3D face recognition with geometrically localized surface shape indexes

Hyoungchul Shin

The graduate school

Yonsei University

Department of Graduate Program in Biometrics

3D face recognition with geometrically localized surface shape indexes

A Dissertation

Submitted to the Department of

Graduate Program in Biometrics

And the Graduate School of Yonsei University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Hyoungchul Shin

2005

This certifies that the masters thesis
of Hyoungchul Shin is approved.

_____

Thesis Supervisor: [Kwanghoon Sohn]

_____

Committee Member : [Sangyoun Lee]

_____

Committee Member : [Toh Kar Ann]

The Graduate School
Yonsei University
December 2005

감사의 글

용기 내어 다시 시작한 학문의 길이기에 석사학위를 받고 졸업하는 것이 아직은 부족함이 많아 아쉽게 느껴집니다. 2년간의 짧은 기간이었지만 지난 세월의 소중한 땀과 노력이 모여 작은 결실을 맺게 되었습니다. 아쉬움과 부족함이 많은 논문이지만, 소중한 결실을 맺을 수 있도록 이끌어 주신 많은 분들께 감사의 말씀을 전하려 합니다.

학부시절부터 다시 대학원 생활을 하게 된 지금까지 학문적 지도와 삶에 대한 가르침을 아끼지 않으시고 삶에 힘이 되는 용기와 다정한 격려를 주신 손 광훈 교수님께 진심으로 감사 드립니다. 아울러 좋은 논문이 될 수 있도록 많은 지도와 격려를 주신 이 상윤 교수님, Toh Kar Ann 교수님께 진심으로 감사의 말씀을 전합니다. 생체인식 협동과정을 만드시고 열심히 공부할 수 있도록 보살펴 주신 김 재희 교수님과 저의 대학 생활을 지켜봐 주시며 많은 가르침을 주신 박 한규 교수님, 윤 영중 교수님, 김 동구 교수님, 한 상국 교수님, 서 종수 교수님께 진심으로 감사 드립니다.

연구실 생활에 빨리 적응할 수 있도록 도와주신 연구실 선·후배 및 동기 분들께 감사 드립니다. 연구실 맏형이자 학부시절부터 카리스마가 넘쳐서 대학원 생활 내내 큰 의지가 되어준 환종이형, 모범적인 연구실 생활을 보여준 남을 이해 할 줄 아는 친구 정은이, 연구실에서는 대선배지만 학부동기로서 많은 부분을 채워준 한성이, 친형처럼 대해주고 때로는 흐트러지지 않도록 충고도 해준 멋쟁이 용태, 장난기 많고 정 많은 생활력 좋은 생체인식 센터의 기둥 욱일이, 연구실 생활의 활력소가 되는 작은 즐거움을 함께했던 홍규형, 착하고 자상한 명수형, 듬직하고 편안한 영범이형, 힘든 1학기 생활을 잘 보낼 수 있도록 도와준 남을 배려할 줄 알고 썰렁한 농담이 일품인 용욱이, 똑똑하고 책임감 강한 개강파티의 숙적 천재99

방장 동보, 항상 여유있고 웃는 모습이 보기 좋은 주현이형, 착하고 부지런한 장기마니아 형갑이, 친형처럼 다정하지만 대결은 피하고 싶은 항상 노력하는 현진이형, 항상 챙겨주고 마음도 잘 맞는 살림꾼 신우, 말수가 적지만 가끔 던지는 한마디가 날카로운 은근한 도전자 정동이, 재미있는 말이 생각나면 말하고 싶어 표정이 변하는 귀여운 동현이, 착하고 이해심 많아서 현진이형에게 항상 당할 것 같은 동병상련 오윤이형, 부족한 선배를 잘 도와주는 헤드폰 록커 봉조, 똘똘하고 융통성 있는 멋 부릴 줄 아는 도우너 동주, 착하고 쇼맨십 좋고 의사표현이 단순한 해맑은 재원이, 정적에 감싸이기 쉬운 연구실을 활기차게 해주는 이쁜이 지현이, 놀기도 잘하고 공부도 열심히 해서 데뷔하면 매니저 하고 싶은 투덜이 상운이, 여행을 좋아하고 독특한 멋을 즐길 줄 아는 신입생 지영이, 안타깝게 아직 입학을 못했지만 정직하고 바른 생각을 닮고 싶은 대성할 체육특기생 민성이에게 고마움을 전합니다. 또한 졸업해서 같이 많은 시간을 같이하지 못했지만 기선, 재환이형, 정환, 승철이에게 무궁한 발전이 있기를 기원합니다.

언제 어디서나 항상 서로를 생각하고 힘이 되어주는 진우회 친구들에게도 감사의 마음을 전합니다. 중학교를 제외하고 대학교까지 항상 같이한 서로의 맘을 읽을 수 있는 친구 성남이, 따뜻한 가슴으로 친구들의 모든 일에 나서는 연예 한 번 못해본 의리의 친구 영준이, 항상 관심을 끌고 싶어하는 잔머리 9단 가훈이, 내 피부를 아직도 부러워하는 친절한 병춘씨, 궁금할 때 모든 대답을 해주는 정보통 호진이, 힘들 때 멋진 노래를 불러준 새신랑 상현이, 꿀 같은 결혼생활을 즐기고 있는 예비 애아빠 양무, 첫 애아빠가 될 세 식구의 가장 로맨티스트 재회, 개업만 하면 되는 얼굴 네모난 치과의사 상선이, 진정 즐길 줄 아는 사이판의 왕자 성민이, 분위기 띄우고 사라지지만 사람 향기가 나는 친구 현석이, 10년 만에 갑자기 여자친구를 데려온 내일 장가갈지 모르는 친구 진모, 언제 봐도 반가운 나보다 머리가 작은 두규에게 감사의 말을 전하며 우리의 우정이 항상 변함이 없기를 기원합니다.

지금은 잘 연락하지 못하지만 초등학교 때부터 소중한 우정을 간직하고 있는 정수, 10년간 같은 학교에서 같은 길을 가고 있는 친구 동현이, 항상 보고 싶은 친구 수창이, 멀리서 응원해준 덕주, 대학원의 길을 같이 선택한 선일이에게 감사의 마음을 전합니다. 후배지만 많은 것을 같이 하고 힘이 되어준 홍성흔을 닮은 꺽다리 현원이와 항상 친형처럼 따르는 회사 후배 신환이, 성민이, 호진이, 동규, 경화, 건동이에게도 감사의 마음을 전하고 싶습니다. 멀리 있지만 잊지 않고 소식을 전해주는 얼굴만큼 마음도 예쁜 동생 미화와 같은 시기에 대학원 생활을 하면서 즐겁게 해준 영오, 얼마 전에 제대 하고 더욱 믿음직스러워진 대선이와 진혁이 그리고 항상 고마운 마음을 전할 길이 없었던 이모와 이모부에게 진심으로 감사의 마음을 전하고 싶습니다. 이 밖에도 미처 말하지 못한 소중한 사회 경험을 같이한 회사 분들과 저를 아는 모든 분들께 죄송한 마음과 감사의 마음을 함께 전합니다.

자주 찾아가지도 못하고 고맙다는 말 한마디 표현 못했지만 항상 제멋대로인 동생을 이해하고 마음속에 든든한 버팀목이 되어 응원해주는 형과 우리 가족을 전보다 밝고 즐겁게 만들어 주신 형수님께도 감사의 마음을 전합니다. 무엇보다도 지난 30년간 모든 정성을 다하셔서 오직 부족한 자식들을 위해 희생하신 사랑하는 부모님과 그리고 무척이나 저를 예뻐해 주셨던 하늘에 계신 할머님께 이 논문을 바칩니다. 못난 아들의 잘못된 행동과 투정을 다 받아주시고 감싸주셨던 어머니와 아버지께 부족하나마 이 논문이 작은 기쁨과 행복이 되었으면 합니다.

2005년 12월 29일

신 형 철 올림

Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two-dimensional
3D	Three-dimensional
DB	Database
ECSVD	Error Compensated Singular Value Decomposition
NPP	Nose Peak Point
NBRP	Nose Bridge Point
NBP	Nose Base Point
CPE	Center Point between Eyebrows
NEP	Nose End Point
NREP	Nose Right End Point
NLEP	Nose Left End Point
EOP	Eye Outside corner Point
EIP	Eye Inside corner Point
ERIP	Eye Right Inside Point
ELIP	Eye Left Inside Point
ICA	Independent Component Analysis
FRT	Face Recognition Technology
SFS	Shape From Shading
EGI	Extended Gaussian Image
ICP	Iterative Closest Point
FRVT	Face Recognition Vendor Test
SVM	Support Vector Machine
BERC	Biometrics Engineering Research Center

Abstract

3D face recognition using geometrically localized surface shape indexes

Hyoungchul Shin
Graduate Program in Biometrics
The Graduate school
Yonsei University

In this dissertation, we propose a pose invariant three-dimensional (3D) face recognition method using distinctive facial features. A face has its structural components like the eyes, nose and mouth. The positions and the shapes of the facial components are very important characteristics of a face. We extract invariant facial feature points on those components using the facial geometry from a normalized face data and calculate relative features using these feature points. We also calculate a shape index on each facial feature point to represent curvature characteristics of facial components. When facial shape index and facial feature points with relative features are used separately, face recognition rates are 83% and 89% at first top rank by the weighted distance matching on average for seven different poses for 300 different people, respectively. However, the recognition rate goes up to 96.7% when they are used together.

Proposed feature vector can be applied to various conventional classifiers because it has a fixed dimension. We propose weighted vector distance matching. We also applied proposed feature vector to support vector machine (SVM) and independent component analysis (ICA).

We have 97.6% recognition rate from our proposed weighted distance matching, 98.6% at first top rank by the SVM and 97.3% by the ICA on average for seven different poses for 300 different people. Although SVM shows the highest recognition rate, the weighted vector distance matching shows a satisfactory recognition rate without any training process. Moreover, the proposed method can be used with incomplete feature vector. When some features are failed to be extracted, the proposed recognition algorithm can still process the incomplete information. But it is still valid for the proposed recognition algorithm. From the experimental results, we have effectively utilized facial shape indexes, geometrical feature points and its relational features for pose invariant face recognition.

Key words: 3D face recognition, facial shape indexes, weighted vector distance matching, support vector machine (SVM), independent component analysis (ICA)

Chapter 1.

Introduction

1.1 Objective of this research

For the past decade, face recognition technologies have made great progress using 2D images. The human face is the most easily acquired biometric feature. The less intrusive nature of the process of acquiring a face image is the primary reason why face recognition based systems are preferred over other biometric systems.

Most face recognition research has been focused on visible spectrum images. The primary reason for using them was because they were easy to acquire. Although 2D face recognition has been shown to have reasonable performance under controlled inner and outer environments, there are still many unsolved problems in varying environments such as pose, illumination and expression. Thus, a face recognition system should be robust to such changes to be used in the real world [1-5].

With the development of a 3D acquisition system, 3D face capture is becoming faster and cheaper. Hence, face recognition based on 3D information draws much attention due to the need

to those limitations incurred by using 2D images. Early work on 3D face recognition was launched a decade ago, and a few approaches have been reported about face recognition using 3D data which were acquired by a 3D sensor and a stereo-based system [6]. Having a 3D system removes many problems associated with lighting and pose that can affect 2D systems. In particular, 3D technology is reported to be effective when a person changes his facial expression or when the subject puts on weight or grows a beard. The solution of 3D face recognition can be more accurate than 2D face recognition as it is able to concentrate on more fundamental features of the face, such as the bone structure around the eyes and nose. These features remains almost completely invariant. Even when you throw away the data coming from the mouth area, 3D technology still captures more than enough information to make a match. However, the quality of data from a 3D sensor is highly dependent on 3D devices that one uses. While most of previous works using a 3D device are based on structured light, there exist some missing data points due to occlusion or improperly reflected regions (e.g., dark regions such as eyebrows). On the other hand, very good quality images of human faces are provided by the 3D range scanner of Cyberware [7]. The system is based on a laser range finder and rotation platform.

The main purpose of this dissertation is to develop a pose invariant 3D face recognition method based on geometrically localized shape indexes by using two different 3D sensors, which have different resolution and a different number of data points for probe and gallery,

respectively. Although 3D facial recognition technology has yet to make significant progress in the commercial world, its promise for high accuracy and better performance than 2D recognition under difficult conditions make it a technology that could attract a great deal of attention over the coming years.

1.2 Overview of the proposed scheme

In this dissertation, we propose a robust 3D face recognition system based on geometrically localized facial shape indexes acquired from two different 3D sensors. The proposed 3D face recognition system is divided into three parts, which are the data acquisition stage, the feature extraction stage, and the recognition stage. Our main goals for developing the 3D face recognition system are as follows.

Firstly, for 3D face data acquisition, it is important for us to explore the data acquisition process of probe and gallery images. When we acquire the face data from a 3D sensor, we need a preprocessing step such as extraction of a white bathing cap from the head region and noise filtering of the obtained face images. Thus, the acquired probe and gallery images are in the established 3D normalized space, which provides good qualities for extracting facial feature points.

Secondly, the proposed 3D facial feature extraction method using the geometric characteristics is described. We efficiently select maximum and minimum points, which are

denoted by the vertical curvature through the NPP (Nose Peak Point). Using the extracted feature points on the vertical curvature, we extract the other feature points such as eye inner points, nose end points along the curvatures on the horizontal direction. On each extracted feature point, we calculate the facial shape indexes. We also propose the relative features that are obtained by using the relations among the previously extracted feature points. Extracted feature points are normalized based on the 3D head pose estimation by Error Compensated Singular Value Decomposition (EC-SVD). EC-SVD is proposed by H. Song et al. They defined an Error Compensated Singular Value Decomposition (EC-SVD) for estimating a head pose with facial features. They estimated the initial 3D head pose using the SVD method, and perform a refinement procedure to compensate for the remaining errors. Less than 1.6 degree error on average for each axis has been achieved for 3D head pose estimation [8].

Lastly, for face recognition, we perform weighted vector distance matching, support vector machine (SVM) and independent component analysis (ICA) for comparison. The extracted feature vectors are supplied for the input of each classifier.

