesis ini menangani cabaran untuk meningkatkan Kuasa Perwakilan (RP) model berasaskan PCA selaras dengan keputusan yang menggalakkan diperolehi daripada kajian empirikal yang dijalankan. Satu set data latihan terhad digunakan dalam usaha untuk meningkatkan ketepatan pembinaan semula 3D. Mengenai kajian empirikal, ia mengkaji kesan faktor-faktor luar biasa (iaitu saiz set latihan dan kepelbagaian contoh-contoh latihan terpilih) ke atas RP model wajah 3D berasaskan PCA. Satu algoritma pembinaan semula wajah 3D terselaras telah diperiksa untuk memahami bagaimana faktor-faktor biasa seperti matriks rombakan, bilangan titik-titik sifat dan parameter rombakan λ menjejaskan ketepatan pembinaan semula wajah 3D berdasarkan model PCA.

Satu model penyesuaian berasaskan PCA adalah dicadangkan untuk meningkatkan RP model pembinaan semula wajah 3D dengan mengubah bentuk satu set contoh dalam dataset latihan. Dengan mengabungkan sampel yang telah diubahsuai bersama-sama dengan sampel latihan asal, ia telah menunjukkan bahawa peningkatan dalam RP dapat dicapai. Ujian pengesahan menyeluruh telah dijalankan untuk menunjukkan bahawa model yang dicadangkan

dengan ketara telah meningkatkan RP model piawai berasaskan PCA dengan mengurangkan ralat pembinaan semula bentuk wajah. Tambahan pula, ia telah dibuktikan kewajarannya bahawa model penyesuaian berasaskan PCA mampu membina semula imej wajah 3D dengan mengekalkan ekspresi muka, walaupun sampel latihan hanya mengandungi ekspresi neutral.

Untuk mengoptimumkan pemilihan parameter rombakan (λ), model berpandukan jarak dicadangkan untuk menentukan nilai λ yang sesuai secara automatik, dan oleh itu, ia bertindak balas ke atas keperluan peringkat pemadanan. Cadangan model berpandukan jarak dinilai dengan membandingkan λ yang ditentukan secara automatik dengan satu nilai pra-hitung terbaik. Selain itu, beberapa contoh wajah 3D yang dibentuk semula diperlihatkan secara visual untuk menjelaskan keteguhan model berasaskan jarak yang dicadangkan. Kemudian, dengan meledingkan tekstur 2D kepada wajah yang telah di bentuk semula, pembentukan wajah 3D dapat dihasilkan. Untuk ledingan tekstur, pengubahsuaian wajah 2D dapat dilatih dari tekstur model dengan menggunakan tanda-tanda muka.

Penilaian akhir dilakukan untuk menunjukkan bahawa keseluruhan sistem cadangan yang terdiri daripada model penyesuaian berasaskan PCA dan model berpandukan jarak mengatasi beberapa pendekatan terkini dari segi kecekapan. Tambahan pula, ia menunjukkan bahawa model yang dicadangkan boleh menyumbang kepada beberapa kajian yang berkaitan dalam bidang pembinaan semula imej.

ADAPTIVE PCA-BASED MODELS TO RECONSTRUCT 3D FACES FROM SINGLE 2D IMAGES

ABSTRACT

Example-based statistical face models using Principle Component Analysis (PCA) have been widely used for 3D face reconstruction and face recognition. The main concern of this thesis is to improve the accuracy and the efficiency of the PCA-based 3D face shape reconstruction. More precisely, this thesis addresses the challenge of increasing the Representational Power (RP) of the PCA-based model in accordance with the encouraging results of the conducted empirical study. A limited set of training data is utilized towards enhancing the accuracy of 3D reconstruction. Concerning the empirical study, it examines the effect of phenomenal factors (i.e. size of the training set and the variation of the selected training examples) on the RP of 3D PCA-based face models. A regularized 3D face reconstruction algorithm has also been examined to find out how common factors such as the regularization matrix, the number of feature points, and the regularization parameter λ affect the accuracy of the 3D face reconstruction based on the PCA model.

Importantly, an adaptive PCA-based model is proposed to increase the RP of the 3D face reconstruction model by deforming a set of examples in the training dataset. By adding these deformed samples together with the original training samples, it has been shown that the improvement in the RP can be achieved. Comprehensive experimental validations have been carried out to demonstrate that the proposed model considerably improves the RP of the standard PCA-based model with reduced face shape reconstruction errors. Furthermore, it has been

justified that the adaptive PCA-based model is capable of reconstructing 3D face images by retaining facial expressions, although the training samples contained only neutral expression.

To optimize the selection of regularization parameter (λ), a distance-based model is proposed to automatically find an appropriate value of λ, and therefore, it responds to the requirements of the fitting stage. The proposed distance-based model is evaluated by comparing the automatically determined λ with the pre-calculated best one. Moreover, examples of reconstructed 3D face shapes are visualized to clarify the robustness of the proposed distance-based model. Then by warping the 2D texture to the reconstructed face shape 3D face reconstruction is achieved. For the texture warping, the 2D face deformation is learned from the model texture using a set of facial landmarks.

Finally experimental evaluations have been demonstrated to show that the overall proposed system which comprises of the adaptive PCA-based model and the distance-based model outperforms some of the recent approaches in terms of efficiency. Furthermore, it is shown that the proposed models could contribute to several related studies in the field of image reconstruction.

CHAPTER 1

INTRODUCTION

1.1 Introduction

The primary objective of 3D facial reconstruction systems is to recover the three dimensional shape of individuals from their 2D pictures or video sequences. One of the major challenges in 3D facial modeling is the accurate reconstruction of 3D faces from given 2D face images. The use of 3D faces in image processing applications has received substantial attention during the last decades. The need for 3D face reconstruction has grown in such crucial applications such as virtual reality simulations, face recognition (Elyan and Ugail, 2007; Avilaq and Rezaie, 2013) and plastic surgery simulations (Bottino et al., 2012).

