GAZİANTEP UNIVERSITY GRADUATE
SCHOOL OF NATURAL & APPLIED SCIENCES

SHAPE-BASED HAND RECOGNITION

M. Sc. THESIS
IN
ELECTRICAL & ELECTRONICS ENGINEERING

BY
GULDEN ÖZBAY
JUNE 2008

Shape-Based Hand Recognition

M.Sc. Thesis
in
Electrical & Electronics Engineering
University of Gaziantep

Supervisor
Asst. Prof. Dr. Nurdal WATSUJI

by

Gülden ÖZBAY

June 2008

T. C.
GAZİANTEP UNIVERSITY
GRADUATE SCHOOL OF
NATURAL & APPLIED SCIENCES
ELECTRICAL & ELECTRONICS ENGINEERING

Name of the thesis: Shape-based hand recognition
Name of the student: Gülden Özbay
Exam date: June 13, 2008

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Sadettin ÖZYAZICI
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Gülay TOHUMOĞLU
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Nurdal WATSUJİ
Supervisor

Examining Committee Members	signature
....	______________________
....	______________________
....	______________________
....	______________________
....	______________________

ABSTRACT

SHAPE-BASED HAND RECOGNITION

ÖZBAY, Gulden
M.Sc. in Electrical & Electronics Eng.
Supervisor: Asst. Prof. Dr. Nurdal WATSUJI
June 2008, 79 pages

In this thesis, a new algorithm for personal recognition has been proposed. Personal recognition has been made using geometric features of the hand of the individuals. Hand geometry is a biometric that identifies individuals by the shape of their hands. In this study, a smart, simple and an effective algorithm is presented for both verification and identification of the people.

The proposed algorithm consists of three major parts: Segmentation for obtaining the hand contours, feature extraction and recognition. Segmentation part separates hand region from the background and extracts the hand and finger contours. Feature extractor finds the lengths of the fingers, finger widths at different heights of the fingers, and the width of the palm as a feature vector. Finally recognition part compares the feature vector obtained from the input image with the database and gives the result for identification or verification purposes.

This algorithm uses right hand images of the individuals for both verification and identification purposes. The necessary codes for the proposed algorithm were written in Matlab software. To see the overall performance, the proposed algorithm has been tested over a large number of images. It was obtained that the identification rate for this algorithm is 97.44 % and verification rate is 98.72 % giving the least error.

Key Words: biometrics, segmentation, hand geometry, smearing, template matching, recognition

ÖZET

ŞEKİL TABANLI EL TANIMA

ÖZBAY Gülden
Yüksek Lisans Tezi, Elektrik-Elektronik Mühendisliği
Tez Yöneticisi: Yrd.Doç.Dr. Nurdal WATSUJİ
Haziran 2008, 79 sayfa

Bu tezde, otomatik kişi tanımada kullanılan yeni bir algoritma önerilmektedir. Kişi tespiti, kişilerin ellerinin geometrik özellikleri kullanılarak yapılır. El geometrisi, kişi tanımada kişilerin el şekillerini kullanan bir biyometrik özelliktir. Bu çalışmada, hem kişi doğrulama hem de kişi tanımlama amaçları için kullanışlı, basit ve etkili bir algoritma sunulmuştur.

Çalışmada önerilen algoritma üç temel bölümden oluşmaktadır : El kenar bölgesinin yani dış hatlarının çıkartılması için bölümleme, özelliklerin elde edilmesi ve tanıma. Bölümleme, arka plan ile el bölgesini ayırır, el ve parmak hatlarını çıkartır. Özellik elde etme kısmı; parmakların uzunlukları, çeşitli yüksekliklerde parmak genişlikleri ve avuç içi genişliğini özellik vektörü olarak bulur. Son kısım tanıma kısmı ise, giriş görüntüsünden elde edilen özellik vektörü ile veritabanını karşılaştırır ve doğrulama veya kişi tanımlama amacı için sonucu verir.

Bu çalışma, hem doğrulama hem de kişi tanımlama için kişilerin sağ el görüntülerini kullanmaktadır. Önerilen algoritma için gerekli olan kodlar Matlab programı ile yazılmıştır. Sistemin performansını tam olarak görebilmek için, önerilen algoritma çok miktarda görüntü ile test edilmiştir. Sonuçta, % 97.44 kişi tanımlama oranı ve % 98.72 doğrulama oranı elde edilmiştir.

Anahtar Kelimeler: biyometri, bölümleme, el geometrisi, lekeleme, şablon eşleştirme, tanıma

ACKNOWLEDGMENTS

I express sincere appreciation to my supervisor, Asst. Prof. Dr. Nurdal WATSUJI, for her guidance, constructive proofreading, and many fruitful discussions. I thank her for helping me during writing this thesis.

I wish also to acknowledge my parents for always supporting my choices in life and my colleague at GMYO for their support and encouragement.

I would like to express my gratitude to my teachers at the Department of Electrical and Electronics Engineering.

And finally, special thanks to my husband, Serkan OZBAY, for sharing the most valuable times with me.

CONTENTS

LIST OF FIGURES page

LIST OF TABLES

LIST OF SYMBOLS

f (x, y)	Two dimensional light intensity function
T	Threshold value
D	Dilation operation
E	Erosion operation
G	Partial derivative
∇	Gradient of a function
D	Distance function
C (m, n)	Cross correlation function
p	Probability function
$\overline{X}$	Average (mean) value

CHAPTER 1

INTRODUCTION

The problem of identifying a person is becoming increasingly important in modern network society. Nowadays we have increasingly electronically connected information society, so accurate identification is becoming very important. Personal identification is associating an identity with an individual. People have many passwords, PIN numbers, account numbers, and other security codes to use in some applications and services for identification. On the other hand, these could easily be forgotten, stolen or duplicated. Many systems such as communication, transportation, ATM machines, and commercial operations need reliable user identification.

Biometric recognition is an automatic recognition of individuals with their physiological or behavioral characteristics. Fingerprint, face, voice, hand geometry, iris, retina, handwriting, gait are the body characteristics of the biometric science. Nowadays these kinds of systems have many application areas such as criminal law enforcement, automatic door opener, or entering a shopping mall. A firm uses biometrics to regulate access to buildings or to control information. Governments are contemplating the inclusion of biometric identifiers in passports, driver's licenses, and possibly a future national ID card.

The method of identification using biometry offers several advantages over traditional methods involving ID cards (tokens) or PIN numbers (passwords) for various reasons: firstly the person to be identified is required to be physically present at the point-of-identification; secondly identification based on biometric techniques obviates the need to remember a password or carry a token. With the increased integration of computers and internet into our everyday lives, it is necessary to protect sensitive and personal data. Biometric techniques can potentially prevent unauthorized access to ATMs, cellular phones, laptops, and computer networks.

In a biometric system, we can recognize a person with two different ways. These are identification and verification (also called authentication). Verification system involves confirming or denying a person's claimed identity. In identification system, the system has to recognize a person from a list of users in the template database.

In the identification mode, the system tries to find an answer to the question “Who am I?” by comparing the captured biometric features of the candidate with all the templates stored in the database. In the verification mode, an answer is “Am I who I claim I am?” and it is making a single comparison with the claimed identity’s template [1]. There are three main modules in a biometric system: enrollment module, identification module and verification module as seen in Figure 1.1 [2]. We can enroll individuals into the biometric system by enrollment module. During the enrollment step, biometric characteristic of an individual is scanned by a biometric sensor to acquire a digital representation of the characteristic. This digital data is processed by a feature extractor. Acquired data is a template set for this system. Template set is the system database which is the central of the biometric system.

Verification system is prepared to solve the problem of confirming or denying a person’s claimed identity. The system validates a person’s identity by comparing the captured biometric data with her own biometric template(s) stored in the system database. The system conducts a one-to-one comparison to determine whether the claim is true or not (e.g. does this biometric data belong to her/him?). Verification is typically used for positive recognition, where the aim is to prevent multiple people usage from using the same identity [2].

Identification system searches the templates in the database for matching so the system conducts one-to-many matching comparison to establish an individual’s identity (or fails if the subject is not enrolled in the system database) without the subject having to claim an identity (e.g., “Whose biometric data is this?”). Identification is a critical component in negative recognition applications where the system establishes whether the person is who she/he (implicitly or explicitly) denies to be. The purpose of negative recognition is to prevent a single person from using multiple identities [2].

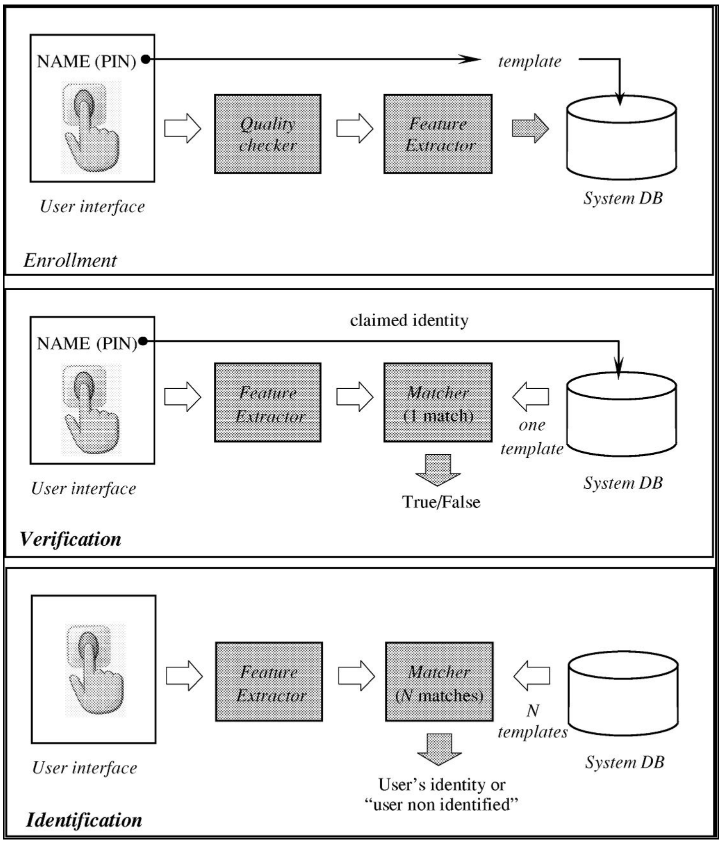


Figure 1.1 Block diagrams of biometrics systems

Biological measurement requires some properties:

- Universality, all members of the candidates must have the characteristic
- Uniqueness, no two persons should be the same in terms of the characteristic
- Permanence, member's characteristics must be unchangeable with time
- Collectability, the characteristic can be measured quantitatively.

The other important properties are:

- Performance, the system is able to meet accuracy, speed, robustness and cost requirements satisfactorily
- Acceptability, people are willing to accept the biometric system
- Circumvention, the system is robust against counter measures.

Design of a biometrics-based identification system could essentially be reduced to the design of a pattern recognition system. The conventional pattern recognition system designers have adopted a sequential phase-by-phase modular architecture as shown in Figure 1.2.

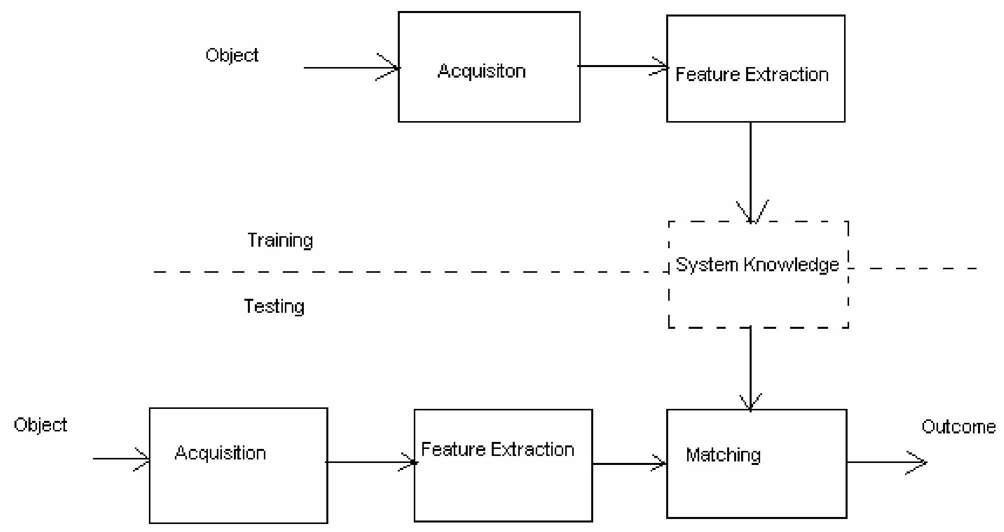


Figure1.2. Architecture of a typical pattern recognition system

Although it is generally known in the research community that more integrated, parallel, active system architectures involving feedback/feed-forward control have a number of advantages, these concepts have not yet been fully exploited in commercial biometrics-based systems. Researchers in pattern recognition have realized that effectively resolving these issues is very difficult and there is a need to constrain the environment to engineer feasible solutions.

A number of biometrics have been proposed, researched, and evaluated for recognition applications. There are 14 different biometric techniques that are either widely used or under investigation, including fingerprint, face, iris, DNA, ear, voice, gait, keystroke dynamics, signature, infrared facial and hand vein thermograms, odor, retinal scan, palmprint, and hand and finger geometry. We can explain briefly these biometric technologies below:

- Fingerprints: Among all the biometric techniques, fingerprint-based identification is the oldest method which has been successfully used in numerous applications. It is known that everyone has unique, immutable fingerprints.

There is an archeological evidence that fingerprints as a form of identification have been used at least since 7000 to 6000 BC by the ancient Assyrians and Chinese. In the late 1990s, the introduction of inexpensive fingerprint capture devices and the development of fast, reliable matching algorithms has set the stage for the expansion of fingerprint matching to personal use.

A fingerprint is made of a series of ridges and furrows on the surface of the finger. Its formation is determined during the fetal period, and it doesn't change throughout life except in cases of some injuries or scratches. Even twins have different patterns. The uniqueness of a fingerprint can be determined by the pattern of ridges and furrows as well as the minutiae points. Minutiae points are local ridge characteristics that occur at either a ridge bifurcation or a ridge ending. These properties make fingerprint an perfect biometric in terms of uniqueness and permanence.

Fingerprint patterns carry sufficient information to allow large-scale identification, making fingerprint technology suitable for many-user applications. One problem with the current fingerprint recognition systems is that they require a large amount of computational resources. With the development of solid-state sensors, the marginal cost of incorporating a fingerprint-based biometric system has become affordable in many applications. Nowadays, we can see this system on PCs, cellular phones.

- Face recognition: Face recognition is a non-intrusive method, and facial images are probably the most common biometric characteristic used by humans to make a personal recognition. This method can be enforcement for mug-shot identification. Some successful applications of face recognition system are driver's licenses and credit cards, gateways to limited access areas, surveillance of crowd behavior.

Two primary approaches to the identification based on face recognition are the following: (i) Transform approach: the universe of face image domain is represented using a set of orthonormal basis vectors. Currently, the most popular basis vectors are eigenfaces: each eigenface is derived from the covariance analysis of the face image population; two faces are considered to be identical if they are sufficiently "close" in the eigenface feature space. A number of variants of such an approach exist. (ii) Attribute-based approach: facial attributes like nose, eyes, etc. are extracted from the face image and the invariance of geometric properties among the face landmark features is used for recognizing features [6].

The general problem of face recognition, that is the ability to recognize faces independent from pose and light in a cluttered environment, is still far from being solved. Therefore, current face recognition systems impose a number of restrictions on how the facial images are acquired (e.g., simple background, uniform and fixed illumination). In order for face recognition to be widely adopted, they should be able to automatically detect whether there exists a face in the acquired image, locate the face if there is one, and recognize the face from a general viewpoint. The main challenges of face recognition today are handling rotation in depth and broad lighting changes together with personal appearance changes.

- Iris: The iris is the color part of the eye that is bounded by pupil and sclera (white of the eye). It has a complex structure and a rich texture. Visual texture of the human iris, formed during fetal development and stabilizes during the first two years of life, is determined by the chaotic morphogenetic processes during embryonic development and is posited to be unique for each person and each eye [7].

The complex iris texture carries very distinctive information useful for personal recognition. The accuracy and speed of currently deployed iris-based recognition systems are promising and point to the feasibility of large-scale identification systems based on iris information. Even twins have different irises, because like fingerprints, these patterns result from developmental processes and stabilize in early childhood; staying unchanged throughout life [2]. It is impossible to copy or modify a person's iris pattern. Speed of the algorithms for feature extraction enables iris recognition based systems to be used in large-scale identification applications with high accuracy. Image capture phase is the important part of the system. For acquired perfect quality image, setup must be designed carefully.

Further, it is rather easy to detect artificial irises (e.g., designer contact lenses). Although, the early iris-based recognition systems required considerable user participation and were expensive, the newer systems have become more user-friendly and cost-effective.