1.3 Outline of this dissertation

The remainder of this dissertation is organized as follows. Chapter 2 describes existing methods of 2D and 3D face recognition, including tabular comparison of their performances. Chapter 3 describes the proposed face recognition system. Firstly, the representation of 3D

faces for the probe and gallery images is explained. Then the preprocessing procedure is described. This is followed by a description of the extraction methods for 3D facial features. Finally, the proposed 3D face recognition methods using weighted vector distance matching, SVM and ICA are explained in detail. In chapter 4, experimental results are analyzed to explain the efficiency of the proposed algorithm. Finally, chapter 5 concludes by suggesting future directions in which the method can be extended.

Chapter 2.

Research background

We describe the background of our research for face recognition in this chapter. Also, we shall review the existing techniques for 2D and 3D face recognition.

2.1. Problems in 2D face recognition

Face Recognition Technology (FRT) is a research area spanning several disciplines such as image processing, pattern recognition, and computer vision. There are many applications which range from matching of photographs to real time matching of surveillance video. Depending on the specific application, FRT has different levels of difficulty and requires wide range of techniques. In 1995, a review paper by Chellappa et al. gives a through survey of FRT at that time [1-2]. During the past few years, FRT is still under rapid evolution. In addition, face recognition has been developed based on 2D and 3D images as illustrated in Fig. 2.1. A lot of researchers are continuing to develop 2D and 3D face recognition systems.

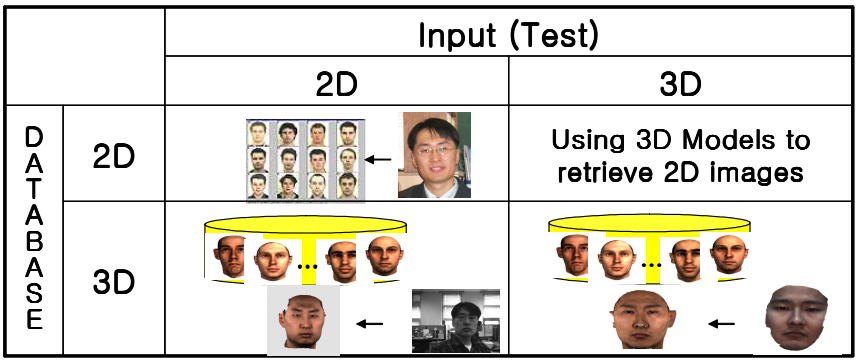


Figure 2. 1 Topology for face recognition

Though many FRTs have been proposed, robust face recognition is still a difficult problem. The recent FERET test has revealed that there are at least two major challenges [4-5]:

- The illumination variation problem
- The pose variation problem

Either one or both problems may cause serious performance degradation in most existing systems. Unfortunately, these problems happen in many real world applications, such as surveillance video. In the following, we will discuss some existing solutions for these problems.

The general face recognition problem can be formulated as follows: Given single image or sequence of images, recognize the person in image using a database. Solving the problem consists of the following steps: 1) face detection, 2) face normalization, and 3) inquire database.

2.1.1. The illumination problem

Images of the same face appear differently due to change in lighting. If the change induced by

illumination is larger than the difference between individuals, systems would not be able to recognize the input image. To handle the illumination problem, researchers have proposed various methods. It has been suggested that one can reduce variation by discarding the most important eigenface. In addition, it is verified that discarding the first few eigenfaces seems to work reasonably well [11]. However, it causes the system performance degradation for input images taken under frontal illumination.

Different image representations and distance measures are evaluated [12]. One important conclusion which this paper draws is that none of these method is sufficient by itself to overcome the illumination variations. More recently, a new image comparison method was proposed by Jacobs et al. [13]. However, this measure is not strictly illumination-invariant because the measure changes for a pair of images of the same object when the illumination changes.

An illumination subspace for a person has been constructed for a fixed view point [14]. Under fixed view point, recognition result could be illumination–invariant. One drawback to use this method is that we need many images per person to construct the basis images of illumination subspace.

Atick et al. suggest using Principal Component Analysis (PCA) to solve parametric shape-from-shading (SFS) problem [15]. The idea is quite simple. They reconstruct 3D face surface from single image using computer vision techniques. Then compute the frontal view

image under frontal illumination. Very good results are demonstrated. Actually, there are a lot of issues regarding how to reconstruct 3D surface from single image.

2.1.2. The pose problem

The system performance drops significantly when pose variations are present in input images. Basically, the existing solution can be divided into three types: 1) multiple images per person are required in both training stage and recognition stage, 2) single image based methods, 3) multiple images per person are used in training stage but only one database image per person is available in recognition stage. The third type is the most popular one.

A. Multiple images approaches

An illumination-based image synthesis method has been proposed for handling both pose and illumination problems [16]. This method is based on illumination cone to deal with illumination variation. For variations due to rotation, it needs to completely resolve the GBR (generalized-bias-relief) ambiguity when reconstructing 3D surface.

B. Single image based approaches

Gabor wavelet based feature extraction is proposed for face recognition and is robust to small-angle rotation [17]. There are many papers on invariant features in computer vision

literature. There are little literatures explaining on how to use this technology for face recognition. Recent work sheds some light in this direction [18]. For synthesizing face images under different lighting or expression, 3D facial models have been explored in [19]. Due to its complexity and computation cost it is hard to apply this technology to face recognition.

C. Hybrid approaches

A lot of hybrid approaches have been proposed. It is probably the most practical solution to date. Three representative methods are reviewed in this report: 1) linear class based method [20], 2) graph matching based method [21], 3) view-based eigenface method [9]. The image synthesis method is based on the assumption of linear 3D object classes and extension of linearity to images [20]. A robust face recognition scheme based on EBGM is proposed [21]. It demonstrated substantial improvement in face recognition under rotation. In addition, it was fully automatic, including face localization, landmark detection and graph matching scheme. The drawback of this method is the requirement of accurate landmark localization which is not easy when illumination variations are present. The popular eigenface approach has been modified to achieve pose-invariance [9]. This method constructs eigenfaces for each pose. More recently, a general framework called bilinear model has been proposed [22]. The methods in this category have some common drawbacks: 1) they need many images per person to cover possible poses. 2) The illumination problem is separated from the pose problem.

2.2. 3D face recognition

Although the first attempts at 3D face recognition are over a decade old, not many papers have been published on this topic. The purpose of this chapter is to summarize and critique existing literature on 3D face recognition. Traditionally, method for face recognition have been broadly classified into two categories: the “appearance-based” methods, which treat the face as a global entity, and “feature-based methods” which locate individual facial features and use spatial relationships between them as a measure of facial similarity. This chapter surveys the existing approaches belonging to both these categories and presents a tabular comparison (see Table 2.1).

Principal Components Analysis (PCA) was first used for the purpose of face recognition with 2D images in the paper by Turk and Pentland [10]. The technique has been applied to recognition from 3D data by Hesher and Srivastava [23]. Their database consists of 222 range-images of 37 different people. The different images of one and the same person have 6 different facial expressions. The range-images are normalized for pose changes by first detecting the nasal bridge and then aligning it with the Y-axis. An eigenspace is then created from the “normalized” range-images and used to project the images onto a lower dimensional space. Using exactly one gallery image per person, a face recognition rate of 83% is obtained.

PCA has also been used by Tsalakanidou et al [24] on a set of 295 frontal 3D images, each belonging to a different person. They choose one range-image each for 40 different people to

build an eigenspace for training. Their test set consists of artificially rotated range-images of all the 295 people in the database, varying the angle of rotation around the Y-axis from 2 to 10 degrees. For the 2-degree rotation case, they claim a recognition rate of 93%, but the recognition rate drops to 85% for rotations larger than 10 degrees.

Yet another study using PCA on 3D data has been reported by Achermann et al [25]. They have used the PCA technique to build an eigenspace out of 5 poses for 24 different people. Their method has been tested on 5 different poses each for the same group of people. The poses of the test images seem to lie in between the different training poses. The authors report a recognition rate of 100% on their data set using PCA with 5 training images per person. They have also applied the method of Hidden Markov Models on exactly the same data set and report recognition results of 89.17% for the Hidden Markov Models' method using 5 training images per person.

None of the above experiments specifies the time-span between the collection of the training and testing images for the same person. The inclusion of sufficient time gaps between the collection of training and testing images is a vital component of the well-known FERET protocol for face recognition [5]. Furthermore, in the work by Tsalakanidou et al [24], the range image database consisted of only one image per person, thereby making the training and test source data nearly identical. The test images were actually created by synthetically manipulating the training images and therefore do not represent the natural variations in the

appearance of a human face over a period of time. The method of facial normalization adopted by Hesher et al [23] consists merely of alignment of the nasal ridge with the Y-axis. However, this does not adequately compensate for changes in yaw, as it is possible for the nasal line to be aligned with the Y-axis even when the face has undergone yaw rotations.

Chang et al [26] reported the largest study on 3D face recognition till date, which is based on a total of 951 range-images of 277 different people. Using a single gallery image per person, and multiple probes, each taken at different time intervals as compared to the gallery, they have obtained a face recognition rate of 92.8% by performing PCA using just the shape information. They have also examined the effect of spatial resolution (in X, Y and Z directions) on the accuracy of recognition. However, they perform manual facial pose normalization by aligning the line joining the centers of the eyes with the X-axis, and the line joining the base of the nose and the chin with the Y-axis. Manual normalization is not feasible in a real system, besides being prone to human error in marking feature points.

Tsalakanidou [24] and Chang [26], [59] claimed a better recognition rate when 3D and the corresponding 2D face data were combined, resulting in a multi-modal recognition system. In both studies the recognition rates using just only 3D information were higher than the recognition rates obtained by using just only the 2D (texture) information.

Tanaka et al [27] calculated the maximum and minimum principal curvature maps from the depth maps of faces. From these curvature maps, they extract the facial ridge and valley lines.

The former are a set of vectors that correspond to local maxima in the values of the minimum principal curvature. The latter are a set of vectors that correspond to local minima in the values of the maximum principal curvature. From the knowledge of the ridge and valley lines, they construct extended Gaussian images (EGI) for the face by mapping each of the principal curvature vectors onto two different unit spheres, one for the ridge lines and the other for the valley lines. Matching between model and test range images is performed using Fisher's spherical correlation, a rotation-invariant similarity measure, between the respective ridge and valley EGI. This algorithm has been tested on a total of 37 range images, with each image belonging to a different person and 100% accuracy has been reported. The variation between training and test images in terms of head pose and time difference in acquisition has again been left unspecified. Moreover, extraction of the ridge and valley lines requires the curvature maps to be thresholded. This is a clear disadvantage because there is no explicit rule to obtain an ideal threshold, and the location of the ridge and valley lines are very sensitive to the chosen value.

Lee and Milios obtained convex regions from the facial surface using curvature relationships to represent distinct facial regions [28]. Each convex region is represented by an EGI by performing a one-to-one mapping between points in those regions and points on the unit sphere that have the same surface normal. The similarity between two convex regions is evaluated by correlating their Extended Gaussian images. To establish the correspondence between two faces, a graph-matching algorithm is employed to correlate the set of only the convex regions in the

two faces (ignoring the non-convex regions). It is assumed that the convex regions of the face are more insensitive to changes in facial expression than the non-convex regions. Hence their method has some degree of expression invariance. However, they have tested their algorithm on range-images of only 6 people and no results have been explicitly reported.

Feature-based methods aim to locate salient facial features such as the eyes, nose and mouth using geometrical or statistical techniques. Commonly, surface properties such as curvature are used to localize facial features by segmenting the facial surface into concave and convex regions and making use of prior knowledge of facial morphology [29-30]. For instance, the eyes are detected as concavities (which correspond to positive values of both mean and Gaussian curvature) near the base of the nose. Alternatively, the eyebrows can be detected as distinct ridge-lines near the nasal base. The mouth corners can also be detected as symmetrical concavities near the base of the nose. After locating salient facial landmarks, feature vectors are created based on spatial relationships between these landmarks. These spatial relationships could be in the form of distances between two or more points, areas of certain regions, or the values of the angles between three or more salient feature-points. Gordon [29] created a feature-vector of 10 different distance values to represent a face, whereas Moreno et al created an 86-valued feature vector [30]. In [30], they basically segment the face into 8 different regions and two distinct lines, and their feature-vector includes the area of each region and the distance between the center of mass of the different regions as well as angular measures. In both [29]

and [30], each feature is given an importance value or weight, which is obtained from its discriminatory value as determined by Fisher's criterion [31]. The similarity between gallery and probe images is calculated as the similarity between the corresponding weighted feature-vectors. Gordon [29] reports a recognition rate of 91.7% on a dataset of 25 people, whereas Moreno et al [30] report a rate of 78% on a dataset of 420 range-images of 60 individuals in two different poses (looking up and down) and with five different expressions. Again, neither of these methods has explicitly taken into account the factor of time variation between gallery and probe images, nor have they given details about the pose difference between the training and test images. A major disadvantage of these methods is that location of accurate feature-points (as well as points such as centroids of facial regions) is highly susceptible to noise, especially because curvature is a second derivative. This leads to errors in the localization of facial features, which are further increased with even small pose changes that can cause partial occlusion of some features, for instance downward facial tilts that partially conceal the eyes. Hence, the feature-based methods described in [29] and [30] lack robustness.

Lee et al [32] performed face recognition by locating the nasal tip in the depth map, followed by extraction of facial contour lines at a series of different depth values. They have reported a rank-five recognition rate of 94% on a very small dataset. This method is clearly sensitive to the discretization in the depth values. It would also not be robust in cases where range images of a person were obtained with scanners with different depth resolutions.

The concept of point signatures has also been used for face recognition in recent work by Wang, Chua and Ho [33]. They manually select four fiducial points on the facial surface from a set of training images and calculate the point signatures over 3 by 3 neighborhoods surrounding those fiducial points (i.e., 9 point signature vectors). These signature vectors are then concatenated to yield a single feature vector. The maximum recognition rate with three training images and three test images per person is reported to be around 85%. The different images collected for each person show some variation in terms of facial expressions.