For example, in biometric identification, face recognition rate could be significantly improved by incorporating the 3D face shape with 2D face images (Hu et al., 2004). Face Recognition Vendor Test 2006 has shown that 2D face recognition can achieve high accuracy under controlled conditions, e.g. when the testing face samples are frontal. However, when face pose changes largely, the performance of existing methods drop drastically, and therefore 2D is a restricted environment (Li and Jain, 2011). In order to resolve the controlled restriction of 2D, 3D faces can be used. By rotating the reconstructed 3D face to different views, pose virtual face images are generated to enlarge the training set of face recognition (Wang et al., 2011). Applications that are more known to common people are in 3D games and movie industries.

For the 3D face reconstruction, the accurate reconstruction of a person's 3D face model from his/her 2D face images still remains as an open challenge. Until now, with the aid of most popular commercially available tools, 3D facial models are obtained not directly from images but by laser-scanning of people's faces (Zhang et al., 2006; Li and Jain, 2011). This technique has the following limitations:

1. These scanners are usually expensive and are targeted to work in tightly controlled environments.

2. Laser-scan based reconstruction could not be applied in certain scenarios; for example, a person's face had been damaged during an accident and his/her face needs to be reconstructed in order to assist plastic surgery. In this situation, laser-scan cannot be applied and the face can be reconstructed by using computational techniques with the aid of available photos of the person, which were taken prior to the accident.

A 3D face can either be reconstructed from a single image or from multiple images. This study focuses on the problem of reconstructing 3D face shapes from single 2D images. This technology, which is only applicable in controlled environments, does not require setting up multiple cameras to capture the objects simultaneously.

The PCA-based model proposed by Blanz and Vetter (1999) with relatively small sample size (100 faces) has primarily been used for face recognition and obtained reasonable results (Blanz and Vetter, 2003). Furthermore, it does not require the generation of synthetic views from 2D input images. Instead, the recognition was based on the model coefficients which represent intrinsic shape and texture of faces. Although in some statistical modeling methods both shape and texture are modeled separately using PCA (e.g. 3DMM), it has been suggested that shapes are more amenable to PCA based modeling than texture because textures are subject to vast variation when compared to shape based features (Jiang et al., 2005). Therefore, the models intended in this contribution are based on modeling of shapes. When shapes are considered, the reconstruction of 3D face shapes from 2D images using shape models is relatively simple. A popular method for reconstructing a 3D face from a 2D image is a regularization based reconstruction where a few feature points are selected as observations for reconstruction (Jiang et al., 2005). Alternatively the regularized algorithm which uses 3D Morphable Model to reconstruct the 3D face shape from facial 2D points has also been presented in (Blanz and Vetter, 2002) and (Blanz et al., 2004). The results based on this method do not go beyond the representational power of the model. Even if a 3D face model is trained with more examples or a different dataset to generate a better representation of the true face, the generated face remains within the boundaries of the PCA-model.

To utilize the prior information modeled by PCA, an extended version of Tikhonov regularization is used to estimate the model parameters by solving the inverse problem of 3D face reconstruction. However, most of the regularization methods that uses prior knowledge tend to smooth the reconstructed image (Agarwal, 2003). To prevent the successive smoothness of the solution, the optimal selection of regularization parameters is highly considered (Zhu et al., 2013).

To fill the gaps in PCA-based model and Tikhonov regularization, this study addresses two significant issues. First, the insufficient representational power of the PCA-based model and its capability of depicting new 3D faces. Second, Tikhonov regularization does not support the automatic selection of regularization parameter. Consequently, this study aims at improving the accuracy as well as the efficiency of the PCA-based 3D face reconstruction from single images.

1.2 Motivation

The need for 3D face reconstruction has grown in various applications such as virtual reality simulations, face recognition (Elyan and Ugail, 2007; Avilaq and Rezaie, 2013) and plastic surgery simulations (Bottino et al., 2012).

As mentioned in the previous section, face recognition rate could be significantly improved by incorporating 3D face shapes with 2D face images (Hu et al., 2004). Furthermore, the training set of face recognition can be enlarged by generating pose virtual images of different views (Wang et al., 2011). In addition, the increase in spending money on face plastic surgery (Adamson and Galli, 2009) triggered new studies in 3D face reconstruction from 2D images. For example, in 2008, the American Society for Aesthetic Plastic Surgery (ASPS) reported a 162% increase of the facial plastic surgeries in ten years whereas more than one million facial plastic surgeries were performed (Statistics, 2009). According to recent statistics released by ASPS (Statistics, 2102), 209,000 maxillofacial surgery procedures were performed in USA in 2012. This reveals that a 7% increase is seen since 2011.

Therefore, it is important to focus on developing techniques that can improve the accuracy and efficiency of current 3D reconstruction systems. Many researchers have attempted to solve problems related to 3D faces reconstruction from single 2D images but their methods have limitations. These limitations will be described in Chapter 2.

1.3 Problem Statement

The problem of 3D facial modeling remains as a partially solved problem in the fields of computer vision and graphics. The purpose of this study is to reconstruct 3D faces from single 2D images. The advancement of 3D scanning technology has led to the creation of more accurate 3D face exemplar models (Luximon et al., 2012). Example-based modeling allows more realistic face reconstruction than other methods (Widanagamaachchi and Dharmaratne, 2008; Levine and Yu, 2009). In the simplest form, example-based 3D face reconstruction methods have two main stages: The model building stage and the model fitting stage. In this study, PCA-based 3D face model and regularized algorithm are used for model building and model fitting respectively.

For PCA-based modeling, however, the quality of reconstructed faces is affected by the selected examples. The two common factors that are generally concerned with such models are the size of the training dataset and the selection of different examples in the training set. For example, Kemelmacher-Shlizerman and Basri (2011), and Gonzalez-Mora et al. (2010) emphasized that learning a generic 3D face model requires large number of 3D faces.