- DNA: DNA (Deoxyribo Nucleic Acid) is the one-dimensional ultimate unique code for one's individuality - except for the fact that identical twins have the identical DNA pattern. It is, however, currently used mostly in the context of forensic applications for identification. Three issues limit the utility of this biometrics for other applications: (i) contamination and sensitivity: it is easy to steal a piece of DNA from an unsuspecting subject to be subsequently abused for an ulterior purpose; (ii) automatic real-time identification issues: the present technology for genetic matching is not geared for online unobtrusive identifications. Most of the human DNA is identical for the entire human species and only some relatively small number of specific locations on DNA exhibit individual variation. These variations are manifested either in the number of repetitions of a block of base sequence or in the minor non-functional perturbations of the base sequence (sequence polymorphism); (iii) privacy issues: information about susceptibilities of a person to certain diseases could be gained from the DNA pattern and there is a concern that the unintended abuse of genetic code information may result in discrimination in e.g., hiring practices [2].

- Ear: It is known that the shape of the ear and the structure of the cartilaginous tissue of the pinna are distinctive. The features of an ear are not expected to be unique to each individual. The ear recognition approaches are based on matching vectors of distances of salient points on the pinna from a landmark location on the ear. No commercial systems are available yet and authentication of individual identity based on ear recognition is still a research topic [6].

- Voice: Voice identification has attractions because of its importance in human communication. We expect to pick up the phone and be able to recognize someone by his or her voice after only a few words. Voice verification is considered a combination of behavioral and physiological biometric because the physical shape of the throat and larynx determines the voice pattern, although the speaker can change it. The system use different characteristics, such as cadence, pitch, and tone to verify the user's identity.

A voice signal available for authentication is typically degraded in quality by the microphone, communication channel, and digitizer characteristics. Before extracting features, the amplitude of the input signal may be normalized and decomposed into several band-pass frequency channels. The features extracted from each band may be either time-domain or frequency domain features [6].

It is an advantage to telephone-based applications where identity has to be verified remotely and no other biometric can be used such as online banking and electronic commerce. Voice recognition systems up to now have been only reliable for verification applications, but with the introduction of new algorithms and techniques, some identification systems having reasonable performance began to appear in the market. The technology is user friendly and does not require extensive user cooperation.

The system is sensitive to variations in the microphone and transmission channel, as well as variations in voice due to illness, emotion or aging. Besides, some people seem to be extraordinarily skilled in mimicking others. Background noise is another source of performance degradation in some circumstances.

- Gait: Gait is the peculiar way one walks and is a complex spatio-temporal behavioral biometrics. Gait is not supposed to be unique to each individual, but is sufficiently characteristic to allow identity authentication. Gait is a behavioral biometric and may not stay invariant especially over a large period of time, due to large fluctuations of body weight, major shift in the body weight (e.g., waddling gait during pregnancy , major injuries involving joints or brain (e.g., cerebellar lesions in Parkinson disease), or due to inebriety (e.g., drunken gait) [6].

Acquisition of gait is similar to acquiring a facial picture and, hence, may be an acceptable biometric. Since gait-based systems use the video-sequence footage of a walking person to measure several different movements of each articulate joint, it is input intensive and computationally expensive [6].

- Keystroke Dynamics: It is known that each person types on a keyboard in a different way. This biometrics is not unique to each individual but it offers sufficient discriminatory information to permit identity authentication. The keystrokes of a person using a system could be monitored unobtrusively as that person is keying in other information [6]. Keystroke dynamic features are based on time durations between the keystrokes. Some variants of identity authentication use features based on inter-key delays as well as dwell times – how long a person holds down a key. Typical matching approaches use neural network architecture to associate identity with the keystroke dynamics features. Some commercial systems are already appearing in the market.

- Signature: Each person has a unique style of handwriting so signature biometric system can be utilized for the purpose of verification and identification.

Two methods are used for signature verification. First method is static (off-line) and the second one is dynamic (on-line). Static signature verification uses scanned type of the signatures, extracting only the geometric features of the signature. The signature is written as important as the static shape of the finished signature. Dynamic signature verification systems possess special hardware to obtain further information such as pen tip location, velocity and acceleration profiles, pen angle and contact pressure. This method gives additional data, and we can detect the problem easier. And also this system is very robust against forgery.

The accuracy of verification of the system is not enough for large-scale identification applications. But it is used in document authentication and transaction authorization applications.

- Infrared Facial and Hand Vein Thermograms: Human body radiates heat and the pattern of heat radiation is a characteristic of each individual body.

An image is acquired from an infrared sensor to indicate the heat of the body parts. These images are called thermograms. Any part of the body could be used for identification. The technology could be used for covert identification solutions and could distinguish between identical twins. It is also claimed to provide enabling technology for identifying people under the influence of drugs: the radiation patterns contain signature of each narcotic drug [6]. A thermogram-based system does not require contact and is noninvasive, but image acquisition is challenging in uncontrolled environments, where heat emanating surfaces (e.g., room heaters and vehicle exhaust pipes) are present in the vicinity of the body. A related technology using near infrared imaging is used to scan the back of a clenched fist to determine hand vein structure. Infrared sensors are prohibitively expensive which is a factor inhibiting wide spread use of the thermograms [2].

- Odor: It is known that each object exudes an odor that is characteristic of its chemical composition and this could be used for distinguishing various objects. A whiff of air surrounding an object is blown over an array of chemical sensors, each sensitive to a certain group of (aromatic) compounds. A component of the odor emitted by a human (or any animal) body is distinctive to a particular individual. It is not clear if the invariance in the body odor could be detected despite deodorant smells, and varying chemical composition of the surrounding environment [8].

- Retinal Scan: The veins beneath retinal surface form a unique and stable pattern that carries distinctive information like the iris. Up to now, retinal pattern recognition systems have been used in high security applications like prisons, and retinal scan has been considered as the most secure and accurate biometric technique. High identification rates enable use of retinal scan systems in large scale.

The image capture requires a person to peep into an eye-piece and focus on a specific spot in the visual field so that a predetermined part of the retinal vasculature could be imaged [6]. Special hardware needed for imaging is expensive, the process requires significant user cooperation for reliable operation and gazing into a light source reduces down the user acceptability. These are the disadvantages of the system. Because of these reasons, the system is used usually in high security applications.

- Palmprint: The palms have pattern of ridges and valleys as the fingerprints. The area of the palm is much larger than the area of a finger and, as a result, palmprints are expected to be even more distinctive than the fingerprints. Since palmprint scanners need to capture a large area, they are bulkier and more expensive than the fingerprint sensors. Human palms also contain additional distinctive features such as principal lines and wrinkles that can be captured even with a lower resolution scanner, which would be cheaper [6].

- Hand and Finger Geometry: Hand geometry recognition systems are based on a number of measurements taken from the human hand, including its shape, size of palm, and lengths and widths of the fingers. Commercial hand geometry-based verification systems have been installed in hundreds of locations around the world [6].

The lengths and widths of the fingers and other hand shape attributes extracted from captured image of a hand are used in this system. Environmental factors such as dry weather or individual anomalies such as dry skin do not appear to have any negative effects on the verification accuracy of hand geometry-based systems. But hand geometry-based recognition systems cannot be scaled up for systems requiring identification of an individual from a large population [6].

The comparison of these biometric systems is given in Table 1.1 [6].

Table 1.1 Comparison of various biometric technologies, high, medium, and low are denoted by H, M, and L respectively

Biometric Identifier	**Universality**	**Distinctiveness**	**Permanence**	**Collectability**	**Performance**	**Acceptability**	**Circumvention**
DNA	H	H	H	L	H	L	L
Ear	M	M	H	M	M	H	M
Face	H	L	M	H	L	H	H
Facial Thermogram	H	H	L	H	M	H	L
Fingerprint	M	H	H	M	H	M	M
Gait	M	L	L	H	L	H	M
Hand geometry	M	M	M	H	M	M	M
Hand vein	M	M	M	M	M	M	L
Iris	H	H	H	M	H	L	L
Keystroke	L	L	L	M	L	M	M
Odor	H	H	H	L	L	M	L
Palmprint	M	H	H	M	H	M	M
Retina	H	H	M	L	H	L	L
Signature	L	L	L	H	L	H	H
Voice	M	L	L	M	L	H	H

CHAPTER 2

LITERATURE SURVEY

The emerging field of biometric technology addresses the automated identification of individuals, based on their physiological and behavioral traits. The broad category of human authentication schemes, denoted as biometrics, encompasses many techniques from computer vision and pattern recognition. The personal attributes used in a biometric identification system can be physiological, such as facial features, fingerprints, iris, retinal scans, hand and finger geometry; or behavioral, traits idiosyncratic of an individual, such as voice print, gait, signature, and keystroke style. Depending on the complexity or the security level of the application, one will opt to use one or more of these personal characteristics, possibly under a multimodal fusion scheme for performance enhancing [2].

Ancient Egyptians used body measurements to classify and identify people. Today's hand geometry scanners use infrared optics and microprocessor technology to quickly and accurately record and compare hand dimensions. Several hand geometry verification technologies have evolved during this century. They range from electro-mechanical devices to the solid state electronic scanners being manufactured today [6].

The hand modality has a number of advantages in that it is user-friendly, hand-imaging devices are not intrusive, and template storage costs are small. Several systems have been developed using the hand modality. They are based on the hand silhouette, hand geometry, finger biometry or palmprints [9].

Hand geometry systems have the longest implementation history of all biometric modalities. The U.S. Patent office issued patents to Robert P. Miller in the late 1960's and early 1970's for a device that measures hand characteristics, and unique features for comparison and ID verification [10]. Miller's machines were highly mechanical and manufactured under the name "Identimation." Several other companies launched development and manufacturing efforts during the 70's and early 80's [6]. In the mid-1980's, David Sidlauskas developed and patented an electronic hand scanning device [11] and established the Recognition Systems, Inc. of Campbell, California in 1986, and the first commercial hand geometry recognition systems became available the next year.

The first applications for hand scanners were as access control components. Government and nuclear facilities used them to protect their facilities [12].

The 1996 Olympic Games implemented hand geometry systems to control and protect physical access to the Olympic Village. Many companies implement hand geometry systems in parallel with time clocks for time and attendance purposes. Walt Disney World has used a similar "finger" geometry technology system for several years to expedite and facilitate entrance to the park and to identify guests as season ticket holders to prevent season ticket fraud.

A photo of a hand recognition system is shown in Figure 2.1.

Figure 2.1: A hand recognition system for access control

The availability of low cost, high speed processors and solid state electronics made it possible to produce hand scanners at a cost that made them affordable in the commercial access control market. At first, systems providers installed hand scanners in the stand-alone mode. The products contained basic access control functions such as time zones, alarm inputs and outputs, duress, and request for exit functions. Then end users became more sophisticated and demanded more elaborate engineered versions of the hand scanner. Today, hand scanners perform a variety of functions including access control, employee time recording and point-of sale applications.

There are some initial works about hand-based biometric system in literature. It is possible to extract two types of biometric indicators from hand images: (i) Hand geometry features which include hand's land, its width, thickness, geometrical composition, shape and geometry of the fingers and shapes of the palm etc. and (ii) palmprint features which are composed of principal lines, wrinkles, minutiae, delta points etc. In conventinal systems, pegs are used extensively to fix the hands placement during image acquisition. Nevertheless systems based on pegs have some problems; pegs deform definitely the shape of the hand, secondly, the fingers may be placed differently in different instants. These problems reduce the overall system accuracy in the feature extraction and further analysis. That is why, new trends are towards peg-free systems, and M. Arif, N. Vincent, and T. Brouard [13] present a detailed feature extraction and fusion based methodology that uses a peg free simple hand image acquisition and all of the possible hand geometry features. Earlier efforts have used combinations of these features for recognition with varying degrees of success but no individual results have been presented.

R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos [14] select 25 features, such as finger widths at different latitudes, finger and palm heights, finger deviations and the angles of the interfinger valleys with the horizontal, and model them with Gaussian mixtures.

A. K. Jain, A. Ross, and S. Pankanti [15] have used a peg-based imaging scheme and obtained 16 features, which include length and width of the fingers, aspect ratio of the palm to fingers, and thickness of the hand. The prototype system they developed was tested in a verification experiment for web access over for a group of ten people [16].

The other research consists of the lengths and widths of the fingers and the width of a palm. Hands place freely without the need for pegs to fix the hand placement. Systems using six different distance functions for verification and identification [17].

In addition to geometric features such as finger widths at various positions and palm size, finger shapes can be also used. These shapes are represented with fourth degree implicit polynomials, and the resulting sixteen features are compared with the Mahalanobis distance [18]. A work utilizes both hand geometry and palm print information as in Kumar et al. [19], which use decision level fusion. Unlike other bimodal biometric systems, the users do not have to undergo the inconvenience of using two different sensors since the palmprint and hand geometry features can be acquired from the same image, using a digital camera, at the same time. Each of these gray level images is aligned and then used to extract palmprint and hand geometry features. These features are then examined for their individual and combined performances.

A global hand shape-based approach for person identification and verification has been proposed in the other work. The algorithm has the normalization of the deformable hand shape. Hand shape normalization involves the registration of fingers by separate rotations to standard orientations as well as the rotation and translation of the whole hand. The system is based on the images of the right hands of the subjects, captured by a flatbed scanner in an unconstrained pose at 45 dpi. In a preprocessing stage of the algorithm, the silhouettes of hand images are registered to a fixed pose, which involves both rotation and translation of the hand and, separately, of the individual fingers. Two feature sets have been comparatively assessed, Hausdorff distance of the hand contours and independent component features of the hand silhouette images. Hand-based recognition is a viable secure access control scheme because of both the classification and the verification performances are found to be very satisfactory as it was shown that, at least for groups of about five hundred subjects [20].

Boreki G. and Zimmer A. [21] work a complete access control system based on a biometric code formed by invariant geometric features extracted from the hand's image, a hardware key and a vital sign detector. The feature extraction process is based on the analysis of the curvature profile of the image, making the system invariant to the rotation and translation of the hand. This makes unnecessary the use of any kind of restriction devices such as pins or pegs to position the hand.

CHAPTER 3

IMAGE PROCESSING BASICS FOR HAND RECOGNITION

Digital image processing is the use of some computer algorithms to perform image processing on digital images. Digital image processing methods are carried out mainly for two purposes: improvement of visual appearances of images for human interpretation, and preparation images for measurement of structures or features that exist in images.

This chapter presents the fundamentals of image processing techniques and algorithms used for hand recognition algorithms starting from digital image definition to pre-processing techniques, image enhancement techniques, image segmentation and object recognition.

3.1 Digital Image Fundamentals

An “image” can be defined as a two-dimensional (2D) light intensity function $f(x, y)$, where x and y denote spatial coordinates and the amplitude of the function f, at any coordinates (x, y) is called the brightness (or gray level) of the image at that coordinate point.

A digital image “g” in a 2D discrete space is obtained from an analog image $f(x,y)$ in a 2D continuous space through the process that is frequently referred to as “digitization”. Digitizing the coordinate values of the analog (continuous) image is called “sampling” and digitizing the amplitude values of the continuous image is called “quantization”.

The two-dimensional continuous image $f(x,y)$ is divided into M rows and N columns. The intersection of a row and a column is called “pixel”. Then a discrete value is assigned to each pixel. A digital image that has been discretized both in spatial coordinates and in brightness may be considered as a matrix whose row and column indices identify a point in the image and the corresponding matrix element value identifies the gray level at that point.

Digitization process requires the decision about the values of M (number of rows), N (number of columns) and the number of discrete gray levels allowed for each pixel. The number of gray levels, L, is normally an integer power of 2 because of sampling, processing and data storage considerations ($L=2^k$). The number, b, of bits required to store a digitized image is

$$b=M \times N \times k \tag{3.1}$$

When a digital image has 2^k gray levels, it is commonly practice to refer to the image as a “k-bit image” [22].

There are standard values for the various parameters encountered in digital image processing. These values can be caused by video standards, by algorithmic requirements, or by the desire to keep digital circuitry simple. Table 3.1 gives some commonly encountered values [23].

Table 3.1: Common values of digital image parameters

Parameter	Symbol	Typical Values
Rows	M	256,512,525,625,1024,1035
Columns	N	256,512,768,1024,1320
Gray Levels	L	2,64,256,1024,4096,16384

These parameters determine the resolution of the image. Image resolution describes the detail that an image holds. Basically, spatial resolution is the smallest discernible detail in an image and a more widely used definition of spatial resolution is simply the smallest number of discernible line pairs per unit distance. Likewise, gray-level resolution is defined as the smallest discernible change in gray level.

3.2 Image Operations

The image operations that can be applied to digital images to transform an input image into an output image may be categorized into three classes: Point, local and global operations.

Point operation is an operation that the output value at a specific coordinate in an image is dependent only on the input value at that same coordinate. In local operation, the output value at a specific coordinate is dependent on the input values in the neighborhood of that same coordinate. In the case of global operation in an image, the output value at a specific coordinate is dependent on all the values in the input image.

Neighborhood operations have an important role in digital image processing. Neighborhood operations are those that combine a small area or neighborhood of pixels to generate an output pixel. It is therefore important to know the various neighborhoods that can be used to process an image. Local operations produce an output pixel value based upon the pixel values in the neighborhood.