Lu, Colbry and Jain have used an ICP-based method for facial surface registration in [34] and [35]. They have employed a feature-based algorithm followed by a hybrid ICP algorithm that alternates in successive iterations between the method proposed by Besl & McKay [36] and the method proposed by Chen and Medioni [37]. In this way they are able to make use of the advantages of both algorithms: the greater speed of the algorithm by Besl and McKay [36], and the greater accuracy of the method by Chen and Medioni [37]. Their hybrid ICP algorithm has been tested on a database of 18 different individuals with frontal gallery images and probe images involving pose and expression variations. A probe image is registered with each of the 18 gallery images and the gallery giving the lowest residual error is the one that is considered to be the best match. Using the residual error alone, they obtain a recognition rate of 79.3%. They improve this recognition rate to 84% by further incorporating information such as shape index and texture.

V. Blanz and T. Vetter utilized 3D morphable model for face recognition [38]. They tested these on the publicly available CMU-PIE and FERET database by fitting a morphable model of 3D faces to images. Using a single frontal gallery image per individual, a recognition rate of 95% has been reported under pose and illumination variation.

H. Song et al proposed 3D face recognition using facial curvature shape indexes [60]. They extracted invariant facial features based on the facial geometry. They extracted distinctive facial shape indexes based on facial curvature characteristics, and they perform dynamic programming and support vector machine for face recognition. In addition, 96.8% and 98.5% face recognition rate have been achieved based on facial shape indexes with dynamic programming (DP) and with support vector machine (SVM) for 300 individuals with seven different poses, respectively.

Y. Lee et al proposed a 3D face recognition system based on geometrically localized facial features. They utilized 3D frontal face images acquired in the BERC (Biometrics Engineering Research Center) face database under the frontal face for their experiments. The resulting recognition rate of DP algorithm was 95 % for 20 people. On the other hand, face recognition result using a SVM is 96 % for 100 people [61].

2.3 Summary of the 3D face recognition methods

This chapter presents a survey of existing methods in the 3D face recognition literature. The results for the different methods are summarized in Table 1. The general trend is that 3D face recognition methods outperform 2D methods. For instance, in studies where corresponding 2D and 3D images of the same set of people were obtained, 3D methods always yielded better results [24], [26], [59]. The performance of the state of the art in 2D face recognition technology can be assessed by means of the FERET protocol and the face recognition vendor test (FRVT), which was administered in 2002 [39]. From the FRVT 2002 report, the best existing 2D face recognition system yielded a recognition rate of 85% on a database of 800 individuals and suffered a decrease of 2% whenever the size of the database was doubled. A simple extrapolation allows us to conjecture that the performance of this system on a database of 200 individuals would be around 89%. On the other hand, the largest 3D face recognition system (developed by Chang [26]) yields a performance of about 92.8% on a database of 200 individuals, thereby outdoing the best existing 2D face recognition method. A combination of 2D and 3D methods has been reported to yield much higher rates than either 2D or 3D alone [24], [26], [33] (also see Table 2.1). However, it should be noted that the focus of this thesis is to make use of only 3D information, ignoring texture completely. This thesis is based on geometrically localized facial features and its shape indexes [60], [61].

Table 2.1 Summary of research on 3D face recognition

	Number of persons	**Number of training images**	**Number of test images**	**Reported performance**	**Pose normalization**
Ref. [23] (2003)	37	1	222	83%	Automated
Ref. [24] (2003)	295	1	120	93% for 2-3 degree rotaion, to 85% for 10 degree rotation	Not done
Ref. [25] (1997)	24	5	120	100%	Automated
Ref. [26] (2003)	200	5	870	92.8%	Manual
Ref. [27] (2003)	37	1	37	100%	Not done
Ref. [28] (1990)	6	1	6	Not reported	Not done
Ref. [29] (1991)	26	26	24	100%	Automated
Ref. [30] (2003)	60	1	420	78%	Automated
Ref. [32] (2003)	35	1	70	94% at rank 5	Automated
Ref. [33] (2002)	50	1	250	91%	Automated
Ref. [34] (2004)	18	1	63	79.37%	Automated
Ref. [60] (2004)	300	1	2100	98.5%	Automated
Ref. [61] (2004)	100	1	100	96%	Automated

Chapter 3.

Proposed 3D face recognition system

3.1 System overview

In this section, we present an overview of pose invariant face recognition system based on an enhanced geometrical feature vector. For our system, we utilize two different 3D sensors. The data from structured light based stereo cameras are used as a probe image. And 3D full laser scanned face data are used for stored images. We can acquire and process a face data in real time based on the structured light whereas we store face images with high quality obtained from a laser scanner. It is important to explore the data acquisition process of probe and stored images under the pose varying environments. Fig. 3.1 briefly presents the whole procedure to build a facial feature extraction with 3D head pose estimation, and finally accomplish face recognition. It is important to explore the capability of 3D face recognition based on 3D techniques.

First, the data acquisition of probe and stored images and their preprocessing procedures are described. When we acquire the data from a 3D sensor, we need preprocessing steps such as noise filtering and filling holes. 3D facial feature extraction technique using the facial geometry is also described. First, we define general facial features which are representative of the human face. In addition to these features, we propose our feature extraction procedures,

and illustrate relative features and shape index calculation using the extracted facial feature points.

In order to perform face recognition, we propose a weighted distance matching. We shall compare with conventional classifiers, such as SVM and ICA. The proposed matching method measures the similarity of the weighted distance between two feature vectors. The weights are determined by comparing Fisher's coefficients of each feature. Proposed matching method is simple and efficient because it does not need a training procedure. On top of the above proposed method, another method is performed based on support vector machine. We utilize extracted facial feature vectors for matching features. Support vector machine is a classifier which can separate classes and maximize the margin between them. The other recognition method performed in this dissertation is independent component analysis (ICA). ICA is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals. ICA defines a generative model for the observed multivariate data, which is typically given as a large database of samples. In the model, the data variables are assumed to be linear mixtures of some unknown latent variables, when the mixing system is also unknown. The latent variables are assumed non-gaussian and mutually independent and they are called the independent components of the observed data. These independent components, also called sources or factors, can be found by ICA. Face recognition results based on the proposed methods are compared with these conventional face recognition methods.

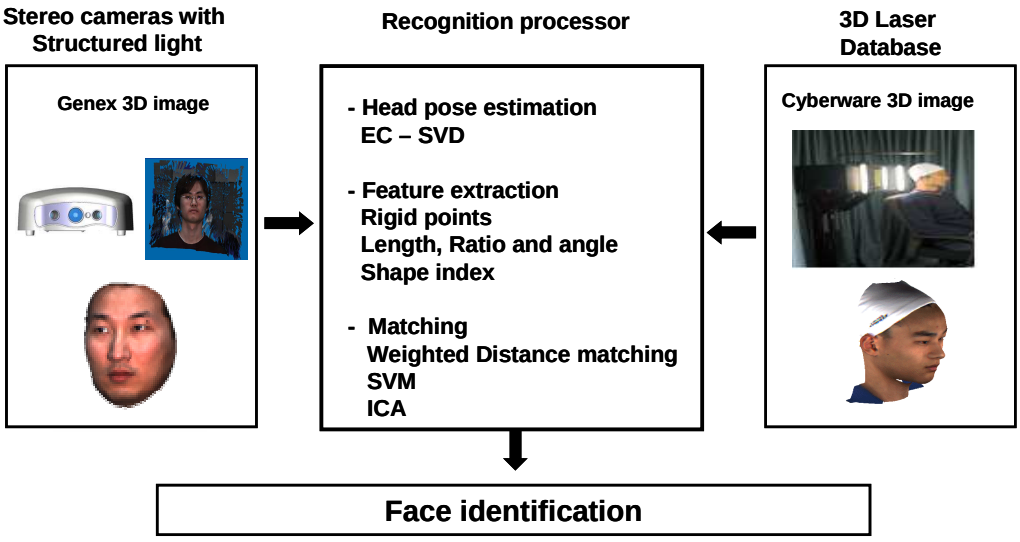


Figure 3. 1 The block diagram of the proposed system

3.2. 3D face data acquisition and preprocessing

3.2.1 3D sensor specification

We adopt two different 3D sensors, the Genex 3D FaceCam System [40] and the 3D Cyberware Model 3030PS/RGB laser scanner [7]. The detailed specification of both devices is tabulated on Table 3.1. The former is used as query device, and the latter as 3D face database collection. The Genex 3D FaceCam System is a novel 3D surface profile measurement system capable of acquiring at high speed full frame dynamic 3D images of faces with complex surface geometry. A “Full Frame 3D image” means that the value of each pixel in an acquired digital image represents the accurate distance from the camera’s focal point to the corresponding point on the face’s surface.

Table 3. 1. Specifications of input and database device

Specs	Input	3D database
	 3D Face-Cam (Genex)	 3D Laser Scanner (Cyberware)
Resolution	640 by 480	512 by 512
Field of View	510mmX356 mmX356mm	Theta 360, Max 340mm(Y)X300mm
3D formats	GTI, STL, PNT	OBJ, PLY,STL
Data	Structured light	Laser scan
Acquisition time	400msec (3D modeling : 15 secs)	90 secs
Vertex points	Max 300,000	Max 100,000

The Cartesian coordinates for all visible points on the face surface are provided by a single 3D image. This technology is based on a structured light process that projects known patterns of light onto the measuring scene. Structured light vision techniques greatly simplify the process of triangulating the position of a point in space. Using this technique, we can perform real time navigation. Structured light vision provides information about at least one aspect of the geometry in front of the camera; thus the triangulation equations are much easier to solve and less computationally expensive.

As for acquiring 3D face database, we utilize the 3D Cyberware Model 3030PS/RGB laser scanner, which is developed and patented by Cyberware's proven 3D digitizing technology. The 3D scan head incorporates a rugged, self-contained optical range-finding system, whose dynamic range accommodates varying lighting conditions and surface properties. Entirely software controlled, the scanner requires no user adjustments during normal use. During operation, the scanner shines a safe, low-intensity laser on an object to create a lighted profile. A high-quality video sensor captures this profile from two viewpoints. The system can digitize thousands of these profiles in a few seconds to capture the shape of the entire object. Simultaneously, a second video sensor in the scanner acquires color information. The scanning process captures an array of digitized points, with each point represented by the Cartesian coordinates for shape and 24-bit RGB coordinates for color. The scanner is designed to scan the head and face of live subjects quickly, comfortably, and safely. Because the scanner moves the digitizer while the subject remains stationary, it works well in many applications involving subjects that are inconvenient to move during digitizing.

3.2.2 Representation of 3D Faces

Normalization of the 3D face model is necessary not only for the head pose estimation but also for the face recognition and especially for the estimation of 3D head poses. Generally, 3D

face data are accompanied by complex 3D rotations for each X, Y, and Z axis. 3D face data in the database include more than 70,000 vertex points for each person and corresponding texture image. However, geometrical coordinates of the vertex points of each person are in a wide and different range (variation up to 100,000 for each axis). It takes a great computational effort to calculate them and different scales exist on each face.

Thus, we normalize the face data to make all the 3D face on the same object/face space. Different scale factors may occur when we acquire input data, so they should be fixed into a normalized face space through this normalization step. We define a 3D face space that normalizes the face representation for head pose estimation and face recognition. Given a 3D face *F*, which is defined by a point set of Cartesian coordinates, we consider the range of coordinates on the X, Y, and Z axis being infinite. However, we normalize these input data to a defined range for each axis, which we denote as *3D normalized face space.* Fig. 3.2 shows the example of a 3D normalized face space that we have established. All the faces that we consider are in this normalized face space and are proportionally located based on the original face data in the limited range of $[-\sigma,\sigma],[-\varepsilon,\varepsilon],[0,Z]$ for the X, Y, and Z axis, respectively. This will be discussed more in detail in chapter 3. First we normalize 3D faces with depth information (Z value) and then proportionally adjust the X and Y range. We acquire the limited range for each axis as follows:

$$F(x_i,y_i,z_i)=\left(\frac{F_x-F_{\min\_x}}{F_{\max\_x}-F_{\min\_x}}\times\sigma,\frac{F_y-F_{\min\_y}}{F_{\max\_y}-F_{\min\_y}}\times\varepsilon,\frac{F_z-F_{\min\_z}}{F_{\max\_z}-F_{\min\_z}}\times Z\right), \tag{3.1}$$

where $F(x_i,y_i,z_i)$ is normalized input data point, F_x, F_y and F_z are input data points for each axis. $F_{\max\_x}$, $F_{\max\_y}$, and $F_{\max\_z}$ are maximum values, and $F_{\min\_x}$, $F_{\min\_y}$, and $F_{\min\_z}$ are minimum values for the each X, Y and Z axis, respectively.

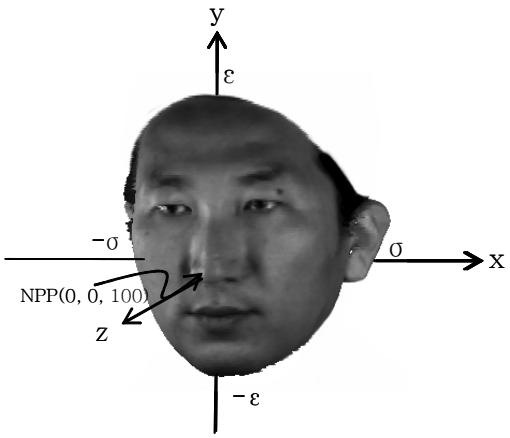


Figure 3. 2 Normalization of 3D face

3.3. 3D facial feature extraction

3.3.1 General features of a face

The main purpose of this section is to describe the generic facial feature points which are geometrically localized on a face and to analyze the characteristics of defined facial feature points. In particular, the following features are defined as generic features of a face [41]:

- nose bridge
- nose base
- nose ridge
- eye corner cavities: inner and outer

- convex center of the eye
- eye socket boundary
- boundary surrounding nose

Typical locations of some of these features are indicated for reference in Fig. 3.3. Each of these face descriptors is defined in terms of a high level set of relationships among depth and curvature features. Most features defined above are located at the nose and the eye regions of a face, because although most people have eye, nose and mouth regions commonly which comprise a face, but these features are mostly representative feature points of individual faces which are well defined for discrimination even under the pose variations [42].