For PCA model fitting, reconstructing 3D faces from 2D images is a linear inverse problem which gives rise to ill-posed linear system. To obtain a meaningful approximate solution, regularization can be employed (Mallik et al., 2012). One of the most appropriate regularization methods is the Tikhonov regularization (Jing et al., 2009). The common factors that generally affect Tikhonov regularization are the regularization Tikhonov matrix (stabilizing item), the number of feature points, regularization parameters, and noise. The performance of regularized approximations can only be controlled through the selection of a regularization parameter (Lu and Pereverzev, 2008; Zhu et al., 2013). Choosing too large regularization parameter causes the solution to be over-smoothed. Otherwise, a too small regularization parameter leads to overfitting. In other words, the regularization parameter balances the tradeoff between the excessive smoothing of the reconstruction and the data misfit.

The overarching research problem of this study is to improve the accuracy as well as the efficiency of the PCA-based reconstruction. This could be further divided into two sub problems as follows:

1. The RP of a PCA-based 3D learning face model is not adequate enough to represent or reconstruct an accurate 3D face shape.
2. An appropriate value of the regularization parameter λ is not known in advance.

To overcome the first problem, an adaptive PCA-based model is proposed to improve the accuracy of reconstruction through enhancing the representational power of the standard PCA-based model. As for as the second problem is concerned, a distance-based model is developed to automatically select an appropriate regularization parameter for the Tikhonov regularization method.

1.4 Research Objectives

The overarching aim of this research relates to improving the accuracy as well as the efficiency of 3D face shape reconstruction from their single 2D images. To achieve this goal, this study seeks to fulfill the following objectives:

1. To formulate the relationship between sample size and the Representational Power of the PCA-based 3D face shape model.
2. To enhance the Representational Power of the PCA-based model in order to increase its capability in depicting new 3D faces of given face images.
3. To categorize the effect of specific factors (feature points, regularization parameter, and Tikhonov matrices) on regularization based 3D face reconstruction using PCA.
4. To propose a distance-based model model to find an appropriate regularization parameter for an optimal and plausible solution.

1.5 Research Scope and Limitation

There are various limitations with respect to the data, which may affect the representational power of the statistical learning model. These limitations include:

1. Sample Size: The available sample size (100 3D faces) is particularly small and restricted to faces of only middle aged people. This sample size may not be sufficient to build a powerful reference model.
2. 3D face examples available in the dataset are neutral faces and do not represent any common facial expressions. This means that there are some difficulties which hinder the reconstruction of expressional face images using the learning model.

1.6 Research Contributions

This study introduces the following significant contributions to the body of knowledge:

1. Formulating the functional relationship between sample size and the Representational Power (RP) of the model.
2. A novel adaptive model to increase the RP of the statistical PCA face shape model for 3D face shapes reconstruction.
3. Examining the effect of common factors such as the regularization matrix, the number of feature points, and the regularization parameter λ on PCA-based 3D face reconstruction using Tikhonov regularization.
4. A novel distance-based model to automatically find an optimal regularization parameter that produces a plausible solution.

1.7 Outline of the Thesis

The thesis is organized as follows:

Chapter 2 reviews the literature on 3D face reconstruction from 2D images. It first covers 3D face reconstruction from images and focuses on approaches of reconstruction from single 2D images. Second, it reviews the various statistical learning-based methods. Third, it explains the statistical 3D face modeling and emphasizes PCA-based models. Finally, it reports a variety of techniques used for 3D face reconstruction.

Chapter 3 introduces the research methodology and the proposed framework of the PCA-based 3D face reconstruction system. The 3D face reconstruction from a limited number of feature points using Tikhonov regularization is discussed. Finally, the problems of low accuracy and efficiency of the PCA-based 3D face reconstruction are addressed.

Chapter 4 encompasses an empirical study on the RP of 3D PCA-based face models using USF Human ID 3D database. A series of experiments are designed to examine the effect of training set on the RP of the model. The common factors that generally affect Tikhonov regularization are also studied. These factors are the regularization Tikhonov matrix (stabilizing item), the number of feature points and regularization parameters. The experimental results of these factors are reported in this chapter.

Chapter 5 introduces a novel adaptive PCA-based model to increase the RP of the model in order to improve its capability in depicting new 3D face shapes of given input face images. The adaptive PCA-based model is used to reconstruct a 3D shape face from the input 2D face image. The technique used to deform the training data is explained. The experimental results of the proposed model are also discussed. Finally, the proposed model is evaluated and compared with the standard PCA-based model.

Chapter 6 introduces a novel distance-based model to automatically find an appropriate regularization parameter for an optimal 3D face shape reconstruction. An evaluation and a comparison are carried out on the results of the distance-based model compared with best solution in terms of accuracy. Moreover, the comparison has been carried out on the results of the overall system (adaptive model and distance-based model) with existing methods in terms of efficiency.

Chapter 7 concludes the thesis by presenting the thesis summary and future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter examines the literature on 3D faces reconstruction from single 2D images. Particularly, this chapter emphasizes statistical 3D face modeling using examples. It also intends to look into the limitations of the existing techniques of the 3D faces reconstruction.

The literature review is divided into four main sections. The first section covers a review of 3D face reconstruction from 2D images. The second section covers a variety of statistical learning-based methods. The third section explains the PCA-based 3D face modeling and concentrates on PCA-based models. Finally, techniques used for 3D face reconstruction are reported.

2.2 3D Face Reconstruction

Reconstruction of 3D faces is an important issue in the fields of computer vision. 3D facial reconstruction systems are to recover the three dimensional shape of individuals from their 2D pictures or video sequences. The need for 3D face reconstruction has grown in applications such as virtual reality simulations, face recognition (Elyan and Ugail, 2007; Fanany et al., 2002; Avilaq and Rezaie, 2013) and plastic surgery simulations (Bottino et al., 2012). For Example, in biometric identification, face recognition rate could be significantly improved by incorporating the 3D face shape with 2D face images (Hu et al., 2004).

3D faces can be reconstructed based on single images, or multiple images. Higher level abstraction taxonomy for 2D image-based 3D face reconstruction techniques can be seen in Figure 2.1.