A pixel p at coordinates (x,y) has four horizontal and vertical neighbors whose coordinates are given by

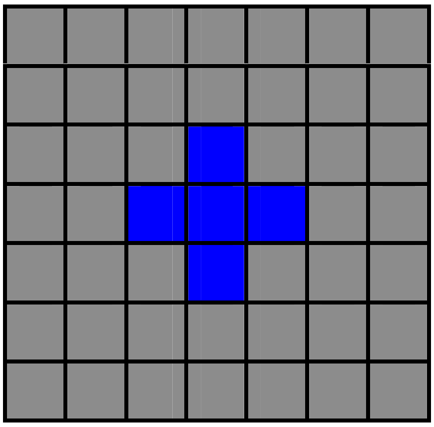
$$(x+1,y)\ ,\ (x-1,y)\ ,\ (x,y+1)\ ,\ (x,y-1)$$

The set of pixels are called the “4-neighbors” of pixel p. The pixel p at (x,y) coordinates has also four diagonal neighbors whose coordinates are given by

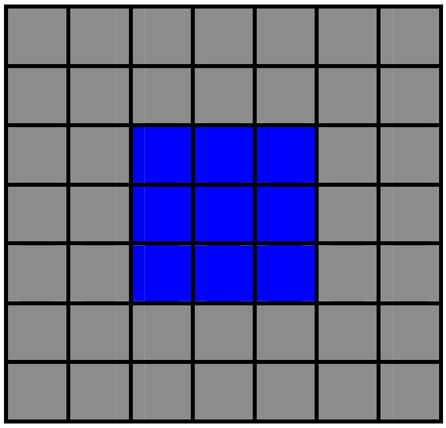
$$(x+1,y+1)\ ,\ (x+1,y-1)\ ,\ (x-1,y+1)\ ,\ (x-1,y-1)$$

These four diagonal neighbors, together with the 4-neighbors, are called “8-neighbors” of pixel p. A neighborhood that contains 4-neighbors are called 4-connected neighborhood and a neighborhood that contains 8-neighbors is called

8-connected neighborhood. 4-connected neighborhood and the 8-connected neighborhood are illustrated in Figure 3.1.



(a)



(b)

Figure 3.1: Neighborhoods, (a) 4-connected (b) 8-connected

3.3 Color Images and Color Spaces (Models)

Natural color images, as opposed to computer-generated images, usually originate from a color scanner or a color video camera. These devices incorporate sensors that are spectrally sensitive to the red, green, and blue portions of the light spectrum. The color sensors typically generate red, green, and blue color signals that are linearly proportional to the amount of red, green, and blue light detected by each sensor. Linear RGB images are the basis for the generation of the various color space image representations. A color space (color model) is an abstract mathematical model for representing colors in terms of intensity values. In other words a color space specifies how color information is represented. There are various color models in use. Mostly used color models are RGB-based color models, and CMY-based color models and gray spaces.

RGB-based color model is a model in which red, green and blue light are added together in various ways to reproduce a broad array of colors. RGB-based color spaces are the most commonly used color spaces in computer graphics, primarily because they are directly supported by most color displays and scanners. The name of the model comes from the initials of the three additive primary colors, red, green, and blue. RGB-based color models include RGB and HLS, HSV models. In RGB, any color is some mixture of three primary colors: red, green and blue. Figure 3.2 shows the "space" defined by RGB signals: it is a Cartesian cubic space, since the red, green, and blue signals are independent and can be added to produce any color within the cube.

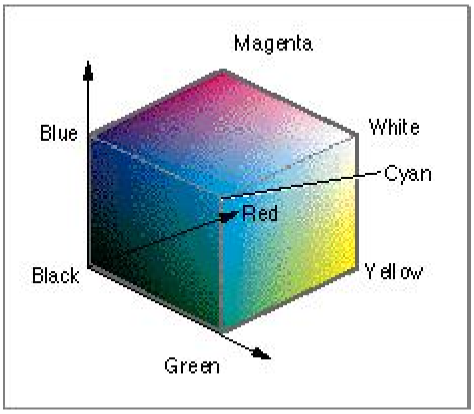
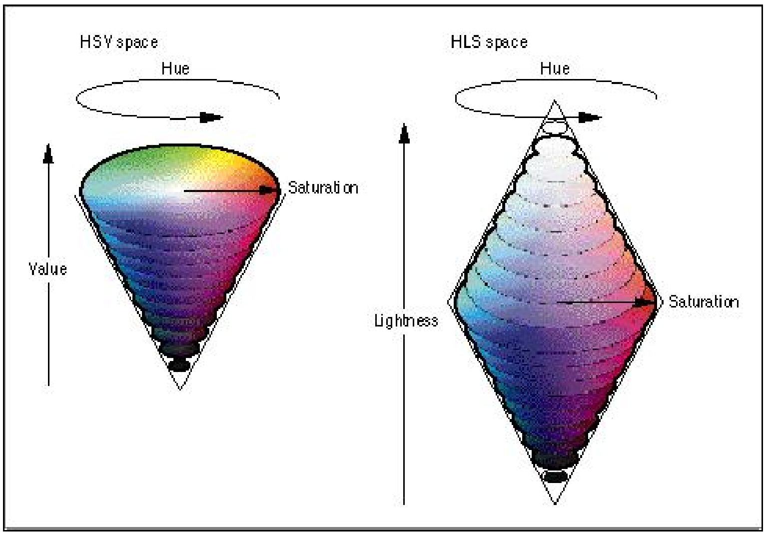


Figure 3.2: RGB Color Space

HLS model describes colors with the terms hue, lightness and saturation, and HSV stands for hue, saturation and value. These color spaces may be thought as cones shown in Figure 3.3.



.

Figure 3.3: HSV and HLS color spaces

In those color models, hue component is the angular measurement, saturation component describes color intensity and value component (in HSV space) and the lightness component (in HLS space) describe brightness or luminance.

CMY (Cyan, Magenta and Yellow)-based color spaces are most commonly used in color printing systems and include CMY and CMYK (Cyan, Magenta, Yellow and blacK) color models. They are subtractive color models. The CMY models work by partially or entirely masking certain colors on the typically white background (that is, absorbing particular wavelengths of light). Such a model is called subtractive because inks “subtract” brightness from white.

Gray spaces or models have a single component, ranging from black to white as shown in Figure 3.4. Gray spaces are used for black-and-white and grayscale display and printing.

Figure 3.4: Gray Space

3.4 Thresholding

Selecting some features within a scene or image is an important prerequisite for most kinds of measurement or for understanding of the scene or image. Thresholding is one of the most effective methods used in such applications called image segmentation. Technically, thresholding is an operation that converts a gray-scale image into a binary image where the two levels are assigned to pixels that are below or above the specified threshold value. Such an image is then usually displayed as a binary or two-level image, using black and white colors. This operation is called thresholding which is also called as binarization.

Automatic thresholding is also an important technique in hand recognition. The basic idea of automatic thresholding is to automatically select an optimal gray-level threshold value for separating objects of interest in an image from the background based on their gray-level distribution.

This technique is based on a simple concept. A parameter "T" called the brightness threshold is chosen and applied to the image g [m, n] as follows:

If g [m, n] >= T   g [m, n] = object =1

Else   g [m, n] = background = 0   (3.2)

This algorithm assumes that we are interested in light objects on a dark background. For dark objects on a light background we would use:

If g [m, n] < T   g [m, n] = object =1

Else   g [m, n] = background = 0   (3.3)

The most important decision in thresholding then becomes: How is the threshold value (T) selected? While there is no universal procedure for threshold selection that is guaranteed to work on all images, there are a variety of alternatives.

One alternative is to use a threshold that is chosen independently from the image, and selecting the threshold mentioned is called "fixed (constant) thresholding". If it is known that one is dealing with very high-contrast images where the objects are very dark and the background is homogeneous and very light, then a constant threshold of 128 on a scale of 0 to 255 might be sufficiently accurate.

On the other hand, there may be no big distinction between the objects and the background in images. Therefore, in such cases the threshold is chosen from the brightness histogram of the region or image that we wish to segment. Selecting the threshold value using the histogram is called "histogram-derived thresholding".

Thresholding techniques may also be categorized into two classes: global thresholding and local thresholding. “Global thresholding” selects a single threshold value from the histogram of the entire image. In other words, when threshold value, T, depends only on gray-level values, the thresholding is called global. “Local thresholding” uses gray-level values in the regions of the image where each region has different brightness distribution. That is, multiple threshold values are chosen in local thresholding; each is optimized for a small region in the image. Global thresholding is simpler and easier to implement but its result relies on uniform illumination conditions. Local thresholding methods can deal with non-uniform illumination conditions, but it causes the system more complicated and slow. For automated applications, where non-uniform illumination is usually not an issue due to controlled lighting conditions as in the case of hand recognition, global thresholding is commonly used for its simplicity and speed.

3.5 Image Enhancement Algorithms

Image enhancement algorithms are the techniques that seek to improve the visual appearance of an image or to convert the image to a better form for analysis by a human or a machine.

There is no unique theory of image enhancement because there is no general standard of image quality that can serve as a design criterion for an image enhancement processor. Consideration is given here to a variety of techniques that may be used in hand recognition algorithms.

3.5.1 Image smoothing

Smoothing operations are used basically for diminishing spurious effect that may be present in a digital image. These algorithms are applied in order to reduce noise and/or to prepare images for further processing such as segmentation.

3.5.1.1 Smearing algorithm

Smearing is a type of smoothing method for eliminating the noisy areas on an image. With the smearing algorithm, the image is processed along vertical and horizontal runs (scan-lines). If an image consists of only 1's and 0's (only black and white) called as binary image and if the number of white pixels is less than a desired threshold or greater than any other desired threshold, white pixels are converted to black or vice versa, and the noisy region is eliminated. The smearing algorithm can be made horizontally or vertically through the image.

This algorithm can be formulated as:

If Number of 'white'/ 'black' pixels < threshold;	Pixels become 'black' / 'white'
Else;	No change
OR	
If Number of 'white' / 'black' pixels > threshold;	Pixels become 'black' / 'white'
Else;	No change (3.4)

3.5.1.2 Neighborhood averaging

Neighborhood averaging is a spatial-domain technique for image smoothing. Given an image *a(m,n)*, the procedure for neighborhood averaging is to generate a smoothed image *b(m,n)* whose gray level at every point *(m,n)* is obtained by averaging the gray-level values of the pixels of *a* contained in a predefined neighborhood of *(m,n)*. The smoothed image is obtained by using the relation below:

$$b(m,n) = \frac{1}{C} \sum_{(x,y)\in S} a(x,y) \qquad (3.7)$$

where S is the set of coordinates of points in the neighborhood of the point (m,n) and C is the total number of points in the neighborhood.

Mean filtering is a filtering that uses neighborhood averaging algorithm. The mean filter is a sliding-window spatial filter which replaces the center value in the window with the average (mean) of all the pixel values in the window (in the neighborhood). The window, or the neighborhood, can be any shape. An example of mean filtering of a 3x3 window of values is shown in Figure 3.5 and Figure 3.6.

5	2	9
3	7	8
6	4	1

Figure 3.5: Unfiltered (raw) data

The center value which is 7 in the unfiltered data is replaced by the mean value of all nine pixel values. (5+2+9+3+7+8+6+4+1) / 9 = 5), that is 5.

*	*	*
*	5	*
*	*	*

Figure 3.6: Filtered with mean filtering

3.5.1.3 Ordering-based filtering

One of the principal disadvantages of neighborhood algorithm is that it blurs edges and other sharp details. To overcome this disadvantage, median filtering is used as an alternative approach.

Median filters are filters in which we replace the gray level of each pixel by the median of the gray levels in a neighborhood of that pixel, instead of by the average. This method is particularly effective when the noise pattern consists of strong components, and where the characteristic to be preserved is edge sharpness. In order to perform median filtering in a neighborhood of a pixel we sort the values of the pixel and its neighbors, determine the median, and assign this value to the pixel. For example, in a 3 x 3 neighborhood, the median is the 5th largest value, in a 5 x 5 neighborhood the 13th largest value, and so on.

An example of median filtering of a 3x3 window of values is shown in Figure 3.7 and 3.8.

15	2	192
3	63	44
6	4	1

Figure 3.7: Unfiltered (raw) data

The center value which is 63 in the unfiltered data is replaced by the median value of all nine pixel values. (the pixel values in order: 1,2,3,4,**6**,15,44,63,192), that is 6.

*	*	*
*	6	*
*	*	*

Figure 3.8: Filtered with median filtering

3.5.2 Morphology-based operations

Morphology-based operations are algorithms for processing binary images based on their shapes. These operations take a binary image as an input, and return a binary image as an output. The value of each pixel in the output image is determined by the corresponding input pixel and its neighbors.

Each morphological operation uses a specified neighborhood. The state of any given pixel in the output image is determined by applying a rule to the neighborhood of the corresponding pixel in the input image. The neighborhood is represented by a structuring element, which is a matrix consisting of only 0's and 1's. The center pixel in the structuring element represents the pixel of interest, while the elements in the matrix that are on (i.e., = 1) define the neighborhood.

There are some morphological operations including dilation, erosion, opening and closure operations that can be used in hand recognition algorithms.

3.5.2.1 Dilation

Dilation is a morphological operation which adds pixels to the boundaries of objects (i.e., changes them from off to on). In other words, if any pixel in the input pixel's neighborhood is on, the output pixel is on. Otherwise, the output pixel is off.

Dilation operation can be formulized as:

$$D(A,B) = \bigcup_{\beta \in B} (A+\beta) \tag{3.8}$$

In dilation operation, A can be considered as the input image and B is called a structuring element. Dilation operation can be explained schematically shown in Figure 3.9 and 3.10.

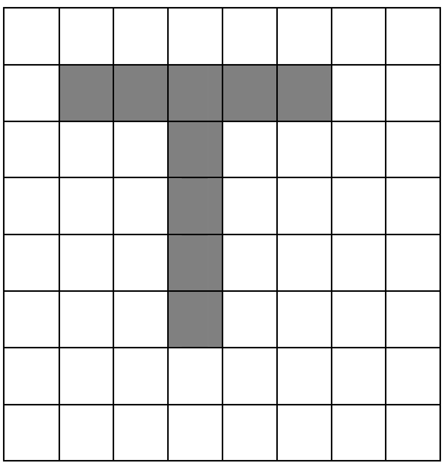


Figure 3.9: Input image

In this figure, original object pixels on the input image, A, are in gray; after dilation operation with 4-neighborhood structuring element, B, the input image becomes as shown below:

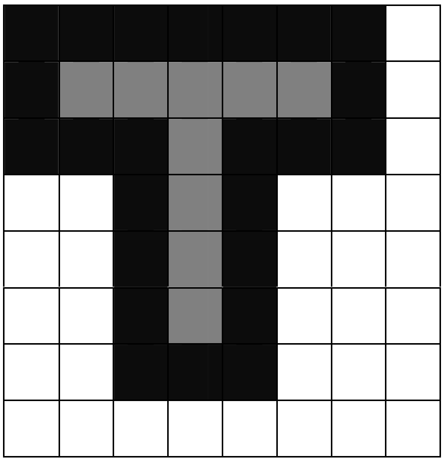


Figure 3.10: Illustration of dilation operation

The black color pixels are added through the dilation operation, so dilation operation causes objects to dilate or grow in size.

3.5.2.2 Erosion

Unlike the dilation operation, erosion operation removes pixels on object boundaries (changes them from on to off). That is to say, for erosion, if every pixel in the input pixel's neighborhood is on, the output pixel is on. Otherwise, the output pixel is off.

Erosion operation can be shown as a mathematical operation below:

$$E(A,B) = \bigcap_{\beta \in B} (A - \beta) \tag{3.9}$$

where $-B=\{-\beta \mid \beta \in B\}$

Like dilation operation, A is an input image and B is a structuring element in erosion operation. The erosion operation can be shown schematically in the figures below.

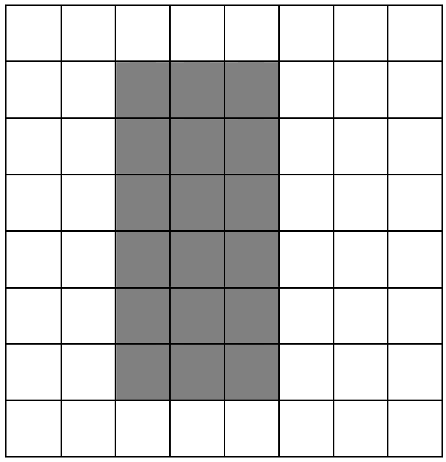


Figure 3.11: Input image

In Figure 3.11, the object pixels are in gray. After erosion operation with an 8-connected neighborhood structuring element, the image becomes as in Figure 3.12.

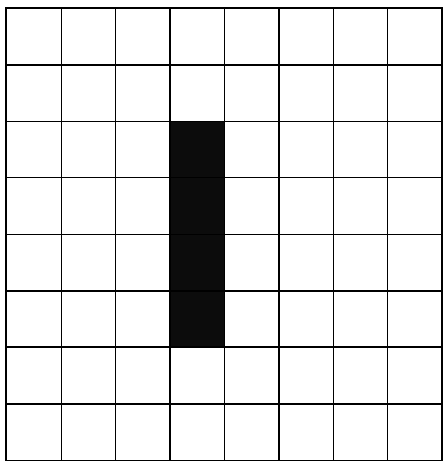


Figure 3.12: Output Image after erosion operation

It can be said from the explanations and figures; erosion operation causes objects to shrink.

