ADDIS ABABA UNIVERISTY
SCHOOL OF GRADUATE STUDIES

Recognition of Isolated Signs in Ethiopian Sign Language

Tefera Gimbi

A THESIS SUBMITTED TO

THE SCHOOL OF GRADUATE STUDIES OF THE ADDIS ABABA UNIVERSITY

IN PARTIAL FULFILLMENT FOR THE DEGREE OF MASTERS OF SCIENCE IN

COMPUTER SCIENCE

JUNE 2014

ADDIS ABABA UNIVERISTY

SCHOOL OF GRADUATE STUDIES

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF COMPUTER SCIENCE

Recognition of Isolated Signs in Ethiopian Sign Language

Tefera Gimbi

ADVISOR:

Yaregal Assabie (PhD)

APPROVED BY

EXAMINING BOARD:

1. Dr. Yaregal Assabie, Advisor ________________________

2. ________________________ ________________________

3. ________________________ ________________________

Acknowledgment

First I would like to thank God for helping me to complete this thesis work.

I would like to express my very great appreciation to Dr. Yaregal Assabie for his valuable and constructive suggestions during the development of this research work. His willingness to give his time so generously has been very much appreciated.

Special thanks should be given to my brother Alemdenekew Gimbi and to my friends Tamrat Amare, Tewodros Tadesse and Tilahun Abdissa for supporting and encouraging me to finish my thesis. I would also like to extend my tanks to all who participated in the data gathering, all friends and classmates who were encouraging me to complete my research work.

Finally, I wish to thank my parents and all family members for their support and encouragement throughout my study.

Table of Contents

List of Figures

List of Tables

Acronyms and Abbreviations

ArSL : Arabic Sign Language

ASL : American Sign Language

DCT : Discrete Cosine Transform

ENAD : Ethiopian National Association for the Deaf

EthSL : Ethiopian Sign Language

HMM : Hidden Markov Model

HSI : Hue Saturation Intensity

RGB : Red Green Blue

SASL : South African Sign Language

TVP : Time Varying Parameters

Abstract

Sign Language is a visual gestural language which is used by deaf people for the purpose of communication. Even if it is widely used in the deaf community, most of the hearing people do not understand it. Due to this communication gap deaf people encounter so many problems in their daily life since they are living with the people who communicate with spoken languages. To narrow this communication gap there should be technological solutions that assist the deaf community. Sign language researches are striving to fill the gap of the communication. The goal of this research is the recognition of Isolated Signs in Ethiopian Sign Language.

The proposed system accepts videos of Isolated signs and get frames in the videos. On each frame, skin color detection algorithm is applied and the equivalent binary image is produced which has white value for skin color and black value for other region. Based on the detected skin regions the hands and head are segmented from other parts of the body since they have very important role in the signing process. Important features are extracted from the segmented body parts and these values are converted into symbols using k-means algorithm. Hidden Markov Models trained using these symbols and Baum Welch algorithm and stored in the database. A Viterbi decoding is applied in the recognition process using the trained HMMs and symbol sequences of the testing Signs which is prepared by the above process.

We evaluated our Isolated Sign Language recognition system and we found a recognition rate of 86.9% using 8 features and 83.5% using only 3 features.

Keywords

Sign Language, Skin Detection, Hidden Markov Model, Recognition

Chapter One: Introduction

1.1 Background

Deafness is the inability to hear. The problem may occur in one or both of the ears. Basically deafness may be inherited or caused by accidents. According to Ethiopian National Association for the Deaf (ENAD) and a report from Central Statistics Authority in 2007, it is estimated that more than 1.5 million deaf people live in Ethiopia [17].

Language is a tool that is used for communication between different people. People express their ideas and thoughts to the community in which they are living. Sign language is a visual-gestural language used in the deaf community [2]. When we say deaf community it includes deaf people, families of deaf people and other people who communicate with deaf people. Sign language is used to facilitate communication and transfer information between the members of the deaf community. It is also used to conduct formal education for the deaf people. Besides these issues, deaf people use sign language to transfer their culture from one generation to the other.

Sign language is a natural language of deaf people that consist of signs, gestures, finger spelling and facial expression. Usually sign languages are identified by the country where they are used e.g., Ethiopian Sign Language (EthSL) [7]. The language is not universal by its nature. We may get variation in sign languages between regions of the same country. Many people think that sign language is a signed version of spoken language, but this is completely a wrong assumption because sign language by itself is a language that has its own grammar and structure.

Signing process is conducted in two ways. The first one is defining specific signs for different objects. There will be a predefined sign to mean father, mother, etc. The other one is using finger spelling. Finger spelling is the representation of written alphabets to signs. According to [2,6] finger spelling is another component of sign language for spelling proper nouns, technical terms, acronyms and words from foreign sign language. The use of finger spelling is limited to hearing people or deaf people who have partial hearing ability. Finger spelling is not preferable by those people who are deaf since their date of birth.

Most of the time signers do not use the finger spelling. They use it when there is no clear sign for the specific situation. Therefore researches focusing on recognition of signs are more appreciated than those researches that are focusing on finger spelling recognition.

Researches in recognition of predefined sign in sign language are divided into two groups. The first one is the recognition of isolated signs. In this type of research the expectation is the recognition of signs which are neither preceded nor followed by other signs. The other group is the recognition of continuous signs in which the signer will create signs continuously without a gap and it is expected to recognize these signs. The challenging task in continuous sign language recognition is identifying the point where one sign is completed and the other one started [6].

Sign language has its own rules, structure and grammar. The following are the basic components of sign language [2].

- Hand shape
- Hand movement
- Orientation
- Location
- Non manual features like facial gestures.

Ethiopian Sign language is originated from American Sign Language with some influence from the Nordic Countries [2,3]. The language also includes local signs which originated from local deaf schools. The language didn't get the chance to be developed and standardize like spoken languages [4]. Ethiopian finger spelling is developed by ENAD in 1971. ENAD also developed Ethiopian Sign Language dictionary in 2008. In Ethiopian sign language there are distinct finger spelling signs for the 33 'Geez Fidel' and their corresponding letters with vowels [17,18].

There are various researches conducted in the area of sign language. These researches can be broadly divided into two groups. The groups are

- Sign language synthesis (translation)
- Sign language recognition

Sign language translation focuses on the conversion of input from different medium such as from text and voice to sign language. This needs the processing of the input and generating of signs which can represent the input [36].

Sign language recognition deals with the capturing and analyzing of signs which are generated from the signer and produce information about the given sign. Different types of techniques are applied to process the captured image or video and detect a specific sign that the signer needs to show[22].

Since this research focuses on the recognition of sign language more emphasis will be given for the recognition process. Previous researches on the area of sign language recognition were based on instrumented gloves. The performer of the sign language should wear gloves which have electronic component to detect the sign and movement of hands. But people usually need their natural environment. Moreover there is a challenge because it is difficult to find such gloves easily. This leads the research path to vision based systems which provide natural environment to the performer. Vision based sign language recognition uses camera to get input signs. In some researches colored gloves which have not any electronic component were used as a marker. This research will also use the vision based sign recognition method to get the input signs. The challenges of vision based recognition systems are accurate detection and feature extraction of interesting body parts such as hand and head and intersection of these parts. Current researches are striving to capture signs without having any marker gloves to have natural environment for the signer [21,24].

Even if various researches are conducted for different sign languages, there are very limited works in Ethiopian Sign Language. Most of the research works on Ethiopian Sign Language focused on the conversion of Amharic text into equivalent signs in Ethiopian Sign Language. There is also a research to recognize Ethiopian finger spelling to Amharic text.

1.2 Statement of the Problem

Sign language is a primary tool for the deaf people to communicate with the community around them. But most hearing people do not have the knowledge of sign language. This situation leads to creation of a huge communication gap between the deaf and hearing people. The problems are expressed as follows:

- A communication gap to interact with the hearing people to accomplish their daily tasks. Most of the deaf people have the problem of writing and speaking. This problem is more visible on those deaf people who lost their hearing ability since birth. The inability of writing and speaking widens the communication gap with the hearing people [3].

- Problems that are encountered in the formal education process that includes
    - Lack of sufficient interpreters for each school in which deaf students are enrolled [1,2].
    - The knowledge of the interpreter is limited to translate the subject matter [1].
- The number of schools that provide sign language translation service are very limited [1,2].
- Most education programs in the electronic media are targeted for hearing people. This is a problem to get necessary and up-to-date information [1].
- The above problems hinder the deaf people from getting equal opportunity in education as well as in participation in community services [1,2].
- The problems are also visible in health institutions. Since health professionals need accurate input from the patient to conduct examination, there should be some way to mediate the deaf patients and the health professional [1].

These communication gaps can be narrowed by the help of interpreters. But this solution is not applicable in all situations because

- There are very limited interpreters [1,2].
- There are some issues such as court cases that must be kept secret from the interpreters [2].

Therefore some technological solutions should be prepared to solve the communication gap between the hearing and deaf people. Thus the research question to be answered in this research is:

- Can we have a system that automatically recognizes isolated signs in Ethiopian Sign Language and translate it to the text from?

1.3 Objectives

General Objective

The general objective of this study is to design a model which can interpret isolated signs in Ethiopian Sign Language and give equivalent text to a given sign.

Specific Objectives

The specific objectives of the study are

- Study the structure of Ethiopian Sign Language.
- Find suitable existing model to extract a sign from a given video.
- Develop a method to convert the extracted sign to equivalent text.
- Develop a prototype which demonstrates the translation of the given sign to equivalent text.
- Test the model.

1.4 Methods

To achieve the objectives of the study, the following methods will be used.

Literature Review

In order to get sufficient knowledge on the area of sign language, different documents will be reviewed. The documents will include books, previous research works, materials from Internet, different articles and other training manuals.

Researches which are conducted for the recognition of other sign languages such as American Sign Language and Arabic Sign Language will also be used in this research work.

Data Collection

All necessary data which is needed to conduct the research will be collected. Various data collection strategies will be followed to acquire the required data. Experts and instructors which are involved in Ethiopian Sign Language study and development will be advised about the data collection. In addition to experts and instructors, native users of sign language who use it in their daily life will be incorporated.

Experimentation and Testing

The developed system will be tested as a whole and per each signer to test its effectiveness. This will help us to see the strengths and weakness of the system.

1.5 Scope and Limitations

The study of sign language recognition focuses on the recognition of finger spelling or the recognition of signs in different sign languages.

This research work focuses only on the recognition of isolated signs in Ethiopian Sign Language. The study will not include the recognition of Amharic Finger Spelling since it was covered by previous studies. The study will not cover the study of continuous Ethiopian sign language. Furthermore the research work will not cover the recognition of non manual feature which is facial expression.

1.6 Application of Results

This research will extend the effort exerted in solving the communication problem of the deaf people with the hearing people. The research will contribute its own part in recognition of Ethiopian Sign Language. The benefits of this research are

- It will be very handy tool for the development of Ethiopian Sign Language Translation system.
- It can be used as learning material for Ethiopian Sign Language.
- It will assist the formal education process.

1.7 Thesis Organization

The rest of the thesis is organized as follows. Chapter 2 describes the overall structure of Ethiopian Sign language and the sign language recognition. Chapter 3 discusses the related researches on Ethiopian Sign language and other sign languages. This chapter will indicate the gap in the researches. In Chapter 4 the detailed description of the proposed model will be presented. Chapter 5 elaborates the implementation of the proposed model. Chapter 6 discuses the evaluation of the proposed system. Chapter 7 is the last chapter which gives conclusions and recommendations for future works.

Chapter 2: Literature Review

This chapter discusses about deafness, deaf communications and the overall structure of sign language and review about Ethiopian Sign language. It also gives review on sign language recognition.

2.1 Overview of Language

Language is a means of communication between human beings [8]. It can be presented either in spoken or in written format. Language in spoken format are expressed verbally and produced by the complex dynamic interaction of articulatory organs and air pressure emanating from the lung. The speech signal generated as a sequence of sound unit carries message of the speaker to the listener [1].

Communication via spoken language can be achieved as long as the speaker and the listener have common communication media. This means the speaker should have the ability to speak and the listener should be capable of hearing.

Sometimes when speech is not convenient, the communication can be effected in written format [1]. Writing is the art of expressing ideas by the use of symbols. These symbols represent letters, words and other language components. To communicate with a language in a written format, skills of writing and reading are mandatory.