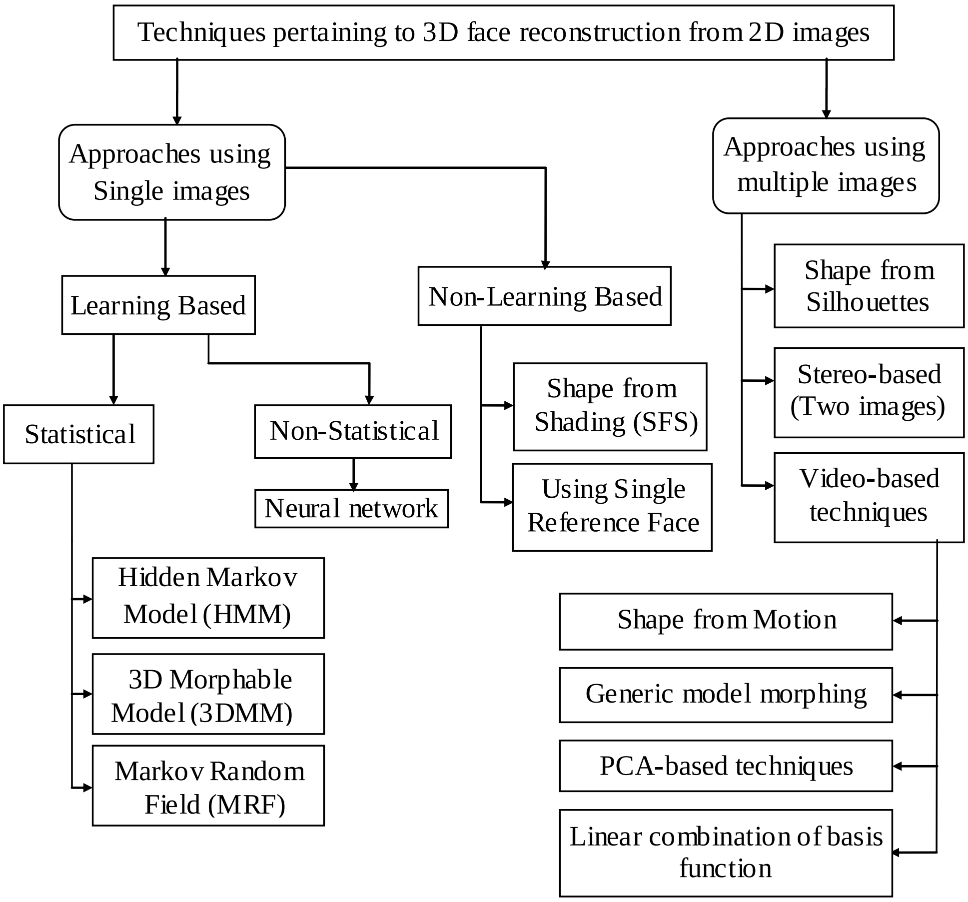

Figure 2.1: Taxonomy of a higher level of abstraction on the 3D face reconstruction techniques from 2D images.

Approaches based on 3D face reconstruction from multiple images include video-based and silhouette-based methods. Video-based techniques are used for 3D face reconstruction from images captured from different viewpoints. There are a diverse number of techniques based on video frames such as reconstruction using generic model morphing (Liu et al., 2001), by using linear basis functions that do not require a generic model (Bregler et al., 2000), and Shape From Motion (SFM) (Amin and Gillies, 2007; Chowdhury et al., 2002) where motion information of feature points from multiple video frames are extracted to obtain 3D reconstruction. However, morphing using SFM needs one or more images taken from different viewpoints and pre-knowledge of the generic shape of the object.

In Silhouette-based methods (Moghaddam et al., 2003; Lee et al., 2003), a 3D face is reconstructed by using multiple face outlines extracted from several face images or video sequences. The silhouettes captured from different angles provide details related to the geometrical structure of the face that can be used to generate a 3D face. Silhouette-based methods have been combined with other 3D reconstruction methods such as statistical methods (Wang et al., 2005; Moghaddam et al., 2003; Lee et al., 2003). This study will focus on the problem of reconstructing 3D face shapes from single 2D images. This technique does not require setting up multiple cameras to capture the objects simultaneously and thereby it is not limited to working in controlled environments.

There are many approaches for the reconstruction of 3D faces from single images. One of the earliest techniques being utilized is Shape-from-Shading (SFS) (Atick et al., 1996; Zhang et al., 1999; Smith and Hancock, 2006), which capitalizes the idea that the depth information is related to the intensity of a face image acquired through a given/chosen reflectance model. SFS estimates the illumination direction in the 2D image to infer the 3D shape of the surface. It has been shown SFS suffers from poor global shape control and being difficult to provide an accurate reflectance model for various environments (Atick et al., 1996).

Recently, a novel method has been proposed by Kemelmacher-Shlizerman and Basri (2011) that combines shading information with generic shape information derived from a single reference model by utilizing the global similarity of faces. This method uses only a single reference model of a different person's face to reconstruct the 3D face shape. It does not need a learning stage to build a model for representing input faces. Their algorithm starts with estimating parameters that best fit the reference model to the input image. These parameters are lighting, pose and depth. Then the estimated depth and spherical harmonics coefficients are used to estimate the albedo. The involved fitting process requires boundary conditions and parameters to be adjusted during the reconstruction process. In addition, not counting on a 3D reference model which keeps shape similarities with the input image may result in inaccurate 3D shape estimation.

There are also conventional (non statistical) learning-based methods, such as neural network (Nandy and Ben-Arie, 2001; Lin et al., 2005) and statistical learning-based methods, such as Hidden Markov Model (HMM) (Nagai et al., 2002), Markov Random Field (MRF) (Saxena et al., 2009), and analysis by synthesis using 3D Morphable Model (3DMM) (Blanz and Vetter, 1999).

Lin et al. (2005) proposed a neural network based adaptive hybrid-reflectance 3D surface reconstruction model. They used the back-propagation learning algorithm to train the neural network. The pixel values of the 2D image have been used as inputs while the output of the neural network was the normal vectors. The normal vectors can then be applied to 3D surface reconstruction by enforcing integrability method.