3.5.2.3 Opening and closing operations

Dilation operation expands an image and erosion operation shrinks an image. There are also some other operations which are just modified version of dilation or erosion or combinations of dilation and erosion used in some applications such as opening and closing operations.

Opening operation consists of erosion followed by dilation operation. The purpose of the operation is to smooth the contour of an object in an image.

As opposed to opening operation, dilation followed by erosion operation is called a closure (closing) operation. Closing of an image with a compact structuring element also smoothes section of contours of objects but as opposed to opening, it eliminates small holes in the objects, and fuses short gaps between objects.

When dialing with binary images, the principal application of morphological operations is extracting image components that are useful in the representation and description of the shapes in the images. Some of the algorithms in morphology are the algorithms for extracting boundaries, region filling, extraction of connected components, thinning, thickening, finding skeleton of a region, pruning, etc.

3.6 Segmentation

In image processing, segmentation is the process of dividing a digital image into multiple regions (sets of pixels). Image segmentation is generally used to locate objects or boundaries in images.

Image segmentation algorithms are generally based on two basic properties of intensity values: discontinuity and similarity. For example, in the case of hand recognition algorithms, the segmentation is based on discontinuity and the segmentation process is to partition a hand contours from the hand image and background using abrupt changes in the intensity in the contours. This is also called edge detection.

3.6.1 Finding discontinuities

Considering the images in digital image processing, it has been seen that there are three basic types of gray-level discontinuities as the points, lines and the edges. The most effective way to detect the discontinuities is to use masking algorithm. Masking algorithm involves the use of appropriate mask for the desired type of discontinuities that is wanted and to run the mask through the target region of the image.

3.6.1.1 Masking

This masking algorithm is made using masks. The algorithm is basically a neighborhood operation. A mask can be thought as a subimage as shown in Figure 3.13, whose coefficients are chosen to detect a given property in an image.

c1	c2	c3
c4	c5	c6
c7	c8	c9

Figure 3.13: A sample mask with 3 x 3 sizes

In implementation of the algorithm, the center of the mask is moved around the image. At each pixel position in the image, every pixel that is contained within the mask area is multiplied by the corresponding mask coefficient. The results of these multiplications are then summed giving the new value of the pixel.

If we consider c1, c2, c3,…, c9 as mask coefficients as in the figure above, we can formulize the operation for a pixel, a(x,y), on a 3x3 neighborhood of (x, y) as the following :

$$\begin{aligned} T[\,a(x, y)\,] = {} & c1 * a(x-1, y-1) + c2 * a(x-1, y) + c3 * a(x-1, y+1) \\ & + c4 * a(x, y-1) + c5 * a(x, y) + c6 * a(x, y+1) + \\ & + c7 * a(x+1, y-1) + c8 * a(x+1, y) + c9 * a(x+1, y+1) \qquad (3.10) \end{aligned}$$

Size of the mask and coefficient values depend on the application needed. Larger masks are formed in a similar manner.

3.6.1.2 Detection of edges

The edge in an image is the sudden intensity change from dark to light or from light to dark pixel intensity. In other words, edges are the places in an image that correspond the object boundaries. To find those pixels that belong to the borders of the objects is called as edge detection.

To find edges looking for places in the image where the intensity changes rapidly uses one of these two criteria:

- Places where the first derivative of the intensity is larger in magnitude than some threshold.
- Places where the second derivative of the intensity has a zero crossing.

First-order derivatives can be computed using the gradient operators and second-derivatives can be obtained using the Laplacian.

The gradient of an image f(x,y) at location (x,y) is defined as:

$$\nabla f = [G_x ; G_y] = \left[\frac{\partial f}{\partial x} ; \frac{\partial f}{\partial y}\right] \tag{3.11}$$

and the magnitude of the gradient denoted by,

$$\text{mag}(\nabla f) = \left[G_x^2 + G_y^2\right]^{1/2} \tag{3.12}$$

Computation of the gradient of an image is based on obtaining the partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ at every pixel location.

A 3x3 mask example of edge detection that uses gradient operation is given below in Figure 3.14. This mask is also called Prewitt mask.

c1	c2	c3
c4	c5	c6
c7	c8	c9

(a)

-1	-1	-1
0	0	0
1	1	1

(b)

-1	0	1
-1	0	1
-1	0	1

(c)

Figure 3.14: Prewitt mask (a) A 3x3 region of an image, (b) and (c) Prewitt masks

These masks give the partial derivatives as:

$$\begin{aligned} G_x &= (c7 + c8 + c9) - (c1 + c2 + c3) \\ G_y &= (c3 + c6 + c9) - (c1 + c4 + c7) \end{aligned} \tag{3.13}$$

Here the difference between the first and third rows of the 3x3 image region approximates the derivative in x-direction, and the difference between the third and first columns approximates the derivative in y-direction.

The Laplacian of a an image is a second-order derivative defined as:

$$\nabla^2 f = \left[\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right] \tag{3.14}$$

The Laplacian is not used in its original form for edge detection because of being sensitive to noise. So, the Laplacian can be combined with smoothing as a precursor to find edges via zero-crossings. Here is the example of 5x5 Laplacian of Gaussian mask in Figure 3.15 [22].

0	0	-1	0	0
0	-1	-2	-1	0
-1	-2	16	-2	-1
0	-1	-2	-1	0
0	0	-1	0	0

Figure 3.15: A 5x5 mask for edge detection using Laplacian of Gaussian function

3.7 Mathematics-based Operations

Binary images can be processed with some mathematics-based algorithms including arithmetic and logic operations. Logic operations are especially useful for implementation of some processing algorithms.

The principal logic operations used in image processing are AND, OR, and NOT (COMPLEMENT). Logic operations are performed on a pixel by pixel basis between corresponding pixels of two or more images (except NOT, which operates on the pixels of a single image) [22]. For example, the AND operation of two binary images at any location yields 1 only when both the corresponding pixels in the two input images are 1. AND, OR and NOT logic functions are summarized in Table 3.2 and Table 3.3.

Table 3.2: AND and OR logic operations

Input-I	Input-II	AND	OR
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	1

Table 3.3: NOT logic operation

Input	NOT
0	1
1	0

When the processor needs pixel intensity values such that an input image has the opposite pixel intensity values unlike the needed, NOT function is used to take the complement of the input image. AND function is especially useful when the processor needs some intersection points between the input images.

3.8 Segment Labeling

The result of any successful image segmentation is the labeling of each pixel that lies within a specific distinct segment. One means of labeling is to append to each pixel of an image the label number or index of its segment. A more succinct method is to specify the closed contour of each segment. The following algorithms describe the common techniques of contour following.

The contour following approach to image segment representation is commonly called bug following. In the following binary image of Figure 3.16 as an example, a conceptual contour follower begins marching from the white background to the black pixel region indicated by the closed contour. When the contour follower crosses into a black pixel, it makes a left turn and proceeds to the next pixel. If that pixel is black, the follower again turns left, and if the pixel is white, the follower turns right. The procedure continues until the follower returns to the starting point. This simple contour follower may miss spur pixels on a boundary.

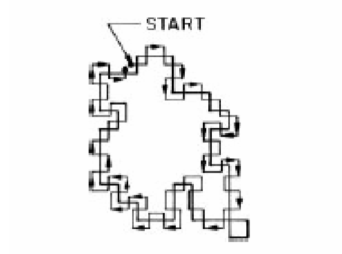


Figure 3.16: Contour Following

The other technique is known as backtracking contour follower. In this algorithm, if the follower makes a white-to-black pixel transition, it returns to its previous starting point and makes a right turn. The follower makes a right turn whenever it makes a white-to-white transition [24].

If there is a difference between the pixels intensity values of the contours and the rest of the image, then the contours pixels can be labeled according to the difference. When the contours are white, these pixels have 1 value, and the rest of the image has the 0 intensity value, black, then the processor can identify the non-zero elements of the image and labels them as contours pixels.

3.9 Feature Extraction

An image feature is a distinguishing primitive characteristic or attribute of an image. Some features are natural in the sense that such features are defined by the visual appearance of an image, while other, artificial features result from specific manipulations of an image. Natural features include the luminance of a region of pixels and gray scale textural regions. Image amplitude histograms and spatial frequency spectra are examples of artificial features [24].

In hand recognition algorithms, features of hand images can be obtained with windowing process and shape analysis such as distance measurement.

3.9.1 Windowing

Most digital signals are generally large that the dataset cannot be manipulated as a whole. Moreover, the features of some regions in an image, not the whole, may be needed, so analysis of these subregions is done as the small subsets of the total data. Windowing is the process of taking a small subset of a larger dataset, for processing and analysis.

In hand recognition algorithms, finding the extremities of the hand features needs windowing as dividing the regions into small subsets and analyzing the small datasets for the interest features of these regions such as peak points of a finger or valley point between two fingers.

3.9.2 Distance measurement

Distance is a real valued function $d\{\ (j_1,k_1),(j_2,k_2)\ \}$ of two image points (j_1,k_1) and (j_2,k_2) satisfying the following properties:

$$d\{\ (j_1,k_1),(j_2,k_2)\ \}\geq 0 \tag{3.15}$$

$$d\{\ (j_1,k_1),(j_2,k_2)\ \}=d\{\ (j_2,k_2),(j_1,k_1)\ \} \tag{3.16}$$

$$d\{\ (j_1,k_1),(j_2,k_2)\ \}+d\{\ (j_2,k_2),(j_3,k_3)\ \}\ \geq\ d\{\ (j_1,k_1),(j_3,k_3)\ \} \tag{3.17}$$

There are a number of distance functions that satisfy the defining properties. The most common measures encountered in image analysis are the "Euclidean distance",

$$d_E=[\ (j_1-j_2)^2+(k_1-k_2)^2\ ]^{1/2} \tag{3.18}$$

the "magnitude distance",

$$d_M=|\ j_1-j_2\ |+|\ k_1-k_2\ | \tag{3.19}$$

and the "maximum value distance",

$$d_X=MAX\{\ |\ j_1-j_2\ |,|\ k_1-k_2\ |\ \} \tag{3.20}$$

In discrete images, the coordinate differences (j_1-j_2) and (k_1-k_2) are integers but the Euclidean distance is not usually an integer [24].

Not only the distance measurement but also perimeter and area measurements can be used in hand recognition algorithms as the feature extraction.

3.10 Recognition

Pattern recognition, or simply recognition, is the research area that studies the operation and design of systems that recognize patterns in data. Important application areas are image analysis, character recognition, speech analysis, person identification and industrial inspection.

Hand recognition systems recognize the individuals from their hand images using some recognition algorithms. Recognition based on decision-theoretic methods is mostly used recognition algorithm in hand recognition. These methods use quantitative descriptors, such as length, area and texture. Matching is one of the algorithms based on decision-theoretic methods and it is an effective algorithm for recognition of the individuals.

Matching algorithms match each input image or image feature with a library of known vectors or images called templates. With comparing the input image with the templates, the best similarity gives the correct recognition. Minimum distance classifier and matching by correlation are the ways of matching used in hand recognition.

Decision-theoretic approaches to recognition are based on the use of decision functions. Let $x = (x_1, x_2, ..., x_n)$ represent an n-dimensional pattern vector.

In minimum distance classifier, we define the prototype, or template, of each pattern class to be the mean vector of the patterns of that class:

$$m_j = \frac{1}{N_j} \sum_{x \in w_j} x_j \qquad j=1, 2,\ldots,W \tag{3.21}$$

where N_j is the number of pattern vectors from class w_j and the summation is taken over these vectors, and W is the number of pattern classes. Using the Euclidean distance to determine closeness reduces the problem to computing the distance measures:

$$D_j(x) = \| \ x - m_j \ \| \quad \text{j=1,2,...,W} \tag{3.22}$$

We then assign x to class w_j if $D_j(x)$ is the smallest distance. That is, the smallest distance implies the best match [22].

Other distance functions can be also used in minimum distance classifiers.

To measure the similarity and find the best match, a statistical method correlation is also used. Correlation is an effective technique for image recognition. This method measures the correlation coefficient between a number of known vectors with the same size unknown vectors with the highest correlation coefficient between the vectors producing the best match. There are two forms of correlations: auto-correlation and cross-correlation. Auto-correlation function (ACF) involves only one vector and provides information about the structure of the vector or the data. Cross-correlation function (CCF) is a measure of the similarities or shared properties between two vectors. Since there are two vectors as unknown input feature vector and known database vector in this study, cross-correlation is used. In the simplest form, the correlation between $f(x, y)$ and $w(x, y)$ is as the following:

$$c(x, y) = \sum_s \sum_t f(s,t)w(x+s, y+t) \tag{3.23}$$

If we have an image denoted by F(j,k) to be searched and template which is denoted by $W(j,k)$ for 1≤j≤J and 1≤k≤K , then the normalized cross-correlation between the image pair is defined as:

$$C(m,n) = \frac{\sum_j \sum_k F(j,k)W(j-m+(M+1)/2, k-n+(N+1)/2)}{\left[\sum_j \sum_k |F(j,k)|^2\right]^{1/2} \left[\sum_j \sum_k |W(j-m+(M+1)/2, k-n+(N+1)/2|^2\right]^{1/2}} \tag{3.20}$$

for m=1,2,….M and n=1,2,….N, where M and N are odd integers.

C(m,n) is known as the cross-correlation coefficient. Its value always lies between -1 and +1, meaning that

- +1 means 100 % correlation in the same sense
- -1 means 100 % correlation in the opposing sense

A value of 0 signifies zero correlation. This means that the signal pairs are completely independent.

CHAPTER 4

HAND RECOGNITION SYSTEM

In many applications, people need reliable personal recognition systems to verify or identify an individual correctly. Hand recognition is an effective and relatively easy way of personal recognition.

In this research, a smart and simple algorithm is presented. Personal recognition has been made using geometric features of the hand of the individuals. In other words, the proposed algorithm uses hand geometry features for identification and verification purposes.

In this thesis, the proposed system consists of three major units:

- Segmentation
- Feature extraction
- Recognition

Segmentation unit separates hand region from the background and extracts the hand and finger contours. Feature extraction unit finds the lengths of the fingers, finger widths at different heights of the fingers, and the width of the palm as a feature vector. Finally recognition unit compares the feature vector obtained from the input image with the database and gives the result for identification or verification purposes. All these processes are done using Matlab software.

In this chapter, all three units for hand recognition system are explained successively in details. Starting from the input image for the system to the output of the recognition, all the details concerned are given comprehensively.

4.1 Input of the System

In this work, input of the system is the image of a hand of an individual captured by a camera. The captured image is taken with a digital camera from 25-30 centimeters away from the hand of the individual. The digital images for the input to the system are colored images with the size 640x480x3. The system uses the right hands of the individuals. The background for the hand input image is a black flat background. The hand can be placed freely only making sure that the fingers do not touch one another. In this study, illumination is controlled and the images have fixed background; therefore the system works on uniform illumination conditions.

4.2 Segmentation

Once the image is taken as a color image for the input, the system first converts it to the gray-scale intensity image using “rgb2gray” function in Matlab. This command converts RGB images to gray-scale intensity images by eliminating the hue and saturation information while retaining the luminance. A sample input image and the gray-scale intensity image after conversion process are given in Figure 4.1 and Figure 4.2, respectively.

Figure 4.1: A sample image

Figure 4.2: Gray-scale intensity image

The gray-scale intensity image is then converted to a binary image by thresholding. Thresholding is an operation that converts a gray-scale image into a binary image where the two levels are assigned to pixels that are below or above the specified threshold value. This process is also known as binarization. A threshold value is selected for the process and binarization maps the image with the threshold and assigns the 1's (white) for the greater values from the threshold and 0's (black) for the smaller than the threshold giving the binary image as the output.

Automatic thresholding is an important part in image segmentation to automatically select an optimal gray-level threshold value for separating objects of interest in an image from the background based on their gray-level distribution. As in the case of uniform illumination and fixed background, global thresholding that is the selection of a single threshold value for an entire image is suitable. In this study, illumination is controlled and the images have fixed background; therefore, global thresholding is used. Among the many global thresholding techniques, the Otsu method is one of the most preferred methods for the images whose histograms are bimodal. Otsu method selects threshold values that minimizes weighted within-class variance meaning that maximizing the between class-variance of the histogram.