Spoken language requires proper functioning of the articulated organ from the speaker side, appropriate medium of communication and proper functioning of the auditory system of the listener. Skills of writing and reading are also mandatory to communicate with a language in a written format [1]. In situations where these conditions could not be met we need to have a third alternative. One of the examples of this situation is the communication of deaf people who do not have the ability of hearing and with the shortage of the skills of writing and reading. The alternative for these people is using visual language.

2.2 Deaf Communication

Deafness is the inability to hear. The problem of deafness can be inherited or caused by various reasons such as complications during birth, some infectious diseases such as meningitis, use of toxic drugs and exposure to excessive noise [2]. Visual language is a primary means of communication for the deaf people especially for those who lost their hearing ability when they are at the age of three or less. Some deaf people can communicate orally using visual information including leap reading to help them understand what is being said [2,14].

There are two basic communication ways which help deaf people to communicate with the hearing people and with themselves. The first one is communication through oral means (lip reading) which is very useful to communicate with the hearing people. The other one is manual means of communication which is expressed using hands and the upper body parts.

Since both ways have their own weaknesses, it is recommended to use their cumulative benefit for conducting education of deaf people. There are two types of mixing these ways. The first one is total communication which is the use of all available oral and manual means of communications. The other one is Simultaneous communication which is the use of oral and manual communication at the same time. In these times the total communication way is more recommended [10].

2.3 Sign Language

Sign is the movement of one or both hands followed by facial expression and body movements to construct gestures which have meaning [1,4]. A sign constructed from the movement of one hand is known as one handed sign and a sign which is constructed from the movement of the two hands is called two handed sign. In signing facial expressions add important information in the emotional aspect of the sign. The term sign in signing represents a word in spoken language. Most of the time sign represents common words. If the word is uncommon or a word which is out of the collection of words which are in use in the daily communication the signer will be forced to spell it by the use of finger spelling [4].

Sign language is a naturally occurred language which is used for the communication of deaf people. It is developed by communities of deaf people who are leaving in different parts of the world. Sign languages are languages by themselves. Like spoken languages they have vocabularies, grammar and spelling (finger spelling) and have comparable language structure with spoken language. With sign language it is possible to manifest deaf people's culture, sing a song, present poetry and anything that a spoken language can do [2]. There is a misconception that sign languages are the sign version of spoken languages [1,2,4].

Sign language is a visual language which is created from signs that are the results of different hand shapes, movements and facial expressions. Since it is a visual language, the abilities of speaking and hearing are not mandatory to use with the language. Sign languages are indigenous. Many countries have their own sign languages. There is a difference between sign languages which are used within the same county just like the difference of dialect in verbal languages [2,4].

2.3.1 Structure of Sign Language

There is a basic difference between spoken language and sign language in the arrangement of sentences [1]. For example, the structure of a sentence in Amharic language is “Subject” + “Verb” + “Object” but in Ethiopian Sign Language the arrangement of a sentence is “Topic” + “Comment” and “Time” may be added before “Topic” to show the event is past, present or future [1]. The following example will illustrate the difference of sentence arrangement between spoken and sign languages.

- **አጠብኩ መኪና (እኔ መኪና አጠብኩ)**

  In this example the comment “**አጠብኩ**” (I wash) is used to illustrate the topic “**መኪና**” (car). The equivalent Amharic sentence is “**እኔ መኪና አጠብኩ**” (I wash a car).

2.3.2 Finger Spelling

Finger spelling is a manual representation of written language, and it is used as a substitute for speech as a live, or face-to-face, medium of communication [1]. Finger spelling is an important component of sign language which is used for complete sign communication. Proper nouns, technical terms, acronyms, initialized signs, loan signs and words from foreign language are signed using finger spelling since there may not be predefined signs for them [12,15]. Signers use their dominant hand to create a word from a series of manual symbols, one for each spelling. Finger spelling is not a sign language by itself but it is used to represent letters of writing systems and numeral systems using hands only [2]. Due to the relationship with writing systems, finger spelling is used by those people who are exposed to verbal languages. The usual signing area for the finger spelling is in front of the chest.

Many countries have developed their own finger spelling. American Finger spelling is the most widely used and studied. The world federation of deaf accepted American Finger spelling to serve as base for international alphabet and base for the development of other manual alphabets. It has 26 signs of which only two (J and Z) have movement [16].

2.3.3 Sign spoken language

There is understanding problem between the deaf and hearing people. This is because deaf people have no ability to hear spoken language and the hearing people are not able to understand sign language. Sign-spoken language uses signs and finger spelling to depict the syntax of spoken language [1]. Sign spoken language is used to expose deaf people to the syntax of spoken language. It is widely used for educational processes.

2.3.4 Signed spoken language

Signed spoken language is an artificial system to represent precisely the spoken language which doesn't include finger spelling [1]. It uses signs from sign language and artificial signs to show the complete representation of spoken language using signs. The main difference with sign spoken language is the use of these artificial signs.

2.4 Components of Sign Language

Sign language is the result of different components of sign. Each component has its own impact in defining the meaning of the sign. The basic components are hand shape, location, movement, palm orientation and facial expression.

Hand Shape is one of the most important components of sign. It is created by stretching and bending of the fingers and the palm. Hand shape is configuration that assumes the beginning to make sign. Shapes can be created either by one hand or by two hands in which both hands may have different shape at the same time. Manual alphabets are the most important hand shapes. But manual alphabets are not the only hand shapes that are used in sign language. There are also additional shapes in addition to manual alphabets.

Palm Orientation is the direction in which the hand is turned (up, down, left, right). It is useful since we can determine the direction of fingers and the back of the hand [15].

Location is a place where the signing is performed. In ASL 75% of signs are formed around the face and neck area because they are easily seen. Location has the power to alter meaning of a sign. There are signs which have the same component but have different meaning due to the location where signing is performed. e.g., Mother and Father. Besides signs with related meaning are formed near the heart and signs related with cognitive concepts are formed near the head [2].

Movement also conveys valid meaning in sign language. Some of these movements are arcs, straight lines and wavy patterns [1]. Among these movements, hand movements take the major role of sign language. Eye movements and torso movement (movement around the chest) may also be used along with hand movement [2]. The movements may be single, double or repetitive. We should notice these movements during signing because missing them may lead to misunderstanding of signs. Frequency of movements shows frequency of location. For example a sign for "see" is frequently performed, it means that someone is staring at the place indicated by the sign. Frequency also shows whether the noun is singular or plural or the distinction between noun and verb [15].

Facial Expressions are important components of sign language and referred to as non manual signs. Like other components facial expressions have the power of changing meaning during signing. For example for yes or no questions eyebrows are raised, eyes open wide and head and shoulder forwarded but for WH questions eyebrows are lowered, eyes are narrowed and head forwarded with slight tilt to the shoulder [2].

2.5 Ethiopian Sign Language

Ethiopia formally begun to use sign language after the 1960s following the opening of deaf schools by American and Nordic missionaries. The missionaries brought the sign language which was in use in their country. For more than 50 years foreign sign language was assimilated in Ethiopian deaf culture and community. The influence of the graduates of these schools is visible in the development of EthSL [11, 18].

Later Ethiopian sign language was developed which contains finger spelling and sign words developed by Ethiopian deaf community [11]. Initially Ethiopian sign language originated from American Sign Language. Even if EthSL originated from American Sign Language there is also influence of Nordic countries sign language such as Finish Sign language [1, 2]. When Ethiopian sign language was developed signs which originated from the local deaf schools were incorporated [1, 2]. Some of the examples of these words are “Injera”,”Habesha”,”Resa”, etc.

2.5.1 Ethiopian Finger Spelling

Ethiopian finger spelling is developed by ENAD in 1971 and later accepted by Ministry of Education. It is known as Ethiopian manual alphabets and is in use in all deaf schools along with American Finger Spelling [17].

American finger spelling is used for describing English words and concepts and other local languages which use latin alphabets such as Oromiffa. Ethiopian finger spelling is used to express words and concepts of local languages that use Geez alphabet like Amharic and Tigrigna [1].

Ethiopian manual alphabets have 33 signs with their corresponding seven movements [4, 17]. In 2009 additional manual alphabet is included for the letter ቨ (Ve). In Amharic languages the 'Geez' do not add anything on the consonant, similarly in the manual alphabet it will not include any movement. The rest six have their own movements. The hand configuration shows the consonants whereas the movement shows the corresponding vowels [13].

Ethiopian manual alphabet is different from American and European manual alphabet because most of the American and European alphabets are static. But Ethiopian manual alphabet use movements for most of the Amharic alphabet which are not 'Geez' which do not need any movement. Like any other manual alphabet Ethiopian manual alphabet also has shapes which have iconic resemblance to represent some of the alphabets [12, 13]. Figure 2.1 shows Ethiopian Finger Spelling.

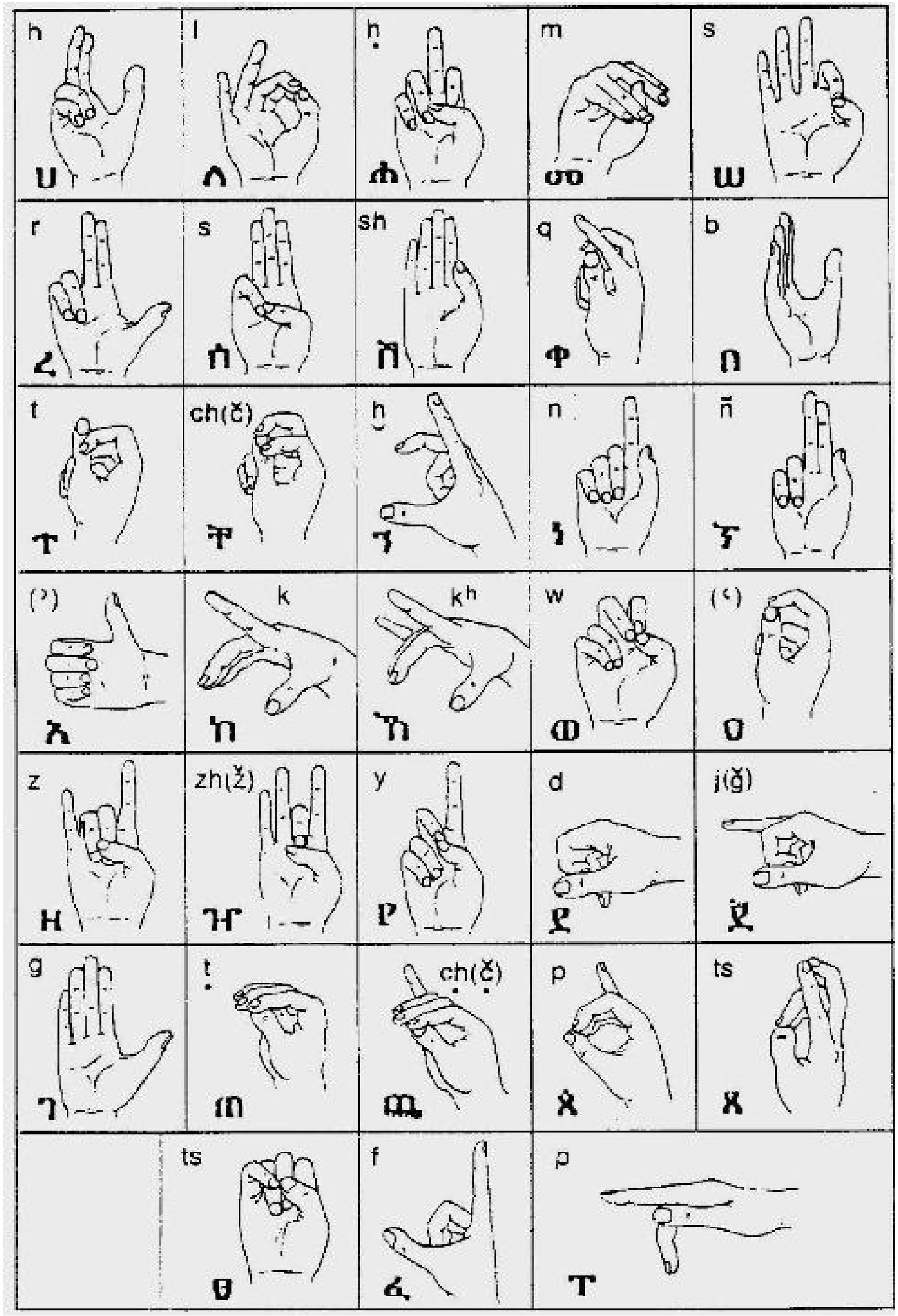


Figure 2.1: Ethiopian Finger Spelling

2.5.2 Signed Amharic

As described in Section 2.3 sign-spoken language uses signs and finger spelling to depict the syntax of spoken language. There is a language which is known as signed English which is different from ASL. It has some components of ASL but it adds additional signs. Similarly signed Amharic is used to facilitate the interaction between the hearing and the deaf people. It is similar to signed English and has an objective of producing signs that correspond to Amharic sentences, in Amharic order. It is used by Ethiopian Television to give information for the deaf people of Ethiopia. Moreover interpreters use it when they sign as the same time as they speak [2].