HMM has been used to model the correspondence between an intensity image and its depth information by learning knowledge of objects from number of samples containing pairs of an intensity image and depth information.

MFR is a generic and state-of-the-art method for unstructured 2D still images. It has been applied to numerous numbers of scenic pictures, including those containing faces. However, the result is not satisfactory because the face area is probably segmented into one plane because of its fairly shading/feature variations. MRF is based on local model to recover the 3D surfaces, which usually suffer from global shape controllability.

Analysis by synthesis is an approach in which the parameters of the 3D statistical model are adjusted to increase the accuracy between the reconstructed face and the 2D face image (Widanagamaachchi and Dharmaratne, 2008). 3DMM has made itself a milestone in 3D face reconstruction area for its realistic results. The strength of the method is that the 3D training faces were obtained using 3D scanners which not only provide the depth information, but also colors/textures. This review will focus on statistical learning methods using 3D examples.

In the last few years, vast research has been performed on 3D face modeling whereas many different approaches for 3D face reconstruction are proposed. However, the accuracy of reconstructed 3D faces still needs some improvements for the real world applications. Consequently, this review will look at existing techniques and focus on the current and common approaches.

The presence of 3D scanning technology lead to create more accurate 3D face model examples (Luximon et al., 2012). Examples based modeling allows more realistically face reconstruction than other methods (Widanagamaachchi and Dharmaratne, 2008; Levine and Yu, 2009). However, all learning based methods suffer from the limitation that the learning model is heavily dependent of the training data. Consequently, the quality of face reconstruction using examples is affected by the chosen examples.

For expressional face images that have not been trained it cannot be reconstructed properly. Some remedy to the problem has been sought by constructing 3D face basis from various expressional face images using Active Appearance Model (AAM) (Zhu et al., 2006), but proved marginal reconstruction accuracy. Other methods dealing with expressions can be found in (Hahnel et al., 2006; Xu and Luo, 2006; Lu and Jain, 2008; Wang and Lai, 2011), but most of these methods are mainly designed for recognition purpose and require different 3D expressional faces to be scanned which is somewhat taxing and not applicable in uncontrolled environments. On the other hand, an incremental Structure from Motion (SFM) approach to learn a generic 3D face model from 2D face images containing different expression is proposed in (Gonzalez-Mora et al., 2010). The main concern of this method is the utilizing of existing 2D face databases to learn a generic 3D face model based on SFM. The proposed technique requires one or a reduced number of input images to reconstruct a 3D face shape.

Furthermore, Kemelmacher-Shlizerman and Basri (2011), and Gonzalez-Mora et al. (2010) urged that learning a generic 3D face model requires large amounts of 3D faces. Moreover, analytical results in Chapter 4 show that the size of training set increase, the more accurate the model can represent a new face. The existing statistical 3D face modeling approaches are summarized in a taxonomy (see Figure 2.2).

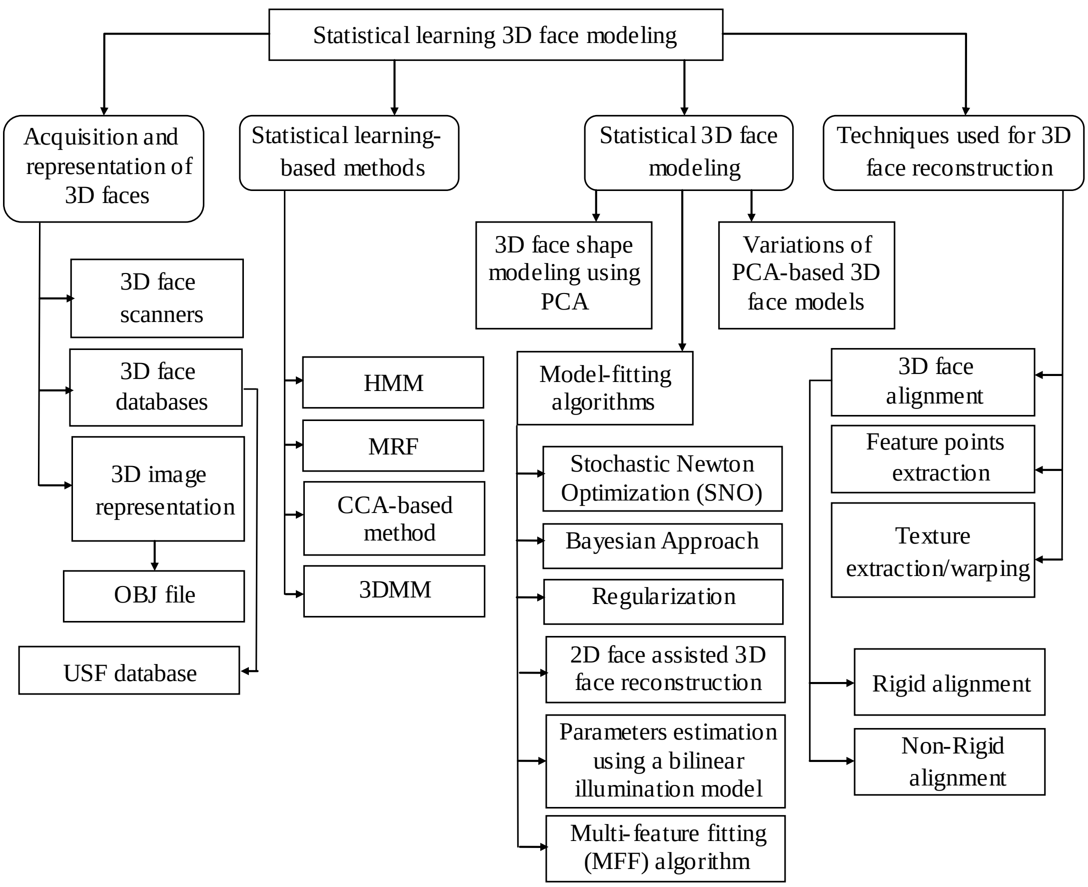

Figure 2.2: Taxonomy of the statistical 3D face modeling approaches covered in this review.

