SEMI-AUTOMATIC CUSTOMIZATION FOR MODELING HUMAN FACE

HASITH PASINDU ABEYSUNDERA
B.S., Computer Engineering, American University of Cyprus, 2009

Submitted to the Graduate School of Science and Engineering
in partial fulfillment of the requirements for the degree of
Master of Science
in
Computer Engineering

IŞIK UNIVERSITY
2011

IŞIK UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

SEMI-AUTOMATIC CUSTOMIZATION FOR MODELING HUMAN FACE

HASITH PASINDU ABEYSUNDERA

APPROVED BY:

Assist. Prof. M. Taner Eskil (Thesis Supervisor)	Işık University	_____
Prof. Ercan Solak	Işık University	_____
Assist. Prof. Devrim Akça	Işık University	_____

APPROVAL DATE: ..../..../....

SEMI-AUTOMATIC CUSTOMIZATION FOR MODELING HUMAN FACE

Abstract

Model-based vision has firmly established its roots as a robust approach in recognizing and locating known traits in rigid objects even under the presence of noise, clutter and occlusion. However the application of such systems has not displayed the same efficiency in modeling non-rigid objects. The dilemma with the prevailing modeling techniques is that that they compensate model specificity to accommodate variability, or the vice versa compromising the robustness of the 3 dimensional model during the image interpretation progression. Face, being a non rigid and a sophisticated structure makes it more arduous to model, using such approaches.

In this study we have presented a novel method in modeling 3 dimensional images employing a generic wireframe and a single 2 dimensional image. Known traits are located in the 3 dimensional space using a variant of ray tracing method. Non-landmark traits are positioned employing a nearest neighbor weighted average customization. Proposed technique has proven its robustness in the experiments conducted employing the Bosphorous database. Furthermore the relative error values attained employing NNWA customization illustrated significantly low values. We compared the obtained results with the ASM and Procrustes Analysis.

İNSAN YÜZÜ MODELLEME İÇİN YARI OTOMATİK ÖZELLEŞTİRME

Özet

Modele dayalı imge işleme katı nesnelerde belirlenmiş nirengi noktalarını işaretleme ve nesne tanımada uygulama bulmuş bir yaklaşımdır. Ancak var olan sistemler aynı başarıyı esnek nesnelerin modellenmesinde gösterememektedirler. Bu yaklaşımlardaki ikilem modeli tespit ederken modelin esnekliğinin, veya esnekliği temsil ederken modelin güvenilirliğinin göz ardı edilmesidir. Esnek ve karmaşık bir yapı olan insan yüzü çoğunlukla bu yaklaşımlarla modellenmeye uygun değildir. Bu çalışmada 2 boyutlu tek bir resim ve 3 boyutlu genel bir telkafes modeli kullanılarak insan yüzünün 3 boyutlu modellenmesini konu alan özgün bir çalışma sunuyoruz. Bu yaklaşımda önceden belirlenmiş nirengi noktaları yüz resminin üzerinde işaretlenir. Nirengi noktalarının 3 boyutlu uzaydaki yeri ışın izleme metodu kullanılarak belirlenir. Nirengi noktaları haricindeki genel telkafes noktaları en yakın komşulara ait yer değiştirmelerin ağırlıklı ortalaması kullanılarak bulunur. Önerilen yaklaşımının güvenilirliği Bosporus 3 boyutlu yüz veri bankası kullanılarak gösterilmiştir. Buna ek olarak yaklaşımın literatürde sıkça kullanılan Procrustes Analizi çözümünden daha hassas sonuç verdiği gösterilmiştir. Önerilen algoritmanın karmaşıklığı literatürdeki diğer algoritmalardan daha düşüktür.

Acknowledgements

First of all I would like to thank my supervisor Asst. Prof. Taner Eskil, for his fatherly guidance and enormous support throughout my graduate studies. The opportunity to work with him was both intellectually rewarding and enjoyable. I consider it a great privilege working with him. Also I would like to express my sincere thanks to my research colleague at PI Lab Kristin Benli for her endless support and advice. I would also like to thank my graduate colleagues at Işık University who made my stay in Şile a very pleasant one. Finally I would like to thank my mother Luckshmi Abeysundera and my brother Janindu Abeysundera for their most valuable support and encouragement throughout my under graduate and graduate studies.

This study was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK) Grant No: 109E061

To my dad. . .

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2-D	**2 D**imension
3-D	**3 D**imension
AAM	**A**ctive **A**ppearance **M**odel
ASM	**A**ctive **S**hape **M**odel
AU	**A**ction **U**nits
COG	**C**enter **O**f **G**ravity
GPA	**G**eneralized **P**rocrustes **A**nalysis
HCI	**H**uman **C**computer **I**nteraction
HIGEM	**HI**gh polygon **GE**eneric **M**odel
ICP	**I**terative **C**losest **P**oint
KNN	**K**- **N**earest **N**eighbors
Mod	**Mod**ulo operation
NNWA	**N**earest **N**eighbor **W**eighted **A**verage
PC	**P**ersonal **C**omputer
PCA	**P**rincipal **C**omponent **A**nalysis
PRP	**P**erspective **R**eference **P**oint
SIFT	**S**cale **I**nvarient **F**eature **T**ransform
SURF	**S**peeded **U**p **R**obust **F**eatures

Chapter 1

Introduction

Over 2500 years ago ancient Greek philosopher Plato introduced the "Theory of Forms". He asserted that non-material information possesses the highest and most fundamental kind of reality. Human mind is very skilled in extracting this non-material information, in other words doing *pattern recognition*. This skill provides humans the valuable capability of abstraction to identify and differentiate objects. He explained this paradigm with the simple fact that an object X is not object Y because there is an ideal form of object X which we employ in identifying all objects of the same kind. Clearly there is an ideal form of object Y as well. He went on to emphasize that these forms are the only true subjects of study that can provide us with genuine information in the notion of a material object.

2500 years later modern science confirms Plato's Theory of Forms. Now we refer to Plato's forms as *schemas*. We are born with an innate ability to learn to recognize and identify objects. The way human mind do this is a highly complicated, astonishing process. Plato did not go as far to elaborate how human mind differentiates between objects of the same kind. He simply stated that there is a basic 'form' of any particular object that human mind represents it with, in other words "the model" to identify it.

For decades one of the main tasks of computer vision was to model and differentiate objects. Perhaps the ultimate and the most challenging goal was to create

a model of a human face. Face, being a non-rigid and deformable structure, is more sophisticated compared to many other objects we encounter. These characteristics of face makes modeling of it an intricate task. A simple movement of a facial muscle can drastically change the appearance of the face and convey a different, yet important message.

Face modeling in computer vision terminology is to create an epitome that can be exploited in simulating any facial behavior or expression. The goal of the face modeling studies is to develop an automated modeling schema. To be practical, this schema must systematically find facial features and correspondences to an actual face with minimal user intervention.

Modeling a face requires comprehensive knowledge of face anatomy. Human facial anatomy is flexible enough to convey thousands of different messages through contraction of facial muscles in varying degrees and combinations. These messages provide clues to our emotional state, our short-term feelings about our immediate environment, our mental health and even our personality or mood. Building a model that can simulate all these behaviors has proven to be a daunting task. Common approach in designing face models is to employ polygons. Vertices in each polygon represent a point on the skin of the face. The wireframe model that is being used in our research is designed following a similar technique.

In the modern information era, machine interaction with humans became ever so valuable. One very promising way of acquiring this information is to study a subject's behavior or emotions. There is no better way of unintrusively attaining this information than facial expression analysis. It is for that reason, analysis and modeling of human face has attracted many researchers in recent years [1, 2, 3, 4, 5].

Face detection and face recognition are few of the most popular applications of face modeling. A face model provides substantial information when compared with an image. Once a 3D model of the subject is acquired, vast number of synthetic training images under varying pose, expression and illumination conditions

can be effortlessly rendered. These variations in the synthetically rendered images of a subject can be used to expand the training dataset of a face recognition system. Therefore it drastically increases the accuracy and performance in face recognition [6].

Applications of face modeling are not limited to facial expression analysis and face recognition. Human face modeling has diverse range of applications including but not limited to medical purposes [2, 7, 8, 9, 10], computer animation [4, 11, 12, 13, 14, 15], video surveillance [3, 16], lip reading [17] and virtual reality [18, 19, 20, 21]. In many of these cases, especially in medical and security applications, very high precision is expected when designing a face model.

One of the main concerns in plastic surgery is to be able to predict the outcome of the operation. Doctors have not been able to do such predictions in the past. With the current advancements of technology this question may not go unanswered. By obtaining 3D volumetric data from a patient it is possible to generate a realistic model. A surgeon can utilize this model to predict the appearance of a person after performing a surgery.

A critical duty of law enforcement agencies is to identify suspects. Most of the time this is performed through the testimony of the witnesses who describe the suspect's face as detailed as possible. The faces are sketched manually by skilled artists. This is a time consuming as well as ineffective method. Furthermore the quality in re-creation of the face depends on the skill of the sketch artist. The whole process can be made more efficient with 3D face modeling. A properly designed 3D model with enough parameters can generate not only different poses but will be able to preview how the person would look like with different disguise techniques and ambient conditions. This can be employed in forensic investigation as well.

One sector that has been changed drastically by 3D modeling techniques is the entertainment industry. Various new techniques are used to keep the spectators' attention to a movie and arouse their curiosity. Especially in Hollywood, a new

wave of 3D animated movies achieved box office records, “Avatar” , “Toy Story 3” , “How to Train Your Dragon” , “Rise of the Planet of the Apes” to name a few. As the animations become realistic, more processing time needs to be allocated to animation of facial expressions.

Facial animations are also used in generating special characters in cinema. These animated characters are being used in performing dangerous stunts and to give life to situations that would be hard to create with professional actors. One good example was the ”Titanic” movie that was produced as early as 1997. Perhaps the most important advantage of using such animated characters is that there would be no risks involved. Considering the rate of advancement in digital technology, it is conceivable that digital actors will some day entirely replace humans. Yet there are still many hurdles to leap in the details of creating a photo-realistic human.

The game industry is another important application area where face modeling is gaining popularity. In more recent games a very high priority is being given to computer graphics. The whole concept of gaming has been changed in the recent few years. Modern gaming has a storyline underneath and the player is supposed to play a role in the story. Game makers rely on high quality computer graphics to make this happen. The most important concern in rendering faces in computer games is the speed. Most of the time graphic makers rely on high performance graphic cards to accomplish this. With the advances in computer hardware and software, obtaining cinematic quality in face modeling is not far ahead.

Face modeling applications are being utilized in the media sector, one rapidly growing field being the news casting. Face modeling in news casting is suprisingly less processor intensive than many other application areas. News casting is an emotionless task where there is no room for exaggerated facial expressions. In other words it requires a limited set of expressions. Hence such a face model requires a small set of parameters. Furthermore a newscaster would generally be stable in one position, showing only a frontal profile to the viewers. The

critical task here is synchronizing the lip movement. There has been an increasing attention to speech synchronization for virtual models [22, 23]. Less camera and head movements together with the limited set of emotions, has made face modeling a plausible solution in virtual news casting.

Human Computer Interaction (HCI) applications are making the technology experience more pleasant with the help of face modeling. We live in a world where many tasks are done through the help of computers. For instance most of the banking needs are computerized today and a client can do transactions without any human assistance. However most of these tasks require input from users. Such information is acquired using an input device and a simple interface. This can be made more pleasant and convenient by the help of virtual characters. The necessary information can be asked from the user and the user will only have to speak out the requested information. This is still an open field of research and its applications are in their infant stages. The most challenging task in here is to create a meaningful conversation between the man and the machine. There are some pilot projects that are able to continue a virtual dialogue. Perhaps the best example for this is the "Siri"that is introduced with iPhone 4S in 2011.

Law enforcement agencies may also use expression recognition to detect deceptive actions of a suspect. Today such techniques are even being employed by some companies to measure the faithfulness of their employees. Specifically security agencies such as CIA and NSA use such technologies in interrogating suspects as well as measuring the reliability of their agents. These applications also benefit from advancements in face modeling techniques.

Other various tasks like Avatar generation, fatigue detection, virtual mouse and numerous amounts of frivolous applications exploit the concept of face modeling [24].

As illustrated with the aforementioned examples, the existing and potential use of face models and animation span a diverse variety of scientific and artistic applications. Different applications require varying qualities depending on the

application context. For instance a surveillance application would require the face model to be both accurate and real time. However an application in plastic surgery would demand very high precision that could be attained at the expense of time. When it comes to movies and applications of the entertainment industry, accuracy becomes less important and rendering quality and smoother animations come to the forefront. In the computer science discipline, challenges include geometric modeling, rendering, animation, numerical simulations and interaction techniques. The system requirements vary greatly with the targeted applications.

Face modeling is becoming an essential part of many applications ranging from the medical sector (requires high precision) to the entertainment industry (requires high speed rendering). Due to the high demand, face modeling has been widely studied at length during the past years. The key challenge of face modeling from 2D images or video frames is the difficulty in establishing accurate and reliable correspondences between the 2D facial image and a generic face model, since faces are non rigid objects, displaying a high degree of variability in shape, texture and pose. Often attaining accurate results necessitate very high computational costs which can be a crucial point in any computer software.

There are two main approaches that have been commonly exploited by researchers in face modeling: parameterized and statistical. Parameterized face modeling exploits the known features of an object in modeling it. This requires an extensive knowledge of human anatomy in order to fully parameterize a human face. Also it is very important in here to have accurate stress-strain relationships of the muscles. Perhaps one of the most prominent parameterized face modeling approaches is mass-spring-damper systems. A statistical model relies on a set of training images to construct an accurate face model. A precise model requires a large and complete training set.

In this report we propose a novel method in appearance based face modeling to fit a 3D wireframe model onto a set of 2D images or a sequence of frames in a video stream. Different poses or different frames enable construction of different

regions of the face. Our method does not require extensive knowledge of the human anatomy, nor it requires a complete training data. It depends only upon a set of landmark points manually selected by the user. With carefully selected 32 locations we manage to fit our model onto different images with acceptable accuracy. We compared our results with the results obtained by Procrustes Analysis and Active Shape Model (ASM) techniques.

The rest of this report is organized as follows: Chapter 2 will illustrate the research that has been carried out in this field. It will provide an in-depth review of anatomy of face, which is a crucial part in accurate face modeling. Furthermore mass-spring-damper method, Procrustes Analysis and ASM techniques are explained in detail here. In Chapter 3 we will discuss the implementation of our proposed technique. Chapter 4 will illustrate a comparative analysis of the available face modeling techniques. We will compare some of these techniques with our proposed method. We will also discuss about the advantages and disadvantages of our approach. Chapter 5 discusses the improvements that could be made on the proposed system and speculates about the future research in the field, concluding this report.

Chapter 2

Literature Survey

2.1 Facial Anatomy

The objective of our research is to generate a 3 dimensional face model employing a 2 dimensional image of the face. Face modeling requires a comprehensive knowledge of the human face. Fleming and Dobbs [25] presented an in-depth review of face anatomy. In our studies we utilized their research outputs.

2.1.1 The Skull

Human skull consists of two major components: *cranial* and *facial components*. Approximately two thirds of the mass of the human skull is occupied by cranial. Remainder is occupied by the facial components. Both facial and cranial components are crucial in generating an accurate face model. Cranial delineates the structure of the face model while facial components are imperative in the construction of the facial texture. Another very important phenomenon of the human head is that it can be fit into a square that is of same height and depth. When modeling a face this is a helpful feature that can be utilized in preserving the proportions. Flemming and Dobbs [25] identify 12 important features on the human skull [Figure: 2.1].

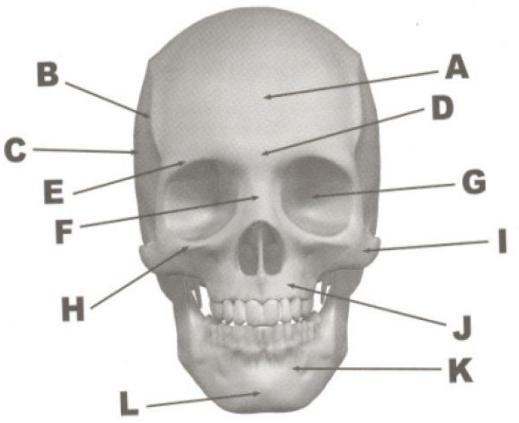

Figure 2.1: Human skull and its features [25]

Table 2.1: Labels of the 24 landmarks on human skull

Skull Feature	Importance
A - Frontal bone	Forms the forehead structure
B - Temporal ridge	Creates the square shaped appearance of the upper skull
C - Parietal bone	Defines the sides of the head
D - Nasion	Frontal bone meets nasal bone
E - Supraorbital margin	Creates the ridge above the eyes
F - Nasal bone	Creates the structure of the nose
G - Orbital cavity	The eye sockets
H - Infraorbital margin	Lower portion of the orbital cavity
I - Zygomatic bone	Creates the structure of the cheeks
J - Maxilla	Upper jaw bone
K - Mandible	Lower jaw, creates the chin structure
L - Mental protuberance	Tip of the lower jawbone

2.1.2 Facial Muscle Structure

Conserving the ratios of the facial trait dimensions is a key element in constructing accurate face models. For instance the distance from the nose tip to mouth is in general much smaller when compared with the distance from the lower lip to chin. Another critical task in designing a face model is to comprehend the behavior of the facial muscles. Effective implementation of the mass spring damping method greatly relies on understanding this phenomenon.