We now illustrate the characteristics of those specific features listed above. Since several features are located on similar regions of a face as mentioned we group these features together,

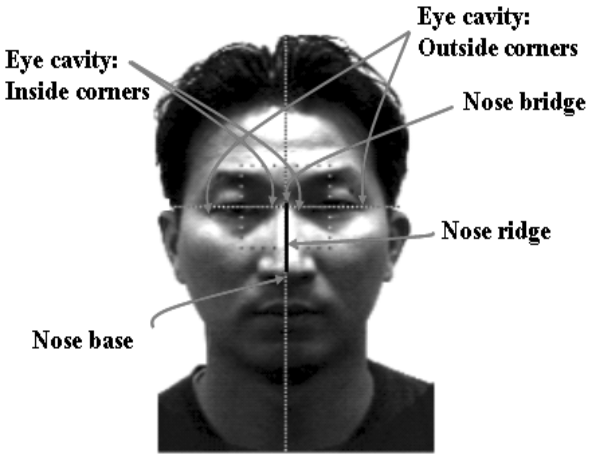


Figure 3. 3 Generic facial features

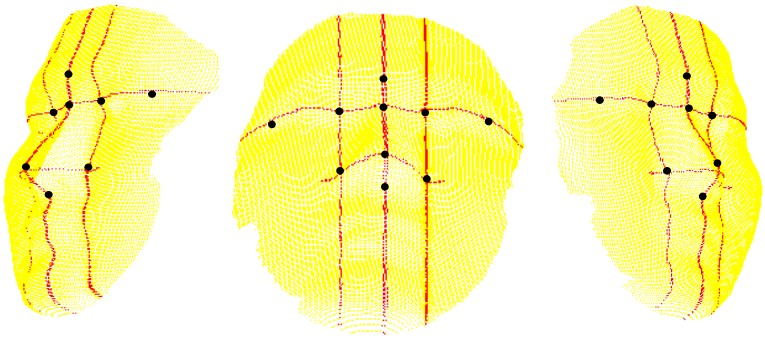
for instance, the nose and eye regions except for mouth region.

A. Features on nose region

The most striking aspect of the face in terms of 3D data is the nose. If the nose can be located reliably, it can be used to put constraints on the location of eyes and other features. There are at least three properties useful in locating the nose. The most obvious property is that it sticks out from the rest of the face. It also has a characteristic roof-like ridge, which is most often in a vertical orientation. Lastly, this ridge falls approximately on the symmetry plane of the face. In our implementation we make use of the first two properties as constraints in the location of the nose where we mark by the location of ridge. We can extract the robust feature points such as the NEP (Nose End Points), the NPP (Nose Peak Point), the NBP (Nose Base Point) and the NBRP (Nose Bridge Point) from the first and the second constraints.

3.3.3 Features on eye region

The search for given feature points can be greatly simplified if the target feature points have a characteristic relationship to the symmetry plane, or if it typically have a symmetrical match on the other side of the symmetry plane as shown in Fig.3.4-(b). We shall discuss these constraints in the context of extracting these particular eye features. We note that this process is general and can be applied to many other cases. We also find out the geometrical characteristics of the eyes, which are related to the NBRP: most people have a concave and convex eye line through the NBRP. In face, we are interested in locating the eye corners or eyeballs using these features, relying a large part on symmetry constraints. Among the more



(a) (b) (c)

Figure 3. 4 The curvature lines of a face: (a) -30° rotated image of a frontal image. (b) A frontal image. (c) 30° rotated image of a frontal face.

consistent curvature features are sets of symmetric concavities which occur at the inside and outside corners of the eyes named as the EIP (Eye Inside corner Point) and the EOP (Eye Outside corner point), respectively.

3.3.4 Feature extraction algorithm

The availability of 3D data can help to reduce the influence of viewpoint through the use of viewpoint invariant geometrical features. We extract facial feature points using 3D geometric information, especially depth information to find the key features such as the nose peak point (NPP). To extract the NPP accurately, we perform a few steps to exactly locate the point.

The Euclidean distance (radius) of the NPP is maximal from the Y axis among the vertex points of the 3D face only when the face is frontal. As the face pose varies, it is difficult to detect the NPP because the depth value of the NPP is not maximal any more. To find the NPP when the pose varies, we select the region from the maximal depth to the depth value lower by

three which is empirically found. We calculate the center of gravity of that selected region and treat as an initial NPP. Then we take a window template in order to calculate the variances of the horizontal and vertical profiles. We find the two points where the minimal variance of the horizontal profiles and the maximum maximal variance of the vertical profiles, respectively. If those two points occur at the same position, we select them as the refined NPP, otherwise we find the center of the two points. We can vertically and almost symmetrically divide the face using the YZ plane which includes the NPP and Y axis, and obtain the face dividing line/curvature. On the face center curve, we use curvature characteristics to extract facial feature points, which are convex and concave points except for the nose peak point. We select ten points which include four eye corner points, a center point between the left and right inner eye corner points, a center point of eye brows, a minimum point of the nose ridge, the nose peak point and two nose base points. Fig. 3.5 shows the extracted facial feature points of a face. When a feature can not be computed (always because of the non-existence of the regions from which it is derived), it is zero valued, except when a symmetric feature exists (then, the non-existent feature is valued like its symmetric feature). Extracted features have been normalized to values in the range of 0 and 1.

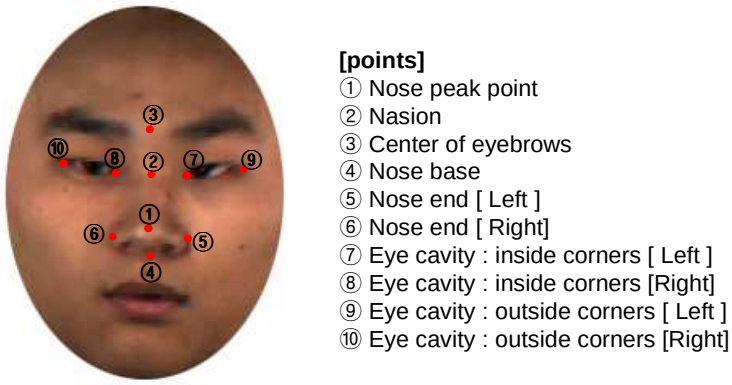


Figure 3. 5 Well extracted generic-facial feature points.

Table 3. 2 Relative feature calculation

	Relational features
Length	**L1=|①-②|, L2=|②-⑤|, L3=|②-⑥|, L4=|①-⑤|, L5=|①-⑥|, L6=|⑤-⑥|, L7=|⑦-⑧|, L8=|③-④|, L9=|⑤-⑦|, L10=|⑥-⑧|**
Ratio	**R1= L1 / L6, R2= L7 / L8, R3= L1 / L8**
Angle	**A1=∠⑤②⑥, A2=∠⑤①⑥, A3=∠⑤⑦②, A4=∠⑥⑧②, A5=∠⑦①⑧,**

3.3.5 Relative features calculation

In this section, we propose some relative features which are composed of previously extracted feature points such as the distance between points and the ratio of two distances and the angle among three feature points. These relative features can well distinguish the individual classes, because relations of 3D data values can produce more distinguishable features.

Each circled number from ① to ⑩ marked in Fig. 3.5 are ten extracted feature points. Each feature point has x, y and z coordinate values. As an example, we define ① feature points as ①(x_1, y_1, z_1). Table3.2. describes the equations to obtain the relative features. L1 to L10 are the relative lengths of facial feature point. R1 to R3 represent the ratio of two distances. In addition, A1 to A5 are the angles among three feature points. For example, L1 is the relative distance between ①(NPP) and ②(NBRP), which is the length of the nose bridge. R1 is the ratio of two distances, L1, L6. L6 denotes the distance between two nose bases. Thus, R1 is the ratio of the length of the nose bridge and the distance between two nose bases. A1 is the angle between L2 and L3. All relative features are carefully chosen to avoid an overlap and to

represent relations of 3D face data values.

3.3.6 Head pose estimation and feature normalization

In most problems of face recognition, the face is considered from frontal view and the effect of head rotation is ignored. Though 3D head pose estimation and normalization in 3D space are important, they have not been considered in the previous research on 3D face recognition. It is obvious that accurate head pose estimation not only enables robust face recognition but also critically determines the performance of the face recognition system. In this section, we describe an adopted 3D head pose estimation algorithm which uses 3D facial features. We use feature point vectors for calculating the initial head pose of the input face based on the Singular Value Decomposition (SVD) method. The SVD method uses a 3D-to-3D feature point correspondence matching scheme to estimate the rotation angle relative to the average frontal 3D face in the database with respect to minimizing the error of all extracted facial feature points [43-44]. Although the error is minimized by the SVD method, it still has some errors which may cause serious problems in face recognition. Error Compensated SVD (EC-SVD) compensates for the rest of the errors which were not yet recovered from the SVD method [8]. The EC-SVD procedure compensates for the error rotation angles for each axis after acquisition of the initial head pose from the SVD method. It geometrically fits the feature points in the 3D normalized face space using an error rotation matrix for each axis. In this method, we build a complete rotation matrix R which consists of six rotation matrices for each axis. They are three rotation matrices from three SVD methods and three error rotation matrices for each axis. We compensate for the error rotation matrix in the order of the X, Y, and Z axis, and this is the reverse order of the forward rotation of an input image. We independently calculate the error rotation matrix for the X and Y axis using the geometrical location of the NPP in the 3D normalized face space, and finally calculate the error angle between a face normal vector which is made up of extracted facial features and the Y axis to

compensate for the error angle of the Z axis. As we estimate the rotation angles based on the EC-SVD for the X, Y and Z axis, respectively, the feature vectors extracted form 3D faces in the database are rotated based on the angles that we have estimated. Then we can generate the pose estimated feature vectors of all faces in the database in the same way as probe data is generated. Therefore, we have a probe feature vector which can be compared with the pose estimated feature vectors which are already in the database.

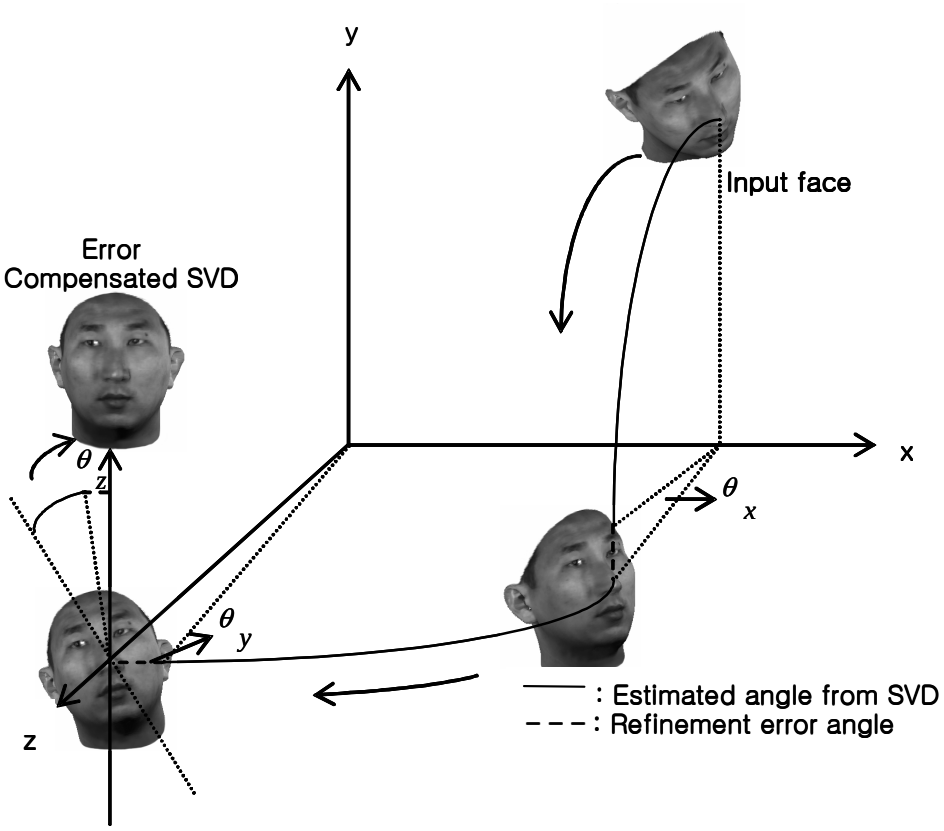


Figure 3. 6 Error Compensated SVD procedure

A. Extracting facial shape indexes

As shown in Fig. 3.5, we first extract feature points related to the position of facial components such as the eyes and nose. In order to estimate curvatures, we fit a quadratic surface to a local M×M window called *Monge patch* and use the least squares method and differential geometry to estimate parameters of the quadratic surface such as surface normal, Gaussian and mean curvatures principal curvatures.

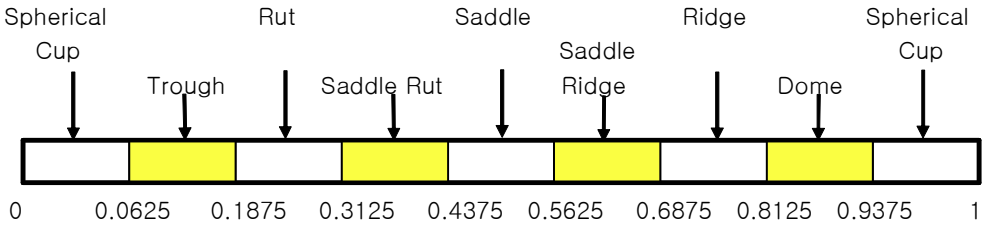


Figure 3. 7 Nine well-known shape types and their locations on the S_i scale

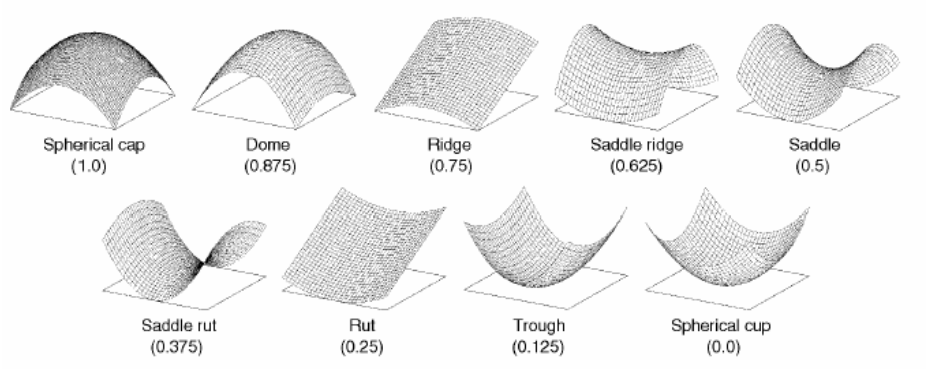


Figure 3. 8 Nine representative shapes on the S_i scale

A quadratic surface to a local M×M window is defined as follows.

$$S = \left\{ (x, y, z) :\ z = f(x, y),\ (x, y) \in D \subseteq R^2 \right\} \tag{3.2}$$

$$f(x, y) = ax^2 + by^2 + cxy + dx + ey + f \tag{3.3}$$

Shape index $S_i(p)$, a quantitative measure of the shape of a surface point *p*, is defined as follows.

$$S_i(p) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)}, \tag{3.4}$$

where $k_1(p)$ and $k_2(p)$ are maximum and minimum principal curvatures, respectively. These shape indexes are in the range of [0, 1]. As we can see from Fig. 3.7 and Fig. 3.8, there are nine well-known shape categories and their locations on the shape index scale [49]. Among those shape indexes, we select the extreme concave and convex points of curvatures as feature points. These feature points are distinctive for recognizing faces because face curvatures are intrinsic. Therefore, we select those shape indexes as feature points, $feature_i(p)$, if a shape index $S_i(p)$ satisfies the following condition.

$$feature_i(p) = \begin{cases} \partial \le S_i(p) < 1, & concavity \\ 0 < S_i(p) \le \beta, & convexity \end{cases}, \tag{3.5}$$

where $0 < \partial, \beta < 1$.