2.3 Statistical Learning-Based Methods (Example-Based Methods)

The strength of example-based methods (e.g. 3DMM) is that the 3D training faces are obtained using 3D scanners which not only provide the depth information, but also the colors/texture. However, all learning based methods suffer from the limitation that the learning model is heavily dependent of the training data. This section provides an overview of different types of statistical learning 3D face models used for 3D face reconstruction purposes. More attention will be focused on 3DMM.

2.3.1 Hidden Markov Model (HMM)

HMM is one of the statistical learning techniques that used to model the corresponding between an intensity image and its depth information. Nagai et al. (2002) proposed an approach called shape from knowledge, which is trained from number of samples containing pairs of an intensity image and depth information. The intensity (appearance model) is represented by knowledge and the 3D shape is represented by the depth information. The method was applied to face and hand images for recognition and reconstruction purposes.

2.3.2 Markov Random Field (MRF)

MRF is used for each small homogeneous patch in the image. It has been applied to model both image depth and the relationships between different parts of the image (Saxena et al., 2009) for estimating the 3D scene structure from single still images. The proposed algorithm in (Saxena et al., 2009) segments the image into small homogeneous patches and uses MRF to infer the 3D position and orientation of each patch.

2.3.3 Canonical Correlation Analysis (CCA)-based method

CCA is used to implement a method for generating face depth maps from a color frontal face image. For training, 150 pairs of face images and their corresponding depth maps were needed to explain the correlation between the training pairs whereas the dimensionality of the trained data was much more than the number of training data (Reiter et al., 2006). An overview of their method is illustrated in Figure 2.3. A reasonable 3D depth map prediction is achieved using a simple matrix multiplication which does not require an iterative optimization method. However, this method is only applicable for frontal 2D face images taken under controlled illumination.

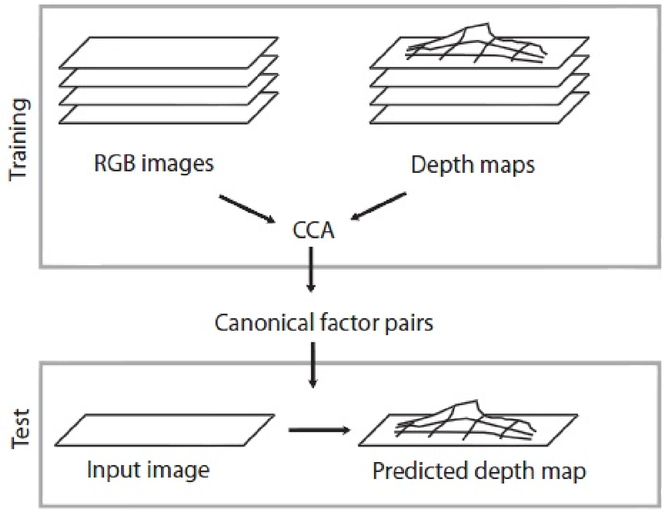

Figure 2.3: Overview of CCA-based algorithm: In the training process, Canonical factors pairs are generated from a set of training examples. During testing, the factors pairs are used for the prediction of a 3D map from texture data. Reiter et al. (2006)

2.3.4 3D Morphable Model (3DMM)

A 3D human face model is represented by its shape and texture where the shape is represented by the 3D coordinates of all vertices in three-dimensional space and texture is represented by the RGB color of each vertex. The 3D morphable model (3DMM), developed by (Blanz and Vetter, 1999) decomposes any 3D face model into a linear combination of shape and texture vectors of set of example faces (100 face models) which obtained by 3D scanners. The linear combination is controlled by shape and texture parameters α and β which can be considered as weights. In order to build the 3D morphable model, Blanz and Vetter performed PCA on shape and texture information, obtained from 100 aligned 3D faces, separately. Section 2.4 illustrates PCA-based statistical 3D face modeling using examples.

The main idea of Blanz and Vetter method is that given a sufficiently large database of 3D face models, any arbitrary face can be generated by morphing between the ones in the database. An analysis-by-synthesis method have been employed for fitting the model to the input 2D image. They used Stochastic Newton Optimization (SNO) algorithm to estimate all model parameters assuming that the pixels are independent and identically distributed (Romdhani, 2005) which is difficult and time-consuming problem.

3DMM has made itself a milestone in 3D face reconstruction area for its realistic results. The strength of the method is that the 3D training faces were obtained using 3D scanners which not only provide the depth information, but also the color/textures. However, in many cases, the produced face is not realistic. This is because the fitting algorithm estimates the shape, texture and image condition from the pixel intensity only (Romdhani, 2005). Section 2.4.3 demonstrates different methods of fitting a morphable model to an input face image.

2.4 PCA-Based 3D Face Modeling

This section provides an overview of parametric deformable models that describe the deformations of the 3D surface/texture. In the literature, the deformable models are usually used with model-fitting algorithms that deform and fit the model to input data.

Building a generic statistical shape model of an object depends on the observation within the training set. The common factors that are generally concerned are the size of the training set and the different choices of the examples in the training set. This section focuses on Principle Component Analysis (PCA)-based statistical 3D face modeling using examples. It is a popular techniques for modeling 3D faces and has been widely used for 3D face reconstruction and face recognition. An earlier idea to use PCA was to employ it for implementing a face recognition system (Turk and Pentland, 1991). In order to build a PCA-based models, PCA is performed on the set of training samples. The eigenvectors and corresponding eigenvalues that define the variability within the training set are computed by applying PCA on the the covariance matrix of training samples (Smith, 2002).

The following section demonstrates 3D face shape modeling using PCA. Modeling texture information is similar.

2.4.1 3D Face Shape Modeling using PCA

The 3D face shape model is a linear combination of eigenvectors obtained by applying PCA decomposition to model shape variability. Each training 3D face shape is represented by the 3D coordinates of all vertices in the triangulated mesh, where shape vectors are given as follows:

$$s_i = (x_{i1}, y_{i1}, z_{i1}, \ldots, x_{in}, y_{in}, z_{in})^T \ , \tag{2.1}$$

where s_i has the dimension $n \times 3$, n is the number of vertices and $i = 1, \ldots, m$ (number of face shapes).