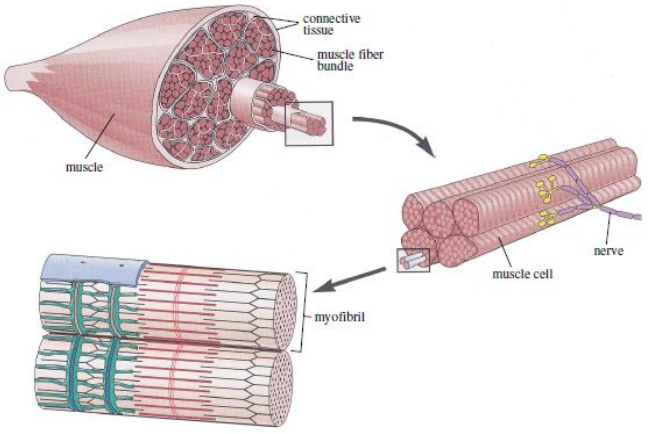

Figure 2.2: Structure of a muscle fiber [26]

Two main variations of muscles can be observed in the human head; *muscles of mastication* and *muscles of expressions*. Muscles of mastication primarily deal with the movement of lower jaw and are utilized in mouth movements. Muscles of expressions deal with generating expressions such as happiness and anger. There is no exact count of the facial muscles due to the fact that some muscles can be viewed as a combination of a cluster of minor muscles [Figure: 2.2]. Therefore different sources provide different figures as to the number of facial muscles.

The analysis of muscle structure and facial anatomy is of great significance in constructing parameterized face models. Contradictory to parameterized methods, statistical methods of face modeling do not rely on the prior knowledge of facial anatomy.

2.2 3D Face Databases

Conventionally modeling of human faces and analysis of facial expressions are performed either on static images or video sequences. 2 dimensional data is capable of providing limited information. Most importantly it does not provide any information about the profile of the object; hence no information about the depths of the feature points are provided. This is a major drawback in the current research efforts.

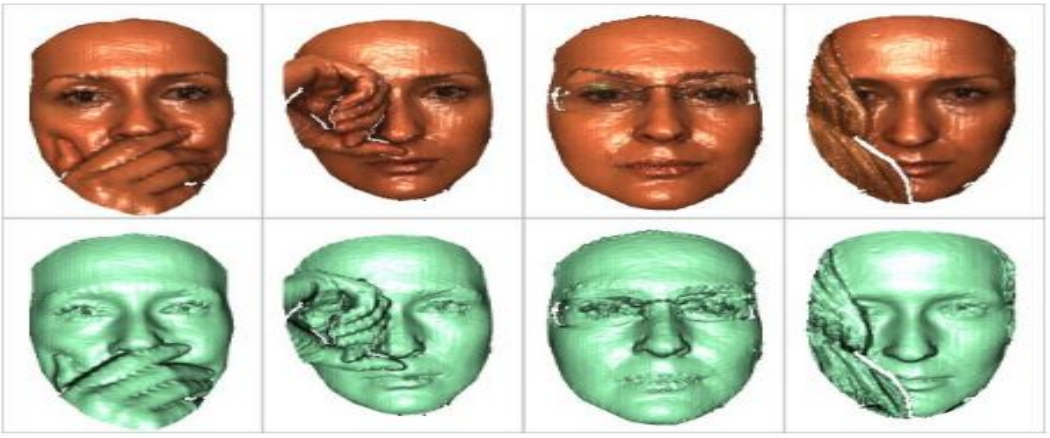

Figure 2.3: Natural occlusion due to yawning, crying, reading glasses and hair

Our research capitalizes on constructing a 3 dimensional face model through the use of a single 2 dimensional image. We can either use a 2 dimensional image or a 2 dimensional video sequence together with a generic wireframe to construct a 3 dimensional face model. We make use of a 3 dimensional data set that serves as ground truth and enable us to evaluate the accuracy of the developed 3 dimensional model. With the researchers' growing interest in face modeling significant amount of different 3 dimensional data sets were made available. A brief description of some of the existing 3 dimensional datasets are provided in [Table:2.2].

In this research we employed Bosphorus 3D Face Database [27] to carry out our experiments and the embodied 3D data to evaluate the accuracy of the proposed technique. The database contains 4666 3-dimensional images captured from 105 different subjects. Each subject has approximately 44 images. Among the 105 subjects 60 of them are men and 45 subjects are women. Majority of the subjects are males between 25 - 35 years of age. 18 of the 60 male subjects have beard or moustaches. Database consists of images with partial occlusion due to the hair, reading glasses or poses that demonstrate crying and yawning [Figure: 2.3]. These partially occluded images were produced maintaining neutral poses.

Even though our research focuses on neutral poses of the subjects, many other expressions are also depicted in the Bosphorous dataset. Among these poses, six

Table 2.2: Face datasets

Database Name	**Subjects**	**Resolution**	**Variations**
BU-3DFE	100	1040 x 1329	Pose Expression
BU-4DFE	101	1040 x 1329	Pose Expression
Honda / UCSD	20	640 x480	Rotation Partial occlusion
UOY	97	1040 x 1329	Pose Expression
Face in Action	200	640 x 480	Pose Illumination
Extended M2VTS	295	720 x 576	Speech Rotation
Max Plank Institute	246	786 x 576	Pose Facial action
VidTIMIT	43	512 x 384	Speech Rotation
Texas	284	720 x 480	Pose Expression
Yale	10	640 x 480	Pose Illumination
PIE	68	512 x 384	Pose Illumination
AR	126	768 x 576	Occlusion Illumination Expression
CAS-PEAL	1040	360 x 480	Pose Illumination Expression
CASIA	123	640 x480	Pose Expression Illumination
EQUINOX HID	91	240 x 320	Speech Expression Illumination
Bosphorous	105	1374 x 1260	Expression Occlusion Pose

Figure 2.4: 6 basic universal emotions

Figure 2.5: Landmark locations provided in the Bosphorous database [27].

basic universal emotions, anger, disgust, fear, happiness, sadness and surprise are available [Figure: 2.4]. The reason behind the exploitation of neutral poses in our research is to simplify creating an initial 3 dimensional model. Since our generic wireframe model is anatomically accurate, it can be exploited in synthesizing different facial expressions once the 3 dimensional model is constructed.

On each image of the database 24 feature points are marked. These feature points are illustrated in [Figure: 2.5] and [Table: 2.3]. The corresponding landmark vertices on the generic face model are presented in [Figure:2.6].

2.3 Data Cloud Manipulation Techniques

Discovering the correspondence points and alignment of shapes is both a critical and an intricate task. Statistical analysis of shapes plays an important role in

Table 2.3: Labels of the 24 landmarks

01. Outer left eye brow	02. Middle of the left eye brow
03. Inner left eye brow	04. Inner right eye brow
05. Middle of the right eye brow	06. Outer right eye brow
07. Outer left eye corner	08. Inner left eye corner
09. Inner right eye corner	10. Outer right eye corner
11. Nose saddle left	12. Nose saddle right
13. Left nose peak	14. Nose tip
15. Right nose peak	16. Left mouth corner
17. Upper lip outer middle	18. Right mouth corner
19. Upper lip inner middle	20. Lower lip inner middle
21. Lower lip outer middle	22. Chin middle
23. Left ear lobe	24. Right ear lobe

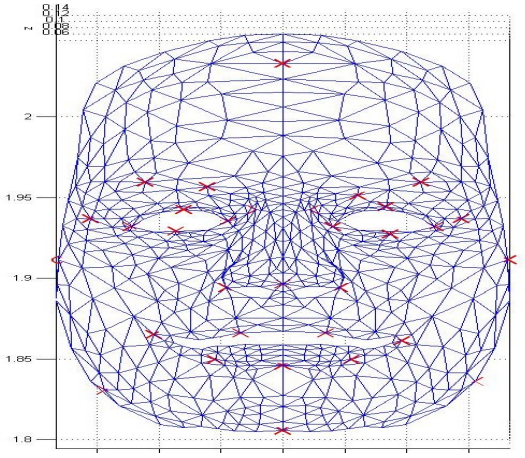

Figure 2.6: Landmark locations employed in our experiment

determining the correspondence of the key traits and determining the validity of the algorithm that is being employed in positioning the landmarks. There are several techniques that have been utilized in shape correspondence and data cloud matching. Among them two are of very high importance: Procrustes Analysis [28] and the Iterative Closest Point (ICP) algorithm [29, 30].

The generic model we employ and the images exploited in this research are of two different scales. Original ICP does not offer a scaling transformation [31]. However today extensions that provide affine scaling in ICP is available. In our research we have made use of Procrustes Analysis both in initial alignment and

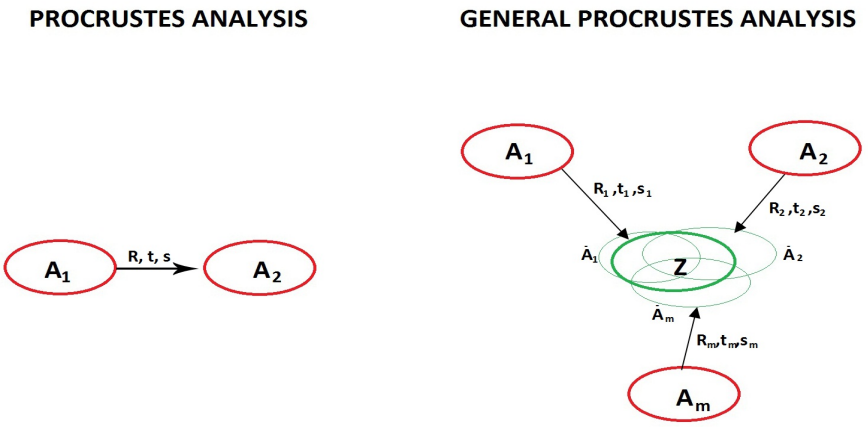

Figure 2.7: Procrustes Analysis and General Procrustes Analysis

performance evaluation of the proposed method. Comprehensive descriptions of Procrsutes Analysis and ICP are provided in the subsequent sections.

2.3.1 Procrustes Analysis

Procrustes analysis [28] is a mathematical technique for superimposing one or more shapes onto another. This is performed by exploitation of isotropic scaling, translation, and rotation. Procrustes Analysis iteratively finds the best fit between two or more shapes outlined by the landmark points. It only allows rigid body transformations on the datasets and the transformations conserve the relative distance between feature points. Procrustes Analysis has many different forms and variations. Of these different variations, General Procrustes Analysis [32], otherwise known as GPA is one of the more commonly exploited techniques in shape correspondence.

The key difference between Procrustes Analysis and General Procrustes Analysis (GPA) is that, GPA employs a set of shapes in the alignment process. Procrustes Analysis performs this task exploiting only two shapes; original and the target [Figure: 2.7]. It utilizes a least square shape metric that involves the alignment of two shapes with one to one correspondence.

The alignment process of the General Procrustes Analysis consists of six fundamental stages [33].

1. Normalize all shapes to unit size and translate their center of masses to origin.
2. Determine mean shape $\mathbf{m} = \sum \mathbf{x_i}$.
3. Align each shape with $\mathbf{m}$ via rotation T_i.
4. Re-calculate $\mathbf{m} = \frac{1}{n} \sum T_i \ (\mathbf{x_i})$.
5. Translate m to origin, and normalize its size.
6. Repeat the alignment process until convergence.

Fundamentally General Procrustes Analysis attempts to obtain the transformation T_i which minimizes the difference between the mean shape $\mathbf{m}$ and target shapes $\mathbf{x_i}$ [Equation: 2.1].

$$D = \sum ||\mathbf{m} - T_i(\mathbf{x_i})||^2 \tag{2.1}$$

and the mean shape is updated as,

$$\mathbf{m} = \frac{1}{n} \sum T_i(\mathbf{x_i}) \tag{2.2}$$

Procrustes Analysis requires same number of input arguments for each shape and provides distance D, rotation component T and scaling component s. This information can be utilized in measuring the similarity of two shapes. In our experiments we exploited Procrustes Analysis employing 32 designated landmark traits and obtained the similarity transformation parameters in the form of scale, rotation and translation. These transformation parameters were then applied on all 612 nodes of the generic wireframe to align the generic 3 dimensional model.

2.3.2 Iterative Closest Point

Iterative Closest Point algorithm [29,30] is an iterative approach to minimize the difference between two clouds of points. At each stage data point correspondences are reconsidered as the solution comes closer to the local minimum error. As any other gradient descent method, ICP performs best when a relatively good starting point is provided. This significantly reduces the possibility of being trapped in a local minimum.

ICP algorithm performs data matching in six simple steps.

1. Determine sample points from the data cloud x.
2. Determine sample points on data cloud y.
3. Calculate the weight of the correspondences [Equation: 2.3].
4. Reject unsuitable point pairs.
5. Assign an error for the current transform.
6. Go to step 2 until the error converges to a minimum.

There are several ways in selecting samples from datasets. In the original ICP Besl and McKay [29] employed all available points. Another commonly used method in selecting samples is to apply uniform subsampling. Masuda *et al.*[34] proposed a random sampling in each iteration to select the sample points.

Once the points are selected the next step is to match points in the data clouds. In the original algorithm closest points are selected as the match. However in another version of the algorithm [30] normal shooting is done. This is to draw a perpendicular line from the sample point to the other data cloud. The first intersection is selected as the match.

Weighting pairs is performed to recognize the resemblance of the objects. Two commonly employed techniques are available for weighting of pairs. The simplest

method is to assign constant weights to all sample points. But more commonly used technique is to assign higher weights to points with lower point to point distance.

$$Weight = 1 - \frac{Dist(P_1, P_2)}{Dist_{max}} \tag{2.3}$$

Again there are numerous methods for rejecting incompatible point pairs. If exist, pairs containing points on the boundaries of the data cloud are generally rejected. Rejection is done using a threshold value for point to point distance. Also a predefined percentage of less incompatible pairs based on a certain metric can be rejected.

Error metric generally is the sum of the squared distances between the corresponding points. Repeatedly sample points are generated and current transformations are applied. The intention of this iterative process is to acquire the transformation that minimizes the error metric.

The main advantage of ICP over Procrustes Analysis is that it does not require same number of sample points as input. Also it does not require the knowledge of correspondence between sample points.

2.4 Face Modeling Techniques

Various image based approaches have been utilized in automatic generation of human face models. These approaches can be distinguished by the type of the image data they employ. A single image, two orthogonal images, a set of images or a video sequence can be exploited in generating face models [5].

There has been considerable amount of research on constructing face models employing multiple images captured from different angles. Stereo vision through two cameras can be utilized in constructing a model. Sometimes as many as five cameras from different angles are exploited in the process of building face

models [1, 5, 8]. Although this approach can produce promising results, it is not very practical for model generation in daily life. His system is capable of automatically generating the correspondence points between images. However it requires a complicated set of apparatus to attain acceptable results.

A sequence of video can be utilized for modifying the model, exploiting a single camera to acquire different poses of the face. In this front there are two popular approaches; statistical modeling and parametric modeling. Subsequent sections will provide an extensive review of these approaches. In our research we are interested in constructing a face model by use of a single image obtained from a single camera.

2.4.1 Parameterized Face Modeling

Parameterized face modeling is one of the more tedious methods of face modeling. In order to generate an accurate face model one should possess a substantial knowledge in human facial anatomy. The main aim of parameterized modeling is to construct a model that would integrate all the key poses of a subject. However as the number of the key poses increases building a model that can replicate these poses becomes a daunting task. This is exactly the situation with the human face. Face is a non-rigid body that is capable of producing many different mimics that would generate distinctive emotional expressions.

Intrigued by the complexity of the task, Parke developed one of the earlier parameterized face models [35]. Parke's model originated as an extension to key-pose animation. His intention was to develop an encapsulated model that could generate a wide range of diverse faces and facial expressions exploiting a small set of input parameters. Parke's model was quite constrained due to the relatively simple techniques he employed. Today the availability of more complex modeling and image synthesis techniques has enabled more complicated parameterized models that permit better facial animations.