3.4 3D face recognition

In this section, we describe the use of feature vectors in face recognition which are extracted from the 3D face database, in order to compare them with that of a probe image. As we estimate the rotation angles based on EC-SVD method for the X, Y and Z axis, respectively, the features extracted from 3D faces in the database are rotated based on the angles that we have estimated. Then we have the pose estimated feature vectors of all faces in the database. That is, we have a probe feature vector which can be compared with the pose estimated feature vectors which are already in the database.

For face recognition, we have three face recognition strategies, such as SVM and ICA and weighted distance matching. We will describe in detail, respectively.

3.4.1 Face recognition using support vector machine (SVM)

In this section, we introduce face recognition algorithm using SVM (Support Vector Machine). A SVM has been recently proposed as a new technique for 3D pattern recognition [51-55]. Intuitively, given a set of points which belong to either of two classes, a SVM finds the hyperplane leaving the largest possible fraction of points of the same class on the same side, while maximizing the distance of either class from the hyperplane. According to [51-53], given fixed but unknown probability distributions, this hyperplane called OSH (Optimal Separating Hyperplane) - minimizes the risk of misclassifying not only the examples in the training set but also the yet-to-be-seen examples of the test set [54, 55].

A. Support Vector Machine

For a two-class classification problem, the goal is to separate the two classes by a function which is induced from available examples. Consider the examples in Fig. 3.9(a), where there are many possible linear classifiers that can separate the data, but there is only one (shown in Fig. 3.9(b)) that maximizes the margin (the distance between the hyperplane and the nearest data point of each class). This linear classifier is termed the OSH. Intuitively, we would expect this boundary to generalize well as opposed to the other possible boundaries as shown in Fig. 3.9(a). Consider the problem of separating the set of training vectors belong to two separate classes, $(X_1, y_1), ...(X_l, y_l)$, where $X_i \in R^n$, $y_i \in \{-1, +1\}$ with a hyperplane

(a) (b)

Figure 3. 9 Classification between two classes using hyperplane: (a) Arbitrary hyperplanes l, m and n. (b) The optimal separating hyperplane with the largest margin identified by the dash lines, passing the two support vectors.

$WX + b = 0$. The set of vectors is said to be optimally separated by the hyperplane if it is separated without error and the margin is maximal. A canonical hyperplane has the constraint for parameters W and b : $\min_{x_i}, y_i(W \cdot X_i + b) = 1$.

A separating hyperplane in canonical form must satisfy the following constraints,

$$y_i[(W \cdot X_i) + b] \geq 1, \quad i = 1, ..., l \tag{3.6}$$

The distance of a point x from the hyperplane is,

$$d(W, b; X) = \frac{|W \cdot X + b|}{||w||}. \tag{3.7}$$

The margin is $\frac{2}{||w||}$ according to its definition. Hence the hyperplane that optimally separates the data is the on that minimizes

$$\Phi(W) = \frac{1}{2}||w||^2 . \tag{3.8}$$

The solution to the optimization problem Eq.(3.8) under the constraints of Eq.(3.6) is given by the saddle point of the Lagrange functional.

$$L(W, b, a) = \frac{1}{2}||w||^2 - \sum_{i=1}^{l} a_i\{y_i[(W \cdot X_i) + b] - 1\}, \tag{3.9}$$

where a_i are the Lagrange multipliers.

The lagrangian has to be minimized with respect to W , b and maximized with respect to $a_i \geq 0$. Classical Lagrangian duality enables the primal problem Eq.(3.9) to be transformed to its dual problem, which is easier to solve. The dual problem is given by,

$$\max_a W(a) = \max_a \{\max_{w,b} L(W, b, a)\} \tag{3.10}$$

The solution to the dual problem is given by

$$\bar{a} = \arg\max_{a} \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} a_i a_j y_i y_j X_i \cdot X_j \tag{3.11}$$

with constraints,

$$a_i \geq 0, \quad i = 1, ..., l \tag{3.12}$$

$$\sum_{i=1}^{l} a_i y_i = 0 \tag{3.13}$$

Solving Eq.(3.11) with constraints Eq.(3.12) and Eq.(3.13) determines the Lagrange multipliers, and the OSH is given by,

$$\overline{W} = \sum_{i=1}^{l} \overline{a}_i y_i X_i , \tag{3.14}$$

$$\bar{b} = -\frac{1}{2} \overline{W} \cdot [X_r + X_s] , \tag{3.15}$$

where X_r and X_s are support vectors, satisfying

$$\overline{a_r}, \overline{a_s} > 0 , \quad y_r = 1, \quad y_s = -1 . \tag{3.16}$$

For a new data point X, the classification is then

$$f(X) = sign(\overline{W} \cdot X + \bar{b}) . \tag{3.17}$$

B. Multi-class recognition

Previous subsection describes the basic theory of SVM for two class classification. A multi-class pattern recognition system can be obtained by combining two class SVMs. Usually there are two schemes for multi-classification. One is the one-against-all strategy to classify between each class and all remaining. The other is the one-against-one strategy to classify

Table 3. 3 One against all SVM for solving the multi-class problem

	Sa	Sb	Sc	Sd
a	+1	-1	-1	-1
b	-1	+1	-1	-1
c	-1	-1	+1	-1
d	-1	-1	-1	+1

between each pair. For one-against-one method, a bottom-up binary tree is constructed for classification. Suppose there are eight classes in the data set, the number denotes the class label. By comparing between each pair, one class label is chosen representing the 'winner' of current two classes. The selected classes (from the lowest level of the binary tree) will come to the upper level for another round of tests. Finally, the unique class will appear on the top of the tree.

For one-against-all method, the training samples are taken with the same label as one class and the others as another class as shown Table 3.2. Then it becomes a two-class problem. The first labeled samples can be classified by SVM_1, and the *i*th labeled ones by SVM_i. For the n class problem, n SVM classifiers are formed and denoted by SVM_i, $i = 1, 2, \cdots n$. As for the testing sample *x*, $d_i(x) = w_i \cdot b_i$ can be obtained by using SVM_i, it belongs to *j*th class if there exists $d_j(x) = \max d_i(x)$

C. Matching by Support Vector Machine (SVM)

In order to utilize facial feature vectors, we feed the extracted facial feature points into SVMs. A SVM requires more feature vectors or intensity values of the gallery images per a

class for training support vectors in order to find the optimal hyperplane, because the more facial feature vectors per class are used, the better recognition rate is performed. However, in our face recognition system, only one facial feature vector of a gallery data per class is used for classification. Advantages of using this SVM are that training support vectors which have a gallery data per class is accomplished in real time and a SVM can find the optimal hyperplane among one and the others by using well defined facial features of 3D data as an input to the SVM.

In order to solve multi-class problem using a SVM which can solve only two-class problem with a linear hyperplane, we induce the one-vs-all strategies as shown in Table 3.2. There are 4-class (a, b, c, d), each S_i is called support vectors. It separates one class from the other classes by positive value +1 and negative value -1.

Given a set of q people and a set of q SVMs, each one associated to one person, the class label y of a face pattern x is computed as follows,

$$y = \begin{cases} n & if\ d_n(x) + t > 0 \\ 0 & if\ d_n(x) + t \le 0 \end{cases}, \tag{3.18}$$

where $d_i(x)$ is computed according to Eq. (3.7) and $d_n(x) = \max\{d_i(x)\}_{i=1}^{q}$.

When a probe image is used for the test, facial shape indexes of a probe image are directly fed into 300 SVMs during the learning, and a SVM having the highest value of output values from each SVM is ether matched or not matched with target class.

In order to find $\bar{a}$, solution of the dual problem as mentioned before, we induce polynomial kernel function $K(x, x_i)$ as defined in Eq.(3.19).

$$K(x, x_i) = (x^T x_i + 1)^p, \tag{3.19}$$

where *p* is degree.

3.4.2 Face recognition using Independent Component Analysis (ICA)

In this section, we introduce the face recognition algorithm using ICA(Independent component analysis). The general framework for independent component analysis was introduced by Jeanny Herault and Christian Jutten in 1986 and was most clearly stated by Pierre Comon in 1994.[56-57] Independent component analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals.

A. Independent component analysis (ICA)

The independence assumption applies to many cases and then the blind ICA separation of a mixed signal gives very good results. It is also used for signals that are not supposed to be generated by a mixing for analysis purposes. The statistical method finds the independent components by maximizing the statistical independence of the estimated components. Non-Gaussianity, motivated by the *central limit theorem*, is one method for measuring the independence of the components. Non-Gaussianity can be measured, for instance, by kurtosis or approximations of negentropy.[57]

Typical algorithms for ICA use centering, whitening and dimensionality reduction as preprocessing steps in order to simplify and reduce the complexity of the problem for the actual iterative algorithm. Whitening and dimension reduction can be achieved with principal component analysis or singular value decomposition. Algorithms for ICA include infomax, FastICA and JADE, but there are many others also.

The ICA method is not able to extract the actual number of source signals. The method is important to blind signal separation and has many practical applications.

B. Mathematical definitions

Linear independent component analysis can be divided into noiseless and noisy cases, where noiseless ICA is a special case of noisy ICA. Nonlinear ICA should be considered as a separate case.

B.1 General definition

The data is represented by the random vector $x = (x_1, \ldots, x_m)$ and the components as the random vector $s = (s_1, \ldots, s_n)$. The task is to transform the observed data x, using a linear static transformation $s = Wx$, into maximally independent components s measured by some function $F(s_1, \ldots, s_n)$ of independence.

B.2 Generative model

B.2.1 Linear noiseless ICA

The components x_i of the observed random vector $x = (x_1, \ldots, x_m)^T$ are generated as a sum of the independent components s_k ,: $k = 1, \ldots, n$

$x_i = a_{i,1}s_1 + \ldots + a_{i,k}s_k + \ldots + a_{i,n}s_n$ () (3.20)

weighted by the mixing weights $a_{i,k}$

The same generative model can be written in vectorial form as

$$x = \sum_{k=1}^{n} a_k s_k \tag{3.21}$$

, where the observed random vector x is represented by the basis vectors. The basis vectors $a_k = (a_{1,k}, \ldots, a_{m,k})^T$ form the columns of the mixing matrix $A = (a_1, \ldots, a_n)$ and the generative formula can be written as

$$x = As, \tag{3.22}$$

where $s = (s_1, \ldots, s_n)^T$.

Given the model and realizations (samples) $x_1, \ldots, x_N$ of the random vector x, the task is to estimate both the mixing matrix A and the sources s.

The original sources s can be recovered by multiplying the observed signals x with the inverse of the mixing matrix $W = A^{-1}$, also known as the unmixing matrix. Here it is assumed that the mixing matrix is square $(n = m)$.

B.2.2 Linear noisy ICA

With the added assumption of zero-mean and uncorrelated Gaussian noise $n \sim N\left(0, diag\left(\sum\right)\right)$, the ICA model takes the form

$$x = As + n. \tag{3.23}$$

B.2.3 Nonlinear ICA

The mixing of the sources does not need to be linear. Using a nonlinear mixing function $f\left(\cdot \mid \theta\right)$ with parameters θ the nonlinear ICA model is

$$x = f(s \mid \theta) + n . \tag{3.23}$$

C. Matching by Independent Component Analysis(ICA)

Like SVM, we feed the extracted facial feature points into ICA. ICA also requires enough feature vectors for training. We apply 300 classes out of 3D face database for the experiments. No dimension reducing by PCA(Principal Component Analysys) is required like 2D face recognition system using ICA. The proposed feature vector has fixed dimension. In order to measure the independence of the components, non-Gaussianity is measured by kurtosis. We induce Kurtosis function $kurt(y)$ as defined in Eq.(3.24).

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 , \tag{3.24}$$

where y is a linear combination of source s, with weights z.

3.4.2 Face recognition using weighted vector distance matching

A. Discriminating power estimation of each feature

Discriminating power estimation of each feature Ψ . has been computed using the Fisher coefficient [11], which represents the ratio of between-class variance to within-class variance, according to the formula $J(x)$

$$J(x) = \frac{\sum_{i=1}^{c} (m_i - m)^2}{\sum_{i=1}^{c} \frac{1}{n_i} \sum_{x=\Psi_i} (x - m_i)^2} \tag{3.25}$$

where c is the number of classes or subjects, Ψ_i is the set of feature values for a class i, n_i is the size of Ψ_i, m_i is the mean of Ψ_i, and m is the global mean of the feature over all classes. There are 300 classes corresponding to the number of distinct subjects.

Although there are seven images per person, when computing the Fisher coefficients, only the 3D facial images having the whole regions and consequently the whole features correctly extracted have been employed. When a feature can not be computed (always be cause of the non-existence of the regions from which it is derived), it is zero valued, except when a symmetric feature exists (then, the non-existent feature is valued like its symmetric feature). Extracted features have been normalized to values in the rank from 0 to 1.