Based on this representation, the mean 3D face among the training set is estimated and the deviation of each training sample from the mean is calculated. PCA is then applied on the covariance matrix. As a result, a new 3D shape s can be generated using the following equation

$$s = s_0 + \sum_{i=1}^{m} \alpha_i e_i \ , \tag{2.2}$$

where s_0 is the mean 3D shape, e_i represent the i^{th} eigenvector of the covariance matrix, α_i is the coefficient of the shape eigenvector e_i and m is the number of significant eigenvectors.

The coefficient of a face shape s_i can be calculated using the following equation

$$\alpha = E^T (s_i - s_0) \ , \tag{2.3}$$

where $E = [e_1, e_2, ..., e_m]$ are the eigenvectors of the covariance matrix. The projected new face shape can be represented by applying Equation (2.2).

Based on this demonstration a 3D face can be represented by a set of shape parameters and a set of texture parameters. These parameters can perform as model descriptors that can be estimated to reconstruct new 3D face from input face image. Figure 2.4 illustrates the building process of PCA-based 3D face model.

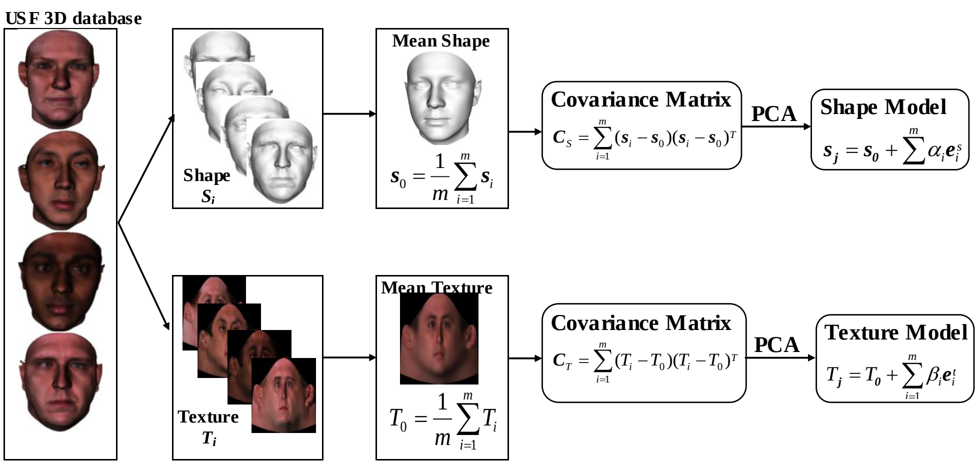

Figure 2.4: Block diagram illustrating a general method for building 3D face model based on PCA.

Accordingly, the 3D reconstruction problem is to search for the best possible fitted shape and texture parameters such that the deformed model optimally matches the input 2D image (Blanz and Vetter, 1999).

2.4.2 Different Applications of PCA-Based 3D Face Models

PCA-based 3D face models have been used in different applications related to 3D faces such as 3D face modeling (Luximon et al., 2012), 3D face reconstruction (Blanz and Vetter, 2002, 1999; Basso and Vetter, 2005), 3D face recognition Blanz and Vetter (2003), age Synthesis Hutton et al. (2001), and face tracking and animation (Blanz et al., 2003). Regarding to 3D face reconstruction, variations of PCA-based models have been used by many researchers. A summary of PCA-based models variation is given in Table 2.1.

Table 2.1: Different Applications of PCA-based 3D face models

Author/Year	Description
Atick et al. (1996)	PCA has been applied on cylindrical coordinates instead of cartesian coordinates. It was used to derive a low-dimensional parametrization of face shape space from 200 laser-scanned 3D faces. Using this representation, an algorithm is developed to solve SFS.
Hutton et al. (2001)	PCA was used for building a dense face surface model from example faces. The surfaces were aligned using thin-plate spline (TPS) based on 9 manually selected points on each training surface
Blanz and Vetter (1999, 2003)	They applied PCA on shape and texture vectors of 100 exemplar faces to build 3D morphable model(3DMM). The shape and texture were processed separately. The main idea is that any 3D face can be modeled as a linear combination of shape and texture vectors of the 100 exemplar faces. The linearity combination is fully controlled by shape and texture parameters, which can be considered as weights. An analysis by synthesis algorithm was employed for fitting the model to the input 2D face image to estimate the model parameters.
Gonzalez-Mora et al. (2010)	A prior shape basis was computed by performing PCA on existing 3D shapes. These shape basis function as prior knowledge about possible solution to regularize the incremental SFM which is proposed to learn a generic 3D face model from large number of 2D images. PCA was used at the first iteration to estimate the deformation basis of the shape.
Luximon et al. (2012)	PCA was applied on 78 males and 78 females separately to model the variation among the 3D face and head shapes. A separate PCA-based male model and PCA-based female model were build using all head vertices whereas the models have the same amount of vertices.

2.4.3 Model-Fitting Algorithms

PCA-based 3D face reconstruction approaches have generally two main stages: The model building stage and the model fitting stage. 3DMM is trained with 3D scan faces, which can be matched with new face images by an optimization method. The estimated parameters can then be used to generate the 3D face of the input image. The optimization is mainly based on the minimization of the intensity difference between the model and the input face. The process of 3DMM fitting to the input face can be done iteratively (Blanz and Vetter, 1999), matrix multiplication (Jiang et al., 2005; Blanz et al., 2004), or any non-linear optimization algorithm as presented in (Lee et al., 2005; Wang et al., 2011)

Table 2.2 describes some fitting algorithms used to estimate the values of shape and texture parameters for generating novel 3D faces.