The ideal parameterized model would be the one that allows any possible face with any possible expression to be replicated by merely selecting an appropriate parameter value set. A parametrization that enables all possible individual faces and all possible expressions and expression transitions is referred to as a *complete* or *universal* parametrization [36].

There are two methods that are being widely employed in constructing facial parameter sets. More commonly employed technique is to observe the surface properties of the face and then to develop a parameter set that would replicate the observed behavior. The second method which is more robust is to understand the underlying anatomical structure in triggering a mimic and develop a parameter set based on the mechanism of the facial expression. A combination of both of these methods can be utilized in creating a model. In this hybrid approach parameters are based on anatomical perception wherever possible and are enhanced as needed through observation.

In parameterized modeling two different kinds of parameter sets are available. These are the expression parameters and conformation parameters. Expression parameters are the parameters that are employed in controlling expressions of the face. Some of the expression parameters found in a facial structure are the eyebrow - eyelid separation, eyelid opening, mouth corner position and upper lip position. Conformation parameters are utilized in controlling the general structure of the face. Jaw width, chin shape, eye size and eye-to-eye separation are a few of the conformation parameters in the human face. To some extent these two sets overlap, but in practice they are considered to be distinctive.

There are three key modeling techniques that require the knowledge of facial structure that triggers muscle mimics: Mass-spring systems, vector representation and layered spring mesh representation. Mass-spring systems disseminate muscle forces in an elastic spring mesh that models skin deformations. Vector representation employs motion fields in delineated regions of influence to observe the facial mesh deformations. Layered spring mesh as its name suggests is an

extended version of mass-spring systems. It exploits multiple layers such as skull, muscle and skin to model the face.

Benchmark in parameterized face modeling is the mass-spring damping method which was first introduced by Platt and Balder [37]. It evolved as a result of a series of experiments on modeling human anatomy. It was initially introduced as a technique to model muscles. Mass-spring system replicates the skin deformations by propagating muscle forces in an elastic mesh. The skin is modeled as a mesh and the muscles are represented by springs attached to nodes of the skin conforming with the anatomical structure of the face. The entire structure is designed to achieve biphasic stress-strain relationship to simulate the dynamics of a real face. Each contraction of a muscle exerts a pressure on the mesh, causing it to deform. These different deformations depict various facial expressions. The forces exerted by the muscles on the skin comply with Hook's law [Equation: 2.4].

$$\mathbf{F} = \mathbf{K}\mathbf{x} \tag{2.4}$$

$\mathbf{F}$ represents the forces exerted on the skin mesh and $\mathbf{x}$ is the displacements of the ends of the springs from their equilibrium states. $\mathbf{K}$ is the spring constant matrix that reflects the stress-strain relationship of the skin or muscles. Platt's later works [37] illustrate a facial model represented as a collection of muscles combined as a block in defined regions of the facial structure. His model contains 38 muscle region blocks connected with a spring system. Later Zhang (2001) exploited mass-spring models in real time animation of facial expressions [38].

Terzopoulos and Waters [39], proposed one of the early facial models that conform to the anatomical structure and dynamics of a human face [Figure: 2.8]. Their model contains three layers that correspond to skin, fatty tissue and muscles. Spring elements were employed in connecting muscle elements to the mesh nodes of each layer.

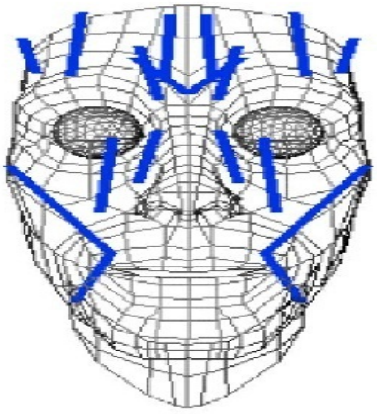

Figure 2.8: The muscle model proposed by Waters [26]

2.4.2 Statistical Face Modeling

Parameterized face modeling requires extensive knowledge about the anatomy of human face. It is a tedious task to construct an accurate parameterized model. Furthermore it is challenging to model the different stress-strain relationships of various muscles. Due to these difficulties resarchers turned to statistics to model human faces. Statistical face modeling relies upon a set of training images in constructing a model. This proved to be both convenient and an efficient method of modeling faces.

Faces can vary widely, but variations of the face can be broken down into two main factors; changes in shape and the texture. Both of these features can also vary among the poses of the same individual due to changes in expressions and camera viewpoints. In developing an appearance based face model statistical techniques can be employed.

To generate a statistical model we rely on obtaining a sufficiently large data set of facial images with a collection of data points defining the correspondences within the set. The positions of the feature points are exploited in defining the shapes, and the pattern intensities are analyzed in developing a texture model.

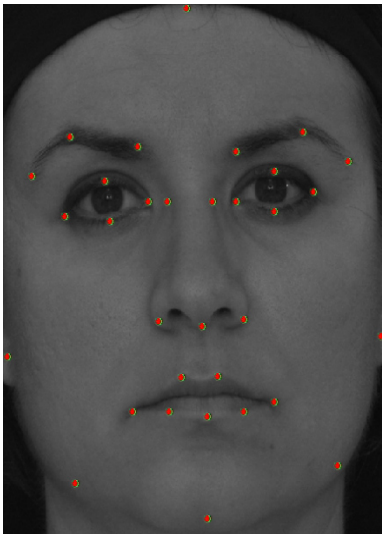

Figure 2.9: 32 landmark points employed in defining the facial features [27]

How one chooses the training data set is crucial in constructing a shape model. The chosen set should cover the types of variations that is to be represented by the model. Another very important point is that the resolution of the data set should be at least as high as the testing data.

Detection of the key traits is done either automatically or manually with human intervention. Since these traits are exploited in correspondence throughout the testing set, a predetermined number of traits needs to be marked in images. We exploited a 66 point facial feature set in our preliminary experiments to define a model. Among them 40 feature points are employed in delineating the contour of the face. Remaining 26 points are exploited in defining the key features of the face such as the eyebrows, eyes, nose and mouth. During our studies we managed to reduce the number of traits to 32 by trial and error [Figure: 2.9].

Landmark selection should be performed cautiously. A good landmark should be conveniently located on facial images. Commonly, marking of landmark locations are done manually. This is a very time consuming task that needs to be carried out throughout the training set. Hence one should be careful to select minimally sufficient number of landmark points that would describe a face. In more recent

studies, semi automatic [40] and automatic methods have been developed to aid the marking process with a reduced number of landmark traits.

Corners of eyebrows, eyes, nose and mouth can be precisely located, which make them excellent feature points. Unfortunately human face does not contain many such features that would provide us sufficient statistics. As a result we have to mark a modest amount of feature points in between the corners [41]. The annotated feature point locations for each image are then put into a vector as the training dataset in order to perform statistical analysis.

It is vital to verify that the dataset is properly aligned before we commence statistical analysis. There are a number of methods that have been utilized in aligning a dataset. Procrustes Analysis is one such method [28]. Another widely employed method for aligning data is the Iterative Closest Point (ICP) algorithm [42]. We covered the details of these algorithms in Section 2.3.

Once the dataset has been aligned, our next target is to model the shape variations. It is important that our shapes comprise of the most indispensable variations. An initial dimension reduction technique is applied to streamline the shape modeling process. One such technique that is commonly applied in dimension reduction is Principal Component Analysis (PCA) [43]. [Figure: 2.10] illustrates 12 different modes of faces that are extracted from a training set. Note that the dimensionality of the model is reduced from twice the number of pixels to only 12.

Statistical model of texture is built using the intensity of color over an image patch. The texture model is constructed abiding a three step process. Initially pre-computation of the pixel positions of the sample in the model reference frame is done. Then employing a warping function landmarks of the mean shape are mapped to the target points. Finally each element in texture model is sampled utilizing the target image at corresponding location using a predefined neighborhood of the feature point [41]. A texture sample always contains a fixed number of pixels, independent of the size of the objects in the target image. Once this is

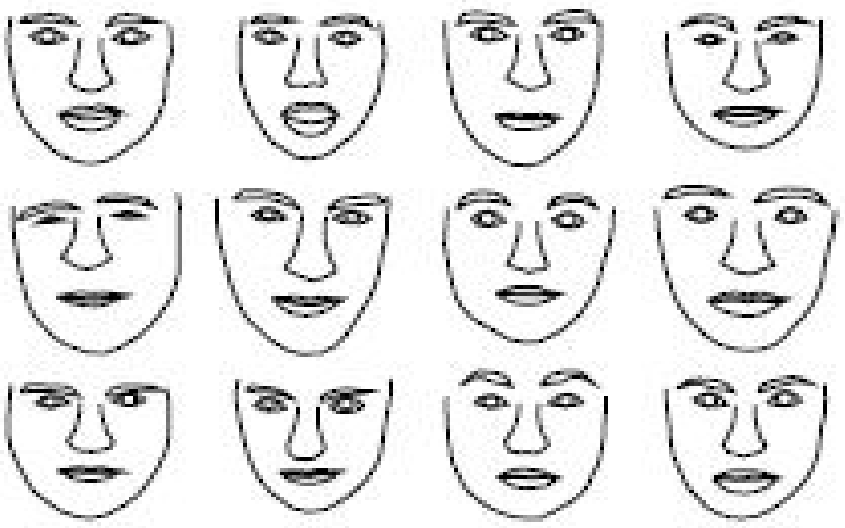

Figure 2.10: Principal components of a face shape model [41]

performed it is then normalized to remove global lighting effects. Generation of a statistical model of texture is depicted in [Figure: 2.11].

Shape and texture models can be coalesced in order to obtain a combined model of appearance. Shape and texture are in general correlated. Therefore again PCA is applied to reduce dimensionality before generating a combined model.

2.4.2.1 Active Shape Model

Active Shape Model (ASM) [44 , 45] can be considered as the pinnacle of research in statistical face modeling. We employ ASM in our research as a benchmark to assess the performance of our proposed method. ASM relies on selection of a reliable set of training data with adequate number of identified feature points [45]. Given a rough initial approximation, an instance of the model will be fit onto the image of the object. A set of shape parameters are utilized in defining the shape of the object in an object centered co-ordinate frame. Then an iterative approach is used in fitting the model onto the object.

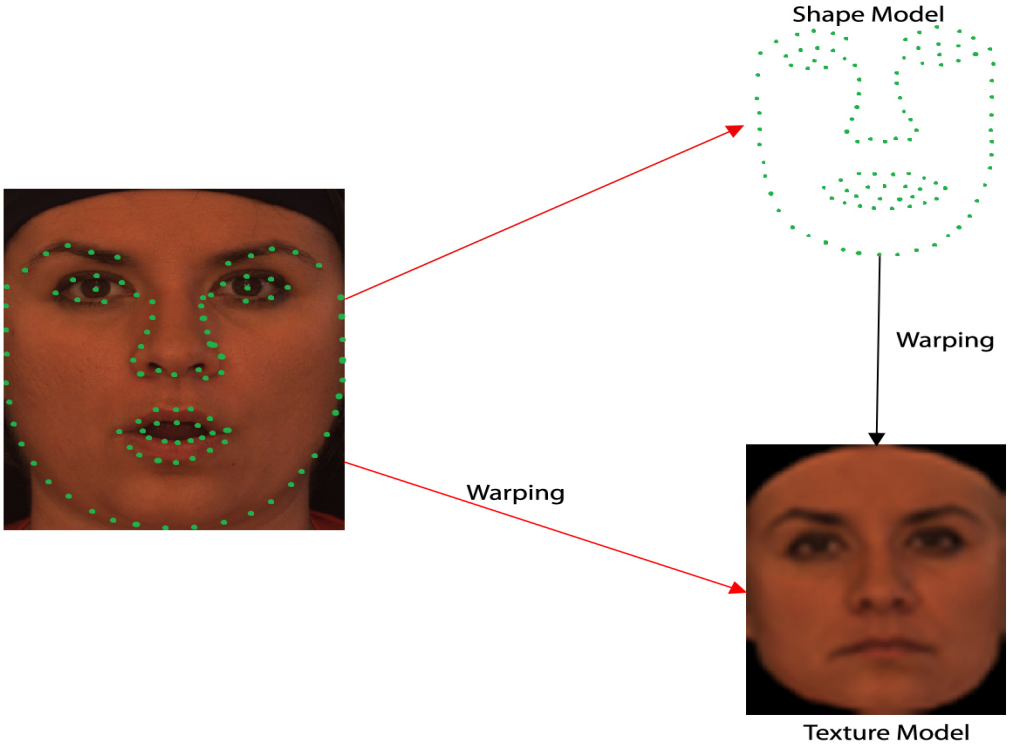

Figure 2.11: Extraction of shape-free texture model

ASM relies on an adequately large training set of images to construct a model that is flexible enough to cover the different variations of an object. An object shape is epitomized utilizing a series of labeled points or landmarks. In order to demonstrate the overall shape and the details of the object, the number of feature points should be sufficiently large. Labeled training set S encompasses N shapes each comprising of n landmarks.

Each shape is represented by a matrix of feature point coordinates as in [Equation: 2.5].

$$\mathbf{X_i} = [(x_{i1}, y_{i1}), (x_{i2}, y_{i2}), ...(x_{in}, y_{in})]^T \tag{2.5}$$

where (x_{ij}, y_{ij}) represents the j^{th} landmark coordinate of the i^{th} shape.

In order to simplify the calculations, we employ PCA and reduce the dimensions of the shape space. In many applications it can be assumed that the first few

principal components accounts for a sufficient percentage of the total variance of the original data.

We can express the difference between an observation and the mean of all observation as a linear combination of the principle components, since this *dissimilarity vector* will also lie in the $2n$ dimensional space spanned by principal components. *Dissimilarity vector* between x_i and the mean vector $\bar{x}$ can be represented as in [Equation: 2.6].

$$\mathbf{dx_i} = \mathbf{x_i} - \bar{\mathbf{x}} \tag{2.6}$$

where,

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x_i} \tag{2.7}$$

In these equations $\mathbf{x_i}$ and $\mathbf{dx_i}$ are 2n $\times$ 1 vectors that are obtained by reshaping the shape matrix.

Representing the difference $\mathbf{dx_i}$ as a linear combination of the principal components we acquire [Equation: 2.8].

$$\mathbf{dx_i} = b_{i1}\mathbf{a_1} + b_{i2}\mathbf{a_2} + b_{i3}\mathbf{a_3} + ...b_{i(2n)}\mathbf{a_{i(2n)}} \tag{2.8}$$

Where, eigenvectors are represented with $\mathbf{a}$ and the scalar weights that construct $\mathbf{dx_i}$ are represented by b_i. But since $x_i = \bar{\mathbf{x}} + \mathbf{dx_i}$, this yields [Equation: 2.9].

$$\mathbf{x_i} = \bar{\mathbf{x}} + \mathbf{a}b_i \tag{2.9}$$

Assuming that the first t principal components represent a sufficiently high percentage of the total variance of the original data, we can further simplify the [Equation: 2.9] to yield [Equation: 2.10].

$$\mathbf{x_i} = \bar{\mathbf{x}} + \mathbf{A}b \tag{2.10}$$

where,

$$\mathbf{b} = [b_1\ b_2\ b_3\ b_4\ ....b_t]^T \tag{2.11}$$

and,

$$\mathbf{A} = [\mathbf{a_1}\ \mathbf{a_2}\ \mathbf{a_3}\ \mathbf{a_4}\ ....\mathbf{a_t}]^T \tag{2.12}$$

This model has $(2n - t)$ fewer dimensions with respect to the original shape space. Yet it still accounts for a considerable amount of variability in the dataset. Moreover these eigenvectors characterizes the specific variability of the class the shapes belong to.

Once the training process is completed, acquired mean shape is exploited in the stage of modeling the candidate subjects. Abstract algorithm of the shape alignment process is illustrated in [Figure: 2.12].

Comprehensive algorithm exploited in the modeling process is described below:

1. Initialize shape parameters, b to zero.
2. Generate the model points using $\mathbf{s} = \bar{\mathbf{x}} + \mathbf{A}b$
3. Find pose parameters (x_t, y_t, s, θ) to align observed shape $\mathbf{y}$ with the current model s.