B. Matching by weighted vector distance matching

The weighted summation of distances between a probe feature and each gallery feature is measured to perform matching process. The matching score S_i for a class i becomes

$$S_i = \sum_{x=\Psi}^{n} w(x) \times d(x) \tag{3.26}$$

,where Ψ is the set of feature values, n is the size of feature set Ψ, $w(x)$ is a weight value for a feature, $d(x)$ is a Euclidean distance of feature element x of probe and gallery feature vector. The weight $w(x)$ is decided by estimated discriminating power $J(x)$.

$$w(x) = \frac{J(x)}{\frac{1}{n}\sum_{x=\Psi}^{n} J(x)} \tag{3.27}$$

,where Ψ is the set of feature values and n is the size of feature set Therefore, a feature that has good discriminating power plays a more powerful role in a matching procedure.

Chapter 4.

Experiments

To verify the performance of the proposed 3D face recognition system using two different 3D data acquisition devices, we generate 3D face data, point cloud data and range image data as probe data by a structured light device and gallery images by a laser scanner respectively (named as Biometrics Engineering Research Center (BERC) face database) [58]. However, we need mandatory preprocessing and normalization procedures of obtained 3D face data in our experiments.

We first describe our experimental environments in section 4.1. We show the acquisition of 3D face data for probe and gallery after preprocessing and normalization in section 4.2. In section 4.3, we illustrate feature extraction results. In section 4.4, the recognition results of weighted vector distance matching with geometrically localized facial feature vector are described and these results are compared with the experimental results of SVM and ICA.

4.1 Experimental environments

In our experiment we utilize Visual studio 6.0 C++ programming tool to implement our face recognition system and OpenGL programming for rendering 3D face data.

We acquire 3D probe images by the structured light device, 3D FaceCam and 3D face database by using the laser scanner, Cyberware Model 3030/RGB as shown in Table 4.1. In order to reduce the illumination effect, we experiment at the restricted environments in the BERC. We use a white bathing cap on the hair as to avoid the large number of 3D vertex points and to extract the face region from the head.

Table 4. 1 Experimental environments of PC and programming.

CPU		Pentium 4 3.0GHz
Memory		512MB
Operating system		Window 2000 professional
Experiment tools	Rendering 3D data	OpenGL programming
	Face recognition algorithm	Visual studio 6 C++

4.2 3D face acquisition

In our experiments, we use BERC face database which contains one gallery image and seven pose-changed probe images from 300 people. As the first preprocessing stage of obtaining 3D face data suitable for face recognition, we eliminate the white bathing cap on hair region, which is unnecessary part for face recognition, and extract only the face region as shown in Fig 4.1.

After normalizing 3D face data, we generate the resulting 3D face images. These 3D images which have only vertex points which are presented as depth information. As shown in Fig. 4.2, we acquire 3D face models and point cloud data for different head poses. We acquire point cloud data of 7 head poses per person, such as frontal, ±15 and ±30 for the Y axis, ±15 for the X axis as probe images. In other words, a total of 2100 images are tested in order to recognize faces in pose varying environments. From here onwards, we will deal with these 3D face data for extracting features and recognizing faces.

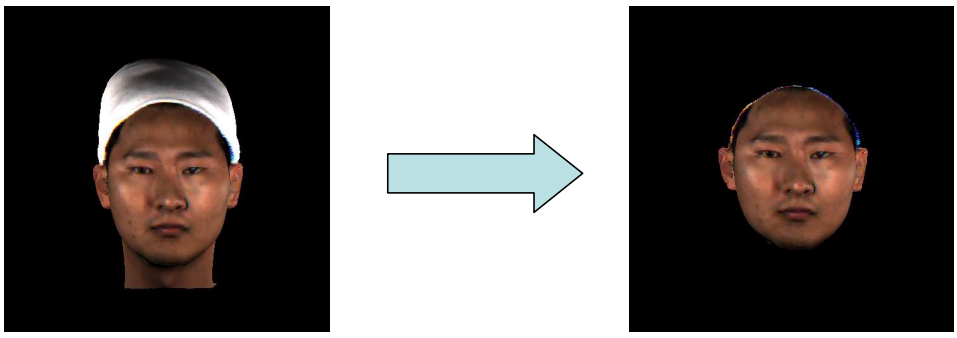
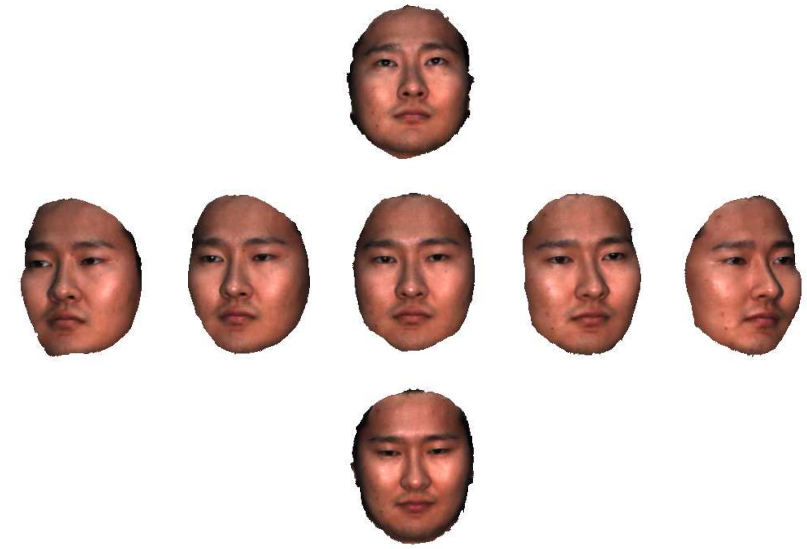
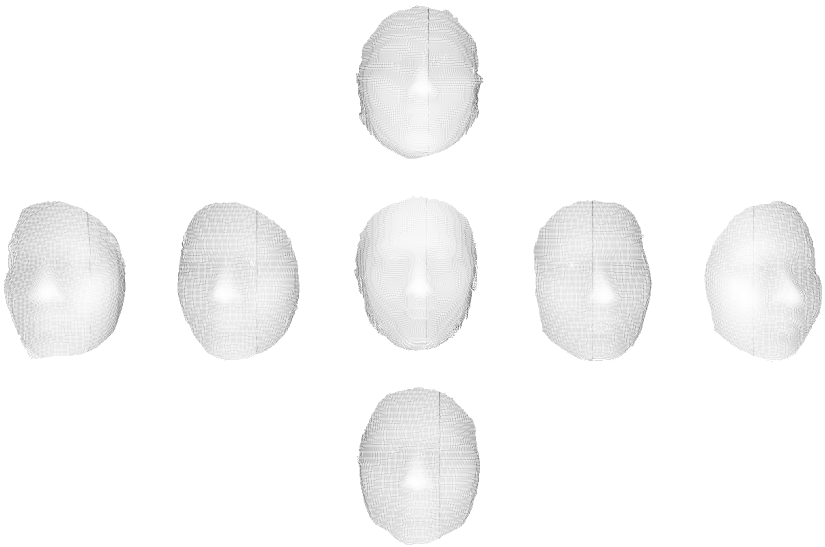


Figure 4. 1 Eliminating the hair region



(a)



(b)

Figure 4. 2 3D point cloud models and range images acquired from 3D FaceCam ; (a) point cloud models, (b) range images

4.3 Experimental results of feature extraction

We have defined general facial features of a human face and have described the proposed feature extraction in chapter 3. The Experimental results of the proposed feature extraction algorithm are reported in this section. Table 4.2 shows the feature extraction rate. Almost all of the features are well extracted when feature1 (NPP) is successfully extracted. However, feature 9 and feature 10 (eye cavity: outside corner points) are not extracted well when the data have pose variation with ±30 degrees. We select the relational features that are not related to feature 9 and feature 10 so that relational features extraction are not affected by them. The average feature extraction rate is 99.5%.

Table 4.3 shows the comparisons of x, y and z values of ten features about a gallery image and a probe image for the same subject. x, y and z values of database (029) as figured in Table 4.3 (a) are very similar to probe (029) in Table 4.3 (b), but probe (045) in Table 4.3 (b) is different and distinguishable from database (029). We normalize all the face data for NPP to be located on 0, 0 and 100 for x, y and z coordinates respectively. This means that all the face data are in the same space of the view port.

Table 4. 2 Feature extraction rate

Data	Probe							Gallery
	Right		Center	Left		Up	Dn	
Pose	**-30**	**-15**	**0**	**+15**	**+30**	**-15**	**+15**	**full**
Feature 1	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Feature 2	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Feature 3	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Feature 4	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Feature 5	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Feature 6	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Feature 7	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Feature 8	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Feature 9	**98.6%**	**99.3%**	**99.6%**	**98.3%**	**85.3%**	**99.6%**	**99.6%**	**99.6%**
Feature 10	**86.6%**	**97.3%**	**99.6%**	**99%**	**99.3%**	**99.6%**	**99.6%**	**99.6%**

Table 4. 3 Extracted feature points without value normalization

(a) Extracted facial feature points of database data.

Features	Database data(029)		
	x	***y***	***z***
Feature 1	0.738	60.11	86.41
Feature 2	0.390	39.14	81.64
Feature 3	0	0	0
Feature 4	0.734	-10.48	86.35
Feature 5	21.21	0	78.29
Feature 6	-20.24	0	76.81
Feature 7	14.34	39.14	74.68
Feature 8	-14.29	39.14	73.94
Feature 9	39.24	37.62	74.23
Feature 10	-40.52	38.01	74.18

(b) The comparison of DB data with the different probe data

Features	Probe(029)			Probe(045)		
	x	***y***	***z***	***x***	***y***	***z***
Feature 1	0.490	60.34	86.21	0.6009	55.7911	78.6374
Feature 2	0.434	39.264	79.93	-0.6437	37.2333	76.4823
Feature 3	0	0	100	0	0	100
Feature 4	0.7052	-10.25	87.77	0.83728	-13.681	80.1466
Feature 5	21.222	-0.028	78.09	17.4555	0.19370	74.8431
Feature 6	-20.63	-0.497	75.72	-17.853	-0.4307	73.6493
Feature 7	14.096	39.116	74.91	11.4168	37.6236	70.2066
Feature 8	-14.57	38.970	72.12	-11.648	37.046	70.0277
Feature 9	39.12	38.14	73.93	32.2131	37.0026	71.0396
Feature 10	-40.21	38.22	73.02	-33.183	37.751	69.5762

4.4 Experimental results of face recognition

4.4.1 Weighted Vector Distance matching

A. Discriminating power estimation

The extracted feature vector consists of 38 features. Fig. 4.3 shows the computed discriminating power of each feature. It is obvious that the first five features have much stronger discrimination power than the others. Based on the estimated discriminating power, weight values are decided. Fig 4.4 shows the weight values. Weight values are normalized one of the estimated discrimination power.

Table 4.4 relates the Fisher coefficients corresponding to each feature rank position in the Fisher coefficient ordered list. R, L, S and P denote a ratio, a length, a shape index and point, respectively. Experiment shows that ratio feature 1 (R1) is the most powerful one. R1 is the ratio of the length of the nose bridge and the distance between two nose bases. Its Fisher coefficient is 144.4.

B. Recognition result of weighted vector distance matching

In order to measure the distance between two feature vectors, the dimensions of feature vectors should agree. However, the dimensions of feature vectors do not always agree because of a failure of the feature extraction. The proposed matching method overcomes disagreement of vector dimension by ignoring the missing elements. Fig 4.5 shows the results of matching experiment. The experiment was performed in three ways with three different feature sets. One set uses only feature points and relative features. Another set uses only shape features.

The remaining set uses the whole features, feature point, relative features and shape indexes. The identification method of experiments follows the three-step process. One's feature set is compared to the others in the system's database and a similarity score is developed for each comparison. These similarity scores are then numerically ranked so that the highest similarity score is first. In an ideal operation, the highest similarity score is the comparison of that one's recently acquired feature set with that one's feature in the database. The percentage of times that the highest similarity score is the correct match for all individuals is referred to as the "top match score." A plot of top match score versus probability of correct identification is called a Cumulative Match Score. Fig. 4.5 and Fig. 4.8 show recognition result using Cumulative Match Score. According to Fig. 4.5, the last experiment gives the best performance. Its top rank recognition rate is 97.6%. As we mentioned before, both geometrical position of facial components and their shapes are important characteristics that distinguish a face from the others. They are also non-independent features. Using both of them can represent a face distinctively.