2.5.3 Explanation of EthSL grammar

As explained earlier in Section 2.4 the five basic components of EthSL are shape, movement, palm orientation, location and facial expression. In addition to these components the following points express the grammatical structure of EthSL.

Dominant hand

Depending of the person the left or the right hand may be the dominant hand. The dominant hand can be the right hand if the signer is right handed or the left hand if the signer is left handed. Signs which are created by one hand are usually performed by a dominant hand. There are three types of signs depending on the use of hands [26]. These are

- One –handed signs:- use only dominant hand.
- Two –handed symmetrical signs- use both hands and both hands move the same way.
- Two-handed non-symmetrical signs- use both hands, dominant hand moves and the other remains stationary.

Signing Area

Signing area is an area where signing takes place. It starts from the area around the head, shoulders and down to the waist. The signing should not go out of this space unless we are speaking for the large audience [26].

Direction

Direction is very important on signing since it will change meaning. It will provide to provide information about the subject and object, e.g., to say ’give’ use the give sign from you to the other body and to say ‘take’ use the same sign with the opposite direction [26].

Tense

In EthSL signs for present are signed in front of the body, past moving the body backward and future by moving body forward [26].

Intensity

In spoken Amharic intensity is shown by adding words. In EthSL intensity is expressed by increasing the speed of signing for the given sign or by adding facial expressions. For example by increasing speed while signing the word ‘Walk’ we can indicate fast walking [26].

Iconic Signs

Iconic signs are those signs that look alike the object they represent. They can be recognized by people that do not know basic sign language [26].

Initialized Sign

Are signs that contain the shape of the first letter of the word. To describe the word the signer is expected to sign the first letter of the word [26].

Person Ending

Person ending signs are those signs that are used to show person’s occupation or nationality. Person ending is performed by combining the sign and pointing to in front of the chest and moving down [26]. Eg. Teach – Teacher, Ethiopia –Ethiopian

Gender

Location shows the gender of some signs. Most male signs are formed on the forehead and female signs in the chin or cheek area[26]. Eg. Father and mother.

Plurals and possessives

Plurals can be expressed by repeating signs or by including signs that indicate number such as "many" after the specific sign. The other one is forming the sign and then pointing to number of locations in the signing are. A possessiveness is rarely used because context is usually used [26].

Negatives

To indicate negative ideas in sign language the signer will include the word "not" after the sign. The other option is shaking the head back and forth while signing [26].

Punctuation

Punctuation is skipped from EthSL since facial expression suffice, if necessary punctuation is created by tracing the punctuation mark by the pointing finger in the air.

Numbers

For each symbols that represent numbers there are predefined signs [26].

Repeating signs

Besides plural forms, in EthSL repeating is used to show continuous action. Continuous action is performed by creating sign with a repeated slow circular movement [2]. If we need to show recurrent action the sign is created with several quick repeated movements [2].

2.6 Sign Language Recognition

In order to improve communication between the deaf and the hearing people, research in automatic sign language recognition is needed. Research in Sign Language Recognition is 30 years behind the research of speech recognition. This is because two dimensional videos are significantly more complex than the one dimensional speech signals. The other reason is the sign language by itself is not fully explored [19].

The major difficulty in sign language recognition when compared to speech recognition is to recognize different attributes of the signer (i.e., hand movements and head movement, facial expressions and body poses) simultaneously.

Sign language recognition systems are classified as signer dependent and signer independent. In the signer dependent recognition systems the signers for the testing phase are those who were involved in the training phase. But in the signer independent systems it is not mandatory to use signers who are involved in training phase for the testing phase.

The other classification of sign language recognition system is based on the input capturing mechanism. Sign language recognition systems which are developed previously used electromagnetic gloves or data gloves which will give the position, movement and shape of the hand to the system. Systems which use data gloves will get more accurate data about the current situation of the hand. The use of electronic gloves conflicts with recognizing signs in a natural context and is very difficult to run in real time [5].

Recently researchers developed Sign Language recognition systems using computer vision techniques. This is very convenient for the signer since the signing process is conducted in a natural context [5]. The input for the recognition systems is captured using video cameras. In some researches the signer will wear colored gloves and other researches try to detect the important signing body parts such as the face and the hands without any indicators. The position of the camera is an important factor that we should give attention. We should set the camera in a position which enable us to capture all body parts that are involved in the signing process. Some researchers use cameras which are capable of getting depth information in the captured video. This will give more information about the sign since some signs could not be expressed in the usual horizontal and vertical directions.

The other point that should be considered during the capturing of videos in systems that use computer vision as an input is the distance of the signer from the camera and the background of the signer. If we have videos of a signer with a same signer but with different distance from the camera, we will get different results for the same sign. Therefore we should think about normalization of the images in the video if we have variable distance of signers from the camera.

Background of the signer is a very important issue in vision based sign language recognition. While we extract body parts that we need for representing a sign, we should be able to identify parts of the foreground and background in the system. There may be colors which resemble skin color that mislead the recognition process by including some unwanted areas as body parts. Not only the color of the background we should also give attention whether the background is static or moving. Moving backgrounds are more difficult by their nature since we can not speculate what color and shape of the background we encounter.

In recognition of signs without using any gloves, skin segmentation will be one major task. To segment a skin color from other parts of the image that we are processing, knowledge of the color space makes our duty smooth.

2.6.1 Color Space Models

A color space is a specification of a coordinate system [27]. Most color models that are in use today focus either towards hardware interpretation or towards applications where color manipulation is a key. Color spaces that are generally used include the Red, Green, Blue or RGB model, the Cyan, Magenta, Yellow, Key or CMYK model and finally the Hue, Saturation, Intensity or HSI model and a subspace within that system where each color is represented by a single point [5,27].

RGB Color Model

RGB is probably the most commonly hardware-oriented model used today. It is used in color monitors and a wide range of color video cameras. Each color in the RGB model appears in its primary spectral components of red, green and blue. Using 24 bit color, each RGB component value ranges from 0, being the lowest intensity that can be displayed on a monitor, to 255 being the highest. All three components are used to produce a single color [5,27]. Figure 2.2 [27] represents an example of the RGB color space. When all three components are combined at their highest intensity, white is produced.

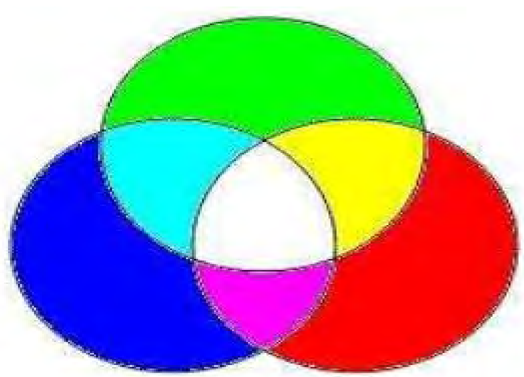


Figure 2.2: RGB primaries

CMYK Color Model

Cyan, Magenta, and Yellow are the secondary colors of light and are the primary colors of pigments [27]. Figure 2.2 shows how CMY component colors are created by combining half of any two primary colors, e.g., when red and blue are combined in equal amounts, Magenta is formed. The K, also known as Key, refers to the black color component. It is formed when CMY colors are combined. The CMYK color space is typically used in color printers and copiers.

The HSI Color Model

The HSI model describes color in terms of hue, saturation and brightness. The hue attribute describes a pure color. Hue is associated with the prevailing wavelength in a mixture of light waves which represent the dominant color as viewed by an observer. Saturation measures to what degree pure color is diluted with white light. Collectively hue and saturation is called chromaticity. Color can be characterized by its color sensation or brightness and chromaticity.

The intensity element expresses color sensation by using brightness as a subjective descriptor. The HSI model owes its usefulness to two important facts. First, by separating the intensity component from color carrying information (hue and saturation), the HSI color model is ideally suited for algorithmic image manipulation that is based on color description [5,27]. Second, the chromaticity component is related to the way in which human beings perceive color [27]. The HSI model is ideally suited for skin segmentation, however hardware uses the RGB color space to communicate with monitors. Since the HSI color space is not natively used by hardware, the RGB color space is converted to the HSI color space.

2.6.2 Feature Extraction

When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input [31].

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or classification algorithm which overfits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

Best results are achieved when an expert constructs a set of application-dependent features. Nevertheless, if no such expert knowledge is available general dimensionality reduction techniques may help [31].There are many features that we can extract from an image. Feature extraction from an image also depends on the area of study to which we need to extract features.

Representation of shapes is broadly divided into two groups. These groups are

- Contour based and
- Region Based

The classification is based on whether shape features are extracted from the contour only or are extracted from the whole shape region. Under each class, the different methods are further divided into structural approaches and global approaches. This sub-class is based on whether the shape is represented as a whole or represented by segments/sections (primitives). These approaches can be further distinguished into space domain and transform domain, based on whether the shape features are derived from the spatial domain or from the transformed domain [33].

Common simple global descriptors are area, circularity, eccentricity, major axis orientation, and bending energy. These simple global descriptors usually can only discriminate shapes with large differences, therefore, they are usually used as filters to eliminate false hits or combined with other shape descriptors to discriminate shapes. They are not suitable to be standalone shape descriptors [33].

Figure 2.3 [33] shows the higherarchy of shape representation and description techniques.

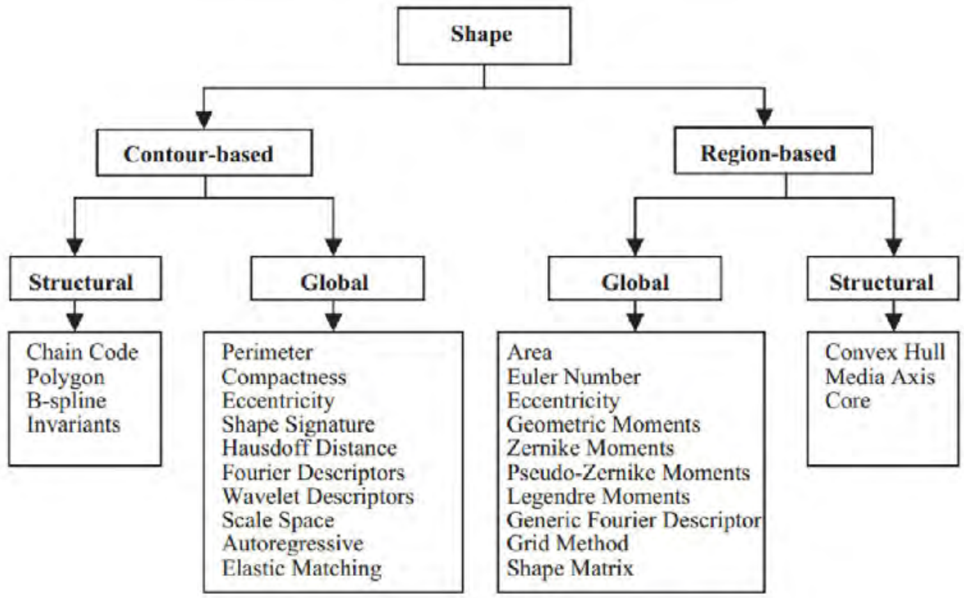


Figure 2.3: Classification of shape representation and description techniques

The feature extraction for sign language recognition focuses on features of the moving hands and head. Important points that could describe these body parts should be extracted efficiently. In addition to this the recognition algorithms for image classifications requires a numerical description of the image to be recognized.