Otsu method can be summarized as the following [25]: An image can be represented by a 2D gray-level intensity function f(x,y). The value of f(x,y) is the gray-level, ranging from 0 to L-1, where L is the number of distinct gray-levels. Let the number of pixels with gray-level i be n_i and n be the total number of pixels in a given image, the probability of occurrence of gray-level i is defined as:

$$p_i = \frac{n_i}{n} \tag{4.1}$$

The average gray-level of the entire image is computed as:

$$\mu_T = \sum_{i=0}^{L-1} i p_i \tag{4.2}$$

In the case of single thresholding, the pixels of an image are divided into two classes C_1 = {0,1, . . . , t} and C_2 ={t + 1, t + 2, . . . , L-1}, where t is the threshold value. C_1 and C_2 are normally corresponding to the foreground (objects of interest) and the background. The probabilities of the two classes are:

$$w_1(t) = \sum_{i=0}^{t} p_i \tag{4.3}$$

$$w_2(t) = \sum_{i=t+1}^{L-1} p_i \tag{4.4}$$

The mean gray-level of the two classes can be computed as:

$$\mu_1(t) = \sum_{i=0}^{t} i p_i / w_1(t) \tag{4.5}$$

$$\mu_2(t) = \sum_{i=t+1}^{L-1} i p_i / w_2(t) \tag{4.6}$$

Using discriminant analysis, Otsu showed that the optimal threshold t* can be determined by maximizing the between-class variance; that is:

$$t^* = \underset{0 \le t < L}{Max}\{\ q_B^2(t)\ \} \tag{4.7}$$

where the between-class variance q_B is defined as:

$$q_B^2(t) = w_1(t)(\mu_1(t) - \mu_T)^2 + w_2(t)(\mu_2(t) - \mu_T)^2 \tag{4.8}$$

In this study, the histograms of the images are used for selecting the threshold value. The value which is the lowest point occurs after the first peak (mode) is selected as the threshold value as shown in Figure 4.3.

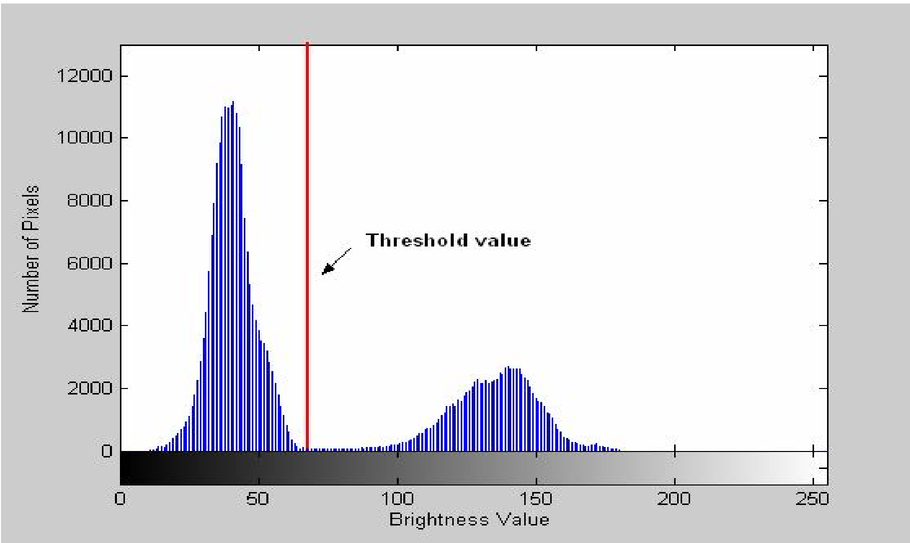


Figure 4.3: Histogram and threshold value

Since the background has uniform color in this study, the proposed method gave better results than Otsu threshold. This proposed method has been tested comparatively with Otsu method and it was found that it is the best algorithm for the application where uniform illumination and uniform color background exist.

The thresholding results for some hand images are given below in Figures 4.4, 4.5, 4.6, 4.7, 4.8, and 4.9. Otsu method and the proposed algorithm method results are shown respectively.

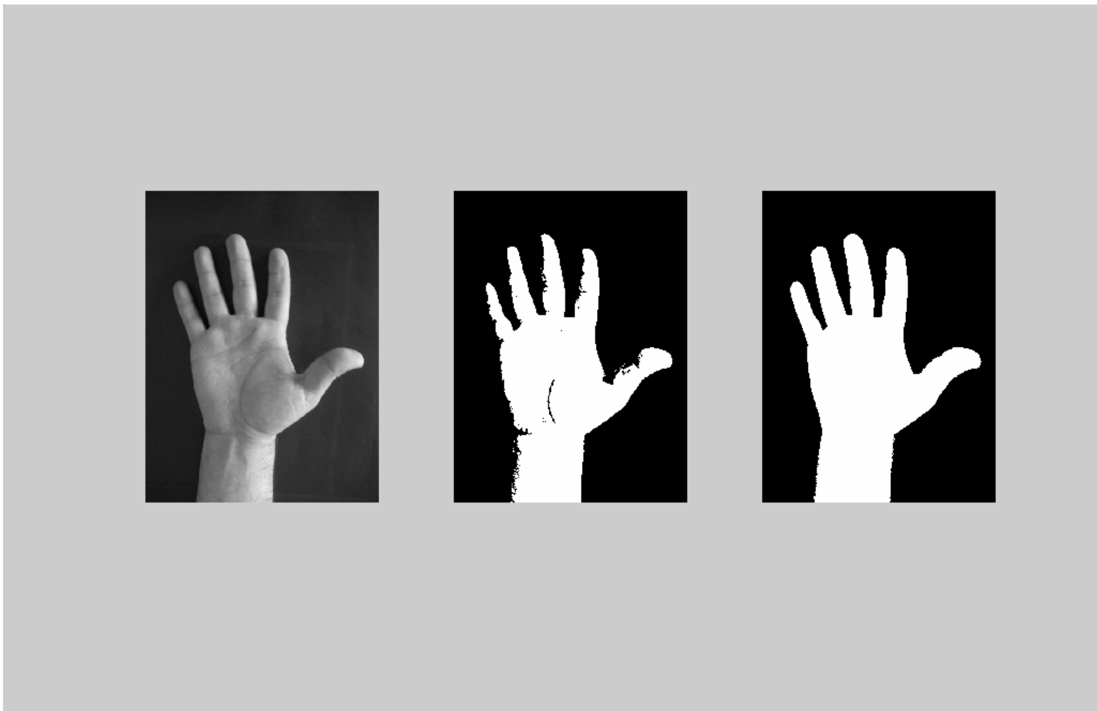


Figure 4.4: Input image, Otsu threshold result, and proposed algorithm result

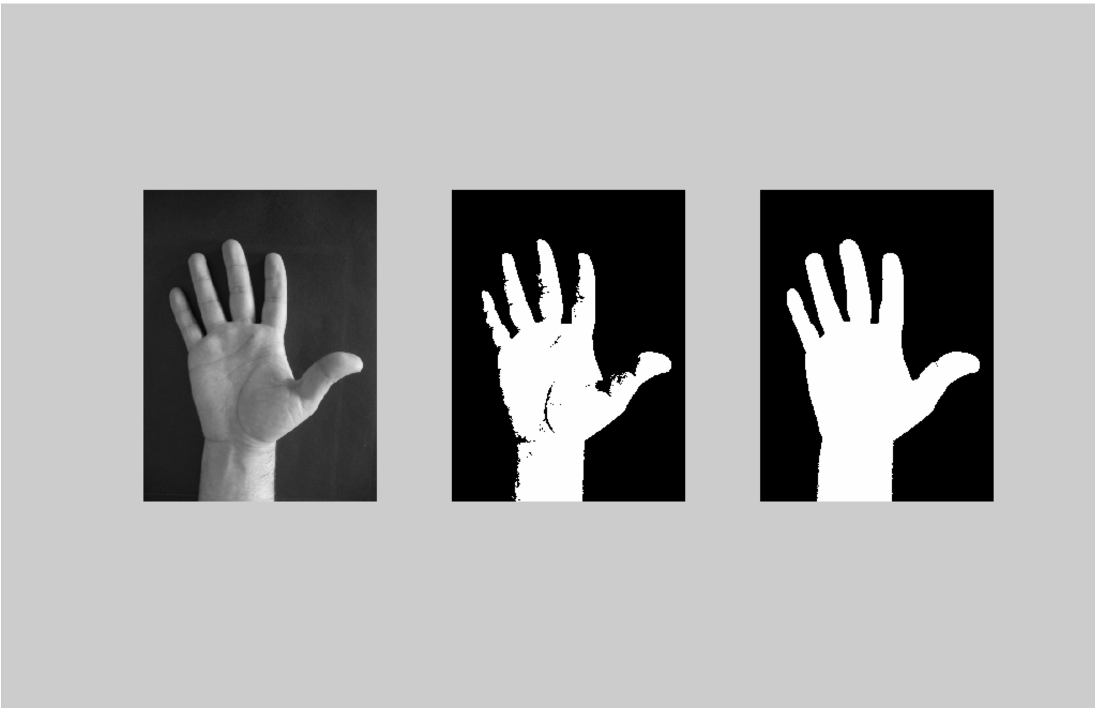


Figure 4.5: Input image, Otsu threshold result, and proposed algorithm result

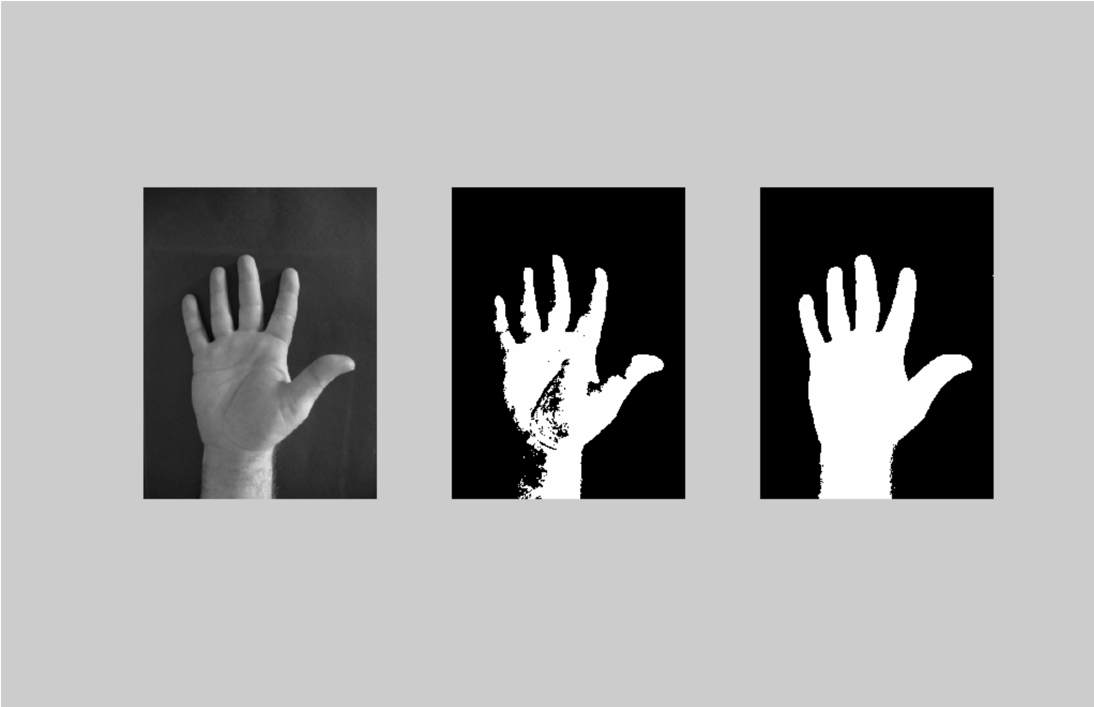


Figure 4.6: Input image, Otsu threshold result, and proposed algorithm result

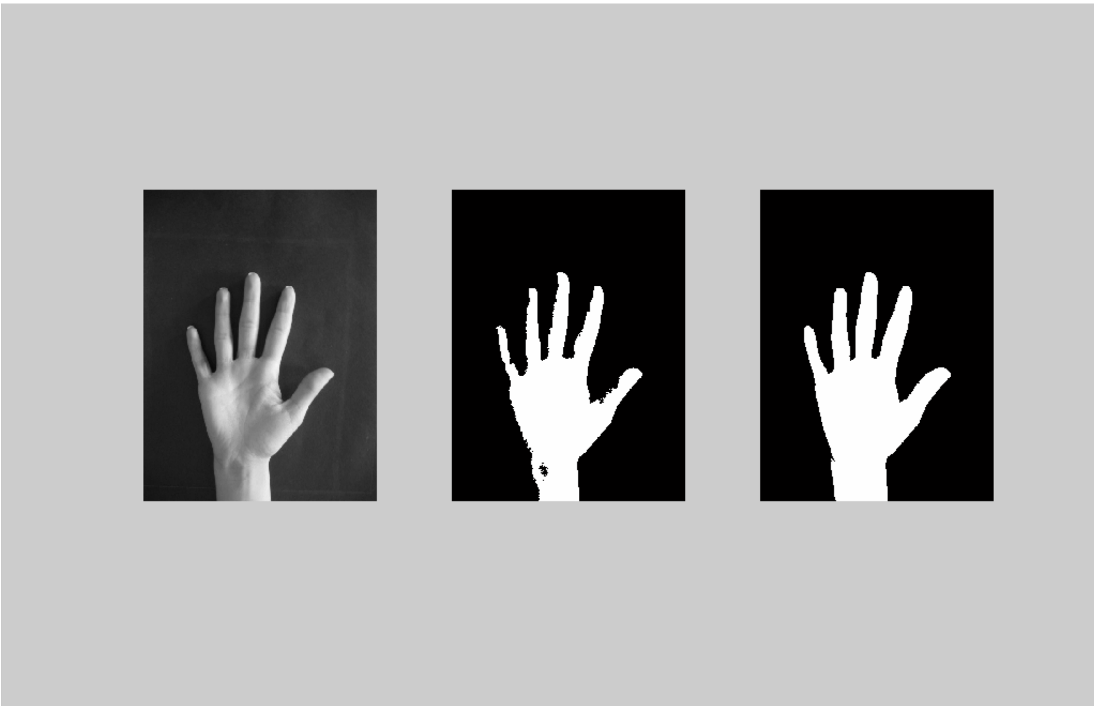


Figure 4.7: Input image, Otsu threshold result, and proposed algorithm result

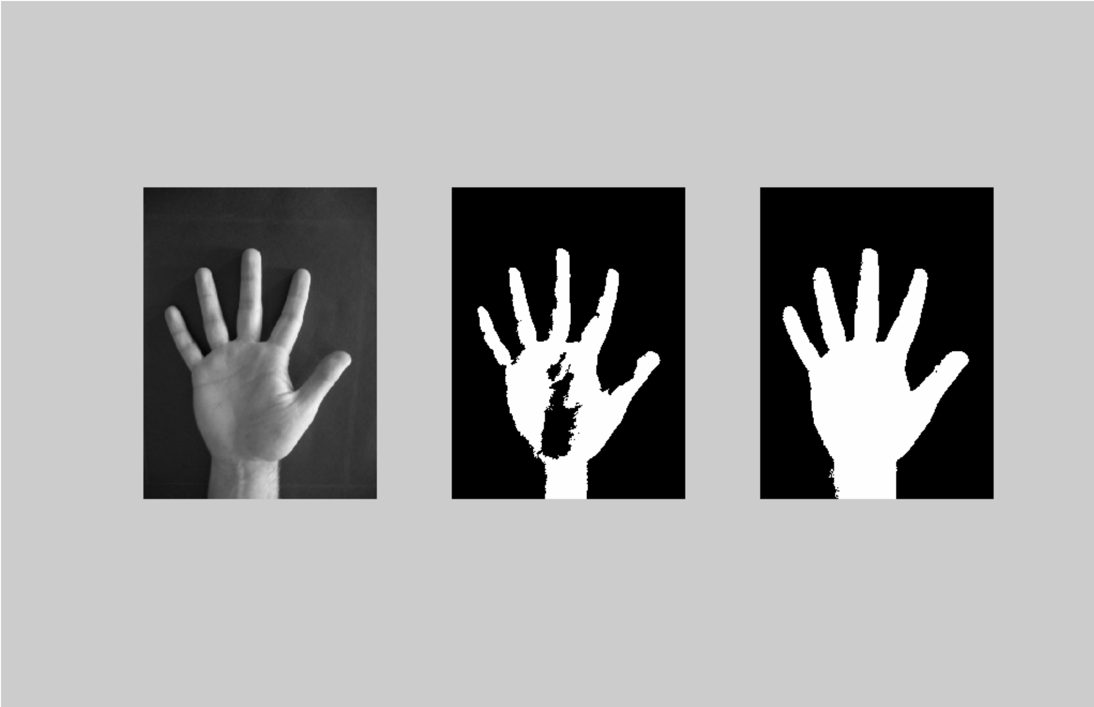


Figure 4.8: Input image, Otsu threshold result, and proposed algorithm result

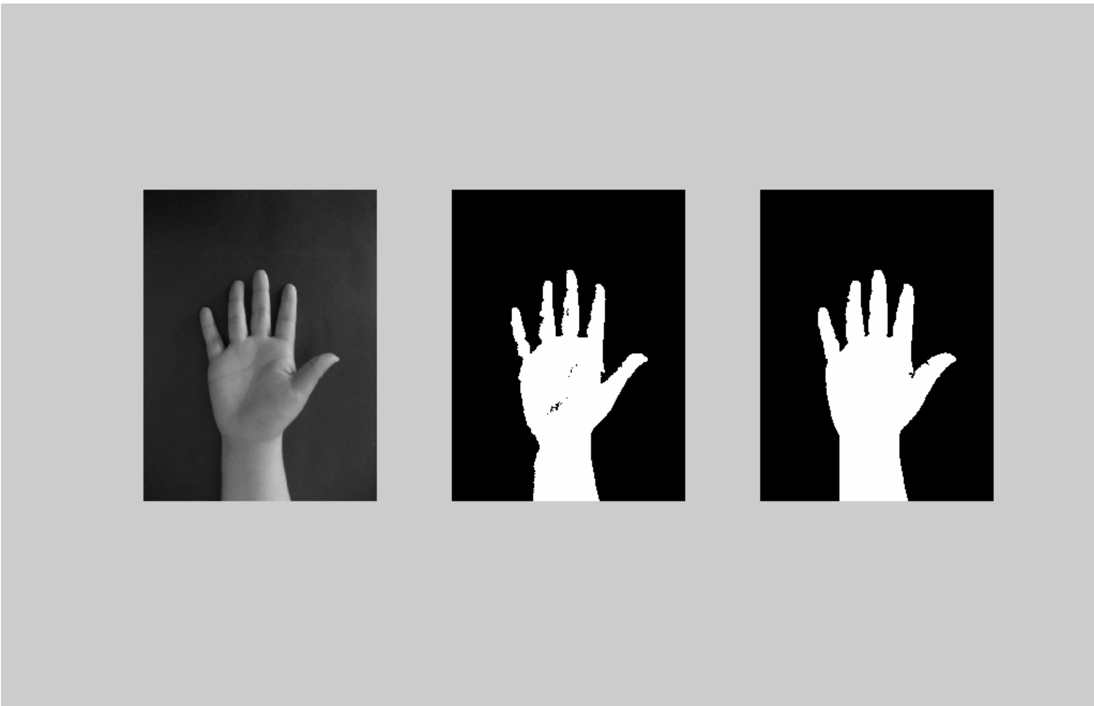


Figure 4.9: Input image, Otsu threshold result, and proposed algorithm result

After binarization process, the hand region is separated from the background and hand region is converted to white while background is black. The binary image is then smoothed with smearing algorithm. Smearing is a smoothing method for eliminating the noisy areas on an image. In this algorithm, image is processed along vertical and horizontal runs as explained in Chapter 3. In this study, there have been some small black regions inside the hand region, and these are eliminated with smearing. In other words, these black regions are converted to white to obtain fully separated image as a hand region and background. The smearing algorithm is made horizontally or vertically through the image, and threshold value is selected as 5 for both horizontal and vertical smearing.