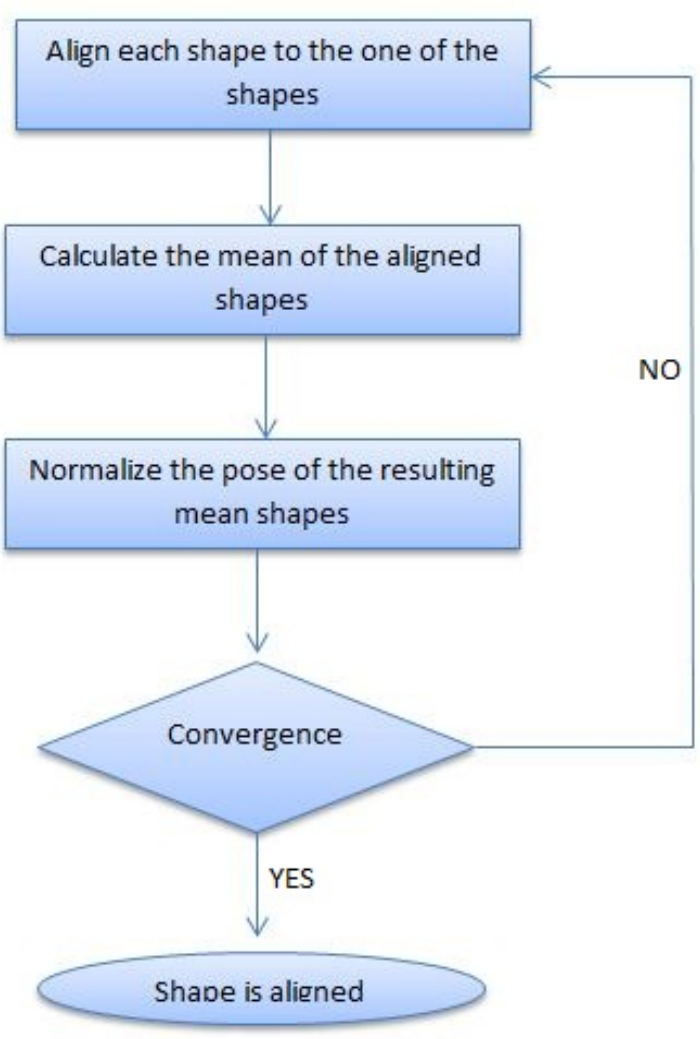

Figure 2.12: Shape alignment algorithm

$$\mathbf{y}^{'} = T_{x_t,y_t,s,\theta}(\mathbf{y}) \tag{2.13}$$

4. Update the model parameters to match to shape model with $\mathbf{y}^{'}$ using [Equation: 2.10].

$$\mathbf{b} = \mathbf{A}^{\mathbf{T}}(\mathbf{y}^{'} - \bar{\mathbf{x}}) \tag{2.14}$$

5. If not converged return to step 2.

The transformation process exploits Procrustes Analysis in the alignment stage. Procrustes Analysis follows a three step progression. Initially the image is positioned on top of the model aligning the Center Of Gravities (COG). Then scaling and rotation is performed on the model progressively to obtain the best fit for the object. A more detailed description of Procrustes Analysis is provided in Section 2.3.1.

ASM can be implemented in a multi-resolution framework to increase its efficiency and robustness. A faster algorithm is attained by first modeling the object in a coarse image and then refining the shape model with a series of finer resolution images. Since this follows a cascade structure more expensive computations are performed at the later stages making this both an efficient and a reliable algorithm. In order to do this the subsequent levels in the pyramid are created by smoothing and subsampling.

The main hindrance of ASM is that it solely depends on the shape parameters in modeling an object. Image texture embodies substantially more information about the object characteristics when compared to the shape parameters. Motivated by this fact, Active Appearence Model (AAM) evolved after a substantial amount of research conducted in this field [45]. AAM utilizes both shape parameters and texture parameters in model adaptation.

Chapter 3

Semi-Automatic Customization

We propose a semi automatic wireframe fitting technique to represent the appearance of a subject with a generic model. Previously a similar approach was proposed by Krindis and Pitas [46]. They utilized a 2 dimensional mesh and manually labeled corresponding positions of the face image and wireframe model. Our method differs from Krindis's since we extend customization to non landmark vertices through a semi-automatic technique in landmark positioning. Our algorithm employed in wireframe fitting can be sketched as follows.

1. Select feature points on image and landmark vertices on the generic 3 dimensional model.
2. Find the center of gravity of the target image and the generic 3 dimensional model.
3. Translate the model in 3 dimensional space to align the projection of its center with the center of the image.
4. Scale the model to fit its projection to image feature points.
5. Determine and update the coordinates of the landmark vertices using ray tracing through the target image.
6. Apply distance based nearest neighbor weighted average algorithm to non landmark nodes to fit the model to the target image.

Starting from the introduction of our 3 dimensional generic face model we will comprehensively elucidate the steps of our algorithm in the subsequent sections.

3.1 A New Generic Face Model: HIGEM

In our research we propose a novel method for constructing a 3 dimensional face model exploiting a single 2 dimensional image. We require a 3 dimensional dataset to conduct Procrustes alignment experiments and employ it as ground truth in evaluating the performance of our technique. For this purpose we employed Bosphorus 3 dimensional face dataset. The details of this database were covered in Section 2.2. In face modeling domain there are two key approaches pursued by the researchers; parameterized and statistical techniques. Parameterized modeling benefits from the prior knowledge of facial anatomy. Statistical modeling exploits statistical techniques as its name suggests. An in depth review of these techniques is provided in Section 2.4.

We take a different path in creation of an accurate 3 dimensional face model employing static images. Our proposed method does not require prior knowledge of human anatomy as parameterized modeling and does not rely on a training dataset as statistical modeling techniques. It combines the robustness of the parameterized modeling and the convenience of the statistical modeling methods.

In the context of this research we developed HIgh polygon GEneric Model (HIGEM). HIGEM is a generic wireframe model that conforms to human face anatomy. Research works conducted in this area has made very little effort in engineering such a generic model. Candide 3 introduced by Ahlberg [47], with 180 vertices is perhaps one of the most sophisticated models available prior to our work. Prevailing generic models were inadequate and we required a more sophisticated model that would enable us to replicate human face more precisely.

Our developed model comprises of 612 nodes [Figure: 3.1]. These nodes come together to generate 1128 polygons, which are known as *faces* in computer graphics

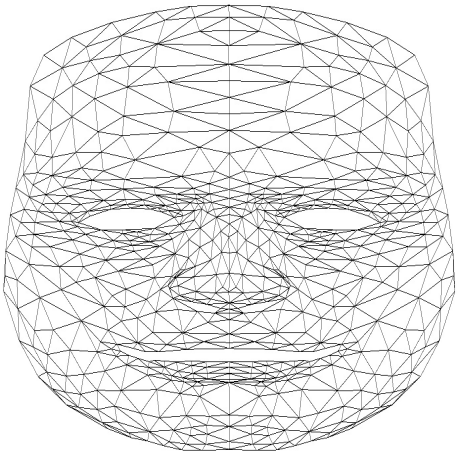

Figure 3.1: HIgh polygon GEneric Model: HIGEM

terminology. Our model incorporates anatomically accurate muscle forces and the edges of the wireframe are modeled as springs, permitting it to replicate diverse facial mimics. This feature enables us to utilize the model in facial animation.

3.2 Selection of Landmark Locations

Our proposed technique relies on the key trait locations in generating an accurate 3 dimensional face model of a target image. Therefore precise selection of these traits is critical.

We manually selected feature points on 104 2 dimensional images of the Bosphorous dataset. Theoretically, the reliability of the proposed method improves as the number of traits increases. Marking a substantial number of features on a facial image proved to be a daunting and time consuming task. Therefore it is important to select the minimal number of landmark traits that is sufficient in defining the human face.

Initially we selected 42 landmark vertices on the generic wireframe model that delineate the shape of a face [Figure: 3.2]. Selection of the landmark vertices is performed only once in accordance with human face anatomy. They are carefully selected to replicate the muscle movements that aid in constructing diverse facial expressions. We selected eyebrows, contours that define eyes and mouth, nose

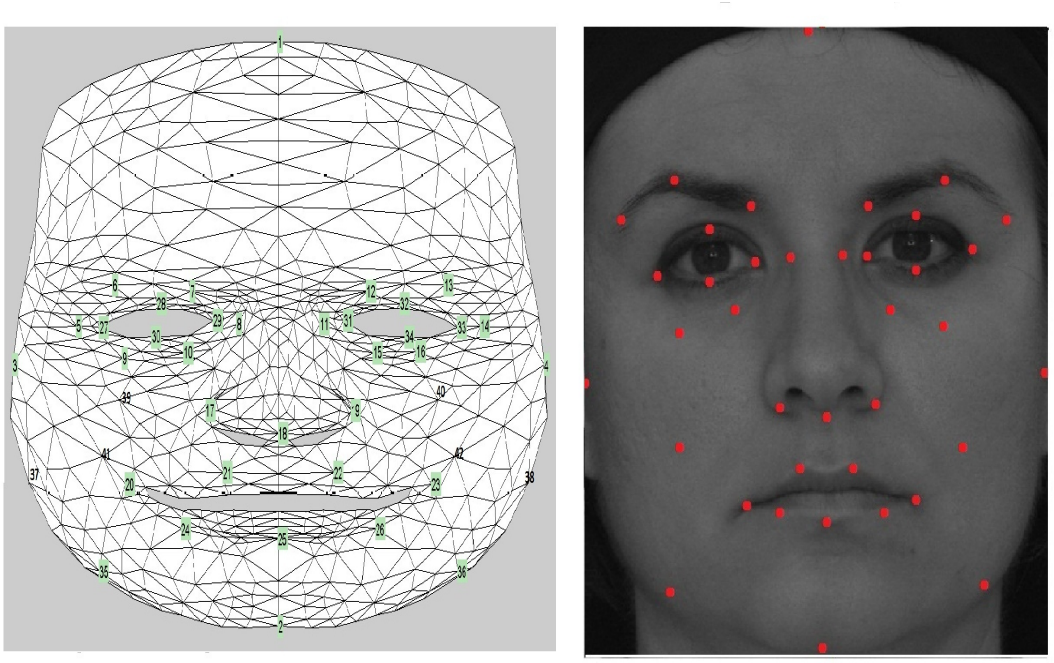

Figure 3.2: Generic wireframe model and a sample face image with 42 key traits

tip and nose saddles as some of our landmarks. Also we employed 6 landmarks to define the contour of the face. Height and width of the target face image is determined by its bounding box. Width of the target images is determined employing the points that represent the two ears of the face.

These landmark points are fixed, in other words they are selected only once on the generic face model. Through trial and error we determined 32 as a sufficient number of landmarks to define a human face [Figure: 3.3]. This analysis is deferred until the introduction of the performance criteria. It provides us the opportunity to remove the key traits on cheeks which were extremely hard to locate accurately on both face image and its 3 dimensional data cloud.

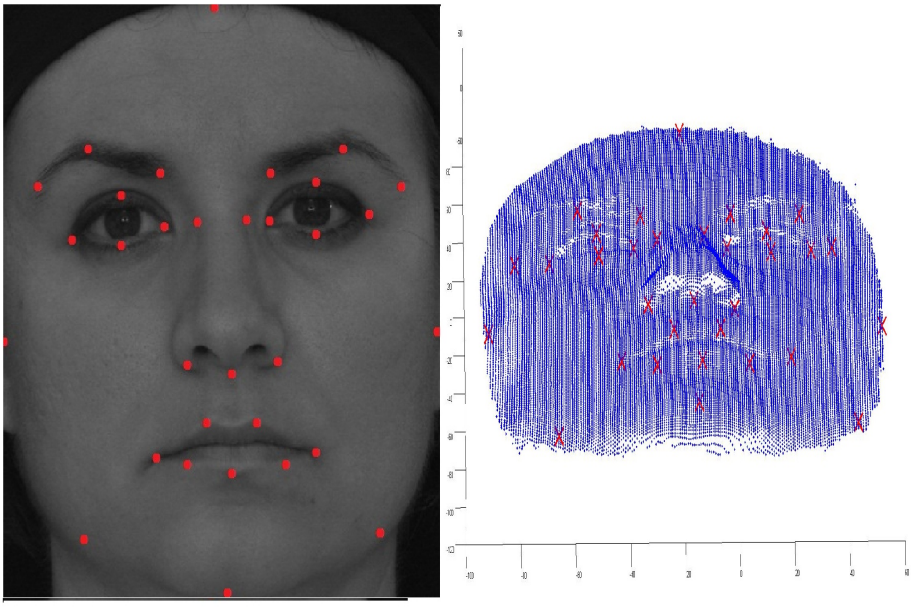

Figure 3.3: Generic wireframe model with 32 landmark vertices, a sample face image and its data cloud that serves as the ground truth

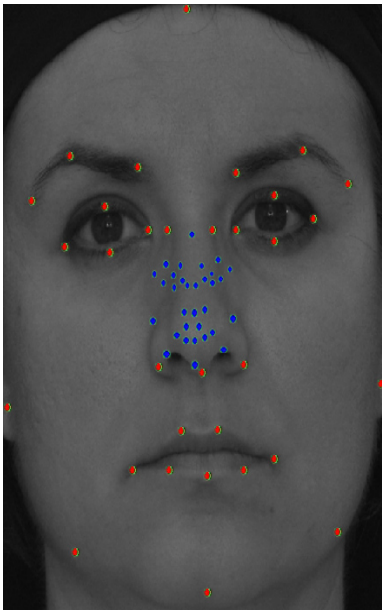

Figure 3.4: Centroid alignment of wireframe model and image

3.3 Nearest Neghibor Weighted Average Customization

3.3.1 Model Alignment

The input to the model alignment stage is the generic face model with marked landmark vertices and a face image. Once the feature points are marked on the image, the next step is to estimate the 3 dimensional coordinates whose projections are the 2 dimensional facial points. This is performed by ray tracing from the image plane to the 3 dimensional wire frame model coordinates. Before performing this operation generic wireframe model and the target image should be properly aligned and scaled. The initial alignment is done by positioning the generic wire frame in 3 dimensional coordinates so that the projection of the center of gravity of the model landmarks collides with that of target image [Figure: 3.4].

Scaling in x, y and z axes is done separately [Figure: 3.5]. The height and width of the face image are acquired by use of the feature points. We employ these height and width parameters to scale the model in x and y directions. Scaling in the z direction is performed employing the same scaling factor as the x direction. This is done relying on the studies of Fleming and Dobbs [25]. They observe that the human head can be fit in to a rectangular box that is of same width and depth.

3.3.2 Estimation of Landmark Vertex Coordinates.

We utilize ray tracing for the purpose of back projecting the 2 dimensional feature points into the 3 dimensional coordinate space. Our implementation of ray tracing is similar to the ray tracing method employed in computer graphics applications. Ray tracing in computer graphics is to generate an image by tracing the path of the light through the pixels in the image plane. Furthermore it enables to simulate any effects of the virtual objects in the path of the beam onto the target pixel.

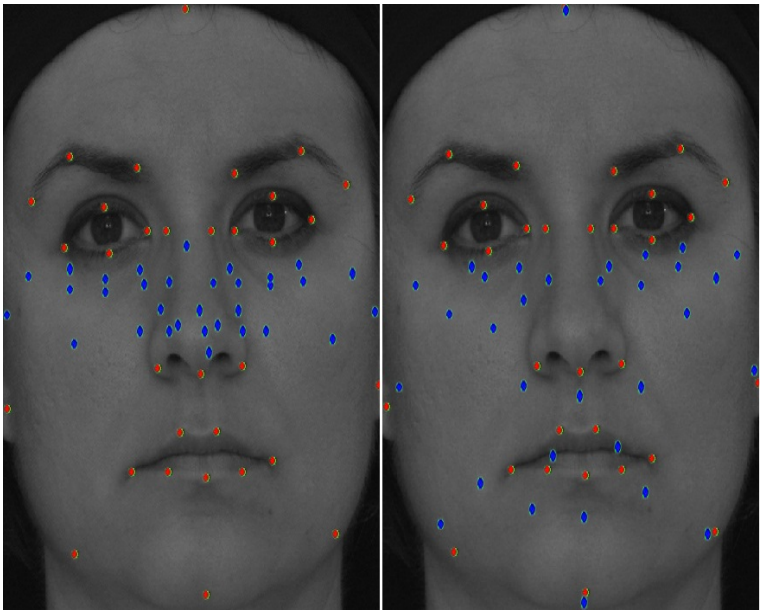

Figure 3.5: Horizontal and vertical scaling of the generic wireframe model

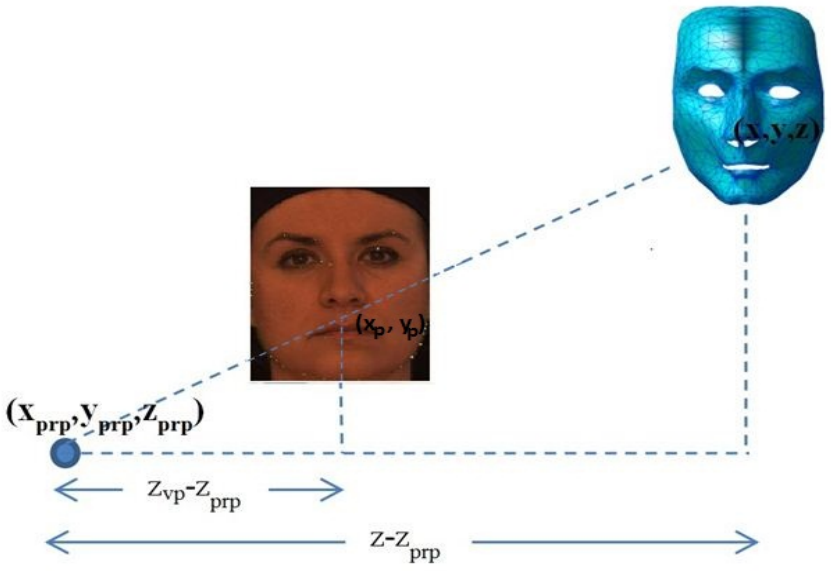

Figure 3.6: Perspective projection

The essence of our research lies in the exploitation of the ray tracing method in model fitting progression. We aim to mold the generic wireframe model in order to replicate the target image in 3 dimensional coordinate axes. Feature points marked prior to this stage are employed in this process to estimate the 3 dimensional coordinates of the corresponding landmark vertices.