The features that we extract should be [32]

- As complete as possible to represent the content of the information items.
- The features should be represented and stored compactly. The size of the feature vector should not be too large.
- The computation of distance between descriptors should be simple, otherwise the execution time would be too long.

2.6.3 Models for Division of Signs into Subunits

In speech recognition there are components called phonemes which are the smallest element in speech. Similarly there is an attempt to classify signs based on subunits. Subunits are the equivalent term in sign language recognition for phonemes in speech. Instead of modeling the entire signs it is better to model them using subunits. The number of subunits should be chosen in such a way that any sign can be composed with sub units. The advantages are [19]

- The amount of data necessary for training will be reduced as every sign consists of limited number of subunits.
- A further enlargement of the vocabulary is achieved by composing a new sign through concatenation of existing sub unit models.
- The general vocabulary size can be enlarged.

Division of a sign into subunits is the major task in modeling the recognition system. The following models are widely known for subdivision of signs into subunits.

Stokoe Model

William Stokoe was one of the first people to shatter the then commonly held belief that signs are unanalyzable entities. He realized that signs can indeed be broken down into smaller parts , and used this observation for devising a transcription system. This transcription system assumes that signs can be broken down into three parameters, which consist of the location of the sign (tabula or tab), the hand shape (designator or dez), and the movement (signation or sig). Stokoe developed a lexicon for American Sign Language by means of the above mentioned types of cheremes. Cheremes are also the equivalent of phonemes in speech recognition. The lexicon consists of nearly 2500 entries, where signs are coded in altogether 55 different cheremes (12 ‘tab’, 19 ‘dez’ and 24 different ‘sig’) [19,35].

The employed cheremes seem to qualify as subunits for a recognition system. However, their practical employment in a recognition system turns out to be difficult. Even though Stokoe’s lexicon is still in use today and consists of many entries, not all signs are included in this lexicon. Also most of Stokoe’s cheremes are performed in parallel, whereas a recognition system expects subunits in subsequent order. Furthermore, none of the signations cheremes (encoding the movement of a performed sign) are necessary for a recognition system, as movements are modeled by HMMs. Hence, Stokoe’s lexicon is a very good linguistic breakdown of signs into cheremes [19].

Movement-Hold Model

As cited in [35] Liddell and Johnson argued convincingly against Stokoe’s assumption that there was no sequential contrast in ASL. They went even further and made sequential contrast the basis of ASL phonology; that is, instead of emphasizing the simultaneous occurrence of phonemes in ASL, they emphasized sequences of phoneme segments. Such models are called segmental models.

Liddell and Johnson describe two major classes of segments in their Movement Hold model, which they call movements and holds. Movements are defined as those segments during which some aspect of the sign's configuration changes, such as a change in hand shape, a hand movement, or a change in hand orientation. Holds are defined as those segments during which all aspects of the sign's configuration remain unchanged; that is, the hands remain stationary for a brief period of time.

Signs are made up of sequences of movements and holds. A common sequence is HMH (a hold followed by a movement followed by another hold, such as GOOD). Movement segments have features that describe the type of movement (straight, round, sharply angled), as well as the plane and intensity of movement. In addition, attached to each segment is a bundle of articulatory features that describe the hand configuration, orientation, location, and local movements [35]. Figure 2.4 [35] shows a HMH pattern of the sign for Good. The sign consists of a hold at the chin, followed by a movement down and away from the body (left picture), followed by a hold contacting the weak hand (right picture).

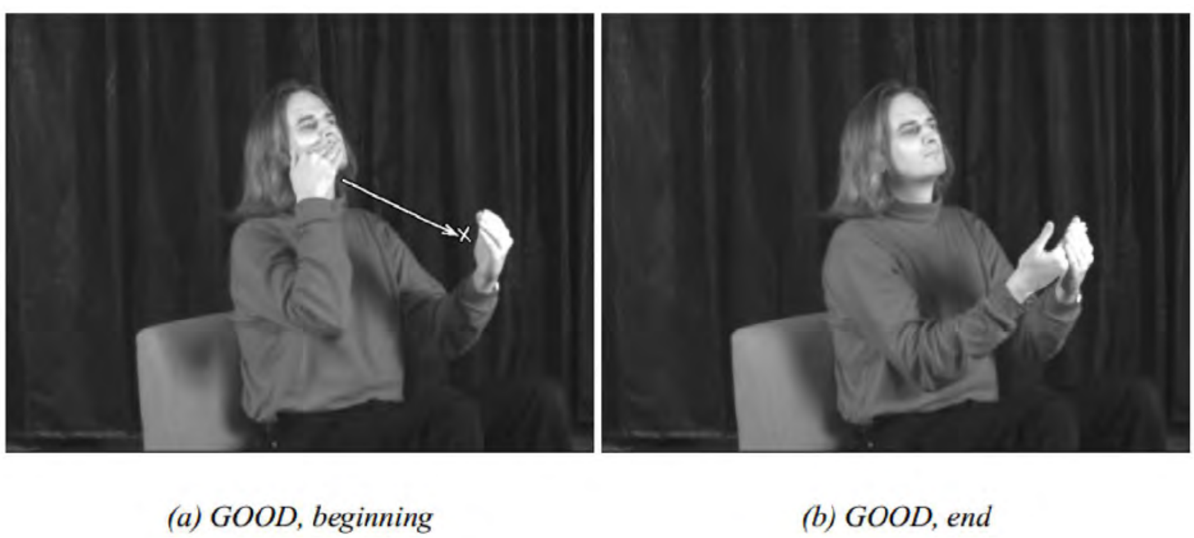


Figure 2.4: HMH pattern

2.6.4 Models for Sign Language Recognition

Hidden Markov Model and Artificial Neural Networks are the two most important models that are widely used for the training phase of the vision based recognition system [37]. For those recognition systems that use Hidden Markov Model, determining the available states is a critical task. We may have various state numbers depending on the type of the problem. The Movement-Hold model that we described above is a good model in defining the states in the recognition of sign language [19].

Chapter 3: Related Work

In this Chapter related works on sign language recognition will be discussed. The review focused on recognition of Ethiopian, Arabic, Indian, South African, Taiwanese and American sign languages. Some of the recognition processes were conducted on static images while the others were conducted on videos.

3.1 Finger Spelling Recognition For Ethiopian Sign Language

The work in [4] discusses vision based finger spelling recognition for Ethiopian Sign Language. This work is a continuation of a previous research work [9] which was able to recognize finger spelling for Ethiopian sign language without considering vowels. The work in [4] extends the study for the recognition of Ethiopian finger spelling with their corresponding vowels. The additions of vowels in the finger spelling will add movements to the basic finger shapes of the consonants.

The work used distance based selection technique to select frames from the input videos. After selecting the frames the next step was preprocessing the selected frames to detect the hand from the other parts of the body. In this process image processing techniques such as converting RGB to Gray scale, adjusting brightness and removing noise with the help of median filter were conducted. Global tresholding was also applied to segment components. Grouping neighborhood was the other technique to select the largest component which was considered as a signed hand. By calculating the distance between the center of masses of the detected signs in two consecutive frames, it was possible to identify the motion of the hand which was very useful in determining the corresponding vowel of the finger spelling. Finite state automata was the preferred technique for the process of recognition of the finger spelling.

The signers wore white gloves with long sleeved black shirts to segment hands from other parts of the body parts using image processing algorithms. The videos were captured in a constant background. Four signers participated in performing 238 Ethiopian finger spellings. The overall performance of the work was 90.76%.

3.2 Sign Language Recognition for non-Ethiopian Sign Languages

The work in [20] uses various methods to recognize South African Sign Language (SASL). The first method was the use of HSI color space over the RGB color space to conduct skin segmentation since the signers didn't wear any marker gloves. The skin segmentation were used to identify the moving hands with respect to the body and from the output of the segmentation process feature vectors were constructed which were later used for the purpose of classification.

The work elaborated that non adaptive approach is more suitable for background modeling. Combination of frame dereferencing techniques were used to eliminate background region. By centering with in video frame (normalizing), the system was able to protect adverse effects of recognition of videos from moving recording devices such as cell phones. The system achieved comparable results in recognizing seen and unseen data using normalization. Hidden Markov Model (HMM) was the tool used for classification purpose.

The work was conducted on 20 SASL gestures which were randomly chosen form English phrase book. Thirty deaf and hard of hearing students participated in signing using short sleeved and long sleeved shirts to ensure robustness. A total of 60 recordings were conducted per SASL gesture. The average result was around 93% and when more and more trainings conducted better results were found.

The idea of the work in [21] started from McConnel's idea of orientation histograms. The scope of the project is to create a method to recognize hand gestures by using histograms of local orientation. Orientation Histograms were used as a feature vector for gesture classification and interpolation. The work used Artificial Neural Network for recognition process.

An image database was built using hand drawn and digitized images. Photographs were also used as they are most realistic approaches. The background was made uniform for those images which have variable background.

Eight training sets each one of which containing three images were used. The number of test images varies for each gesture. There is no reason for keeping those on a constant number. Some images can tolerate much more variance and images from new databases and they can be tested extensively, while other images are restricted to fewer testing images. The training set was converted into feature vector and during the recognition process, the image to be recognized was also converted to feature vector and compared with the feature vector of the training set.

The work was able to show that Histogram orientation has an advantage of being robust under different lightning conditions. The other one was translation invariance which has shown that the position of the hand didn't affect the content of the feature vector.

The work in [22] aimed in converting videos of Indian sign language to text and voice. The work used the power of image processing and Artificial Intelligence techniques. Powerful image processing techniques such as frame differencing based tracking, edge detection, wavelet transform, and image fusion were used to segment shapes in videos. It also used Elliptical Fourier descriptors for shape feature extraction and principal component analysis for feature set optimization and reduction. Database of extracted features were compared with input video of the signer using a trained fuzzy inference system. Matlab was the tool for preparing user friendly Graphical user interface.

Video acquisition process was conducted under controlled environment with good lighting and definite camera displacement. The process was also performed under different lighting conditions.

Head and hands were segmented separately and the areas of segmented head and hand position were found. Then the system applied DCT to extract features and finally a neural network was used to classify the gesturers. The work used Canny operator with double tresholding to mark true edges and Elliptical Fourier descriptors as shape extractors.

Ten signers were participated with constant backgrounds for 80 signs and 91% overall performance was achieved.

The work in [23] is developed for the recognition of Taiwanese sign language. A statistics based context sensitive model is presented. The proposed system architecture was similar with the architecture of speech recognition.

During signing process a discontinuity may occur. To handle this situation the work in [23] proposed Time Varying Parameters (TVP) detection which solved the ending point problem. It was assumed that a gesture is constructed from a set of postures. A gesture is decomposed as a sequence of postures and these postures were recognized quickly using Hidden Markov Model (HMM).

The posture based continuous gesture recognition system consists of three major functional units which are posture analysis, gesture level match, and sentence level match. 71 most frequently used vocabulary and 30 sample sentences were used. The system was able to recognize gestures in real time and only the postures of right hand used.

The work in [25] focused on identifying the ability of different neural networks for recognition of Arabic Sign language. The paper presented feed forward neural networks and recurrent neural networks along with different architectures. Based on the conducted analysis better efficiency was found using fully recurrent neural networks.

The system used white gloves with different colors for each finger tips. The pictures of these hands were taken in RGB color system and converted into HSI color system. 900 different colored images for 30 different hand gestures were used for training and another 900 were used for testing. To perform these gestures two signers were participated by performing which needed 30 repetitions for each gesture. The feature took into account the finger tips and the wrist along with the orientation between them. The results showed the fully recurrent architecture has performance with the accuracy of 95% for static gesture recognition.

The work in [24] was developed using Hidden Markov Models (HMM) and unobtrusive single view camera to recognize American Sign Language (ASL). The system aimed to recognize full ASL sentences using visual cues.

A forty word lexicon consisting of personal pronouns, verbs, nouns and adjectives was used to create 494 randomly constructed 5 word sentences. The data were divided into 395 sentences for training and 99 sentences for test sets. While signing the 2D position, orientation and eccentricity of bounding ellipses of the hands were tracked in real time with the assistance of solidly colored gloves.