This can be formulized as given below,

$$\begin{array}{l} \text{If number of 'black' pixels} < 5 \quad ; \quad \text{pixels become 'white'} \\ \qquad\qquad\qquad\qquad\quad \text{Else} \;\; ; \quad \text{no change} \end{array} \tag{4.9}$$

The problem that there may be some black regions inside the hand region does not exist in all input images but it gives good result for the images having the problem mentioned, and it does not affect the images that have not the problem. The positive effect of smearing algorithm can be seen in Figure 4.10.

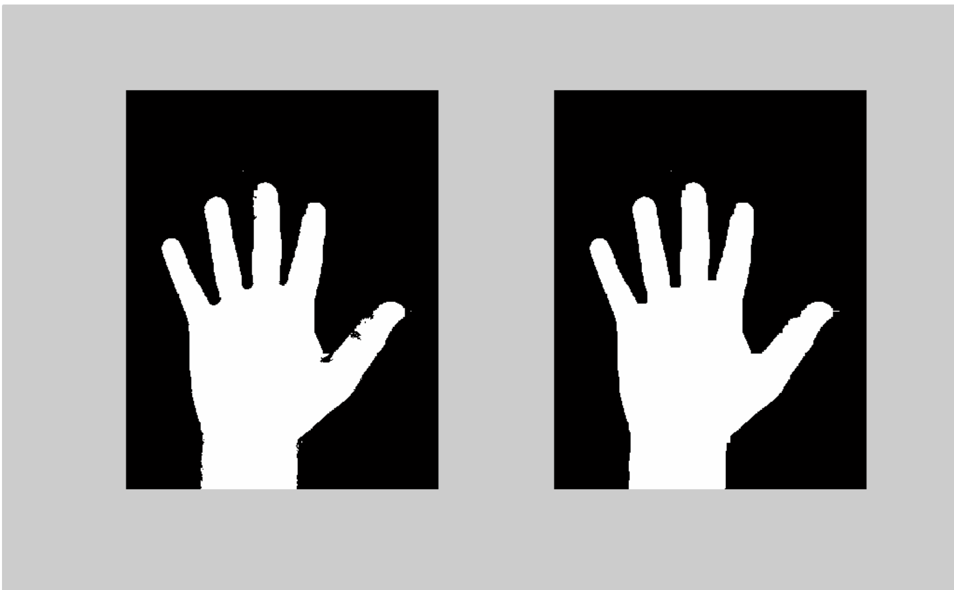


Figure 4.10: Binary images before and after smearing, respectively

After binarization and smearing process, the next step is to obtain the hand contours from the image. This step is also called boundary extraction. The image before this step is shown in Figure 4.11.

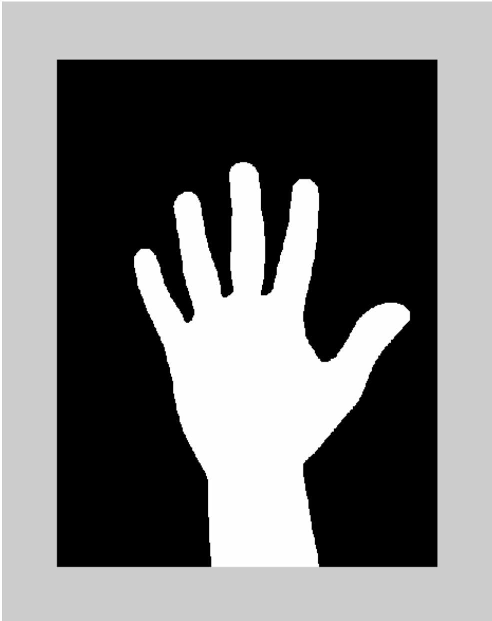


Figure 4.11: Binary input image after thresholding and smearing

The boundary of a set A, denoted by $\beta(A)$, can be obtained by first eroding A by B and then performing the set difference between A and its erosion. That is,

$$\beta(A) = A - (A \Theta B) \tag{4.10}$$

where B is a suitable structuring element and Θ operation is morphological erosion operation [22].

Using similar approach; the boundary, the contours of the hand, is obtained using the morphological erosion, dilation operations and median filtering. The image is first eroded with a 10x10 structuring element which is identity matrix with ones on the diagonal and zeros elsewhere. Then the eroded image is dilated with the same structuring element. After that, the image is filtered with a median filter in a 3x3 neighborhood. Then the filtered image is eroded with a 5x5 structuring element with ones everywhere. Finally the filtered image and the image after last erosion are used to find the contours of the hand. The pixels where the two images have the same intensity pixel values are labeled as the boundary points between the hand and background. For this algorithm, the boundary points are labeled as black, 0 intensity value, and white, 1 intensity value, elsewhere on the image as shown in Figure 4.12 below.

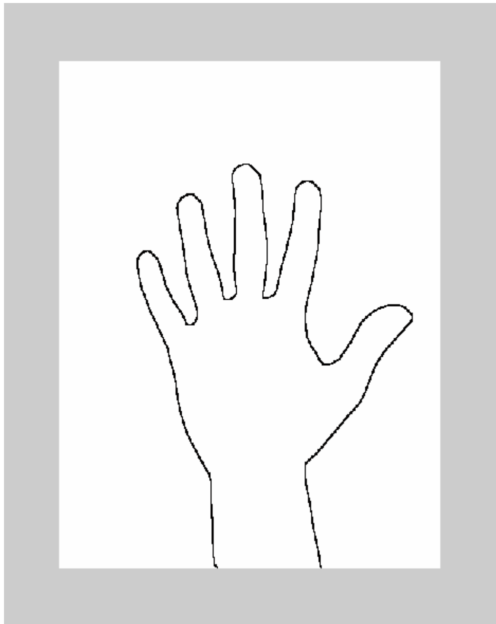


Figure 4.12: The image showing contours of the hand

4.3 Feature Extraction

The purpose of feature extraction is to obtain the characteristics of the hand image. As it is known that each human hand has its own characteristics and it is unique, the feature extraction part obtains these characteristics features for the recognition.

After obtaining the contours of hand image, the next step is to obtain the features of the hand. This study uses the lengths of each finger, widths of each finger at three different heights, and width of the palm as a feature vector. In order to get these features, the proposed algorithm uses the contour pixels and their locations. First, NOT (Complement) operation is implemented. This is because of having the contour points as white and elsewhere black. NOT function takes the complement of the image. That is to say, it converts the white pixels to black, and the black pixels to white as it is shown in Figure 4.13.

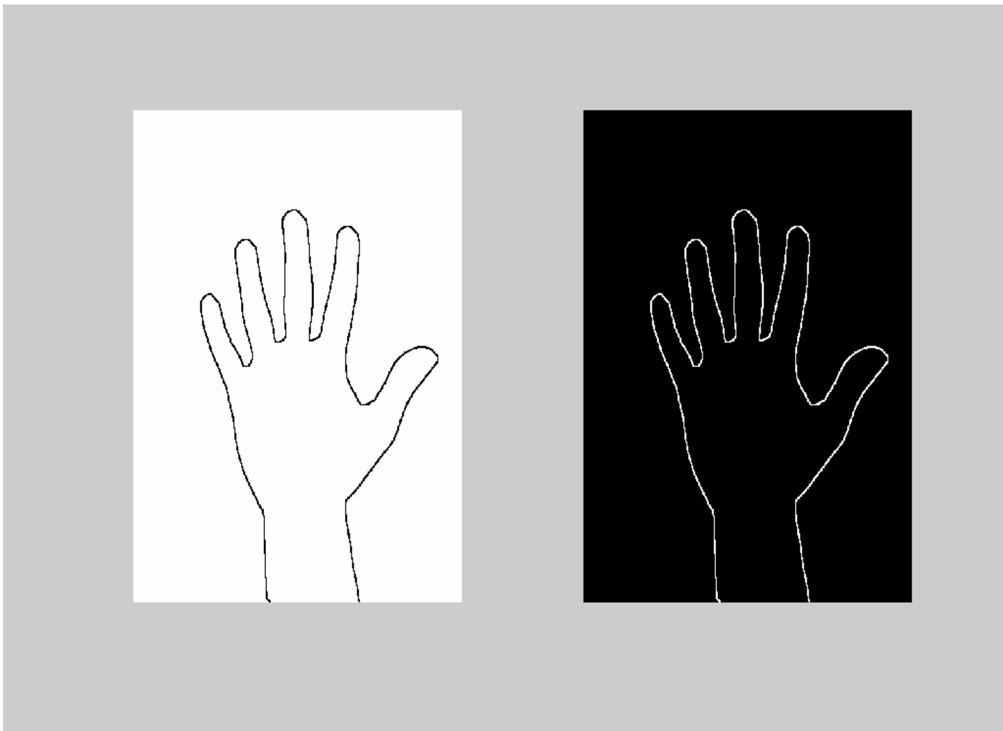


Figure 4.13: NOT (complement) operation

In this study, the locations of contour pixels are found using "find" command in Matlab. This command finds and enrolls the locations of the non-zero elements in an array, matrices or the images. For that reason, contour pixels' intensity values are converted from zeros to ones (black to white) with the complement operation. Obtaining the locations of the contour pixels is extremely important because it enables us to process or deal with only those pixels, not all the pixels in the image. It provides quick processing algorithm.

To obtain the characteristic features of the hand, hand and finger extremities are to be needed. These extremities are fingertips, valley points, and wrist location of the hand that are given in Figure 4.14.

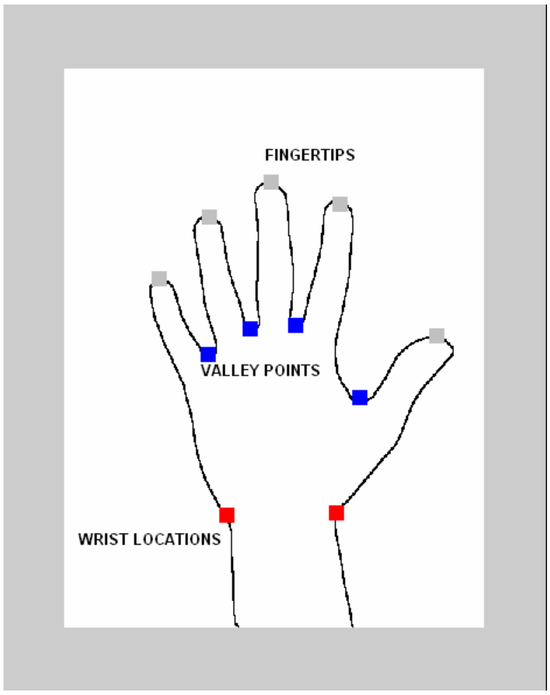


Figure 4.14: Fingertips, valley points and wrist locations of a hand

First wrist locations are found. Then a reference point is selected on the wrist location that is the middle point between the left and right limit points of the wrist.

The next step is to find the fingertips of all fingers. These points have the property that they are the farthest points from the reference point for each finger. In other words, maximum distance values are between the reference point and the fingertips of each finger in the neighborhood of corresponding fingers. In this study, distance measurement is done with the Euclidean distance formulation as explained below:

Let a point on the hand contour be (X, Y), and the reference point on the wrist location be (X_{ref}, Y_{ref}).

The Euclidean distance, D_E, between the point on the contour and the reference point is,

$$D_E = \sqrt{(X - X_{ref})^2 + (Y - Y_{ref})^2} \tag{4.11}$$

The proposed algorithm searches the maximum Euclidean distances for each finger in the corresponding finger neighborhood. The algorithm uses windowing to analyze the subregion (neighborhood of each finger) and calculates all the distance functions in this neighborhood. Then the algorithm finds the maximum Euclidean distance and registers this point on the contour as the fingertip. This algorithm is implemented for all five fingers.

The windowing, the selection of neighborhood size, is done as the following: A subset is selected and the algorithm searches maximum Euclidean distance in this subset. Then the subset is enlarged and maximum distance is searched again in the enlarged subset. If these two maximum Euclidean distance values are equal, the same, then this point is thought as the fingertip. Otherwise, this is not the fingertip, just having the maximum distance value in the selected subset, not in the finger neighborhood.

The similar approach is used for finding the valley points between the fingers. Using the Euclidean distance measurement that is the comparison of distances between the points on the hand contour and the reference point, the valley points between the fingers are found. Calculating the Euclidean distances on the neighborhood, minimums are found giving the valleys between the fingers. In other words, the minimum Euclidean distance is searched in this part and the point having the minimum distance is labeled as a valley point. In order to make the algorithm simpler, the image showing the contours are cut. That is, the bottom regions of the image are cut. This is made for better and quicker process in finding the minimum extremities, valley points between the fingers. This operation is done with the AND operation. Since the peak points, fingertips, and wrist location are known, the selection of the region is made using these parameters as the limit points.

The region that is needed for searching the valleys is obtained by multiplying the input image with the image that has ones in the region of interest and zeros elsewhere. It is known that the multiplication, AND operation, of two binary images at any location yields 1 only when both the corresponding pixels in the two input images are 1. The operation is shown in Figure 4.15. Input image, multiplier image, and the output image after the operation are given respectively.

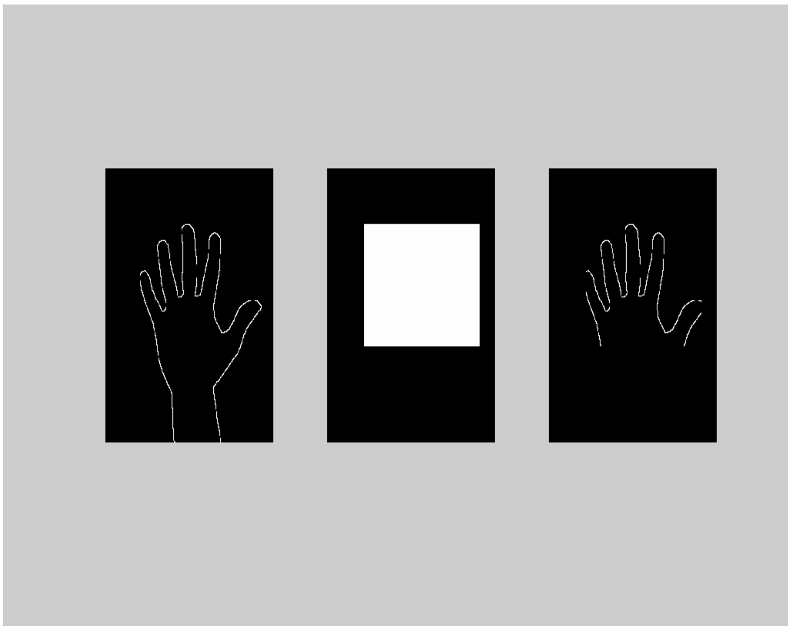


Figure 4.15: AND operation

It is easily seen that there are five fingertips, but four valley points between the fingers. It is known that to find the widths and lengths of the fingers, two valley points are needed for each finger. However, for a thumb, an index, and a little finger; each has only one adjacent valley point. Thus, in this research, the other valley points are assumed to be on the opposite side of the finger with the same distance from the fingertip to the existing valley point. The proposed algorithm searches the equal-distant point on the contour to the fingertip in the corresponding finger. The valley points for thumb, index and little finger are shown in Figure 4.16.