Coordinates of the feature points in the target image x_p and y_p can be quantified exploiting perspective projection [Figure: 3.6].

$$x_p = x(\frac{z_{prp} - z_{vp}}{z_{prp} - z}) + x_{prp}(\frac{z_{vp} - z}{z_{prp} - z}) \tag{3.1}$$

$$y_p = y(\frac{z_{prp} - z_{vp}}{z_{prp} - z}) + y_{prp}(\frac{z_{vp} - z}{z_{prp} - z}) \tag{3.2}$$

Here z_{vp} stands for the z coordinate of the view (camera) plane and (x_{prp} , y_{prp}, z_{prp}) is the projection reference point. We can simplify the perspective projection calculations by choosing the projection reference point on the z axis. This transforms projection reference point coordinates x_{prp} and y_{prp} to zero. Hence it allows us to simplify [Equation: 3.1] and [Equation: 3.2] to obtain the perspective projection form as in [Equation: 3.3].

$$f_p(x, y, z) = (x_p, y_p) = [x(\frac{z_{prp} - z_{vp}}{z_{prp} - z}), y(\frac{z_{prp} - z_{vp}}{z_{prp} - z})] \tag{3.3}$$

Here our observation consists only of feature points in the image plane. We can estimate the 3 dimensional coordinates of these feature points by keeping the depth of the landmark vertices fixed and inverting the transformation given in [Equation: 3.3].

$$x = x_p(\frac{z_{prp} - z}{z_{prp} - z_{vp}}) \tag{3.4}$$

$$y = y_p(\frac{z_{prp} - z}{z_{prp} - z_{vp}}) \tag{3.5}$$

$$f_p^{-1}(x_p, y_p, z) = [x_p(\frac{z_{prp} - z}{z_{prp} - z_{vp}}), y_p(\frac{z_{prp} - z}{z_{prp} - z_{vp}}), z] \tag{3.6}$$

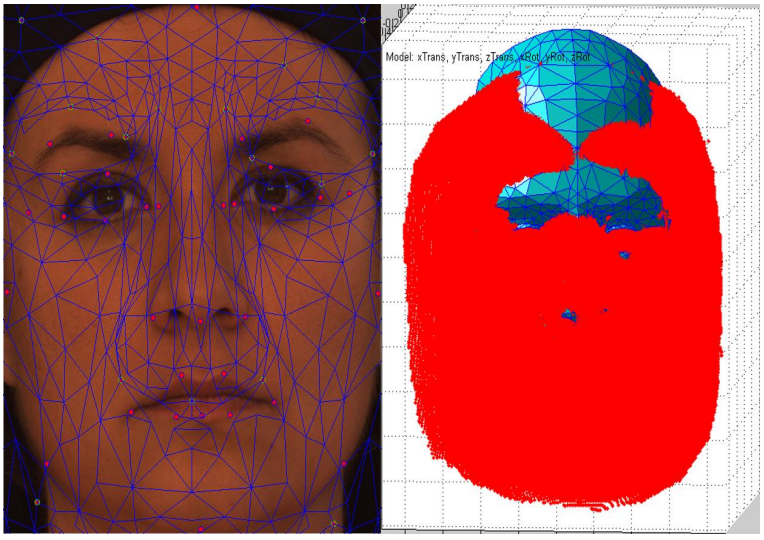

Figure 3.7: After alignment with ray tracing (a) The projection of the generic model onto image plane (b) The generic model and the data cloud

3.3.3 Estimation of Non-landmark Vertex Coordinates

As can be observed in [Figure: 3.7] back projection of feature points to align with landmark vertices is not sufficient to construct an accurate model of the subject's face. Non landmark vertices should also be transformed in accordance with the inverse projected feature points. In the generic face model we have 612 nodes and only 32 of them are denoted as landmark vertices. We update the coordinates of the non landmark vertices using a Nearest Neighbor Weighted Average (NNWA) algorithm.

Customized locations of the landmark vertices were attained using the ray tracing technique. At this point we can exploit the transformations of the landmark vertices to calculate a translation vector for each non landmark vertex. Translation vector for each landmark vertex is calculated simply by subtracting the initial position of the wireframe model with the customized positions of the reshaped model [Equation: 3.7].

$$\Delta v_i^l = v_{i,org}^l - v_{i,custom}^l \tag{3.7}$$

Figure 3.8: Test results on Caltech dataset

The contribution of the landmark vertex to the translation of a non landmark vertex is inversely proportional with the square of the distance in between. We calculated the Euclidean distance d_{ij} between each non landmark vertex and the landmark vertices in the original wireframe model. These distances are utilized in the NNWA algorithm to determine the effect of k-nearest landmark vertices to a non landmark vertex [Equation: 3.8] [Equation: 3.9]. Inverse of the squared distances of the landmark vertices are used as the weight in determining the translation of a non-landmark vertex.

$$T_j = \frac{\sum_{i=1}^{k} \frac{\Delta v_i^l}{d_{ij}^2}}{\sum_{i=1}^{k} \frac{1}{d_{ij}^2}} \tag{3.8}$$

$$v_{j,custom}^{nl} = v_{j,org}^{nl} + T_j \tag{3.9}$$

Our proposed technique was applied on both Caltech and Bosphorous datasets.

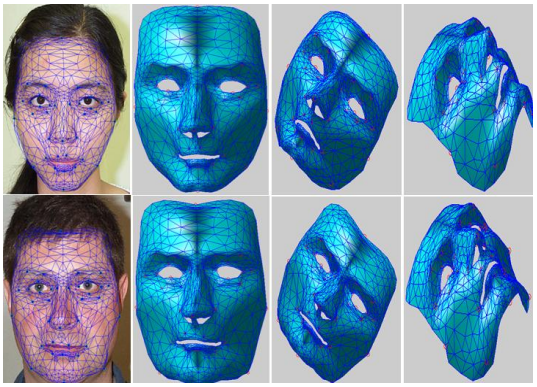

Figure 3.9: Test results of weighted nearest neighbor customization on sample images chosen from Caltech dataset

In the experiments conducted on the Caltech dataset [Figure: 3.8], we observed that the landmark locations of the model overlapped with the landmark locations of the target images. Furthermore as predicted the non landmark vertices of the model translated smoothly on the target image constructing a suitable model of the subject.

As illustrated in [Figure: 3.9], the constructed 3 dimensional model aids us in estimating the subject's face structure.

Caltech is a two dimensional database. Therefore the accuracy of the experiments cannot be verified using the information in the dataset. Hence we extended our experiments using the Bosphorus 3 dimensional database. An in-depth review of this extended research and the performance criteria are presented in detail in Chapter 4.

3.4 Customization through Procrustes Analysis

Procrustes Analysis provides scaling, translation and rotation parameters for the generic model to fit a data cloud. Detailed information in Procrustes Analysis is available in Section 2.3.1.

Perhaps the greatest disadvantage in Procrustes Analysis is its prerequisite to have a 3 dimensional dataset in order to construct a model of a target image. In

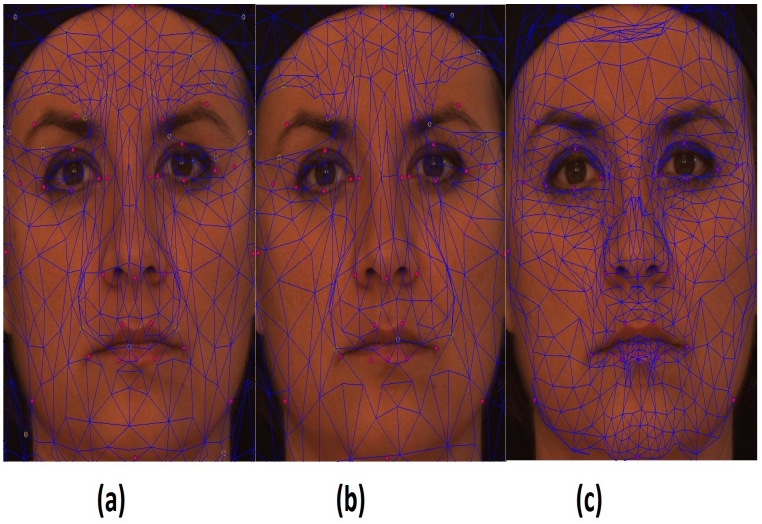

Figure 3.10: Perspective projections of aligned wireframe model, wireframe customized through Procrustes Analysis with the data cloud and weighted nearest neighbor customization

this study we aim to construct a realistic facial model of a subject using a single 2 dimensional image. We exploit the information extracted from the 3 dimensional image database only as ground truth.

The generic wireframe we employ is of neutral pose and illustrates a frontal profile. But in real life examples we often encounter varying facial expressions and orientations of the head. Application of Procrustes Analysis provides the generic model with rotation and scaling parameters that would better suit the target image.

[Figure 3.10] illustrates wireframe fitting with alignment only, with Procrustes Analysis and with weighted nearest neighbor customization, respectively. The performance levels of the respective methods are evident with the data cloud representation [Figure: 3.11].

In the model fitting stage performance of the Procrustes Analysis was measured against the proposed method. Procrustes Analysis requires same number of key traits in both data clouds. We apply Procrustes Analysis using the landmark vertices on the generic wireframe and the corresponding feature points in the data cloud. Once the transformation vector is obtained it is applied on non

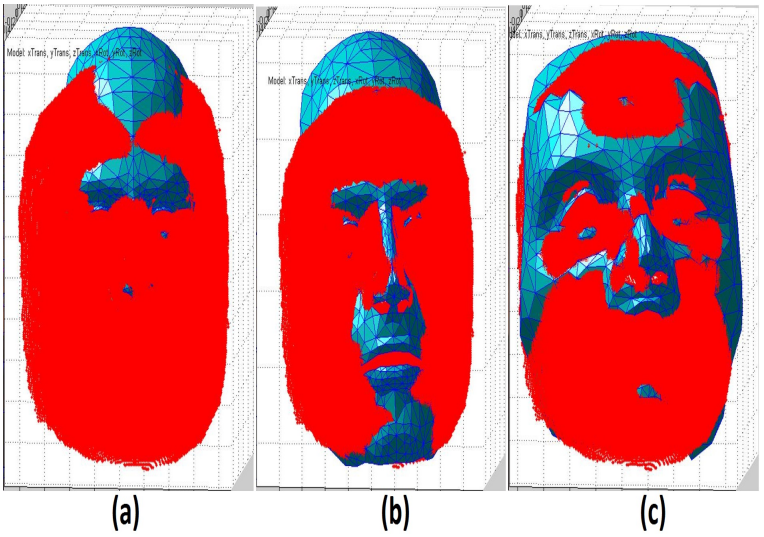

Figure 3.11: Comparison of aligned wireframe model, wireframe customized through Procrustes Analysis with the data cloud and weighted nearest neighbor customization

landmark vertices to attain a realistic 3 dimensional model of the target image. We evaluated relative error values for both Procrustes Analysis and our method. Proposed technique produced substantially better results in comparison with the Procrustes Analysis. Experimental results are described in detail in Section 4.3.

3.5 Customization through Active Shape Model

Active Shape Model is one of the benchmarks in face modeling. ASM relies on selection training data with adequate number of identified feature points in constructing an accurate face model. Given a decent initial approximation, ASM is capable of producing an accurate fit to the target image employing an instance of the model. Original ASM was first introduced by Cootes *et al.* [44]. An in-depth review of the ASM is presented in Section 2.4.2.1.

In our study we implemented ASM for the comparative evaluation of the proposed face modeling technique. The original ASM was designed for 2 dimensional images. However in this study we deal with the customization of a 3 dimensional generic face model. Therefore we implemented an extended version of ASM. A

comprehensive explanation of the implementation is presented in the rest of this chapter.

The 3 dimensional data cloud and the generic wireframe model we use in our study are of two different scales. Initially we normalize both 3 dimensional data cloud and the generic wireframe model. The scaling is performed as to fit the largest axis of the model and the data cloud to the range [-0.5 , 0.5].

Once the model and the data cloud is translated to a common coordinate system PCA is applied to acquire the eigenvectors. The eigenvectors are exploited in determining the modes of variations to be used in our experiments. In our experiments we employed the first 9 eigenvectors that account for 82% of the variation. We also determine the mean model using the locations of the landmark vertices. From this stage on the obtained mean model will serve as the generic wireframe model. Note that the transformation is applied only on the landmark vertices.

After the preprocessing stage we explore an iterative approach similar to the original ASM in acquiring the instance of the generic wireframe model that best fits the target image. The iterative search process is explained in-detail in the subsequent paragraphs.

The next stage is to estimate the Euclidean transformation for the wireframe model for a best fit with the image feature points. Transformation vector $\mathbf{b}$ comprise of translation (t) and rotation (α) parameters along the *x, y and z* axes.

$$\mathbf{b} = [t_x, t_y, t_z, \alpha_x, \alpha_y, \alpha_z] \tag{3.10}$$

Translation in 3 dimensional space can be represented using a 3×1 vector. In the same manner rotation can be represented with a 3×3 matrix. In order to combine the transformations on all 6 degrees of freedom we use homogeneous coordinates and 4×4 transformation matrices.

Once the transformation is applied on the landmark vertices we find their locations in the image plane using perspective projection P. Error $\mathbf{E}$ is defined as the difference between projected landmark vertices and the image feature points $\mathbf{F}$. It is therefore an $n \times 2$ matrix where n is the number of features.

$$\mathbf{E} = P(\mathbf{T}\mathbf{R_x}\mathbf{R_y}\mathbf{R_z}\mathbf{X}) - \mathbf{F} \tag{3.11}$$

We define residual $\mathbf{r}$ as the distance between projected points and the feature points. In other words it is sum squared error $\mathbf{E}$ for each feature point. Thus $\mathbf{r}$ is an $n \times 1$ vector. We now find the Jacobian of residual $\mathbf{r}$ with respect to transformation $\mathbf{b}$.

$$\mathbf{J} = \mathbf{J_r}(\mathbf{b}) \tag{3.12}$$

The Jacobian is an $n \times 6$ matrix which comprises of the gradient of residuals on each degree of freedom. Once the Jacobian is generated the direction of the steepest descent is determined using the Gauss-Newton method. Gauss-Newton method is employed for solving non-linear least square problems.

$$\Delta\mathbf{b} = (\mathbf{J^T}\mathbf{J})^{-1}\mathbf{J^T}\mathbf{r} \tag{3.13}$$

The next step is to refine the step size estimated by the Gauss-Newton method. This is performed iteratively reducing the step size by half until a step size that reduces the error is discovered. This process converges to a minimum depending on the starting point. At this point we can update the transformation parameters.

$$\mathbf{b} \leftarrow \mathbf{b} + \boldsymbol{\Delta}\mathbf{b} \tag{3.14}$$

Since we have an estimation of the Euclidean transformation for the wireframe model we can now customize the wireframe model using the 9 variation modes in other words, the eigenvectors that we obtained from our training set.