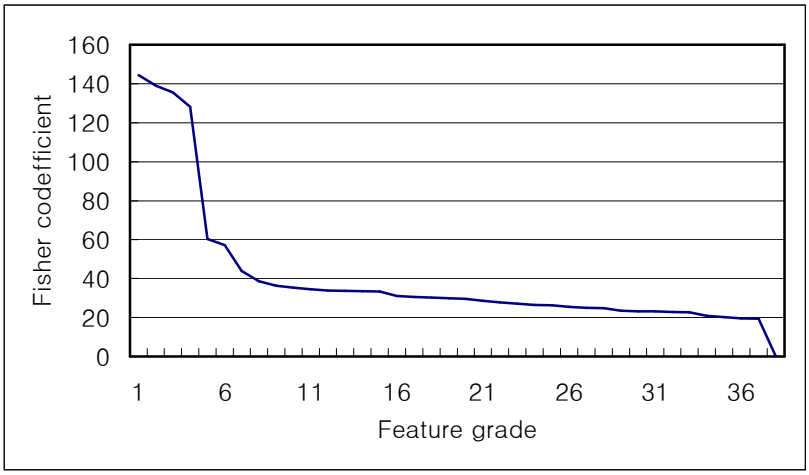


Figure 4. 3 feature power estimation

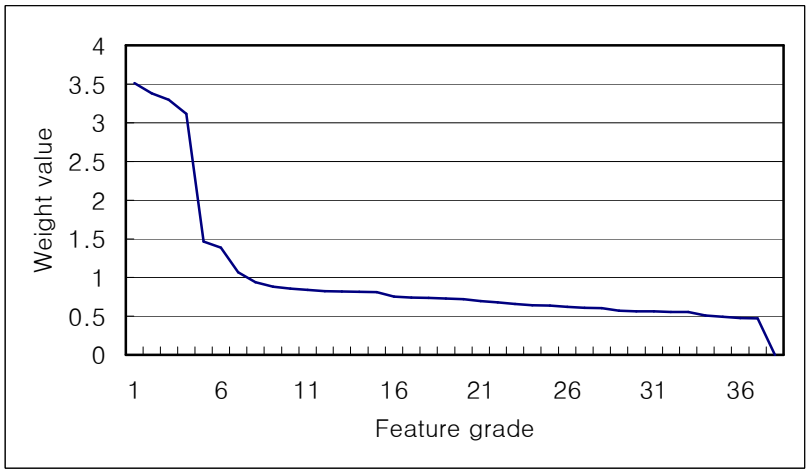


Figure 4. 4 weight values

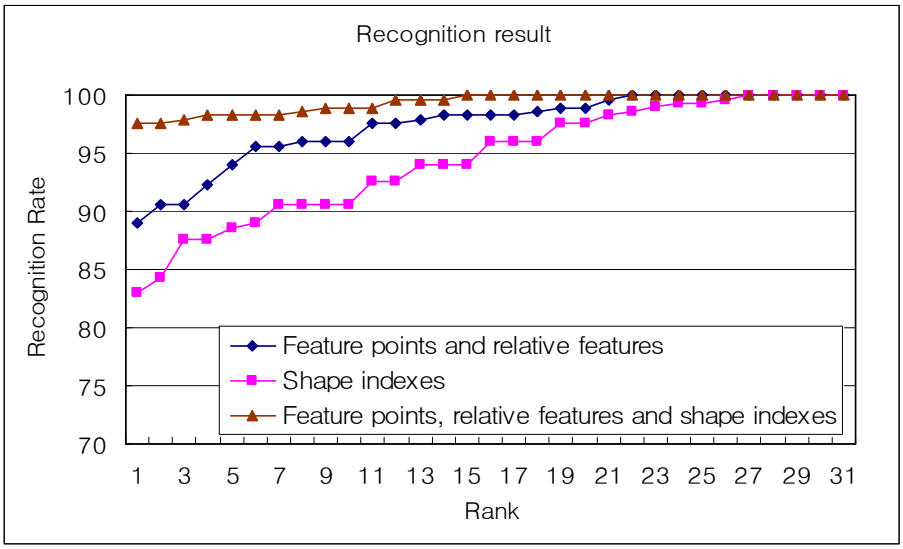


Figure 4. 5 face recognition rate according to feature

Table 4. 4 feature grade position with weight value

Grade	feature	Fisher coeff.	Weight Value	Grade	feature	Fisher coeff.	Weight value
1	R1	144.4	3.510685	**20**	S8	29.6	0.719642
2	R3	139	3.379399	**21**	L2	28.5	0.692898
3	S4	135.6	3.296737	**22**	L7	27.8	0.67588
4	R2	128.2	3.116827	**23**	L5	27.1	0.658861
5	A1	60.4	1.468458	**24**	L4	26.4	0.641843
6	L6	57.1	1.388228	**25**	L8	26.2	0.63698
7	S6	43.8	1.064875	**26**	A2	25.5	0.619962
8	S5	38.5	0.93602	**27**	P6	24.9	0.605374
9	A3	36.2	0.880102	**28**	P5	24.8	0.602943
10	S1	35.2	0.85579	**29**	P4	23.4	0.568906
11	A5	34.5	0.838772	**30**	P7	23.2	0.564044
12	A4	33.8	0.821753	**31**	P8	23.1	0.561612
13	S3	33.7	0.819322	**32**	S9	22.8	0.554319
14	L1	33.4	0.812028	**33**	S10	22.7	0.551887
15	L10	33.3	0.809597	**34**	P3	20.9	0.508125
16	S2	31	0.753679	**35**	P2	20.2	0.491107
17	L9	30.5	0.741523	**36**	P10	19.5	0.474088
18	L3	30.2	0.734229	**37**	P9	19.4	0.471657
19	S7	29.9	0.726935	**38**	P1	0	0

4.4.2 SVM

In this section experimental results for face recognition algorithm with feature-based SVM are shown. We use 300 people from BERC face database as gallery and probe images. We apply feature vectors of 300 people to a probe module of SVMs as mentioned in chapter 3. The dimension of the input of SVM should be fixed. We do not use feature 9 and feature 10 as inputs of SVM because they can not be always located. Eight-feature points, their relative features and their shape indexes are directly fed into the SVMs.

As SVM is one of the famous supervised learning algorithms, we have to align target values according to each class in order to train gallery face images. As mentioned before, the SVM solves the only two-class problem. Therefore, we induce one-vs-all strategy in order to solve the multi-class problem

According to the theory of SVM mentioned in section 3.4.1, we find the Lagrange multipliers for each shape feature by Eq.(3.11). For an example, Lagrange multipliers of the 23^{rd} and the 211^{th} classes, which are randomly selected samples, are shown in Table 4.5. Lagrange multipliers for the 23^{rd} class and other classes can be surely distinguishable in this table. When we classify 23^{rd} class with the other classes by using Lagrange multipliers: a_1 is 1.98165 and the rest of them are around zero as mentioned in Table 4.5, we have to find out the optimal weight scalar value through Eq.(3.14). The resulting weight value $\bar{W}$ is 3.95987. We can also find a bias value by solving equation depicted as $\bar{b} = \bar{w} \cdot x_i - y_i$. The bias value $\bar{b}$ is 0.981652.

In order to classify the first class and the second class by using the optimal hyperplane function, we need to solve Eq.(3.17) as mentioned in section 3.4.1. When the features of probe

face data are directly fed into SVM for testing face recognition, if the outcome of the sine function as defined in Eq.(3.17) is positive value +1, the probe face data is the same as the gallery face data, If the outcome is negative value -1, the probe face data is different from the gallery face data. As for testing our recognition system, when we put the 23^{rd} face and 211^{th} face for probe image into the system, respectively, the same face data among gallery faces has only a positive value and the other classes are negative values as shown in Fig.4.6 and Fig. 4.7.

Table 4. 5 Values of Lagrange multiplier for face class 23 and 211

Class number	Lagrange multiplier (For class 23)	Lagrange multiplier (For class 211)
1	0.021463	0.021423
2	0.021484	0.025631
3	0.020509	0.021943
4	0.019531	0.024484
5	0.021484	0.021272
:	**:**	**:**
20	0.019533	0.025494
21	0.019531	0.025483
22	0.019522	0.021666
23	**1.98165**	0.025484
:	**:**	**:**
210	0.019121	0.023542
211	0.021434	**1.93282**
212	0.020509	0.025531
:	**:**	**:**
298	0.020509	0.024431
299	0.019831	0.024733
300	0.019692	0.023513

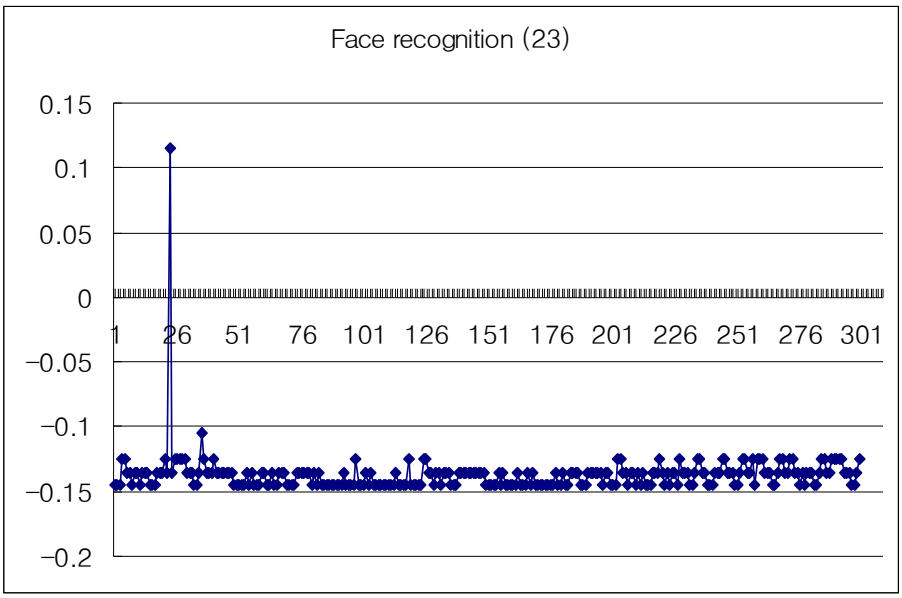


Figure 4. 6 Face recognition result for class 23

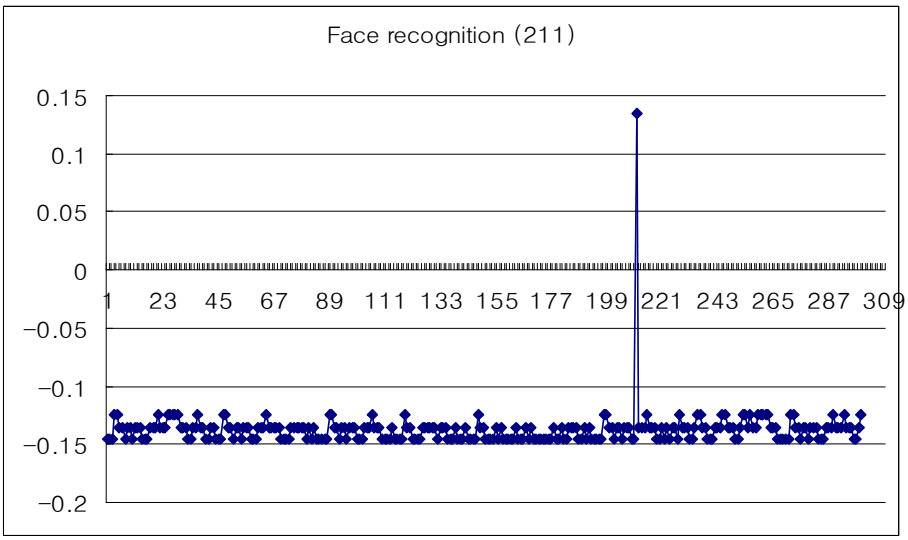


Figure 4. 7 Face recognition result for class 211

4.4.3 ICA

In this section, we show the experimental results for face recognition algorithm using ICA. We use 300 people of BERC face database as gallery and probe images. We apply feature vectors of 300 people to a probe module of ICA. We find the independent components by observing gallery data. The statistical method finds the independent components by maximizing the statistical independence of the estimated components. Non-Gaussianity is one method for measuring the independence of the components. Non-Gaussianity can be measured, for instance, by kurtosis or approximations of negentropy.

The dimension of the input of ICA should be also fixed. We don't use feature 9 and feature 10. Thus eight-feature points, their relative features and their shape indexes are projected onto directly the independent components. We use 30 independent components for experiment. No dimension reducing is process required. The dimension of the feature vector is already compact enough.

4.4.4 Comparing recogniton results

Finally, we compare the experimental results of each face recognition method. As we can see from Fig. 4.8, we have a 97.6% face recognition rate for the weighted distance matching, 98.6% at first top rank by the SVM and 97.3% by the ICA on average for seven different poses for 300 different people. Although SVM shows the highest recognition rate, the weighted vector distance matching shows not too lower recognition rate than SVM. Moreover, it doesn't need training and it is not a two-class classifier like SVM. From the experimental results, we have effectively utilized facial shape indexes, geometrical feature points and its relational features for pose invariant face recognition.

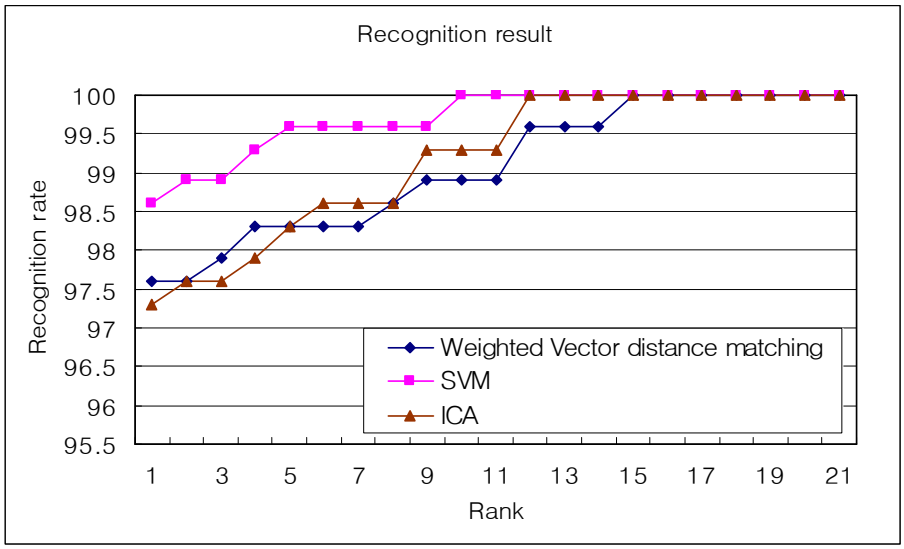


Figure 4. 8 Comparison of face recognition results

Chapter 5.

Conclusion

In this dissertation, we propose a pose invariant 3D face recognition system based on two different 3D face data acquisition devices for probe and gallery data, respectively. We utilize a 3D laser scanner to obtain the gallery data and a structured-light based 3D device for acquiring the probe data. These 3D face data have different number of vertex points, which consist of a face, a large amount of data and background noise. In order to solve these problems we need preprocessing for 3D data. After the preprocessing, we extract ten invariant feature points that are the most representatives of a face. We calculate relative features such as the distance and the ratio between points and the angle among feature points, which consists of previously extracted feature points. We also propose shape index features that represent the shapes of facial components. The relative features and shape indexes can well distinguish the individual classes having the extracted feature points well. In the recognition stage, we induce three different recognition algorithms, the weighted vector distance matching, a feature-based SVM and ICA. We obtain 97.6% face recognition rate for the weighted distance matching, 98.6% at first top rank by the SVM and 97.3% by the ICA on average under seven different head poses, respectively. Since the above results are based on 2100 images from 300 people in the BERC database with pose varying environments, we can see from these experimental results, it is highly acceptable for pose invariant face recognition. As a further study, we need to research a

feature selection method in order to find best feature set among all possible relative features and shape indexes. Moreover, we also need to test other classifiers with proposed feature vector with a fusion of 2D texture information. And 2D texture information can be used to find facial components. Finally, missing/no-missing features should be researched to make proposed method more invariant under pose varying environment.