With a stronger grammar the system achieved 97% recognition and with no grammar an accuracy of 91% was reached . 95% was an average result for the overall system.

3.3 Summary

There are very few researches which are conducted on Ethiopian Sign Language. But most of them are conducted on the conversion of an Amharic text to equivalent sign in Ethiopian sign language. The researches on Ethiopian Sign Language recognition were mainly focused on the recognition of Ethiopian finger spelling. There are no researches which are conducted for the purpose of recognizing a predefined signs in Ethiopian Sign Language. There are similar researches such as the ones listed on Section 3.2 that are conducted to recognize non-Ethiopian Sign Languages. But these researches are highly language dependent and their effectiveness not yet proved to recognize Ethiopian Sign Language.

Chapter Four: Design of the system

The Chapter describes detailed techniques and ways employed to design and model the recognition of isolated signs in Ethiopian sign language. The system architecture describes the overall design of the system and all the components of the system architecture will be discussed in detail.

The system accepts a video of isolated sign of Ethiopian sign language and generates equivalent text as an output.

4.1 System Architecture

The system has five main components:

- Skin color detection
- Hands and Head segmentation
- Extracting features from the segmented body parts
- Training HMM model using the extracted features and
- Recognition

The system architecture in Figure 4.1 shows how these components interact to accomplish the recognition process. The system architecture is mainly divided into two parts. The first part is the training section in which the system will be trained. The second part is the recognition process which will follow similar step with the first one till the feature extraction step and the recognition process will be continued. The system uses different data which was not used during the training process.

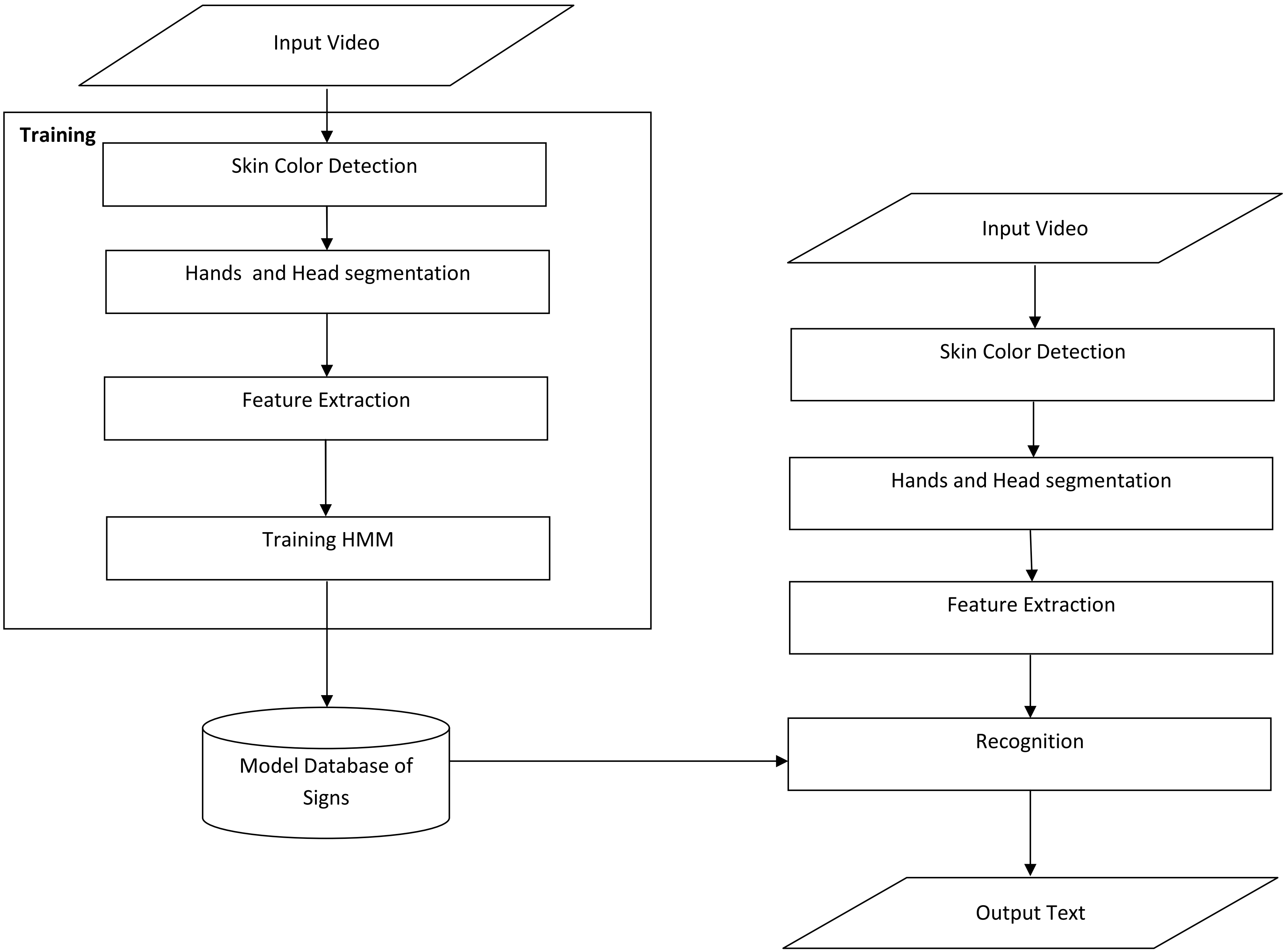


Figure 4.1: System Architecture

4.2 Video Acquisition and frame extraction

The first task before proceeding to any recognition process is the acquisition of the input video in appropriate manner. We captured videos in a way that helped us to extract all necessary information .

To get videos which are suitable to recognition of Sign Language, preparation of the environment in which we capture videos is a vital task. Therefore our video acquisition environment had the following setting.

- Camera stationed in equal distance from all signers to have equivalent signer size in all videos.
- Signers wore long sleeved clothes to assist the hands and head segmentation process.
- Smooth background with a color that doesn't resemble skin color.
- Uniform lighting condition to minimize the effect of illumination in skin color detection.

The skin detection process will be conducted in each of the frames of the videos. Before moving to the skin detection we are expected to extract each frame from the video. At this stage we have a set of sequences of frames for each video and we are going to perform various image processing algorithms in these frames to have the detail information about the given signs. Figure 4.2 shows a frame which is extracted from a video.

Figure 4.2: A frame extracted from the original video

4.3 Skin Color Detection

In the recognition of sign language one of the vital processes is the detection of body parts of the signer which have direct relation in the meaning of signs. To detect these body parts (which are the hands and head) various methods were followed by different researches as discussed in Chapter 2. In this research we used the skin color to segment these body parts.

As indicated in Chapter 2, the HSI model is ideally suited for skin segmentation. However hardware uses the RGB color space to communicate with monitors. Since the HSI color space is not natively used by hardware, the RGB color space was converted to the HSI color space.

We used an algorithm implemented in [28] to detect human skin color. The algorithm sets ranges of values for Hue, Saturation and Intensity in which most of human skin color falls. The algorithm is shown below.

```
1. Get frame I

2. Get HSI value of I    (Hue, Saturation, Intensity)

3. Get the Luma I

  Luma= 0.148* Red(I) - 0.291* Green(I) + 0.439 * Blue(I) + 128;
4. Get the Chroma I

 Chroma =  0.439 * Red(I) - 0.368 * Green(I) -0.071 * Blue(I)+ 128;

5.Get the Size of the image
  w=Width(I);
  h=Height(I);
6.for i=1 to w

7.    for j=1 to h
8.        if 140<= Chroma (i,j)  && Chroma (i,j)<=165 && 140<= Luma  (i,j)  &&
             Luma  (i,j)<=195 && 0.01<=Hue(i,j)  && Hue (i,j)<=0.1

9.              detect(i,j) as skin color
10.       else
11.             detect(i,j) as non skin color
12.       End if
13.   End For
14.End for
```

Algorithm 4.1: An algorithm to detect a human skin color in RGB frame

The algorithm uses the Hue, luminosity and chromaticity values of the frame. The human skin color is studied for variety of people and the range as shown in the algorithm is set. Therefore this algorithm uses the ranges and determines whether the given color falls in the range or not [28].

Since the HSI color space is very important for skin segmentation, we change the video frames to HSI color space as shown in Figure 4.3. Using the results in HSI color space and the algorithm by Mahendran [28], we segment the skin region from other parts of the image.

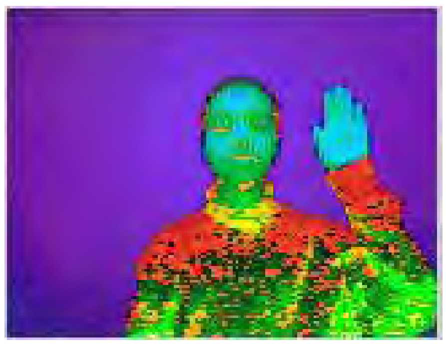


Figure 4.3: A frame in HSI color space

4.4 Hands and Head Segmentation

The Hands and Head Segmentation process has three main steps which are preparation of binary image from the skin segmented frame, smoothing the binary image to remove noise and segmenting the hands and the head using the 8 connected component method. Figure 4.4 shows these steps.

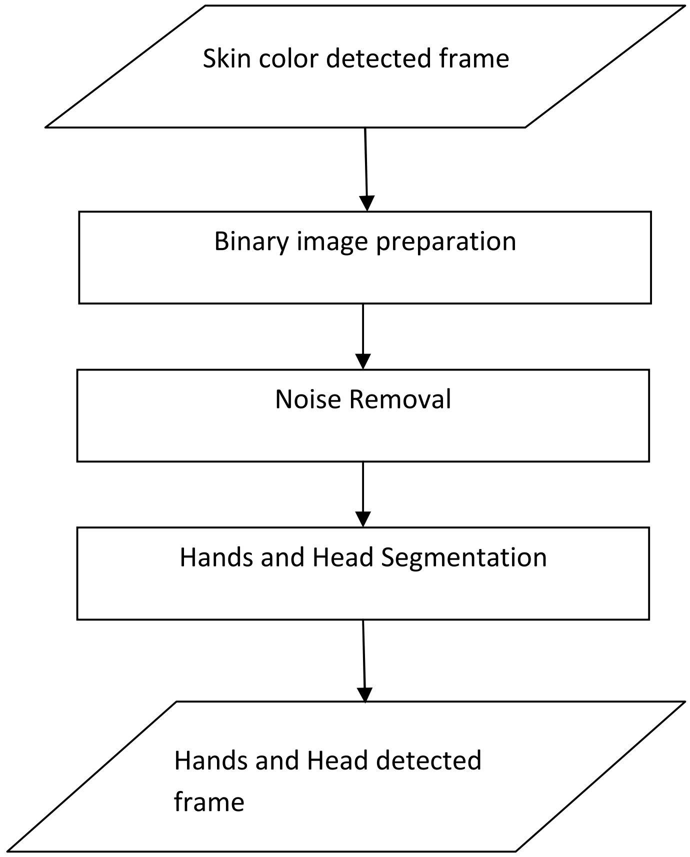


Figure 4.4: Steps for detection of Hands and Head

Binary Image Preparation

At this step our aim is to produce a binary image which will have a white color for the skin regions and black for the rest of the image region. To produce this binary image, first we prepare a binary image which has exactly the same size with a frame under consideration and all pixel values of this image will be set as zero (black). Using the algorithm we used above for skin segmentation, we will check each pixel whether it can be classified as a skin region. If the selected pixel is categorized as a skin region, then we will make the corresponding pixel white in the binary image we prepared ahead. When we finish tracking all the pixels and adjusting pixel value for the binary image, we will have a black and white image showing the skin regions as a white and the remaining parts as black. Algorithm 4.2 shows the steps for binary image preparation.

```
Input: Skin color identified HSI color space image
Output: Skin color identified Binary image
1. Get the frame I
2. Get the size of frame
    w=Width(I)
    h=Length(I)
3. for  i=1 to w
4.            for j=1 to h
5.                    If  I(i,j)= Skin color
6.                        I(i,j)=1
7.           Else
8.                       I(i,j)=0
9.           End If
10.          End for
11.End for
```

Algorithm 4.2: Algorithm for binary image preparation

Noise Removal

During skin color detection, some small components which are not part of the skin may be depicted as skin color. This happened due to illumination at the moment of hand movement. We need to exclude these components from the detected parts. The first way to resolve this issue is the use of filtering. We processed the binary images that we got from skin segmentation activity using median filter. Median filter is chosen because of its noise reduction ability with a better preserving of edges. The filtering process enabled us to remove small components which are not part of the skin. The noise removal process using Median filter is shown in Algorithm 4.3 [29].

```
Input: Image X of size m × n, kernel radius r

Output: Image Y of the same size as X

Initialize kernel histogram H

for i = 1 to m

        for j = 1 to n

              for k = −r to r

                        Remove X(i+k,j−r−1)  from H

             Add X(i+k,j+r) to H

            end for

Y(i,j) ← median(H)

end for

end for
```

Algorithm 4.3: Median Filter

Hands and Head Segmentation

The filtering process by itself is not sufficient to remove all components which were detected as skin. But it is known that the hands and the head are the biggest components in the detected objects. Therefore when we separate the biggest components, the remaining parts are not necessary and can be removed. To achieve the segmentation the 8-connected component method is used.