Using [Equation: 2.10] we represent the customized model $\mathbf{x}$ in terms of eigenvector coefficients β as;

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{A}\beta \tag{3.15}$$

The coordinate vector $\mathbf{x}$ is reshaped into an $n \times 3$ matrix $\mathbf{X}$. Using the estimation of Euclidean transformation we obtained in the previous step we recalculate error $\mathbf{E}$ as in [Equation: 3.11]. Residual $\mathbf{r}$ is calculated once again as the sum squared error for each point. At this stage we calculate the Jacobian by calculating the gradient of $\mathbf{r}$ on each of the 9 variation modes.

$$\mathbf{J} = \mathbf{J}_{\mathbf{r}}(\beta) \tag{3.16}$$

The Jacobian with respect to β is an $n \times 9$ matrix. We again apply the Gauss-Newton method and carry out an iterative search to determine the best eigenvector coefficients.

$$\mathbf{\Delta}\beta = (\mathbf{J}^{\mathbf{T}}\mathbf{J})^{-\mathbf{1}}\mathbf{J}^{\mathbf{T}}\mathbf{r} \tag{3.17}$$

$$\beta \leftarrow \beta + \mathbf{\Delta}\beta \tag{3.18}$$

This entire process is repeated until error converges or decreases below a predefined threshold value. This iterative process repeatedly optimizes the transformation parameters and the eigenvector coefficients. Once the iterations are complete the coordinate matrix $\mathbf{X}$ becomes the customized wireframe model. [Figure: 3.12]

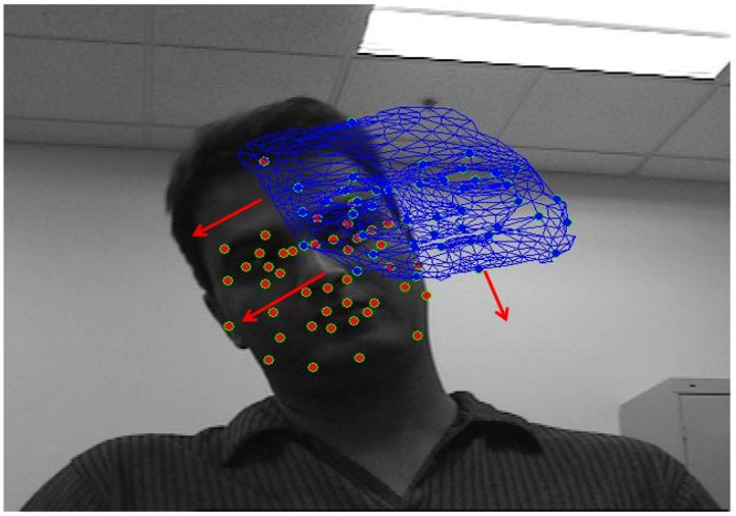

Figure 3.12: Iterative model fitting progression in ASM

demonstrates the iterative model fitting process in our Active Shape Model implementation.

Chapter 4

Comparative Study of Face Modeling Techniques

Several experiments are conducted in order to evaluate the performance of the proposed algorithm in comparison with Procrustes Analysis and ASM. In this stage of our research we exploited the Bosphorous 3D face dataset in order to evaluate the error rate of the algorithms. We carried out the experiments on neutral poses of 104 different subjects.

4.1 Evaluating the Performance of Customization

In real life we often encounter faces with different dimensions. Therefore to bring all the error measurements to a standard base we exploited *relative error* in our experiments. Relative error is one of the most popular approaches in data cloud registration. To simplify the calculations we benefited from the bounding box in quantifying the relative error. For each subject the mean error is divided by the diagonal length of the bounding box belonging to that subject.

We also developed a model coloring strategy to illustrate the error variation on the surface of the model. In our first experiment on visualizing the error, each of the 612 landmark vertices of the model is compared with the corresponding data cloud feature points. Each of the data cloud feature points is assigned red or green color based on the error. Positive and negative errors are displayed in red and green respectively [Figure: 4.1].

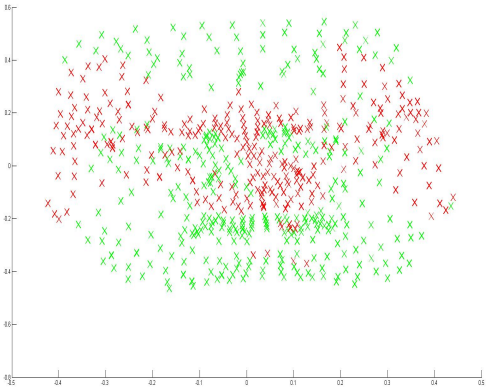

Figure 4.1: Binary coloring strategy to display the error variation

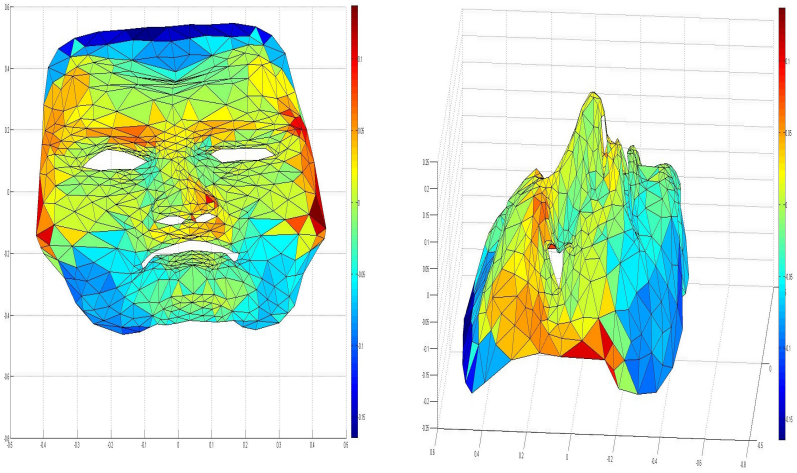

Figure 4.2: Illustration of error variation on a customized model

We observed that this method is not sufficient to observe the error variations in the model. Consequently we modified the *trisurf* command in *Matlab* to facilitate better illustration of the error rate variation through the model. The negative errors are illustrated in shades of blue and the positive errors are displayed in shades of red. The perfect fit of the model with the data cloud is represented in green. The frontal and lateral profiles of a model are illustrated in [Figure: 4.2].

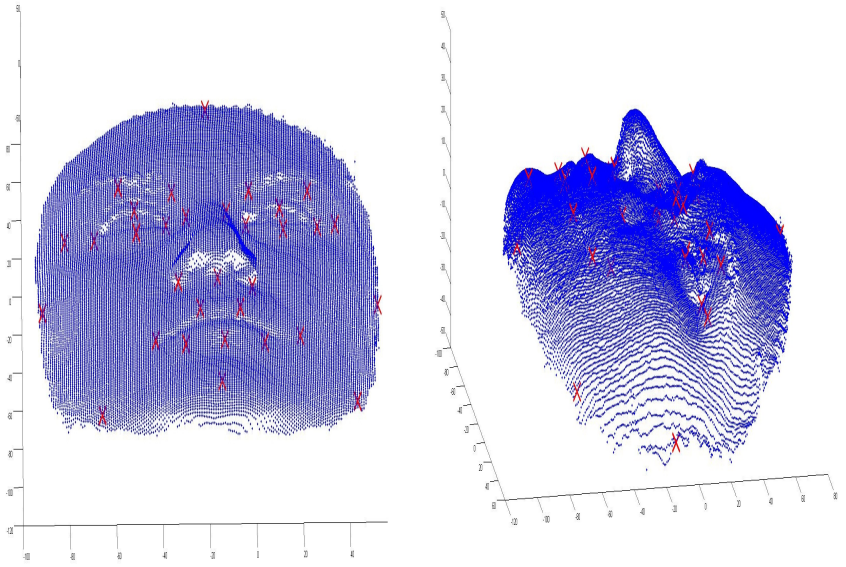

Figure 4.3: Illustration of feature points on the data cloud

4.2 Nearest Neighbor Weighted Average Customization Results

4.2.1 Identifying the Landmark Vertices

Our first experiment was to evaluate the variation of the performance with respect to the number of features used in customization. This was performed to determine the ideal number of landmarks required to define a face accurately. It is strenuous task to mark the feature points on the target images. It is even more exhausting to mark the corresponding vertices on the data cloud. Frontal and lateral views of the feature vertices on a sample data cloud is illustrated in [Figure: 4.3].

The goal of our experiment was to decrease the number of feature points utilized in defining a face. We start our experiment with 42 key traits. These traits were selected using prior knowledge of human anatomy. These key traits are marked on the target image and their corresponding data cloud manually.

We perform this experiment gradually decreasing the number of key traits from 42 to 10. We exploited 15 randomly selected subjects from the Bosphorous dataset

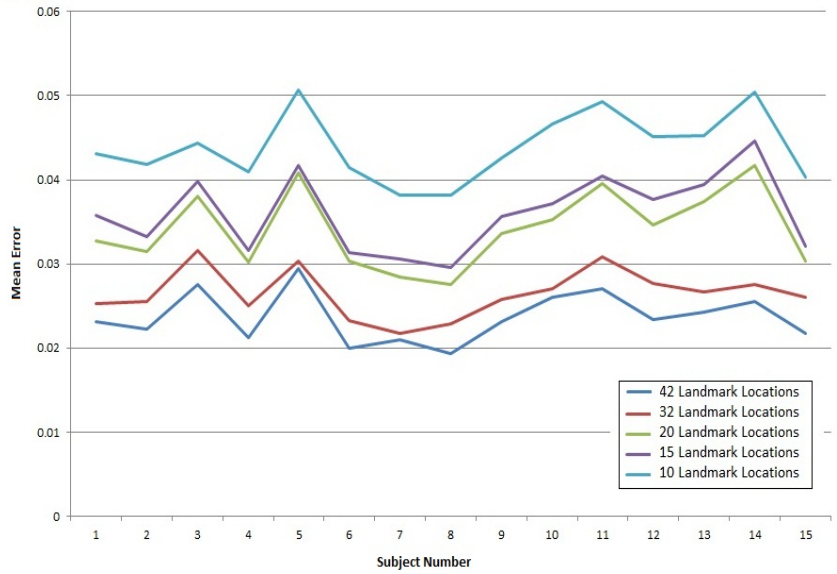

Figure 4.4: Relative error comparison for the proposed method with respect to varying number of key traits

in this experiment. We manually marked feature points on the images and corresponding data clouds for these 15 subjects. We observed the variation of the mean error with respect to the number of key traits allocated in defining the facial images [Figure: 4.4].

As expected we observed that the mean error of the model increases as the number of key traits decreases. There is a tradeoff between the effort required in locating key traits and the accuracy. Taking this fact into consideration we decide to employ 32 landmark traits in the customization process. These key traits are illustrated in [Figure: 4.5]. All of our subsequent experiments are conducted using these key traits.

4.2.2 Choosing the Number of Neighbors

One of the other experiments we performed was to evaluate the effect of the number of neighbors to be used in the Nearest Neighbor Weighted Average (NNWA) customization stage. The customization of the non-landmark vertices depends highly on the number of landmarks utilized as the nearest neighbors. When the

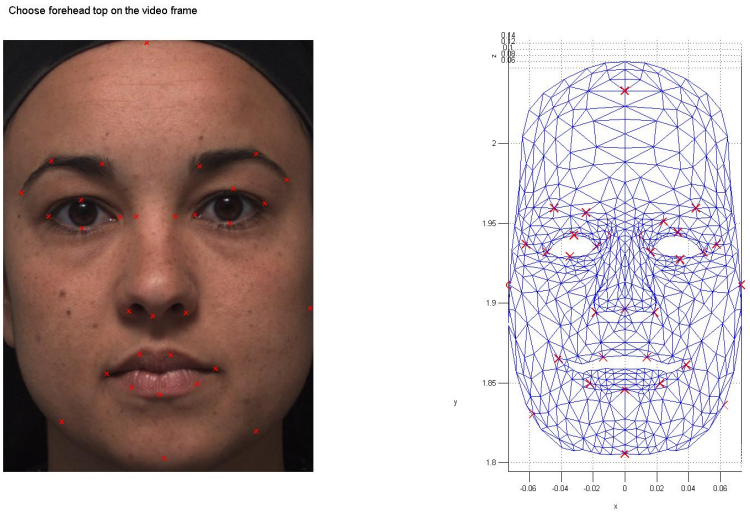

Figure 4.5: Illustration of selected feature points on face image and landmark vertices on HIGEM

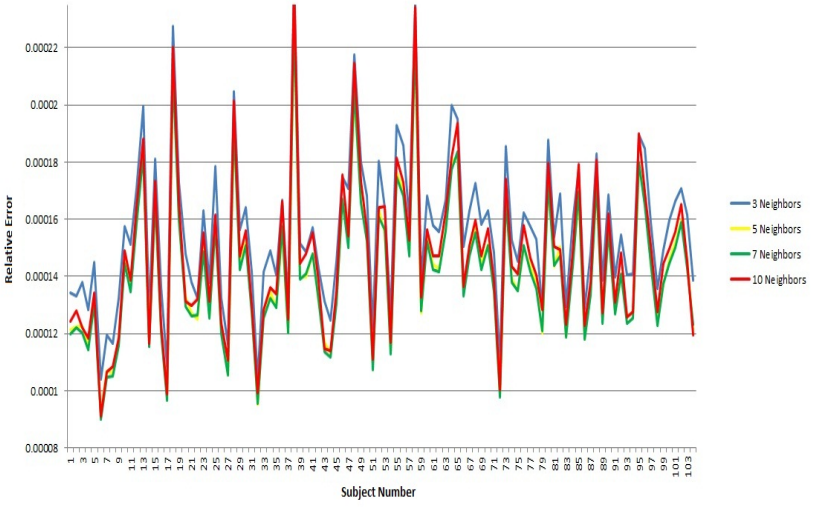

Figure 4.6: Relative error comparison for the proposed method with respect to the varying number of nearest neighbors

number of neighbors increases we encounter an overfitting problem. On the other hand selecting too few neighbors would impede the smooth distribution of landmark vertex translations to the non landmark vertices. The neighbor size should be selected in a manner that would delineate facial muscle groups that can act independently. Relative error magnitudes we observed for the Bosporous dataset with respect to the varying number of neighbors are depicted in [Figure: 4.6].

We observed that the variation rate of relative error with respect to the number

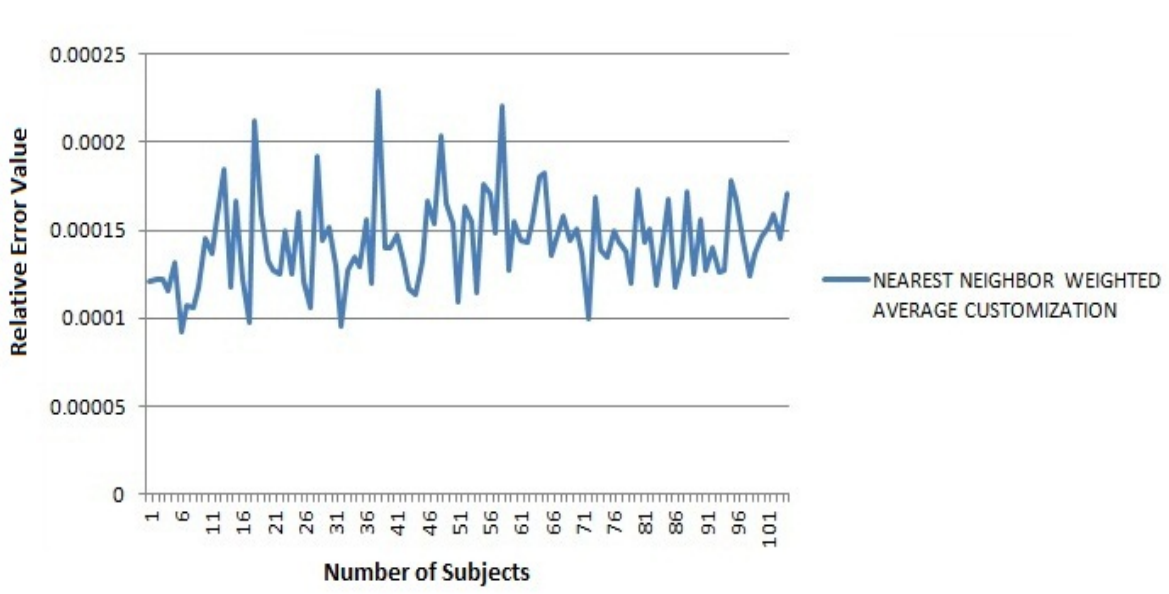

Figure 4.7: Magnitudes of relative error using 5 neighbors for all subjects in the Bosphorous dataset

of nearest neighbors is not very oscillatory. This is partially due to the astute selection of landmarks. Landmarks are selected in a fashion that they would delineate the independent facial muscle regions. There is only a slight increase in the relative error as the number of nearest neighbors decreases. Due to overfitting we occasionally observed a slight increment in relative error when the number of neighbors is increased. From our experiments we deduced that 5 to 7 neighbors can be considered sufficient in NNWA customization.

Computation time increases linearly with the number of neighbors exploited in NNWA customization. Therefore in the customization of the non landmark vertices we have employed 5 neighbors. Using the NNWA customization described in Chapter 3 we acquired the results depicted in [Figure: 4.7]. These error measures are normalized exploiting relative error. The results demonstrate substantially low error magnitudes.