Now the system has a fingertip and two valley points for each finger. These parameters are necessary for calculating the length and width of the fingers. On the other hand, the system needs to other parameters related with these extremities. These are called “finger baselines”. Finger baselines are the lines that are between the valley points on both sides of corresponding fingers so finger baselines are obtained by connecting the valley points on both sides of corresponding fingers.

The finger baselines for each finger are shown in Figure 4.17. Since we have four original valley points and three virtual valley points, we can construct a baseline for each finger.

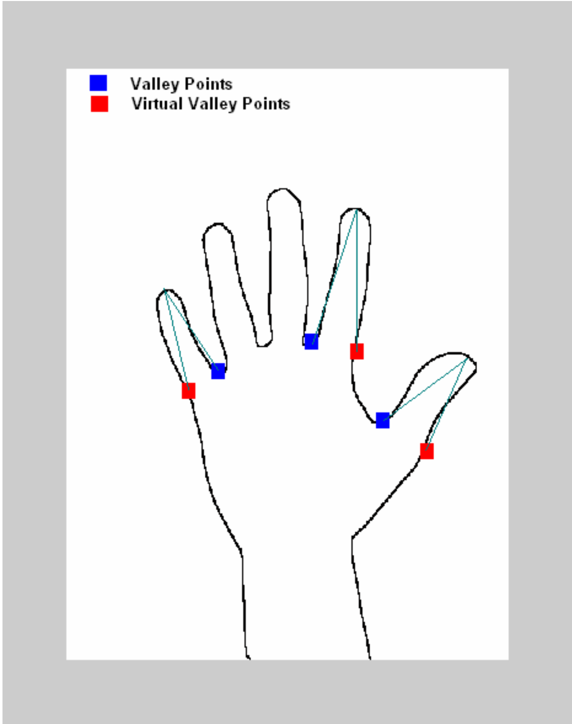


Figure 4.16: Valley points

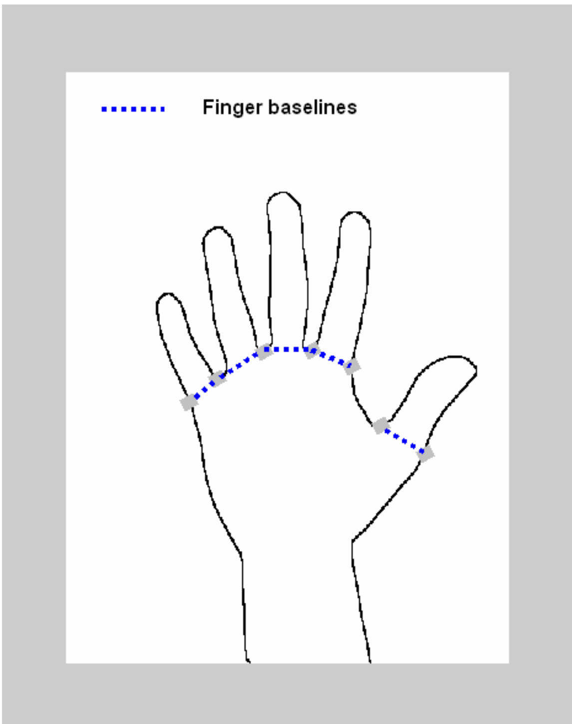


Figure 4.17: Finger baselines

4.3.1 Finger lengths

Proposed algorithm uses finger lengths as the elements of characteristic vector (feature vector). The length of a finger is the measure of distance between the fingertip and baseline of that particular finger. In order to calculate the finger length, the system measures the Euclidean distance from the fingertip point to the middle point of baseline. This calculation is implemented for all fingers and the finger length values are registered as the elements of feature vector. Visual representations of finger lengths are shown in Figure 4.18.

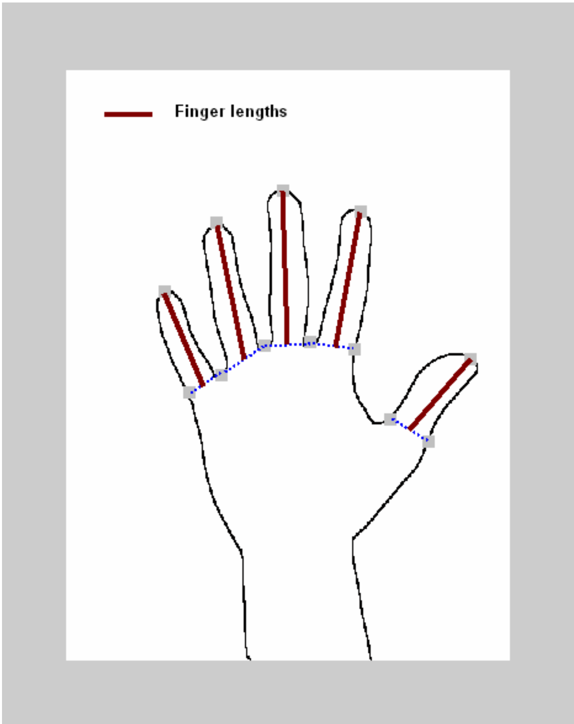


Figure 4.18: Finger lengths

4.3.2 Finger widths

In this study, widths of each finger are also used as the elements of feature vector. As it is known that the widths of the fingers may vary along the fingers, the width measurement can be done in various points along each finger. This approach enables us to obtain more distinct features of the fingers. In this research, finger widths are measured at three different locations along the finger. Finger widths are the widths of the fingers measured from left contour points to right contour points for 3 different locations. The Euclidean distance calculation is used to measure the width distances.

For each finger, the first distance measurement is done on the centre of the finger length, middle point , the second one is on the three-eighth and the third one is on the five-eighth of the each finger lengths as shown in Figure 4.19.

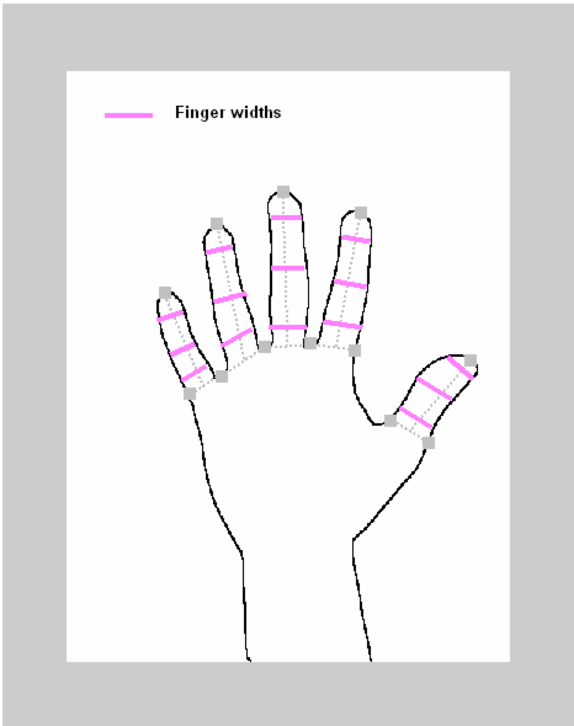


Figure 4.19: Finger widths

4.3.3 Palm width

The width of the palm in an hand is also one of the characteristic features so this study uses the palm width as an element of feature vector. Palm width is the distance measurement from a point on the left contour of the palm region to another point on the right contour of the palm region. As in the finger lengths, and the widths of the finger, the palm width is calculated using the Euclidean distance function. The schematic representation of palm width is shown in Figure 4.20.

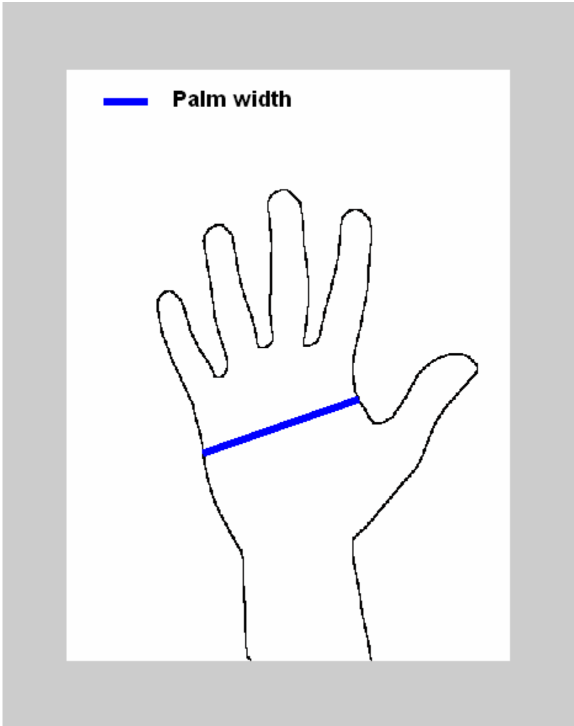


Figure 4.20: Palm width

As a result, proposed algorithm uses totally twenty-one features for the feature vector, including five finger lengths (one for each finger), fifteen finger widths (three widths for each finger) and palm width.

4.4 Recognition

The system registers the extracted features and matches the feature vector with the database for the recognition. The problem of recognizing the identity of a person can be categorized into two major types: verification and identification.

Verification deals with to the problem of confirming or denying a person's claimed identity. In a verification system, an individual claims as one of the authorized users previously enrolled to the system. The system confirms or denies the claimer by matching the individual's feature vector (extracted data) with those of the claimed person which is stored in a database. Thus, a verification system is a one-to-one matching process.

Identification concerns about the problem of establishing a subject's identity. In an identification system, the claimer's feature vector showing some characteristics of the claimer is matched with those of all registered individuals. The system recognizes, or identifies, the individual. Therefore, identification does a one-to-many matching process.

Matching is an effective algorithm for recognition. The feature vector is compared with the ones in a database and the best similarity is measured. With comparing the input image with the ones, templates, in the database, the best similarity gives the correct recognition. Minimum distance classifier and matching by correlation are the ways of matching used in hand recognition as explained in details in Chapter 3.

To make the matching for verification and identification purposes, some distance functions based on minimum distance classifier and correlation function are used in this study. Four distance functions are examined. These functions are given as the following:

- Distance Function-I

$$D_I = \sum_{k=1}^{n} |u_k - d_k| \tag{4.12}$$

- Distance Function-II

$$D_{II} = \sqrt{\sum_{k=1}^{n} (u_k - d_k)^2} \tag{4.13}$$

- Distance Function-III

$$D_{III} = \sum_{k=1}^{n} \frac{|u_k - d_k|}{u_k + d_k} \tag{4.14}$$

- Distance Function-IV

$$D_{IV} = \frac{1}{n}\sum_{k=1}^{n}\frac{\min(u_k, d_k)}{\max(u_k, d_k)} \tag{4.15}$$

where $U = \{u_1, u_2, u_3, ... u_n,\}$ is the feature vector of an unknown individual or a claimer, and $D = \{d_1, d_2, d_3, ... d_n,\}$ is the database vector.

Including the variances of the database vectors, two additional functions are also examined as follows:

- Distance Function-V

$$D_V = \sum_{k=1}^{n}\frac{|u_k - d_k|}{v_k} \tag{4.16}$$

- Distance Function-VI

$$D_{VI} = \sqrt{\sum_{k=1}^{n}\frac{(u_k - d_k)^2}{v_k^2}} \tag{4.17}$$

where $V = \{v_1, v_2, v_3, ... v_n,\}$ is the variance vector having the entries of the variances of each features in the database vector.

Note that variance is a measure of the average distance between each of a set of data points and their mean value; so it is equal to the sum of the squares of the deviation from the mean value. While the average (or mean) is a measure of the center of a group of numbers, the variance is the measure of the spread. Variance is computed as the following:

$$Var(S) = \sum_{i=1}^{N}(S_i - E(S)^2)/N \tag{4.18}$$

where S_i is the ith element of the set S, N is the number of elements in S, and E(S) is the mean over the values of set S.

Cross-correlation function (CCF) is also a measure of the similarities or shared properties between two vectors. Because there are two vectors as unknown input feature vector and known database vector in this study, cross-correlation is used for the matching algorithm. The correlation between two vectors, $f(x, y)$ and $w(x, y)$ can be written as:

$$c(x, y) = \sum_{s} \sum_{t} f(s,t) w(x+s, y+t) \tag{4.19}$$

Correlation is done using "corr2" function in Matlab. This function computes two-dimensional correlation coefficient between an input feature vector and a database vector. The proposed algorithm matches the individual's feature vector with claimed person's feature vector which is stored in the database for the verification (one-to-one matching process), or the system matches the claimer's feature vector with all registered individuals feature vectors in the database for the identification (one-to-many matching process).

CHAPTER 5

EXPERIMENTAL WORKS

The proposed algorithm and its details about the processing the input image through the three main units (segmentation, feature extraction and recognition) have been explained thoroughly in previous chapter. The experimental works about the algorithm and experimental results showing the performance of the proposed algorithm are given in this chapter. The performance criteria definitions are also given in the chapter.

In this thesis, input of the system is the image of a hand of an individual. The input hand images are captured by a digital camera from 25-30 centimeters away from the hand of the individual. The camera used in this study has 640x480 resolutions. The images are color images. The proposed algorithm uses the right hands of the individuals. The background for the hand input image is a black flat background. That is, the system is designed for the hand image on a fixed background. The user can place a hand freely since there is no need to fix the position of the hand with a peg. The hand can be placed freely but only there is a restriction that the fingers do not touch one another. In this study, illumination is controlled and the images have fixed background; therefore the system works on uniform illumination conditions.

The system is tested with an AMD Turion 64x2 laptop computer 1.81 GHz with 2 GB RAM. The operating system is Windows XP. The system is designed in Matlab 6.5. Matlab is the software produced by The MathWorks Company and it is used for performing mathematical calculations, analyzing and visualizing data, and writing new software programs.

The performance of the proposed algorithm has been tested on about 1000 hand images. The images were taken from 54 individuals. Twenty-seven of the individuals (half of all individuals) are male, and the other half is female. The age interval of the individuals used for testing is between the ages 18 and 55. The sexual variety and age variety are adjusted so deliberately that the system gives best testing over the testers.

As it is said before, right hand images are taken from each test user. For the feature vector, this study uses the lengths of each finger, widths of each finger at three different heights, and palm width. That is, proposed algorithm uses totally twenty-one features including five finger lengths (one for each finger), fifteen finger widths (three widths for each finger) and palm width for the feature vector.

Five hand images from each user are used for forming the database feature vectors. The average values (mean) of each feature obtained from five hand images for each user are kept as the database vectors.

The average value (or simply mean) of a set of numbers is the sum of all the members of the set divided by the number of items in the set.

If we denote a set of data as $X = (x_1, x_2, ... x_N)$, the average (mean) of the set is calculated as the following:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}(x_1 + x_2 + ... x_N) \tag{5.1}$$

For this algorithm the number of members in the set is five since five hand images are used. This process is done to generalize the database vectors.

The variances of each extracted features are also calculated and used for recognition purposes. As explained in previous chapter, variance is a measure of how spread out a distribution is. In other words, it is the measure of variability. This algorithm calculates and uses the variances for each feature in the feature vector.

5.1 Performance Evaluation Criteria and Accuracy Calculation

Proposed hand recognition system can be used for two different purposes: verification and identification as mentioned before. Performance of the system therefore must be thought for verification and identification in different ways. As a result, performance evaluation of the algorithm is done differently for identification and verification.

5.1.1 Identification

In identification, an unknown individual feature vector is matched with all the vectors registered in the database and the algorithm determines or makes a decision that the claimer is one of the registered users or not. As a result, the system identifies the claimer with giving the best match. The performance of the algorithm in identification is evaluated by the correct identification rate or by the system's percent error.

Correct identification rate, or accuracy, can be considered as the ratio of the number of true identification to the total number of identification attempts, and it can be expressed in mathematical expression as:

$$Accuracy = \left(\frac{TotalNumberofCorrectIdentification}{TotalNumberofIdentificationAttempts}\right) \times 100\% \tag{5.2}$$

The performance can be also expressed in terms of percent error as given below:

$$ErrorRate = \left(\frac{TotalNumberofIncorrectIdentification}{TotalNumberofIdentificationAttempts}\right) \times 100\% \tag{5.3}$$

5.1.2 Verification

In verification, the process involves matching a given hand to a person previously enrolled in the database. The claimer feature vector is then compared with the feature vector stored in the database associated with the claimed identity and the system decides that the claimer is right or not. A threshold value is selected for the decision. If the matching, or comparison, has the value greater than the threshold, the system decides that the claimer is right. Otherwise, the system decides that the claimer is not right.

Verification performance is measured from the two types of errors; False Rejection Rate (FRR) and False Acceptance Rate (FAR).