The connected component will segment all components with different sizes. Since we are interested only on the biggest areas we will select these biggest regions. To do so we used the area measurement of the segmented regions and we dropped with smaller areas as shown in Algorithm 4.4. Now we are able to find a binary image on which only hands head are segmented from other parts of the body. Figure 4.5 shows the binary image after processing using median filter and selecting the largest components using the connected component method.

Input : Binary Image of skin detected area

Output: Hands and Head segmented binary image

1. Get the binary image from Noise removal step

2. Group the white areas using 8-Connected component method

3. Get each grouped area

// Because most of the unwanted components are less than 100 pixels, we used 100 pixels as a bench mark.

4. If Area(Group)<100 Pixels

5. Change each pixel in the area to white

6.Return the changed binary image

Algorithm 4.4: Algorithm to segment Hands and Head

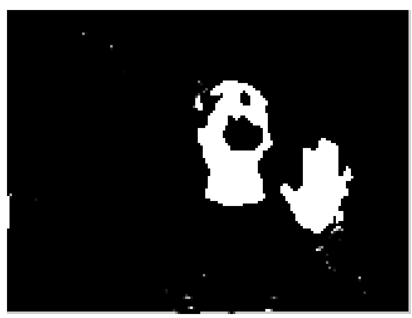


Figure 4.5: A Skin area segmented frame

4.5 Feature Extraction

Features are components of an image which have very important information that could represent the image. The feature extraction for sign language focuses on features of the moving hands and head. Important points that could describe these body parts should be extracted efficiently. In addition to this the recognition algorithms for image classifications require a numerical description of the image to be recognized.

At the end of the feature extraction, every video that we captured for both training and recognition purposes will be converted into set of numbers which represent the image as important features. Generally the representation of shape feature are broadly divided into two groups. This groups are:-

- Contour based and
- Region Based

The classification is based on whether shape features are extracted from the contour only or are extracted from the whole shape region. Both groups help us for representing the shape of the sign since shape is very critical component in defining a sign.

Even though there are ample amount of features to describe the sign under consideration, the following are the preferred ones for our work and their definition based on [34]. The selected features extracted for each regions which are segmented as body parts.

Area: Vector that represents the number of pixels in labeled regions. Algorithm 4.5 used to calculate the areas of each labeled regions computed using the 8 connected component method.

```
Input : Binary Image
Output: Area value of a segmented region in the image
1. Get the Image I
2. Get the size of Image
    w=Width(I)
    h=Length(I)
3. Set Area value 0
   A=0
4. for each segmented region
5.      for  i=1 to w
6.           for j=1 to h
7.                   If  I(i,j)= white
8.                      A=A+1
9.                   End if
10.          End for
11.     End for
12. End for
```

Algorithm 4.5: Algorithm to calculate area of segmented regions

Centroid: M-by-2 matrix of centroid coordinates, where M represents the number of blobs (segmented body parts). The Centroid of each blob is computed using Equation 4.1.

$$X_c = \frac{1}{M}\sum_{i=1}^{n} x_i m_i \quad , \quad Y_c = \frac{1}{M}\sum_{i=1}^{n} y_i m_i \qquad (4.1)$$

where M is the sum of m_i, m_i is the pixel value of segmented region, x_i and y_i are pixel location on the image, n is the total number of pixels.

Bounding Box: M-by-4 matrix of [x y length breadth] bounding box coordinates, where M represents the number of blobs and [x y] represents the upper left corner of the bounding box.

For each blob the bounding box is computed using Algorithm 4.6 [30].

```
Input : Binary Image  of segmented regions
Output: Values of starting position, length and width of the bounding box (m,n,Length,Bridth)
1. Get the Image I
2.For each segmented region
3. Get the size of Image
       w=Rows(segmented region)
       h=Columns(segmented region)
4. m= the minimum value of w
5. n=the minimum value of h
6. m=m-0.5
7.  n=n-0.5
8.  Length= Maximum(w)-(Minimum(w)+1 )
9.  Bridth=Maximum(h)-(Minimum(h)+1 )
10.  End For
```

Algorithm 4.6: Algorithm for computing bounding box for a blob

Major Axis: Vector that represents the lengths of major axes of ellipses. The Major Axis is the axis of minimum inertia passing through the centroid. We can calculate it using central moments as shown in Equations 4.2 and 4.3.

$$tan^2\theta + \frac{\mu_{20}-\mu_{02}}{\mu_{11}}\tan\theta - 1=0 \qquad (4.2)$$

where μ_{20}, μ_{02} and $\mu_{11}$are central moments

The quadratic can be solved for tan θ and

$$Major\ Axis = \arctan(\theta) \qquad (4.3)$$

Minor Axis:-Vector that represents the lengths of minor axes of ellipses. Minor Axis is the axis of maximum inertia passing through the centroid. If maximum axis is known, then the minimum axis computed as the line 90^{o} to the maximum axis. It is calculated using Equation 4.4.

$$Minor\ Axis = Major\ Axis + 1.5714285714 \qquad (4.4)$$

Eccentricity:-Vector that represents the eccentricities of the ellipses. Eccentricity is calculated as given in Equation 4.5

$$Eccentricity = \frac{Major\ Axis}{Minor\ Axis} \qquad (4.5)$$

Orientation:-Vector that represents the angles between the major axes of the ellipses and the x-axis. Orientation is calculated with the formula in Equation 4.6.

$$\theta = \frac{1}{2}arctan\frac{\mu_{02}-\mu_{20}}{\mu_{11}} + (sin\ \mu_{11})\frac{\pi}{4} + \pi n, n = 0,1 \qquad (4.6)$$

Perimeter:-Vector containing an estimate of the perimeter length, in pixels, for each blob. To get the perimeter of the blobs we used Algorithm 4.7

Input: Body parts segmented binary image

Output: Perimeters of the blobs

1. for each segmented regions

2. Get the boundary points of the region

3. Calculate the distance between consecutive boundary points using distance formula

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

4. Perimeter= sum(all calculated distances)

5.End for

Algorithm 4.7: Algorithm to calculate the perimeter of a segmented area

Figure 4.6 shows the Bounding box and Centroid features for the areas which are detected as skin color in the frame.

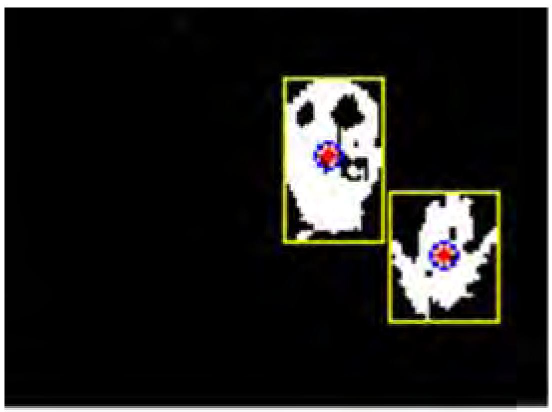


Figure 4.6: Bounding box and Centroid features on the skin detected areas

All the selected feature listed above that are extracted from all videos are stored in the database.

In most sign language recognition researches the basic features (i.e., centroid, area and orientation) that are thought very essential for the recognition process are considered. In this work we need to show the effect of additional features in the recognition process. Therefore we have prepared two parameter sets to represent the features of videos. The first one contains all the eight features listed above. The second one contains only the basic features (i.e., centroid, area and orientation). We will conduct the training and recognition processes for both feature sets and experiment whether we can find satisfactory amount of recognition rate using only the second feature sets or we need to add the other features.

The storing of feature data in the database involves accessing each frame in a given video, getting the required features and then saving the feature data for each frame. The feature data will contain the identification of the sign, the identification of the signer , whether the video under consideration is used for training or testing video and the id indicating the sequence number of the video. The sequence number of the video is used to identify a specific video among other similar videos since we had 15 similar videos for training and another 5 videos for testing purpose signed by a single signer. Table 4.1 shows the feature data extracted and stored in the database.

Table 4.1: Features extracted and saved in the database for the sign "Enat" (The first 16 rows)

SignId	SignName	SignerId	Type	SNo	FrameNo	CentroidX	CentroidY	Area	BoundingBoxX	BoundingBoxY	BoundingBoxWidth	BoundingBoxHeight	MajorAxisLength	MinorAxisLength	Eccentricity	Perimeter	Orientation
1	Enat	1	1	1	1	289.55	131.82	11221.00	240.50	62.50	102.00	140.00	140.75	102.20	0.69	570.82	86.30
1	Enat	1	1	1	2	289.53	131.64	11166.00	240.50	62.50	102.00	140.00	140.06	102.27	0.68	566.58	85.47
1	Enat	1	1	1	3	289.47	131.58	11137.00	240.50	62.50	102.00	140.00	140.28	101.92	0.69	595.20	85.99
1	Enat	1	1	1	4	289.45	131.51	11053.00	240.50	62.50	102.00	139.00	139.56	101.68	0.68	559.06	85.92
1	Enat	1	1	1	4	469.22	466.34	1836.00	428.50	444.50	80.00	36.00	72.35	36.60	0.86	253.58	4.48
1	Enat	1	1	1	5	289.49	131.73	11108.00	240.50	62.50	102.00	140.00	140.40	101.62	0.69	576.13	85.50
1	Enat	1	1	1	5	451.45	457.52	4439.00	386.50	424.50	118.00	56.00	116.94	54.66	0.88	366.17	3.53
1	Enat	1	1	1	6	289.59	131.57	11139.00	240.50	62.50	102.00	140.00	140.66	101.61	0.69	557.16	86.20
1	Enat	1	1	1	6	444.20	443.15	6746.00	358.50	402.50	136.00	78.00	133.11	72.05	0.84	466.45	-16.36
1	Enat	1	1	1	7	289.71	131.45	11135.00	240.50	62.50	102.00	140.00	140.52	101.68	0.69	582.13	85.21
1	Enat	1	1	1	7	436.45	422.29	7275.00	358.50	378.50	136.00	96.00	127.43	79.05	0.78	632.62	-26.79
1	Enat	1	1	1	8	290.17	131.30	11265.00	240.50	60.50	102.00	142.00	140.74	102.43	0.69	551.45	84.87
1	Enat	1	1	1	8	421.75	389.38	6142.00	342.50	352.50	132.00	86.00	125.78	69.62	0.83	549.50	-20.63
1	Enat	1	1	1	9	290.37	131.22	11296.00	240.50	58.50	102.00	144.00	141.33	102.42	0.69	562.52	84.77
1	Enat	1	1	1	9	407.51	354.94	5208.00	330.50	322.50	130.00	76.00	130.30	57.53	0.90	541.61	-17.44
1	Enat	1	1	1	10	290.59	130.99	11238.00	242.50	59.50	100.00	143.00	141.27	101.96	0.69	564.23	84.49

4.6 Training HMM

In order to recognize a sign from a video the system should have the prior knowledge about the sign, which is achieved by training. We used HMM for training and recognition. Training in the context of sign language recognition is taking the features that enable us to represent a sign, converting these feature values to symbols which are suitable to use as an input for the HMM model and training the HMM. For each sign we trained HMM and store it in a database.