[Figure: 4.8] illustrates the results of the proposed method with 5 nearest neighbor customization on a randomly selected subject together with the data cloud. The

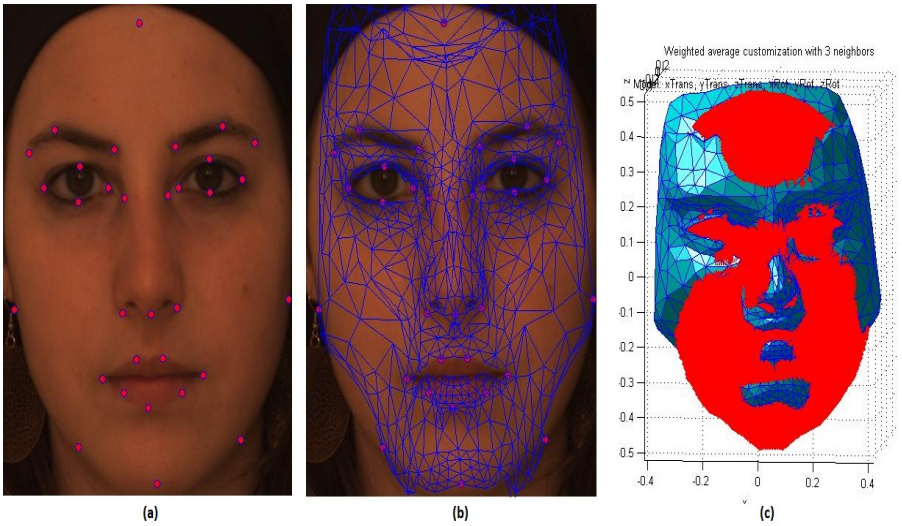

Figure 4.8: Feature points on sample subject, generic wireframe model overlayed on the image, acquired 3 dimensional model and the data cloud (Subject Number 15)

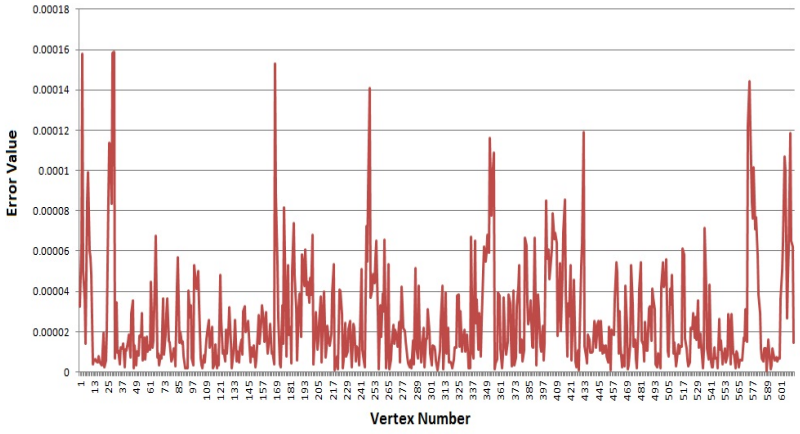

Figure 4.9: Error magnitudes for model vertices when nearest neighbor weighted average customization is applied (Subject Number 15)

graph that demonstrates the error magnitudes for each of the 612 vertices is presented in [Figure: 4.9].

We also generated a histogram to observe the variation of error [Figure: 4.10]. As can be observed the majority of the error values lie in close proximity to the mean error. In other words the variation of the error is quite small.

[Figure: 4.11] demonstrates the distribution of the error on the acquired facial model surface. As expected the regions adjacent to landmark vertices have higher

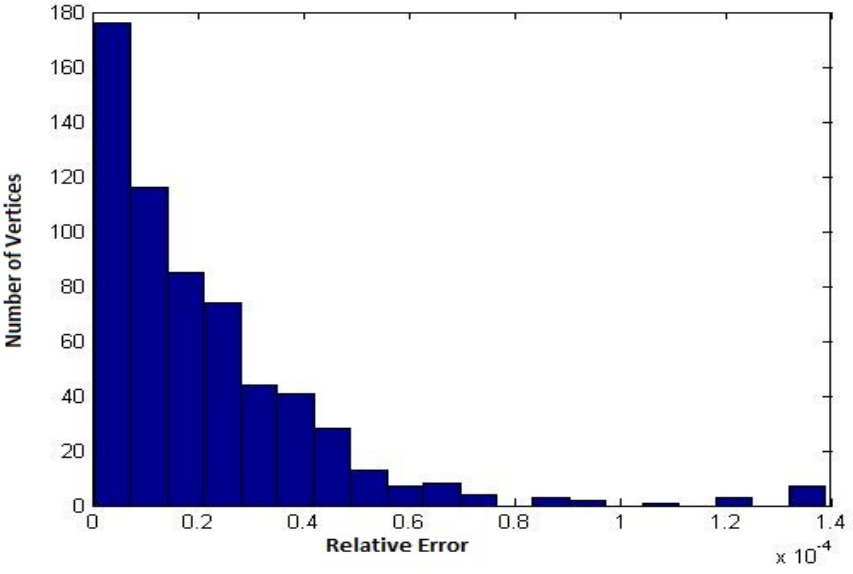

Figure 4.10: Relative error histogram for a sample subject (Subject Number 15)

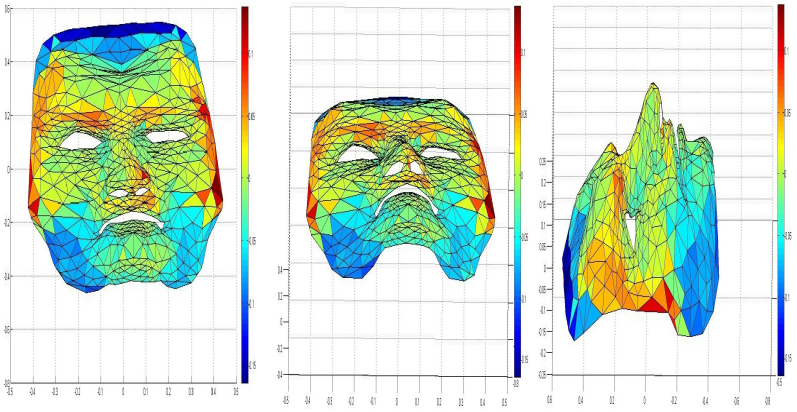

Figure 4.11: Varying view points for the obtained 3 dimensional model for a sample subject (subject number 15)

accuracy rates when compared to the distant regions.

4.3 Procrustes Analysis Results

Procrustes Analysis is employed generally as an alignment technique. Transformation vector produced by Procrustes Analysis can be employed in the 3 dimensional modeling process. We also employed Procrustes Analysis in the relative

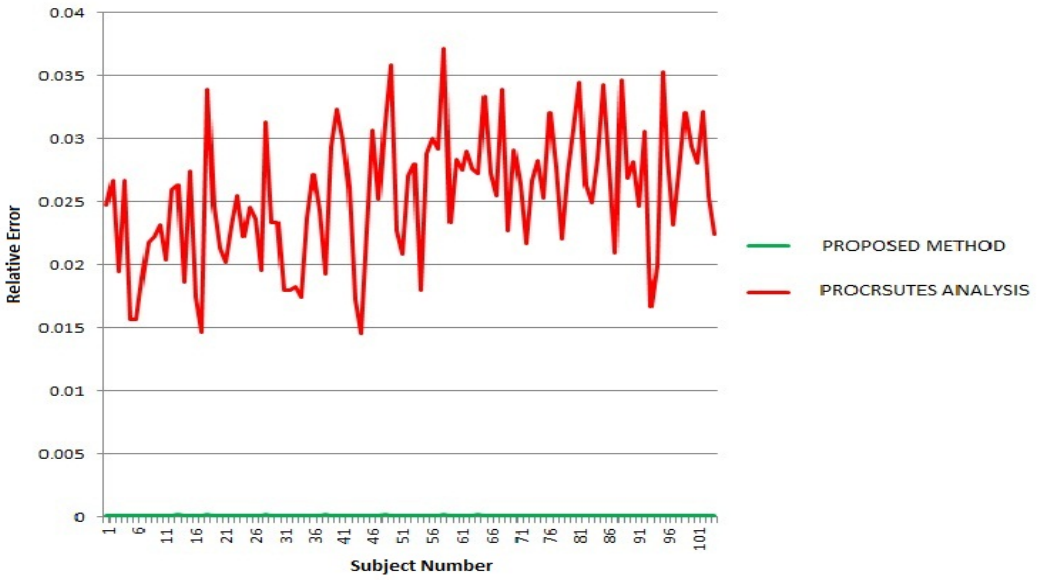

Figure 4.12: Relative error comparison for the proposed method against Procrustes Analysis

error evaluation process.

We exploit Procrustes Analysis only to evaluate the performance of our algorihm. Procrustes Analysis cannot be utilized in the modeling process since it requires a 3 dimensional data cloud. [Figure 4.12] illustrates the comparison of the relative error value variations for the subjects in the datasets for the two methods; the proposed technique and Procrustes Analysis. We observed substantially better results for all the subjects in the dataset when modeling with the proposed method.

4.4 Active Shape Model Results

In the face modeling field there are two prominent methods; parameterized and statistical face modeling. Methods such as mass-spring-damping fall under parameterized techniques. They require extensive knowledge about the anatomy of human face. Due to this reason they have become less popular today. Statistical modeling relies on a training dataset in constructing a face model for a given

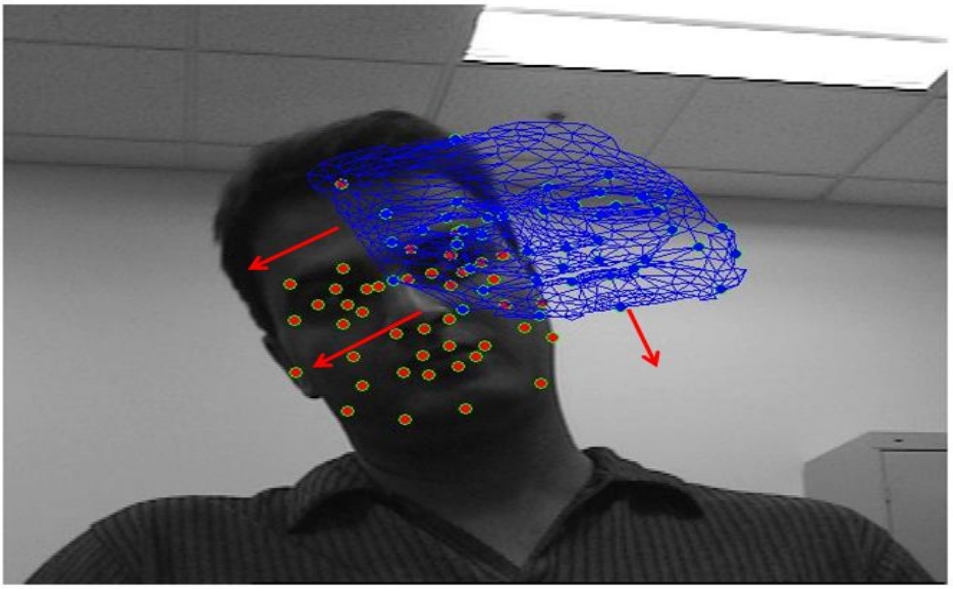

Figure 4.13: ASM - Iterative model fitting process

subject. Despite the fact that the training process is time consuming, statistical face modeling is capable of producing reliable results.

We compared the performance of our proposed technique with Active Shape Model (ASM) and Procrustes Analysis. ASM is originally implemented for 2 dimensional images. In our research we extended ASM to be employed with a 2 dimensional image and a 3 dimensional generic face model. More detailed report of this implementation is presented in Section 3.5.

We implemented ASM to facilitate comparison with the results of our algorithm. Iterative fitting process of ASM is shown in the figure [Figure 4.13]. Initially we compared the results obtained by ASM and Procrustes Analysis. Procrustes Analysis being a model alignment and registration technique illustrated abysmal results in comparison with the ASM.

4.4.1 Comparison of Face Modeling Methods

In the performance evaluation stage of our proposed technique, we utilize the results obtained for the aforementioned three algorithms; ASM, Procrustes Analysis

Figure 4.14: Relative error comparison for NNWA customization and ASM

and finally NNWA customization. Employing NNWA customization we achieved substantially lower error results consistently for all the subjects in the dataset. Procrustes Analysis is generally employed as a data registration and alignment technique. This fact justifies the higher relative error values produced by Procrustes Analysis.

ASM demonstrated significantly better error results in comparison with our proposed technique [Figure 4.14]. This fact alone does not make ASM better than our proposed technique. There are a lot of constraints in customizing a model with ASM. ASM does not guarantee the convergence of the model to the global optimum. It is quite possible for the model to converge in to a local optimum. Furthermore the demonstrated relative error values for ASM are quantified employing only the 32 key trait locations. Whereas NNWA customization provides the customization for the non-landmark vertices of the model as well. Finally when comparing the relative error magnitudes for ASM and proposed technique it is visible that the magnitudes of the relative error values are very low for both of these methods.

Furthermore we compared the complexities of the three algorithms. Procrustes Analysis provides the transformation relying on the distance between the generic model and the target image of the subject. It operates in the complexity of $\theta(n^2)$. ASM is an iterative method which involves calculation of Jacobian matrices and

solution through Gauss-Newton approximation. For that reason time complexity of ASM is substantial. The highlight of our proposed technique is the use of nearest neighbor algorithm. This is performed offline and only once for a generic wireframe model. Ray tracing, positioning of the landmark vertices and weighted average customization of the non-landmark vertices are performed in linear time making the complexity of our algorithm $\theta(n)$ where n is the number of vertices in the generic wireframe model.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

Modeling has history that spans over 2500 years. Earliest idea of modeling was presented by Plato in his *theory of forms*. He articulates that each entity has its own form and this attribute abets in differentiating one object from another. Yet for over centuries face modeling is considered as a disheartening task. This is mainly due to the sophisticated structure of the face and numerous complications that arise in modeling at large as a reason of the non-rigid form that it possesses. This challenging behavior of the task has intrigued many researchers to experiment in this field. Researchers' immense interest in this field is not only due to the challenging nature of the problem but also due to the enriched application areas that it promises.

Parameterized and statistical methods are the two prominent approaches in face modeling. Parameterized Face modeling is greatly dependent on the prior knowledge of facial anatomy including the muscle structure. Construction of an accurate parameterized model is a daunting task due to the fact that a simple movement of a muscle can generate an entirely different yet significant facial expression. There are diverse variations of modeling techniques that utilize the concepts of parameterized face modeling. Among them mass-spring-damping method is the most famous approach.

In statistical modeling as its name suggests, statistical techniques are exploited in representing an object. The most significant advancement was the introduction of Active Shape Model (ASM) that utilizes shape parameters to identify an object. This is an iterative method which involves calculation of Jacobian matrices and solution through Gauss-Newton approximation. For that reason the time complexity of this method is substantial.

Apart from these two methods there have been studies in constructing a 3 dimensional model employing multiple images captured from multiple cameras in precise orientations. The main shortcoming of this approach is the inability to employ it in daily life.

In this report we have presented a novel customization based method for face modeling. The proposed semi-automatic method is capable of generating realistic 3 dimensional face models using 2 dimensional information. Perspective projection and ray tracing techniques are exploited in achieving this task. Carefully identified 32 key traits were utilized in the customization process. These traits, which are choosen conforming with human facial anatomy are marked manually on the image and the data cloud.

The input to the customization process is a face image on which 32 feature points are marked manually. Customization is initially applied on these 32 landmark vertices, then extended to the entire wireframe to generate a face model for a given subject. This process is conducted employing a weighted nearest neighbor approach. Therefore the translation of a non-landmark vertex would be impacted predominantly by the landmark vertices nearest to the point.

Our approach is substantially different from the existing parameterized and statistical modeling strategies. In contrast with the parameterized modeling techniques, this method does not require extensive facial anatomical knowledge. Furthermore it does not require a training dataset as in the statistical modeling

techniques. Unlike the multiple camera approach that requires a complex apparatus our method relies on a single 2 dimensional image to create an acceptable replica of a subject.

We carried out a comparative analysis of our algorithm with Procrustes Analysis and Active Shape Model. In this process we utilize 3 dimensional data clouds embodied in Bosphorous 3D datasets as ground truth. Starting with a marked image of a subject we applied all three algorithms to obtain customized models. Since the scale of the models may vary we exploit a relative error measurement in the evaluation process.

Our algorithm operates in substantially lower time complexity when compared with many of the other algorithms available in this front. The highlight of our proposed technique is the use of the nearest neighbor algorithm. This is performed offline and only once for a generic wireframe model. Ray tracing, positioning of the landmark vertices and weighted average customization of non-landmark vertices are performed in linear time. Overall the complexity of our algorithm is $\theta(n)$ where n is the number of vertices in the generic wireframe model.