False Rejection Rate, FRR, is the measure of the likelihood that the system will incorrectly reject a claimer. A system's FRR simply is stated as the ratio of the number of false rejections divided by the number of verification attempts.

$$FRR = \frac{TotalNumberofFalse\,\mathrm{Re}\,jections}{TotalNumberofVerificationAttempts} \times 100\% \tag{5.4}$$

On the other hand, False Acceptance Rate, FAR, is the measure of the likelihood that the system will incorrectly accept a claimer. FAR is stated as the ratio of the number of false acceptances divided by the number of verification attempts.

$$FAR = \frac{TotalNumberofFalseAccep\tan ces}{TotalNumberofVerificationAttempts} \times 100\% \tag{5.5}$$

That is to say, there are a total of four possible outcomes:

- A genuine individual is accepted.
- A genuine individual is rejected.
- An imposter is rejected.
- An imposter is accepted.

Outcomes ("A genuine individual is accepted", and "an imposter is rejected") are correct while outcomes ("A genuine individual is rejected", and "an imposter is accepted" are incorrect. In performance analysis, we use FRR and FAR to indicate the identification accuracy of a verification system [6].

If the threshold value is selected bigger, or higher, FAR of the system decreases whereas FRR increases. In the case of smaller, or lower, threshold value, FAR of the system increases whereas FRR decreases. This is shown in Figure 5.1.

For a verification system, the optimal performance of the system is the situation where the FRR equals the FAR.

On the other hand, it may not be true to set the threshold value to achieve the system's optimal performance in practice. Setting the system threshold actually depends on the applications. A system that needs high security does not reject a genuine user, and at the same time, it does not accept an imposter either. Therefore, this kind of system needs a threshold that yields high FRR and low FAR. However, some applications may prefer lower security level to gain users' acceptance. The reason is that genuine users might get annoyed if they are rejected. The threshold for such a system can be set to a value that lower FRR and thus, higher the FAR. Such a system must also be aware to take some risks from imposters.

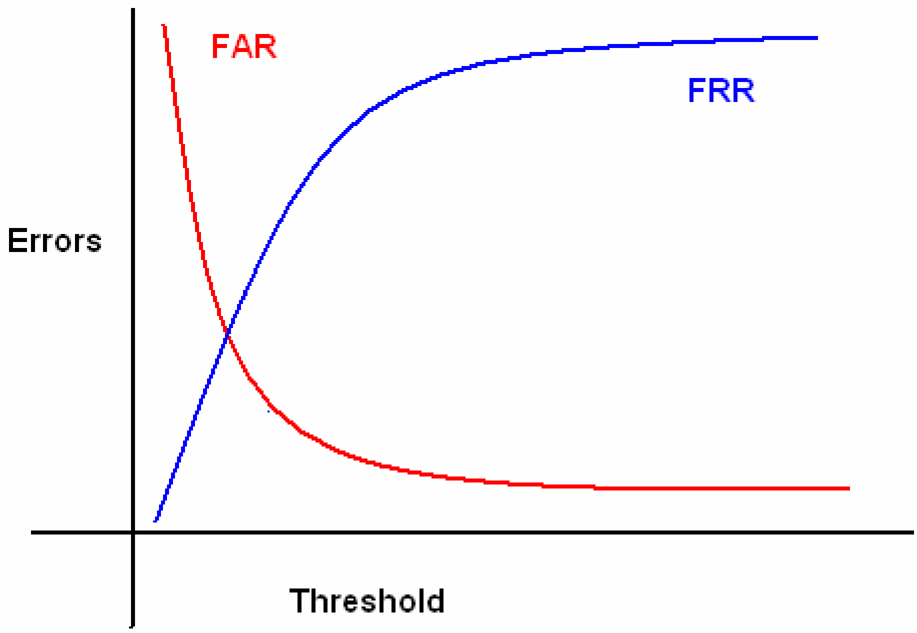


Figure 5.1: Graph of FRR and FAR

5.2 Test Results

As stated before, five hand images from each user are used for forming the database feature vectors. The average values and the variances of each feature for these five hand images are calculated and enrolled as the database feature vectors. The other hand images taken from 54 individuals are used for testing the performance of the proposed algorithm. The algorithm has been tested on both identification and verification tasks.

The proposed algorithm was tested on 983 hand images from 54 individuals. Forty-two hand images are discarded because of the error at extracting the feature vector, so 941 hand images are used. The errors are mainly about finding the extremities (fingertips and valley points) of thumb and little fingers.

In identification, an unknown individual feature vector obtained from an input hand image is matched with all the vectors registered in the database and the algorithm makes a decision that the claimer is one of the registered users or not. The algorithm has been used "corr2" function, and six distance functions defined as:

Table 5.1: Distance functions

Distance Function-I	$D_I = \sum_{k=1}^{n} \lvert u_k - d_k \rvert$
Distance Function-II	$D_{II} = \sqrt{\sum_{k=1}^{n} (u_k - d_k)^2}$
Distance Function-III	$D_{III} = \sum_{k=1}^{n} \frac{\lvert u_k - d_k \rvert}{u_k + d_k}$
Distance Function-IV	$D_{IV} = \frac{1}{n} \sum_{k=1}^{n} \frac{\min(u_k, d_k)}{\max(u_k, d_k)}$
Distance Function-V	$D_V = \sum_{k=1}^{n} \frac{\lvert u_k - d_k \rvert}{v_k}$
Distance Function-VI	$D_{VI} = \sqrt{\sum_{k=1}^{n} \frac{(u_k - d_k)^2}{v_k^2}}$

where $U = \{u_1, u_2, u_3, \ldots u_n,\}$ is the feature vector of an unknown individual or a claimer, $D = \{d_1, d_2, d_3, \ldots d_n,\}$ is the average-valued database vector , and $V = \{v_1, v_2, v_3, \ldots v_n,\}$ is the variance vector having the entries of the variances of each features in the database vector.

The feature vector, U, has 21 elements as given in the vector form below:

U=[L1 L2 L3 L4 L5 W11 W12 W13 W21 W22 W23 W31 W32 W33 W41 W42 W43 W51 W52 W53 PW]

In this vector, L's denote the lengths of each finger (little finger, ring finger, middle finger, index finger, and thumb respectively). W's are the widths of fingers at three different heights (little finger, ring finger, middle finger, index finger, and thumb respectively). For example W11, W12, and W13 are the little finger widths measured on the centre of the finger length, middle point , at the three-eighth and at the five-eighth of the finger length respectively. PW denotes palm width as the last entry.

Identification is made with searching the minimum value in the matching process of Distance Function-I, Distance Function-II, Distance Function-III, Distance Function-V and Distance Function-VI over all database vectors. In other words, the unknown feature, U, is matched with all database vectors, and the database vector having the minimum value is selected as the vector of the unknown person. Minimum-valued matching means minimum difference between the unknown vector and known database vector giving the best match. Ideally, there is no difference in correct identification. Matching has "zero" value in correct identification.

On the other hand, the matching algorithms Distance Function IV and correlation make the decision search the maximum value. U is matched with all database vectors, and the database vector that has the maximum value is selected as the vector of the unknown individual. In an ideal case, correct identification has "one" value. The maximum value which is the near one (100% match) is the best match.

The test results for identification task are given in Table 5.2.

Table 5.2: Identification Performance Test Results

Matching Algorithm	Correct Identification Rate
Distance-I	94.68 %
Distance-II	93.94 %
Distance-III	92.02 %
Distance-IV	94.04 %
Distance-V	56.64 %
Distance-VI	55.15 %
Correlation	91.71 %

Table 5.2 shows that the matching algorithms using the Distance-V and Distance-VI functions have lower correct identification rates, and the algorithms using the Distance-I, Distance-II, Distance-III, Distance-IV distance functions, and the algorithm using correlation function have better correct identification rates.

The matching algorithms using Distance IV function and correlation have good identification rates and they work with the same principle (searching the maximum value as the best match). If these two functions can be combined with their weights, this new function can give better result.

The new function (F_{NEW}) is formed by the equation below:

$$F_{NEW} = 100 \times Correlation + 10 \times Dis\tan ceIV \tag{5.6}$$

In this study, the new function has been tested over the inputs; the performance of this matching algorithm has been found 97.44%. This is the best identification rate in all matching algorithms tested in this work.

In verification task, the algorithm makes the recognition by matching an input hand image's feature vector with a person hand image feature vector previously registered to the database. According to the comparison between the input feature vector as a claimer and registered feature vector, the algorithm decides that the claimer is right or not.

Verification performance can be measured with two forms of errors; FRR and FAR as mentioned before. The FRR is the percent error of a system that rejects genuine users as imposters while FAR is the percent error of a system that accepts imposters as genuine users. The optimal performance of the system is obtained when the FRR equals FAR. If the threshold that the system decides according to is selected as FRR is equal to FAR, then the system's correct verification rate is obtained. The proposed algorithm was tested with eight matching algorithms including the sum of weighted Distance-VI and weighted-correlation function, and the optimal threshold values have been adjusted for each matching algorithm so that the system's FRR and FAR errors are equal. The results are summarized in Table 5.3.

Table 5.3: Verification Performance Test Results

Matching Algorithm	Correct Verification Rate
Distance-I	97.23 %
Distance-II	96.06 %
Distance-III	95.88 %
Distance-IV	97.07 %
Distance-V	82.46 %
Distance-VI	71.18 %
Correlation	95.85 %
F_{NEW} (Sum of weighted-Distance-IV and weighted-correlation)	98.72%

Table 5.3 shows that the best verification rate is obtained with the sum of weighted-Distance-IV and weighted-correlation function having the value of 98.72%. As in the identification case, the matching algorithms using the Distance-V and Distance-VI functions have lower correct identification rates than the other matching algorithm functions.

CONCLUSIONS

In this thesis, we presented a new algorithm for personal recognition. Personal recognition is made using hand recognition. Proposed hand recognition algorithm uses hand images of the individuals to recognize them automatically. Hand recognition algorithm is based on the geometric features of the hand of the individuals. Hand geometry is a biometric that identifies individuals by the shape of their hands.

Input of the system is the image of a hand of an individual captured by a digital camera. The captured image taken from 25-30 centimeters away from the hand of the individual is first processed through segmentation part that separates hand region from the background and extracts the hand and finger contours. Then, feature extraction unit finds the lengths of the fingers, finger widths at three different heights, and the width of the palm as a feature vector. Finally recognition part compares the feature vector obtained from the input image with the database feature vector(s) and gives the result for recognition.

The system uses the right hands of the individuals. The background for the hand input image is a black flat background. The hand can be placed freely only making sure that the fingers do not touch one another. In this study, illumination is controlled and the images have fixed background; therefore the system works on uniform illumination conditions.

Once the image is taken as a color image, the system first converts it to the grayscale image and then grayscale image is converted to the binary image by thresholding in segmentation part. The new method that uses the histograms of the images is implemented for selecting the threshold value automatically. The value which is the lowest point occurs after the first peak (mode) in the histogram of the image is selected as the threshold value. This proposed method has been tested comparatively with other thresholding methods and it was found that it is the best algorithm for the application where uniform illumination and uniform color background exist. After thresholding process, the hand region is separated from the background and hand region is converted to white while background is black. Then the algorithm extracts the hand contour as the last process of segmentation giving the output to the feature extraction unit. Feature extraction finds and registers 21 features as the feature vector. In the last unit, recognition, the algorithm matches the input feature vector with the one(s) previously enrolled to the system. As a result of the matching, the algorithm makes the recognition.

For observing the performance of the system, the proposed algorithm has been tested on about 1000 hand images from 54 individuals. The system has been tested in both identification and verification processes separately. In recognition part, we compared and tested eight different functions for matching algorithm (six distance functions matching algorithms, correlation function matching algorithm, and a weighted combination of a distance and correlation function).

Finally it is proved correct identification rates to be 94.68 percent for Distance-I, 93.94 percent for Distance-II, 92.02 percent for Distance-III, 94.04 percent for Distance-IV, 56.64 percent for Distance-V, 55.15 percent for Distance-VI, 91.71 percent for correlation, and 97.44 percent for the weighted-combination of distance-IV and correlation functions.

It is also proved correct verification rates to be 97.23 percent for Distance-I, 96.06 percent for Distance-II, 95.88 percent for Distance-III, 97.07 percent for Distance-IV, 82.46 percent for Distance-V, 71.18 percent for Distance-VI, 95.85 percent for correlation, and 98.72 percent for the weighted-combination of distance-IV and correlation functions.

In identification and verification processes, the weighted-combination of distance-IV and correlation functions yields the best performance. The identification rate for this algorithm is % 97.44 and verification rate is % 98.72 giving the least error.

REFERENCES

[1] Ratha, N. and A. W.Senior, (2001, March). Tutorial on Automated Biometrics, *Proceeding of International Conference on Advances in Pattern Recognition*, Rio de Janiero, Brazil.

[2] Anil K,Jain, Arun Ross, Salil Probhaker, (2004, January). An Introduction to Biometric Recognition, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.14, No.1.

[3] J. G. Daugman, (1993). High confidence visual recognition of persons by a test of statistical independence, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, No. 11, pp. 1148-1161.

[4] J. P. Campbell, L. A. Alyea, and J. S. Dunn, (1996). Biometric security: Government applications and operations, http://www.biometrics.org/.

[5] J. G. Daugman and G. O. Williams, (1996). A proposed standard for biometric decidability, *in Proc. CardTech/SecureTech Conference*, Atlanta, CA, pp. 223-234.

[6] Anil Jain, Ruud Bolle and Sharath Pankanti, (1998, August 5). *Biometrics: Personal Identification in Networked Society*.

[7] J. G. Daugman, (1993). High confidence visual recognition of persons by a test of statistical independence, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, No. 11, pp. 1148-1161.

[8] T. A. Dickinson, J. White, J. S. Kauer, and D. R. Walt, (1996). A chemical-detecting system based on a cross-reactive optical sensor array, *Nature*, Vol. 382, pp. 697-700.

[9] Geoffroy Fouquier, Laurence Likforman, Jérôme Darbon and Bulent Sankur, (2007, April). The Biosecure Geometry-based System for Hand Modality, *In the proceedings of the 32nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2007)*, Vol. I, pp. 801-804, ISBN: 1-4244-0728-1. Honolulu, Hawaii, USA.

[10] R. P. Miller, (1971). Finger dimension comparison identification system, *US Patent* No. 3576538.

[11] D. P. Sidlauskas, (1988). 3D hand profile identification apparatus, *US Patent* No. 4736203.

[12] J. R.Young and R. W. Hammon, (1981, June). Automatic Palmprint Verification Study, *Rome AirDevelopment Center*, Report No. RADC-TR-81-161, Griffith AF Base, New York.

[13] M. Arif, N. Vincent and T. Brouard. (2006, April). Personal Identification and Verification by Hand Recognition, *International Conference on Engineering of Intelligent Systems (IEEE ICEIS 2006)*, Islamabad (Pakistan).

[14] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos, (2000, October). Biometric identification through hand geometry measurements, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 22, no. 10, pp. 1168–1171.

[15] A. K. Jain, A. Ross, and S. Pankanti, (1999, March). A prototype hand geometry based verification system, *in Proc. 2nd Int. Conf. Audio- and Video-Based Biometric Person Authentication*, pp. 166–171.

[16] A. K. Jain, A. Ross, and S. Prabhakar, (1998). Biometrics-BasedWeb Access, *MSU Tech. Rep.* TR98-33.

[17] Nongluk Covavisaruch, Pipat Prateepamornkul, Puripant Ruchikachorn, and Piyanaat Taksaphan, (2005, November). Personal Verification and Identification Using Hand Geometry. In *Transactions on Computer and Information Technology,* ECTI-CIT, Vol. 1, No.2.

[18] C. Öden, A. Erçil, and B. Büke, (2003). Combining implicit polynomials and geometric features for hand recognition, *Pattern Recognit. Lett.,* Vol. 24, pp. 2145–2152.

[19] Y. A. Kumar, D. C.M.Wong, H. C. Shen, and A. K. Jain, (2003, June). Personal verification using palmprint and hand geometry biometric, *In Proc. 4th Int. Conf. Audio Video-Based Biometric Person Authentication*, Guildford, U.K., 9–11, pp. 668–678.

[20] Erdem Yörük, Ender Konukoğlu, Bülent Sankur, Senior Member, IEEE, and Jérôme Darbon, (2006, July). Shape-Based Hand Recognition, *IEEE Transactions on Image Processing*, Vol. 15, No. 7.

[21] Boreki.Guilherme, Zimmer.Alessandro, (2005). Hand Geometry Feature Extraction Through Curvature Profile Analysis, *XVIII Brazilian Sypmposium on Computer Graphics and Image Processing*.

[22] Gonzalez and Woods, (2002). *Digital Image Processing*, Prentice Hall (Second Edition).

[23] *Image Processing Fundamentals* [Ebook] http://www.ph.tn.tudelft.nl/Courses/FIP/frames/fip.html.

[24] William K. Pratt, (2001). *Digital Image Processing*, John Wiley & Sons, Inc. (Third Edition).

[25] Hui-Fuang Ng, (2006). Automatic thresholding for defect detection, *Pattern Recognition Letters,* No:27, pp. 1644-1649.