The first important factor in representing a sign is tracing the hand trajectory that shows the movement of the hand. This is performed by getting the centroid of the hand in each frame. We used segmented hands and head. Since movement is one of the basic components in the recognition process we should get the trajectory.

The other component which has a significant role in characterizing a sign is the shape of the hand. Two signs may have similar hand movements but we should be able to distinguish them using the hand shapes. Therefore, we took features that are important in describing the hand shape.

We used Eccentricity and Perimeter for contour based shape representation and Area, Major axis and Minor axis of the best fitting ellipse and bounding box for region based shape representations. These features are selected because their cumulative effect is very high in representing shapes. Besides they are very suitable for fast computation [33].

Besides the movement and the shape of the hand the other most important component is the direction of the hand movement. We can have the direction of the moving hand in the signing process using the orientation feature.

In each frame, the hand should be represented by features that describes movement, shape and direction features. In addition to the hand, the features for the head of the signer should be taken in each frame. This is because the hand position should be taken in to consideration with respect to the head.

4.6.1 Vector Quantization

The features that we extracted and stored in our database have various dimensions and value. For a single sign we have multidimensional feature vectors. In order to train our HMM model we should have one dimensional symbol sequence which can be used as an input during the training. The process is called vector quantization. For quantizing our feature values we used the well known k-means clustering algorithm.

The quantization process started with taking all feature vectors from all training data for a specific sign. These features are taken from all signers. After having the whole data for a sign the next task is clustering the data into m clusters using the k-means algorithm which have the same dimensions of the original data.

The size of the clusters was determined empirically. According to our experiment we had a good result when we made the size of m=8. This means, we will have a cluster matrix which has 8 rows each representing feature vector. For each row, we will have a corresponding symbol that is used as an identification of a feature vector. The cluster matrix will be our codebook and the symbol that represents a row in the cluster matrix will be our code word. We will later use this symbol as observation symbol during the training of the HMM mode. Table 4.2 and Table 4.3 show cluster matrices for the sign "Enat" with 8 features and 3 features respectively. Each row of the cluster will be assigned to an Id from 0 to 8 sequentially.

Table 4.2: A cluster matrix for the sign "Enat" with 8 features.

SignId	SignerId	Type	SNo	CentroidX	CentroidY	Area	BoundingBoxX	BoundingBoxY	BoundingBoxWidth	BoundingBoxHeight	MajorAxisLength	MinorAxisLength	Eccentricity	Perimeter	Orientation
1	1	1	1	470.23	370.41	7361.00	392.50	314.50	142.00	116.00	149.79	73.44	0.87	714.46	-34.79
1	1	1	1	344.41	166.71	17180.00	260.50	72.50	198.00	194.00	229.31	121.25	0.85	1149.98	-48.72
1	1	1	1	313.70	195.83	898.00	288.50	168.50	61.00	44.00	68.61	26.79	0.92	243.98	-4.81
1	1	1	1	425.18	316.54	4556.00	344.50	274.50	114.00	82.00	125.30	56.74	0.87	489.66	-17.67
1	1	1	1	355.75	172.97	10267.00	260.50	72.50	190.00	194.00	251.20	120.62	0.88	1985.37	-52.35
1	1	1	1	319.93	148.72	13253.00	256.50	72.50	124.00	154.00	154.35	112.20	0.69	712.04	-82.74
1	1	1	1	295.14	131.00	10974.00	244.50	62.50	104.00	142.00	140.11	101.65	0.69	614.18	84.62
1	1	1	1	320.09	145.16	7782.00	254.50	70.50	122.00	154.00	165.35	116.15	0.72	1534.19	-83.87

Table 4.3: A cluster matrix for the sign "Enat" with 3 features.

SignId	SignerId	Type	SNo	CentroidX	CentroidY	Area	Orientation
1	1	1	1	470.23	370.41	7361.00	392.50
1	1	1	1	344.41	166.71	17180.00	260.50
1	1	1	1	313.70	195.83	898.00	288.50
1	1	1	1	425.18	316.54	4556.00	344.50
1	1	1	1	355.75	172.97	10267.00	260.50
1	1	1	1	319.93	148.72	13253.00	256.50
1	1	1	1	295.14	131.00	10974.00	244.50
1	1	1	1	320.09	145.16	7782.00	254.50

After having our codebook the next task was conversion of the feature data into equivalent symbols in the codebook. To produce the symbols for a given sign, we will take the feature data again which is performed by one of the signers. Next, we will take each feature vector from the feature matrix (a row in the feature matrix) and calculate the distance from the feature vector and the feature vectors that are found in the cluster that we prepared ahead. We will use Euclidean distance algorithm and will take a feature vector in the cluster which has shortest distance with the feature vector that we need to convert to a symbol. Therefore the symbol representing the selected feature vector in the cluster matrix will be the one which will replace the feature data of a given sign. Now if a sign is represented by a set of 100 feature vectors (each of which is representing a frame), we will have a sequence of 100 symbols representing the sign. Similarly we will convert all the training videos for all signers and at the end we have a symbol matrix containing all the training for a given sign. Figure 4.7 shows the quantization process, where n is the number of all frames in all training videos for a sign, k is the number of dimensions, m is the number of clusters and symbols, and j is the number of frames in a single video.

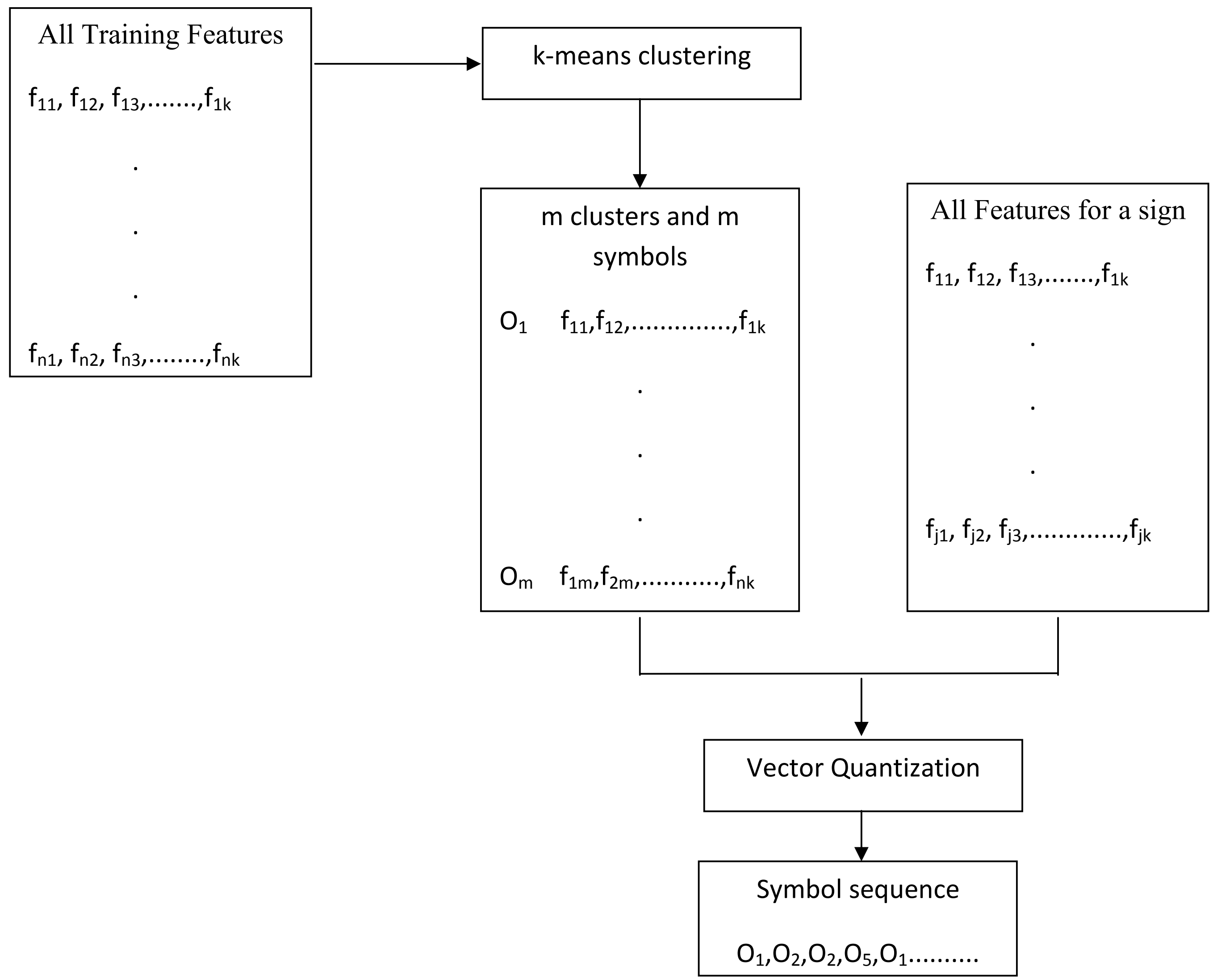


Figure 4.7: Vector quantization process

To change a feature data for the word "Enat" (which is indicated in Table 4.1) into equivalent symbol sequences, we calculated the Euclidean distance of each row in the feature data to the rows of cluster data (which is indicated in Table 4.2). The row of the cluster matrix which has the minimum Euclidean distance from the row of the feature data in consideration is selected. The corresponding symbol representing the row of the cluster matrix represented the row of the feature data. Since we have a cluster matrix with 8 rows, the representation symbols are from 0 to 7. Therefore the feature data indicated in Table 4.1 substituted with the following symbol vector. The vector has 165 symbols because the video for the word "Enat" has 165 frames.

(7,7,7,7,3,7,4,...,8,7,8,7)

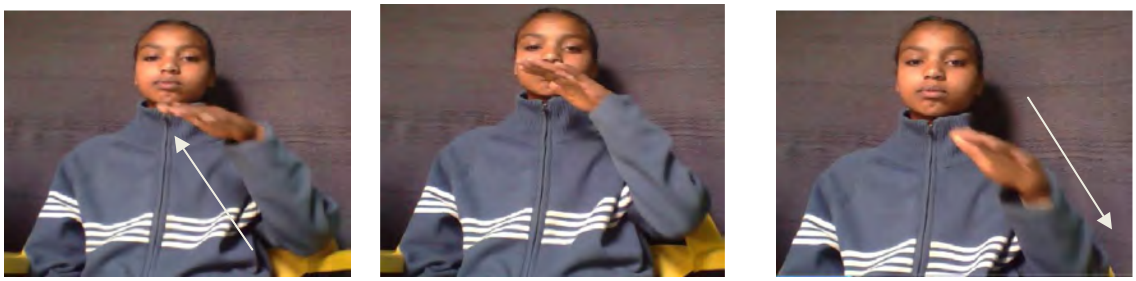
4.6.2 Building HMM

Training the HMM is one of the basic components of the recognition process. Now each sign is represented with a sequence of symbols. For a single training video we have a symbol vector. Therefore, for a single sign, we have a symbol matrix which has 45 rows since we used 45 videos for training a sign. We used the symbol matrix as an input for our HMM to be trained.

In training HMM the basic inputs are the output sequence (which is the symbol matrix in our case), the initial transition and emission matrices. Before the preparation of the initial transition and emission matrices deciding the type of the model and number of states is a very critical task. Depending on the type of problem on hand, various model types and number of states can be selected.

For the purpose of sign language recognition we selected the left right HMM. To decide the number of states, we used the movement hold model described in Chapter 2. When we see most of the signs under consideration, they begin in movement, hold and end with movement again. This property helps us to decide the number of states into 3.

Figure 4.8 shows the different states of the word "Enat" and Figure 4.9 shows its HMM topology.



a. Moving up to the chin b. Holding on the chin c. Moving down to starting position

Figure 4.8: Different states for the word "Enat".

It starts by moving up then there will be hold on the chin and finally it will move down to the starting position.