Our future research plans focus on automating the identification of feature points on face images. Numerous feature detection algorithms such as Speeded Up Robust Feature (SURF) may be considered here.

5.2 Future Work

Our method relies on a very careful selection of feature points on the input image. This is the only stage where direct human interaction is required. Therefore it is the main vulnerability of the proposed method.

Currently we are conducting research to make this a fully automatic face modeling system. Numerous feature detection techniques can be applied for identification of facial feature points. One of our strategies in achieving this task was the exploitation of Scale Invariant Feature Transform (SIFT) features in the automatic

feature point extraction. However we obtained mediocre results by applying this technique. Currently other methods of automatic feature extraction such as the SURF algorithm are being researched.

In the current stage of our research we are capable of tracking a face in a video and update the model accordingly. However we have not performed much experimentation in this front. The greatest obstacle here is the difficulty of acquiring video databases that include 3 dimensional ground truth data. As a future work our algorithm could be applied on a 3 dimensional video dataset. This would enable us to observe how the performance improves when varying poses of the same subject are utilized.

References

1. N. D'Apuzzo. "Motion capture from multi image video sequences." in *Proc. of the XVIIIth Congress of the International Society of Biomechanics,* July. 2001.

2. E. Okada. "Three-Dimensional Facial Simulations and Measurements: Changes of Facial Contour and Units Associated with Facial Expression." *The Journal of Craniofacial Surgery,* vol.12, pp.167-174, 2001.

3. U. Park and A.K. Jain. "3D Model-Based Face Recognition in Video." in *Proc of the 2nd International Conf. on Biometrics,* 2007.

4. Z. Liu *et al.* "Rapid Modeling of Animated Faces from Video." in *Proc. of the 3rd International Conf. on Visual Computing (Visual2000),* pp. 58-67, 2000.

5. N. D'Apuzzo *et al.* "Markerless full body shape and motion capture from video sequences." *International Archives of Photogrammetry and Remote Sensing,* vol.31, pp.256-261, 2002.

6. J. Huang. "Component-based face recognition with 3D morphable models." M.A. thesis, Massachusetts Institute of Technology, USA, 2002.

7. F.A.S. Banda *et al.* "Automatic Generation of Facial DEMs." *International Archives of Photogrammetry and Remote Sensing,* vol.29, pp.893-896, 1992.

8. N. DApuzzo. "Automated Photogrammetric Measurement of Human Faces." *International Archives of Photogrammetry and Remote Sensing,* vol.32, pp.402-407, 1998.

9. R.M. Koch *et al.* "Simulating Facial Surgery Using Finite Element Models." in *Proc of the SIGGRAPH96,* 1996.

10. N. Motegi *et al.* "A Facial Growth Analysis Based on FEM Employing Three Dimensional Surface Measurement by a Rapid Laser Device." in *Proc of the Okajimas Folia Anatomica Japonica,* vol.6, pp.323-328, 1996.

11. V. Blanz and T. Vetter. "A Morphable Model for the Synthesis of 3D Faces." in *Proc of the SIGGRAPH'99,* pp.187-194, 1999.

12. W.S. Lee and N.T. Magnenat. "Fast Head Modeling for Animation." *Image and Vision Computing Journal,* vol.4, pp.355-364, 2000.

13. S.R. Marschner *et al.* Modleing and Rendering for Realistic Facial Animation." in *Proc. of the 11th Eurographics Workshop on Rendering,* 2000.

14. F. Pighin *et al.* "Synthesizing Realistic Facial Expressions from Photographs." in *Proc of the SIGGRAPH'98,* pp.75-84, 1998.

15. R.Sitnik and M.Kujawinska. "Opto-Numerical Methods of Data Acquisition for Computer Graphics and Animation Systems." in *Proc of Three-Dimensional Image Capture and Applications III,* pp.36-43, 2000.

16. S.V. Duhn *et al.* "Three-view surveillance video based face modeling for recognition." in *Proc of the Biometrics Symposium,* 2007.

17. S. Minaku *et al.* "Three-Dimensional Analysis of Lip Movement by 3-D Auto Tracking System." *International Archives of Photogrammetry and Remote Sensing,* vol.30, 1995.

18. Y. Shan *et al.* "Model-Based Bundle Adjustment with Application to Face Modeling." in *Proc. of the 8th International Conf. on Computer Vision (ICCV01),* vol.2, pp.624-651, 2001.

19. P. Fua. "Regularized Bundle-Adjustment to Model Heads from Image Sequences without Calibration Data." *International Journal of Computer Vision,* vol.2, pp.153-171, 2000.

20. D. DeCarlo *et al.* "An Anthropometric Face Model Using Variational Techniques." in *Proc of the SIGGGRAPH'98,* pp.67-74, 1998.

21. A.Borghese and S. Ferrari. "A Portable Modular System for Automatic Acquisition of 3-D Objects." in *IEEE Trans. on Instrumentation and Measurement,* vol.5, pp.1128-1136, 2000.

22. G.Drettakis and R.Scopigno. "Animating lips-sync speech faces with compact key shapes." *Eurographics,* vol.27, No3, 2008.

23. Y.X.Hu *et al.* "Camera and Microphone Array for 3D Audiovisual Face Data Collection." in *The 33rd International Conference on Acoustics, Speech, and Signal Processing (ICASSP2008)*, 2008.

24. T.F.Cootes and P.Kittipanyangam. "Comparing variations on the active appearance model algorithm." in *BMVC2002*, 2002.

25. B. Fleming and D. Dobbs. "Animating facial features and expressions." *Charles River Media.* 1999.

26. K.Kayler. "A head model with anatomical structure for facial modeling and animation." Phd.Thesis, University of Saarlandes, 2003.

27. A. Savran *et.al.* "Bosphorus Database for 3D Face Analysis." in *The First COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008),* 2008.

28. P.H. Schonemann. "A generalized solution of the orthogonal Procrustes problem." *Psychometrika,* vol.31, pp.1-10, 1966.

29. P.J. Besl and H.D. McKay. "A method for registration of 3-D shapes." in *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1992.

30. Y. Chen and G. Medioni. "Object Modeling by Registration of Multiple Range Images." in *Proc. of the 1992 IEEE Intl. Conf. on Robotics and Automation,* pp.2724-2729, 1992.

31. S.Y. Du *et al.* "An extension of the ICP algorithm considering scale factor." in *Proc 14th IEEE International. Conf. on Image Processing (ICIP),* pp.193196, 2007.

32. J. Gower. "Generalized procrustes analysis." *Psychometrika,* vol.40, pp.1-10, 1966.

33. J.Gower and G.B. Dijksterhuis. "Procrustes Problems." *Oxford University Press,* 2004.

34. T.Masuda *et al.* "Registration and Integration of Multiple Range Images for 3-D Model Construction." in *Proc. CVPR,* 1996.

35. F.I. Parke. "Parameterized Models for Facial Animation." in *IEEE Computer Graphics and Applications,* vol.2, pp.61-68, 1982.

36. F.I. Parke and K. Waters. "Computer Facial Animation 2nd Edition." ISBN.978-1-56881-448-3, 2008.

37. S.M. Platt and N. Badler. "Animating facial expression." *Computer Graphics,* vol.15(3), pp.245-252, 1981.

38. Y.Zhang *et al.* "Real-time physically-based facial expression animation using mass-spring system." in *Proc. Computer Graphics International 2001,* pp. 347-350, 2001.

39. D. Terzopoulos and K. Waters. "Physically-based facial modeling, analysis and Animation." *Journal of Visualization and Computer Animation,* vol.1 pp.73-80, 1990.

40. K.S. Benli *et al.* "Semi-Automatic Adaptation of High-Polygon Wireframe Face Models Through Inverse Perspective Projection." in *Proc. of International Symposium on Computer and Information Sciences (ISCIS 2011),* 2011.

41. T.F. Cootes and C.J. Taylor. "Statistical Models of Appearance for Computer Vision." 2004.

42. K.L. Low and A. Lastra. "Reliable and Rapidly-Converging ICP Algorithm Using Multi resolution Smoothing." in *Proc of the Fourth International. Conf. on 3-D Digital Imaging and Modeling (3DIM 2003),* 2003.

43. K.Pearson. "On lines and planes of closest fit to systems of points in space." *Philosophical Magazine,* vol.2, no.11, pp.559-572, 1901.

44. T.F.Cootes *et al.* "Active Shape Models - Their Training and Application." *Computer Vision and Image Understanding,* vol.61, No.1, pp.8-59, 1995.

45. T.F. Cootes *et al.* "Active Appearance Models." in *Proc. European Conference on Computer Vision,* Vol.2, pp.484-498, 1998.

46. S. Krindis and I. Pitas. "Statistical analysis of human facial expressions." *Journal of Information Hiding Multimedia Signal Processing,* vol.1, no.3, pp.241260, 2010.

47. J.Ahlberg. “CANDIDE-3 - an updated parameterized face.” *Report No. LiTH-ISY-R-2326, Dept. of Electrical Engineering, Linkping University, Sweden,* 2001.

References

1. N. D'Apuzzo. "Motion capture from multi image video sequences." in *Proc. of the XVIIIth Congress of the International Society of Biomechanics,* July. 2001.

2. E. Okada. "Three-Dimensional Facial Simulations and Measurements: Changes of Facial Contour and Units Associated with Facial Expression." *The Journal of Craniofacial Surgery,* vol.12, pp.167-174, 2001.

3. U. Park and A.K. Jain. "3D Model-Based Face Recognition in Video." in *Proc of the 2nd International Conf. on Biometrics,* 2007.

4. Z. Liu *et al.* "Rapid Modeling of Animated Faces from Video." in *Proc. of the 3rd International Conf. on Visual Computing (Visual2000),* pp. 58-67, 2000.

5. N. D'Apuzzo *et al.* "Markerless full body shape and motion capture from video sequences." *International Archives of Photogrammetry and Remote Sensing,* vol.31, pp.256-261, 2002.

6. J. Huang. "Component-based face recognition with 3D morphable models." M.A. thesis, Massachusetts Institute of Technology, USA, 2002.

7. F.A.S. Banda *et al.* "Automatic Generation of Facial DEMs." *International Archives of Photogrammetry and Remote Sensing,* vol.29, pp.893-896, 1992.

8. N. DApuzzo. "Automated Photogrammetric Measurement of Human Faces." *International Archives of Photogrammetry and Remote Sensing,* vol.32, pp.402-407, 1998.

9. R.M. Koch *et al.* "Simulating Facial Surgery Using Finite Element Models." in *Proc of the SIGGRAPH96,* 1996.

10. N. Motegi *et al.* "A Facial Growth Analysis Based on FEM Employing Three Dimensional Surface Measurement by a Rapid Laser Device." in *Proc of the Okajimas Folia Anatomica Japonica,* vol.6, pp.323-328, 1996.

11. V. Blanz and T. Vetter. "A Morphable Model for the Synthesis of 3D Faces." in *Proc of the SIGGRAPH'99,* pp.187-194, 1999.

12. W.S. Lee and N.T. Magnenat. "Fast Head Modeling for Animation." *Image and Vision Computing Journal,* vol.4, pp.355-364, 2000.

13. S.R. Marschner *et al.* Modleing and Rendering for Realistic Facial Animation." in *Proc. of the 11th Eurographics Workshop on Rendering,* 2000.

14. F. Pighin *et al.* "Synthesizing Realistic Facial Expressions from Photographs." in *Proc of the SIGGRAPH'98,* pp.75-84, 1998.

15. R.Sitnik and M.Kujawinska. "Opto-Numerical Methods of Data Acquisition for Computer Graphics and Animation Systems." in *Proc of Three-Dimensional Image Capture and Applications III,* pp.36-43, 2000.

16. S.V. Duhn *et al.* "Three-view surveillance video based face modeling for recognition." in *Proc of the Biometrics Symposium,* 2007.

17. S. Minaku *et al.* "Three-Dimensional Analysis of Lip Movement by 3-D Auto Tracking System." *International Archives of Photogrammetry and Remote Sensing,* vol.30, 1995.

18. Y. Shan *et al.* "Model-Based Bundle Adjustment with Application to Face Modeling." in *Proc. of the 8th International Conf. on Computer Vision (ICCV01),* vol.2, pp.624-651, 2001.

19. P. Fua. "Regularized Bundle-Adjustment to Model Heads from Image Sequences without Calibration Data." *International Journal of Computer Vision,* vol.2, pp.153-171, 2000.

20. D. DeCarlo *et al.* "An Anthropometric Face Model Using Variational Techniques." in *Proc of the SIGGGRAPH'98,* pp.67-74, 1998.

21. A.Borghese and S. Ferrari. "A Portable Modular System for Automatic Acquisition of 3-D Objects." in *IEEE Trans. on Instrumentation and Measurement,* vol.5, pp.1128-1136, 2000.

22. G.Drettakis and R.Scopigno. "Animating lips-sync speech faces with compact key shapes." *Eurographics,* vol.27, No3, 2008.

23. Y.X.Hu *et al.* "Camera and Microphone Array for 3D Audiovisual Face Data Collection." in *The 33rd International Conference on Acoustics, Speech, and Signal Processing (ICASSP2008)*, 2008.

24. T.F.Cootes and P.Kittipanyangam. "Comparing variations on the active appearance model algorithm." in *BMVC2002*, 2002.

25. B. Fleming and D. Dobbs. "Animating facial features and expressions." *Charles River Media.* 1999.

26. K.Kayler. "A head model with anatomical structure for facial modeling and animation." Phd.Thesis, University of Saarlandes, 2003.

27. A. Savran *et.al.* "Bosphorus Database for 3D Face Analysis." in *The First COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008),* 2008.

28. P.H. Schonemann. "A generalized solution of the orthogonal Procrustes problem." *Psychometrika,* vol.31, pp.1-10, 1966.

29. P.J. Besl and H.D. McKay. "A method for registration of 3-D shapes." in *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1992.

30. Y. Chen and G. Medioni. "Object Modeling by Registration of Multiple Range Images." in *Proc. of the 1992 IEEE Intl. Conf. on Robotics and Automation,* pp.2724-2729, 1992.

31. S.Y. Du *et al.* "An extension of the ICP algorithm considering scale factor." in *Proc 14th IEEE International. Conf. on Image Processing (ICIP),* pp.193196, 2007.

32. J. Gower. "Generalized procrustes analysis." *Psychometrika,* vol.40, pp.1-10, 1966.

33. J.Gower and G.B. Dijksterhuis. "Procrustes Problems." *Oxford University Press,* 2004.

34. T.Masuda *et al.* "Registration and Integration of Multiple Range Images for 3-D Model Construction." in *Proc. CVPR,* 1996.

35. F.I. Parke. "Parameterized Models for Facial Animation." in *IEEE Computer Graphics and Applications,* vol.2, pp.61-68, 1982.

36. F.I. Parke and K. Waters. "Computer Facial Animation 2nd Edition." ISBN.978-1-56881-448-3, 2008.

37. S.M. Platt and N. Badler. "Animating facial expression." *Computer Graphics,* vol.15(3), pp.245-252, 1981.

38. Y.Zhang *et al.* "Real-time physically-based facial expression animation using mass-spring system." in *Proc. Computer Graphics International 2001,* pp. 347-350, 2001.

39. D. Terzopoulos and K. Waters. "Physically-based facial modeling, analysis and Animation." *Journal of Visualization and Computer Animation,* vol.1 pp.73-80, 1990.

40. K.S. Benli *et al.* "Semi-Automatic Adaptation of High-Polygon Wireframe Face Models Through Inverse Perspective Projection." in *Proc. of International Symposium on Computer and Information Sciences (ISCIS 2011),* 2011.

41. T.F. Cootes and C.J. Taylor. "Statistical Models of Appearance for Computer Vision." 2004.

42. K.L. Low and A. Lastra. "Reliable and Rapidly-Converging ICP Algorithm Using Multi resolution Smoothing." in *Proc of the Fourth International. Conf. on 3-D Digital Imaging and Modeling (3DIM 2003),* 2003.

43. K.Pearson. "On lines and planes of closest fit to systems of points in space." *Philosophical Magazine,* vol.2, no.11, pp.559-572, 1901.

44. T.F.Cootes *et al.* "Active Shape Models - Their Training and Application." *Computer Vision and Image Understanding,* vol.61, No.1, pp.8-59, 1995.

45. T.F. Cootes *et al.* "Active Appearance Models." in *Proc. European Conference on Computer Vision,* Vol.2, pp.484-498, 1998.

46. S. Krindis and I. Pitas. "Statistical analysis of human facial expressions." *Journal of Information Hiding Multimedia Signal Processing,* vol.1, no.3, pp.241260, 2010.

47. J.Ahlberg. “CANDIDE-3 - an updated parameterized face.” *Report No. LiTH-ISY-R-2326, Dept. of Electrical Engineering, Linkping University, Sweden,* 2001.