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” *ACM, Computing Surveys*, Vol. 35, No.4, Dec. 2003.

[2] R. Chellappa, C. L Wilson, and S. Sirohey, “Human and machine recognition of faces: A survey,” *Proc. of the IEEE*, vol.83, pp.705-740, 1995.

[3] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi and M. Bone, “Face Recognition Vendor Test 2002,” *NIST Technical Report*, NIST IR 6965, March 2003.

[4] S. Rizvi, P. Phillips and H. Moon, “The FERET Verification Testing Protocol for Face,” *NIST Technical Report*, NIST IR 6281, October 1998.

[5] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, “The FERET Evaluation Methodology for Face-Recognition Algorithms,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 10, pp.1090-1104, Oct. 2000.

[6] K. W. Bowyer, K. Chang, and P. Flynn, “A Survey Of Approaches To Three-Dimensional Face Recognition,” *Proceedings of the International Conference on Pattern Recognition*, Vol.1, pp.358-361, 2004.

[7] Cyberware, “Head and face 3D color scanner” (Cyberware , Monterey, Calif. 2003), http://www.cyberware.com/products/psInfo.html.

[8] H. Song, S. Lee, J. Kim and K.Sohn, “3D sensor based face recognition,” *Applied Optics*, Vol. 44, No. 5, pp. 677-687, Feb. 2005.

[9] A. Pentland, B. Moghaddam and T. Starner, “View-based and Modular Eigenspaces,” *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp.84-91, 1994.

[10] M. Turk and A. Pentland, “Eigenfaces for Recognition,” *Journal of Cognitive Neuroscience*, Vol. 3, pp. 71-86, 1994.

[11] Belhumeur, P.N. Hespanha, J.P.; Kriegman, D.J., “Eigenfaces vs. Fisherfaces: recognition using class specific linear projection,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19 Issue 7, pp. 711-720, Jul 1997.

[12] Adini, Y., Moses, Y., Ullman, S., **“**Face recognition: the problem of compensating for changes in illumination direction,**”** *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19 Issue 7, pp. 721-732, July 1997.

[13] Jacobs, D.W., Belhumeur, P.N., Basri, R., “Comparing images under variable illumination,” *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 610-617, 1998.

[14] Belhumeur, P.N., Kriegman, D.J., “What is the set of images of an object under all possible lighting conditions?**,”** *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 270 -277,1996.

[15] Joseph J. Atick, Paul A. Griffin, A. Norman Redlich, “Statistical Approach to Shape from Shading: Reconstruction of 3D Face Surfaces from Single 2D Images,” *Neural Computation*, Vol. 8, pp. 1321-1340, 1996.

[16] Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J., “Illumination-based image synthesis: creating novel images of human faces under differing pose and lighting,” *Proceedings of the IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes*, pp.47 -54, 1999.

[17] Manjunath, B.S., Chellappa, R., Von der Malsburg, C., “A feature based approach to face recognition,” *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp.373 -378, 1992.

[18] Alferez, R., Yuan-Fang Wang, “Geometric and illumination invariants for object recognition,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 21, Issue 6, pp. 505-536, Jun. 1999.

[19] Akimoto, T., Suenaga, Y., Wallace, R.S., “Automatic creation of 3D facial models*; Computer Graphics and Applications, IEEE*, Vol. 13, Issue 5, pp. 16-22, Sep. 1993.

[20] Vetter, T. Poggio, T., “Linear object classes and image synthesis from a single example image,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, Issue 7, pp.733-742, Jul 1997.

[21] Wiskott, L., Fellous, J. M., Kuiger, N., von der Malsburg, C., ”Face recognition by elastic bunch graph matching,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, Issue 7, pp. 775-779, Jul 1997.

[22] Freeman, W.T., Tenenbaum, J.B., “Learning bilinear models for two-factor problems,” *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp.554 -560, 1997.

[23] C. Hesher, A. Srivastava, and G. Erlebacher, “A novel technique for face recognition using range images,” *Proceedings of the 7th International Symposium on Signal Processing and Its Applications*, 2003.

[24] F. Tsalakanidou, D. Tzovaras and M. Strintzis, “Use of Depth and Color Eigenfaces for Face Recognition,” *Pattern Recognition Letters*, Vol. 24, pp. 1427- 1435, 2003.

[25] B. Achermann, X. Jiang, and H. Bunke, “Face recognition using range images,” *Proceedings of the International Conference on Virtual Systems and MultiMedia*, Geneva, Switzerland, pp.129-136, 1997.

[26] K. Chang, K. Bowyer and P. Flynn, “Face Recognition Using 2D and 3D Facial Data,” *2003 Multimodal User Authentication Workshop*, pp. 25-32, Dec. 2003.

[27] H. T. Tanaka, M. Ikeda and H. Chiaki, “Curvature-based face surface recognition using spherical correlation,” *Proceedings of the 3th International Conference on Automatic Face and Gesture Recognition*, pp.372-377, 1998.

[28] J.C. Lee and E. Milios, “Matching range image of human faces,” *Proceedings of the International Conference on Computer Vision*, pp. 722-726, 1990.

[29] G. G. Gordon, “Face recognition based on depth maps and surface curvature,” *SPIE Proceedings : Geometric Methods in Computer Vision*, San Diego, CA, Proc. SPIE 1570, 1991.

[30] A. Moreno, A. Sanchez, J. Velez and F. Diaz, “Face Recognition using 3D Surface-extracted Descriptors,” *Proceedings of Irish Machine Vision and Image Processing Conference*, September 2003.

[31]31 R. Duda and P. Hart, “Pattern Classification and Scene Analysis,” *New York: Wiley and Sons*, 1973.

[32] Y. Lee, K. Park, J. Shim and T. Yi, “3D Face Recognition using Statistical Multiple Features for Local Depth Information,” *International Conference on Multimedia and Expo. ICME 03’*, Vol.3, pp. 133-136, 2003.

[33] Y. Wang, C. Chua and Y. Ho, “Facial Feature Detection and Face Recognition from 2D and 3D Images,” *Pattern Recognition Letters*, Vol. 23, pp. 1191-1202, 2002.

[34] X. Lu, D. Colbry and A. Jain, “Three Dimensional Model-Based Face Recognition,” *Proceedings of the International Conference on Pattern Recognition*, pp. 362-366, 2004.

[35] X. Lu, D. Colbry and A. Jain, “Matching 2.5D Scans for Face Recognition,” *Proceedings of the International Conference on Biometric Authentication (ICBA)*, pp. 30-36, 2004.

[36] P. Besl and N. McKay, “A Method for Registration of 3D Shapes,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp. 239-256, 1992.

[37] Y. Chen and G. Medioni, “Object Modeling by Registration of Multiple Range Images,” *Proceedings of the International Conference on Robotics and Automation*, 1991.

[38] V. Blanz, and T. Vetter, “Face recognition based on fitting a 3D morphable model,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063-1074, Sep. 2003.

[39] P. Phillips, P. Grother, R. Michaels, D. Blackburn, E. Tabassi and J. Bone, “FRVT 2002: A Overview and Summary,” http://www.frvt.org/FRVT2002/ documents.htm, Mar. 2003.

[40] Genex Technologies, Inc., “3D FaceCam” (Genex Technologies, Inc. Kensington, Md., 2003), htttp://www.genextech.com/products_services/ rainbow / facecam.html.

[41] Peter L. Hallian “Two-and Three Dimensional patterns of the Face,” A K Peters LTD, pp.202-203.

[42] Maja Pantic, “Automatic Analysis of Facial Expressions: The state of the Art”, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, pp.1424-1445, Dec. 2000

[43] R.M. Haralick, H.N. Joo, C.N. Lee, X. Zhuang, V.G. Vaidya, and M.B. Kim, “Pose estimation from corresponding point data,” *IEEE Trans. on Systems, Man and Cybernetics* vol. 19, no. 6, pp. 1426-1446, 1989.

[44] T.S. Huang, A.N. Netravali, “Motion and structure from feature correspondences: A Review,” *Proceedings of the IEEE*, vol. 82, no. 2, pp. 252-268, 1994.

[45] B.K.P. Horn, “Closed-Form solution of absolute orientation using unit quaternion,” *J. Optical Soc. Am.*, vol. 4, pp. 629-642, Apr. 1987.

[46] B.K.P. Horn, H.M. Hilden, and S. Negahdaripour, “Closed-form solution of absolute orientation using orthonormal matrices,” *J. Optical Soc. Am.*, vol. 5, pp. 1127-1135, 1988.

[47] K.S. Arun, T.S. Huang, and S.D. Blostein, “Least-squares fitting of two 3D point sets,” *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 9, no. 5, pp. 698-700, Sep. 1987.

[48] M. Bichsel and A. Pentland, “Human face recognition and face image set’s topology,” *CVIP:Image Understanding*, vol. 59, pp.254-261, March 1994.

[49] G. G. Gordon, “Face recognition based on depth and curvature features,” *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 808-810, 1992.

[50] C. Dorai and A. K. Jain, “COSMOS-A Representation Scheme for 3D Free-Form Objects,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 10, pp. 1115-1130, Oct. 1997.

[51] H.Murase and S.K. Nayar,”Visual Learning and Recognition of 3-D Object from Appearance,” *Int. J. Comput. Vision*, vol.14, pp.5-24, 1995.

[52] E.Osuna, R.Freund and F.Girosi, “Training Support Vector Machines: an Applications to Face Detection,” *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp.130-136, 1997.

[53] B.Heisele, P.Ho, and T.Poggio.” Face Recognition with Support Vector Machines: Global Versus Component-based approach”, *Proceedings of the International Conference on Computer Vision*, vol.2, pp.688-694, 2001.

[54] Guodong Guo, Stan Z.Li, and Kapuk Chan, “Face recognition by Support Vector Machines,” *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pp.196 – 201, Mar. 2000.

[55] M.Pontil and A.Verri. “Support Vector Machines for 3D object recognition.” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.20, pp.637-646, 1998.

[56] J. H'erault and C. Jutten, "Space or time adaptive signal processing by neural networks models," *in Intern. Conf. on Neural Networks for Computing, Snowbird* (Utah, USA), pp. 206-211, 1986.

[57] P. Comon, "Independent component analysis - a new concept," *Signal Processing* vol. 36, pp.287-314, 1994

[58] BERC, 3D Face database, (Biometrics Engineering Research Center at Yonsei University, 2002), “http://berc.yonsei.ac.kr”.

[59] K. Chang, K. Bowyer and P. Flynn, “An evaluation of Multimodal 2D+3D Face Biometrics,” *PAMI*, vol. 27, pp. 619-624, 2005.

[60] H. Song, S. Lee, J. Kim and K. Sohn, "3D Sensor Based Face Recognition," Applied Optics, vol. 44, no. 5, pp 677-687, Feb. 2005.

[61] Y. Lee, H. Song, and K. Sohn, “Local feature based 3D face recognition,” AVBPA 2005, LNCS , pp 909-918 , July. 2005.

Summary in Korean

얼굴 특징 영역에서의 표면곡률지수를 이용한 3차원 얼굴인식

본 논문에서는 얼굴 표면곡률지수(surface shape indexes)를 이용한 3차원 얼굴인식 기술을 제안한다. 눈, 코, 입과 같은 얼굴 구성요소의 구조적 위치 정보와 추출된 구성요소의 위치에서 얻어진 표면곡률지수를 이용하여 얼굴의 구조적 정보와 구성요소의 형태 정보를 동시에 포함하는 고정된 차원의 특징벡터(feature vectors)를 제공함으로써 개선된 분류성능과 정합의 편리성을 제공하는 3차원 얼굴인식 기술을 제안한다.

입력단과 DB단에서는 각각 획득된 얼굴 데이터의 전처리를 위하여 특징점을 추출하며, 추출된 특징점을 이용하여 머리 부분을 제거하고 입력과 DB얼굴 데이터가 같은 공간 상에 놓이도록 얼굴포즈보정 및 정규화 과정을 수행한다.

특징 추출단계에서는 얼굴의 기하학적 정보를 바탕으로 하여 깊이 정보가 가장 큰 값을 갖는 코끝점을 중심으로 3개의 얼굴 윤곽 곡선들과 10개의 특징점들을 추출하게 된다. 또한, 추출된 10개의 특징점들간의 거리와 각도, 비율 등을 이용한

18개의 상관 특징들을 얻게 된다. 마지막으로 추출된 10개의 특징점 영역에서 얼굴표면곡률지수를 계산하여 얼굴인식을 위한 특징벡터(feature vector)를 구성한다.

본 논문에서는 얼굴인식 방법으로 가중치벡터거리정합 방법을 제안한다. 제안하는 방법의 결과는 기존의 방법인 SVM (Support Vector Machine)과 ICA (Independent Component Analysis)를 이용한 결과와 비교한다.

본 실험에서는 BERC(Biometrics Engineering Research Center)에서 제공하는 300명의 얼굴 데이터를 사용하였다. 가중치벡터거리정합 알고리즘 기반의 얼굴인식의 경우 정합을 수행한 결과 97.6%의 인식률을 보였다. SVM 기반 얼굴 인식의 경우에는 98.6%의 인식률을 보였다. 또한 ICA의 경우는 96.3%의 인식률을 보였다. SVM방식이 가장 좋은 인식률을 보이지만 가중치벡터거리정합 방식은 훈련과정이 없고, SVM이나 ICA보다 등록과정이 간단한 장점이 있다.

또한 가중치벡터거리정합 방법을 사용하여 데이터를 표면곡률지수만으로 구성된 특징벡터와 얼굴 구성요소의 위치와 상대적 정보만으로 이루어진 특징벡터로 나누어 인식을 수행한 결과 각각 83% 와 89%의 인식률을 나타냈다. 그러나 모든 특징을 다 사용하여 인식을 수행한 결과 97.6%의 인식률을 나타냈다. 실험을 통해 얼굴구성요소의 위치뿐만 아니라 형태 정보도 중요한 얼굴의 특징이라는 것을 보여준다.

핵심되는 말: 3차원얼굴인식, 얼굴표면곡률지수, 가중치벡터거리정합, support vector machine (SVM), independent component analysis (ICA)