In the last few decades much interest has been shown in the area of extracting 3D surfaces from observed 2D images by using statistical models. These models can be used as prior information which can be incorporated with a fitting algorithm to estimate the complete 3D face shape from the given information such as a set of facial feature points. One of the most appropriate methods that can be employed is the Tikhonov regularization (Jing et al., 2009). It is a popular and effective method that can be easily incorporated with prior information embedded in a closed-form solution which can be readily obtained by applying PCA on the 3D face shapes. Importantly, reconstruction by means of a Tikhonov regularization method can be computed in one step (non-iterative way), thereby enabling faster 3D reconstructions. Hence, Tikhonov method is an efficient choice for several 3D oriented interactive tools. The following subsection will review a fitting algorithm based on Tikhonov regularization method.

Table 2.2: Model-fitting algorithms

Algorithm/ References	Remarks
Stochastic Newton Optimization (SNO) (Blanz and Vetter, 1999, 2003)	The SNO algorithm has been used to estimate the model parameters for shape and texture. The algorithm was also used to optimize 22 rendering parameters concatenated into a vector which include pose angels, 3D translation, ambient light intensities, directed light intensities, the angles of the directed light, focal length, color contrast, gains and the offsets of color channels.
Bayesian Approach (Blanz and Vetter, 2002)	The proposed method finds the face shape vector with maximum posterior probability, given examples data. The 3D shape face model is estimated from small set of facial points using a morphable face model. The approach based on example-based vector space and on statistical properties of the 3D face data.
Regularization (Jiang et al., 2005; Blanz et al., 2004)	Jiang et al. compute the regularization in an iterative manner to estimate the shape parameters using 84 feature points. The position of the feature vertices on the face were forced for accurate alignment of the feature points on the 2D image. They perform an additional interpolation method (Kriging interpolation) to compute the displacement of the non-feature vertices. Blanz et al. computed the regularization in a direct non iterative way using Singular Value Decomposition (SVD). The regularized algorithm computed an optimal tradeoff between the surface fitting and plausibility in term of prior probability. 17 feature points have been used for shape reconstruction.
2D face assisted 3D face reconstruction (Wang et al., 2011)	The shape and texture parameters of the 3DMM have been estimated to optimize the objective energy function which takes not only the frontal face image intensity, but also 2D face fitting results into account. The 2D face-fitting algorithm is called Random Forest Embedded Active Shape Model. The optimization algorithm uses Levenberg-Marquardt method, which is non-linear optimization problem.
Parameters estimation using a bilinear illumination model (Lee et al., 2005)	Using both morphable model and illumination model, the parameters of the morphable model are estimated given a single photograph. The combined two models leads to a simple fitting method that deal with illumination and complex face reflectance. The pose parameter (rotation and translation) of the face is also estimated using 9 feature points. The illumination model uses higher-order SVD that describes 3D shape and illumination variation. Downhill simplex method, a non-linear minimization algorithm, is employed to optimize the cost function.
Multi-feature fitting (MFF) algorithm (Romdhani, 2005)	Romdhani proved experimentally that the minimization of intensity difference between the face and a model instance does not always give the optimum position of the model. To deal with this problem, he included additional image based and model-based features to the objective function. Image-based features depend on edge information, pixel intensity, the position of anchor points in the input image, and specular highlights. Model-based features are the distance between the model instance and the mean face shap/texture and texture related constraints (acceptable limits of texture). These additional features ensure that the candidate model instances are plausible and the estimated texture lye within a specific range. Romdhani show that MFF algorithm has a smoother cost function than SNO algorithm, which does not require to use a stochastic optimization algorithm. Consequently, much fewer iteration are needed to reach convergence.

2.4.3(a) Tikhonov Regularization-Based Fitting Algorithm

For robust, plausible and stable results, the regularization mechanism needs to find a tradeoff between fitting 3D shape to the given 2D facial landmarks and producing plausible solution in terms of prior knowledge (Blanz et al., 2004; Zhu et al., 2013). The Standard Tikhonov Regularization method (STR), which uses the identity matrix as a regularization matrix, is used to estimate the model parameters by solving the inverse problem and preventing the overfitting. However, the quality of the reconstructed face shapes is very similar to the mean face shape (excessive smoothness) which leads to loss of information about the reconstructed images (Jing et al., 2009).

Furthermore, by using Tikhonov regularization, the problem of choosing an appropriate regularization parameter arises. Choosing too large regularization parameter causes the solution to be over-smoothed. Otherwise, too small regularization parameter leads to overfitting. In other words, the regularization parameter balances the tradeoff between the excessive smoothing of the reconstruction and the data misfit.

There are numerous strategies for determining the regularization parameter (Honerkamp and Weese, 1990). Some mathematical methods such as the discrepancy principle, the Tikhonov prior estimation, the Engl criteria, and Arcangeli criteria method need prior information about the data noise (Jing et al., 2009). In practice, however, such prior information cannot be easily acquired and it is highly impractical to obtain the noise characteristic in real time (Jagannath and Yalavarthy, 2012). Other methods including L-curve and generalized Cross Validation need less prior information but are time consuming. In addition, some factors can influence the parameter selection. These factors include e.g. diffusion of errors in the process of numerical computation, and the random fluctuation of errors in the input data (Jing et al., 2009). Furthermore, these methods have also their limitations. For example, although in the last decade, L-curve gained attention for determining optimal regularization parameters, yet, however, its limitation is of having asymptotic property which means it is non convergent (Agarwal, 2003).

A different strategy is to select the regularization parameter in a straightforward way and setting its value as constant for all images (Ying et al., 2004). For example, in (Jing et al., 2009) the range of the regularization parameter was determined empirically by solving typical cases in advance.

However, empirically determination of regularization parameter leads to an unwanted bias in the solution. Furthermore, it varies for different problems and requires prior information on the target images as well as the noise in the data.

This study uses a different strategy for optimal selection of the regularization parameter. The distance from the average face shape and the reconstructed face shape is employed through an optimization function to control the regularization process. A new approach based on this distance is proposed in order to determine an optimal regularization parameter which ensures that the obtained 3D face shape is plausible and not over-smoothing.

2.5 Techniques used for 3D Face Reconstruction

In this section, general techniques will be demonstrated and discussed mainly used for building PCA-based 3D face models as well as fitting the PCA-based model to new faces. Such techniques are frequently employed in different 3D reconstruction approaches reported in the literature.