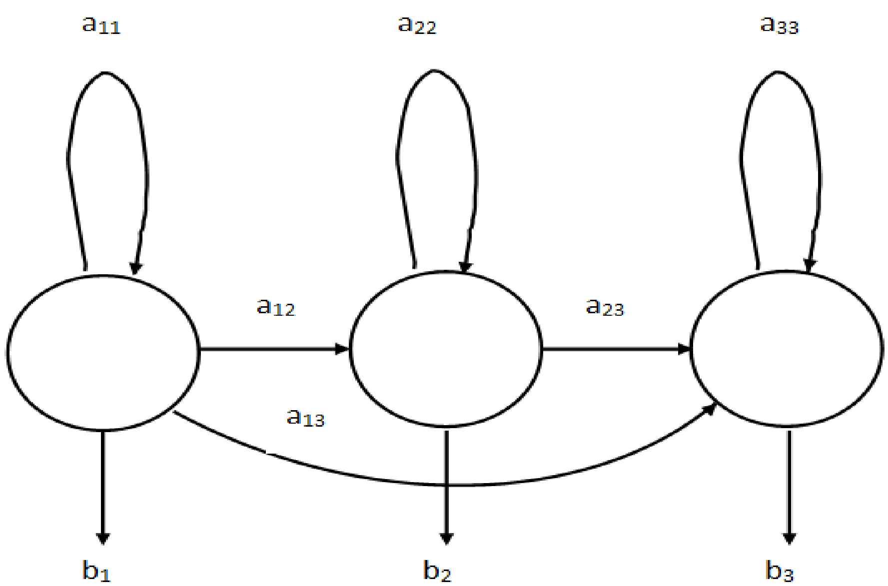


Figure 4.9: A three state HMM topology for the word "Enat"

To train the HMM, in addition to the observed symbol sequence and the decision of the number of states, we need initial state transition and emission matrices. We will prepare these matrices with random numbers. The state transition matrix is an NXN matrix where N is the number of states. The emission matrix is an NXM matrix where N is the number of states and M is the size of the observable symbols which is found empirically during the vector quantization process. Now we can start training our model and the final HMM parameters will be produced by iterative process. We used the Baum Welch algorithm to train the HMM because of it enabled us to compute the parameters of our HMM model. The algorithm updates the parameters of the HMM iteratively until convergence following the procedure below.

The forward procedure : $\alpha_i(t) = p(O_1= o_1, , ,O_t= o_t, Q_t= i| \lambda)$,which is the

probability of seeing the partial sequence $o_1, , , o_t$ and ending up in state i at time t.

We can efficiently calculate $\alpha_i(t)$recursively as:

$$\alpha_i(t) = \pi_i b_i(o_1)$$

$$\alpha_j(t+1) = b_j(o_{t+1}) \sum_{i=1}^{N} \alpha_i(t) \cdot a_{ij} \tag{4.7}$$

The backward procedure: This is the probability of the ending partial sequence $o_t+1, , , o_T$ given

that we started at state i, at time t. We can efficiently calculate $\beta_i(t)$as:

$$\beta_i(T) = 1 \tag{4.8}$$

$$\beta_i(t) = \sum_{j=1}^{N} \beta_j(t+1) a_{ij} b_j(o_{t+1}) \tag{4.9}$$

using α and β, we can calculate the following variables using Equation 4.10 and 4.11:

$$\gamma_i(t) \equiv p(Q_t = i \mid O, \lambda) = \frac{\alpha_i(t) \beta_i(t)}{\sum_{j=1}^{N} \alpha_j(t) \beta_j(t)} \tag{4.10}$$

$$\xi_{ij}(t) \equiv p(Q_t = i, Q_{t+1} = j \mid O, \lambda) = \frac{\alpha_i(t) a_{ij} \beta_j(t+1) b_j(o_{t+1})}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i(t) a_{ij} \beta_j(t+1) b_j(o_{t+1})} \tag{4.11}$$

having γ and ξ, one can define update rules as shown in Equation 4.12 to Equation 4.14:

$$\overline{\pi}_i = \gamma_i(1) \tag{4.12}$$

$$\overline{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_{ij}(t)}{\sum_{t=1}^{T-1} \gamma_i(t)} \tag{4.13}$$

$$\overline{b}_i(k) = \frac{\sum_{t=1}^{T} \delta_{O_t, O_k} \gamma_i(t)}{\sum_{t=1}^{T} \gamma_i(t)} \tag{4.14}$$

Finally we will have our HMM for a given sign. The HMM model will be represented by the state transition and emission matrices produced after training. These matrices will be retained in the database with their respective sign id to use them later for recognition process. Table 4.4 and Table 4.5 show the state transition and emission matrices produced after training for the word "Enat".

Table 4.4: Transition Matrix for the word "Enat"

0.9710	0.0228	0.0063
0.0671	0.9329	0.0000
0.0000	0.0000	1.0000

Table 4.5: Emission Matrix for the word "Enat"

0.0000	0.0000	0.0854	0.0929	0.0060	0.3677	0.2068	0.2412
0.0000	0.4574	0.0753	0.0000	0.3961	0.0409	0.0028	0.0274
1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

4.7 Recognition

Recognition is the process of finding the most probable HMM from a set of HMMs (which were produced during the training phase) that can produce a given observed sequence. For this process we used the Forward and backward algorithms. To compute the likelihood that a model m produced for a given observation sequence, the formula given in Equation 4.15 is used.

$$P(p_1 = O_1, \ldots, p_T = O_T | M) \qquad (4.15)$$

The likelihood can be effectively calculated using dynamic programming by forward algorithm which reproduce the observation through HMM and backward algorithm which back trace the observation through HMM. These algorithms are shown in Equations 4.16 and 4.17.

Forward Algorithm:

$$\alpha_i(t) = P(p_1 = O_1, \ldots, p_t = O_t, q_t = h_i | M) \qquad (4.16)$$

Backward Algorithm:

$$\beta_i(t) = P(p_t = O_t, \ldots, p_T = O_T | q_t = h_i) \qquad (4.17)$$

The same steps will be followed as the training part to have a set of feature vectors that represent the frames and vector quantization process which enabled us to substitute a sequence of feature vector values to a set of distinct symbols. But for testing the recognition system we used different data from the ones that we used for training purpose.

A given symbol sequence for a sign will be checked against each trained HMM and the model with a largest probability will be selected. We prepared the Amharic texts of the signs in jpge image format and save them in their corresponding sign id. After selecting the most probable model, the sign id of the model will be taken and the image with the same id that shows the text of the sign will be displayed.

Chapter 5: Implementation

In this Chapter we will discuss the implementation of isolated sign recognition system for Ethiopian Sign Language. The tools used, the database, the training and testing processes and data used to implement the prototype will be discussed.

5.1 Tools

The following tools are used to develop the prototype.

- Matlab 2012:- to develop the front end component of the prototype. We used all necessary Matlab classes which were helpful in implementing the logics and algorithms in Ethiopian Sign Language Recognition.
- Microsoft SQL server: - To implement the back end of the prototype. We used Microsoft SQL server to create the database and to handle data manipulation activities.

5.2 Data Collection

The input video was captured with a laptop web cam camera which has frame rate of 30 frames per second. The position of the camera is stationed in front of the signer and made static. This position of the camera enabled us to get all the body parts which are necessary to have meaningful signs. This position has equal distance of the signers from the camera which enabled us to get equivalent size of the signers in the video. The signers wore long sleeved cloths that helped us to facilitate the skin detection and hands and head segmentation process.

We used smooth background with a color that doesn't resemble to the skin color. The lighting of the signing area is provided in such a way that it reduces high illumination conditions to have videos with better quality.

Since we are working on isolated signs we made the starting and ending position of the hand uniform. Initially when the signing process is started only the head will be visible and gradually the hands will become visible and began to move to the proper direction and position and back to the initial position. The position and the direction in which the hands will move vary depending on the sign. We captured a separate video for each sign and on average the videos have 1 to 2 second length.

After capturing the video, the first task was removing video parts which are not necessary for the training as well as the recognition process. These parts of the video are usually found at the beginning and end of the videos. We cropped these parts manually since they contain frames which may affect the results of the recognition.

In the data collection three signers were involved. Twenty five signs has been chosen to be part of the research. Out of these 25 signs, 20 of them are one handed signs while the rest 5 are signs which are conducted using two hands.

Each signer is expected to perform a sign 20 times. We used 15 of them for training purpose and the rest five for testing purpose. Therefore for a specific sign we will have 45 training videos and 15 testing videos. We have a total of 1500 videos for all 25 signs. The videos are captured in MP4 format. The signers perform the sign in a constant speed.

Ids are given for each sign and signers. In addition to this we give sequence numbers for videos of a sign. Therefore we organized the captured videos with a signer id, sign id and a sequence number of the video. Table 5.1 shows the words that we collect.

Table 5.1: The collected words

Id	Word	
1	እናት	Mother
2	አባት	Father
3	አክስት	Aunt
4	አጎት	Uncle
5	እኔ	Me
6	ወንድ	Male
7	መብላት	Eat
8	መጠጣት	Drink
9	ማብላት	Feed
10	እንጀራ	"Enjera"(local food)
11	ሎሚ	Lemon

12	ማንጎ	Mango
13	ሽንኩርት	Onion
14	መናደድ	Angry
15	አሪፍ	Cute
16	ቅናት	Jealousy
17	ፖሊስ	Police
18	መናገር	To Talk
19	ማድመጥ	To Hear
20	እውነት	Truth
21	ማመን	Believe
22	መስጠት	Give
23	መቀበል	Take
24	ወተት	Milk
25	እንቁላል	Egg

5.3 System Components

There are six components of the system. These components are responsible to conduct major tasks in the recognition process. We discuss them as follows.

1. Skin Detection and Hands and Head Segmentation component:- This component is responsible to input videos from the folder where we put training as well as testing videos. After taking a video the component will extract every frame. The skin detection algorithm described in Chapter 4 is applied on each frame and we had binary images showing white space for skin region and black for other region.
2. Since the skin detection algorithm we used in Chapter 4 is not sufficient, we used median filtering to remove small components that are detected as skin color since the skin detection algorithm by itself was not enough. After smoothing with median filter we had white blobs that are segmented as skin part, but still they were not skin region. To tackle these problems we used the 8-connected component method to remove the smaller blobs and use the larger ones only. When we analyze the image, the hands and the head

are the largest blobs in the image. Therefore to get these blobs we checked the areas of the blobs and remove those which have lesser areas and preserve the blobs with larger areas. After these processes we will have images that segment the hands and the head parts of the body. These images are used as an input for feature extraction component.

3. Feature Extraction Component:- This component is dedicated to accept the binary images as an input, extract the features listed in Chapter 4 and store it in the database. The component stores the features in an organized way to identify the sign, the signer and the video identification number. Furthermore we preparation of the input to train HMM also conducted in this component. Cluster matrices will be formed for each sign using the k-means algorithm. The component will also prepare symbols which are the indices of the cluster entries. To prepare the cluster matrix for a sign the clustering component uses the data of the sign that was prepared by the Feature Extraction Component. The clusters and the symbols will be stored in separate tables for each sign. The other task of this component is changing the feature data to a sequence of symbols. It uses the cluster matrices and the symbols created along with cluster to change the raw feature data to sequence of symbols. These sequences of symbols will be stored in the database for each sign. The overall implementation of this component is given in Appendices A and B.
4. HMM training Component: - As shown in Appendix C, this component trains the HMM model for each sign and stores the model in the database. The component uses the Baum Welch algorithm and symbol sequence produced by vector quantization component. The HMM model is represented by its transition and emission matrices which were produced in the HMM model generation process.
5. Recognition component: - This component is the one which performs the recognition task. As an input it will take sequence of symbols prepared for the testing data and the HMM models produced by HMM training component. It will check the probability of the model that produces a given symbol sequence from the trained models and gives the model with maximum probability. The component used the Viterbi decoding to compute the maximum probability of a given observation symbol being produced by a given trained HMM. The detail implementation is shown in Appendix D.

5.4 Database

The database is prepared to store every data that is produced in the recognition process. Microsoft SQL server is chosen to implement the database since it has very important features suitable for the development of various applications.

The database contains 8 tables to organize the data appropriately. These tables are for both feature sets which have 8 features and the data with only 3 features. The tables contain feature data produced from Feature Extraction Component, the Clusters and Symbols from the Clustering Component and the Transition and Emission matrices which represent an HMM model.

Figure 5.1 shows the database tables to store the feature data of sign videos. The detail of the sign such as Sign Id, by which signer the sign is conducted (Signer Id), whether the sign is used for training or testing (Type) , the frame number of the video (SNo) and all the selected features with their values is stored in the tables.

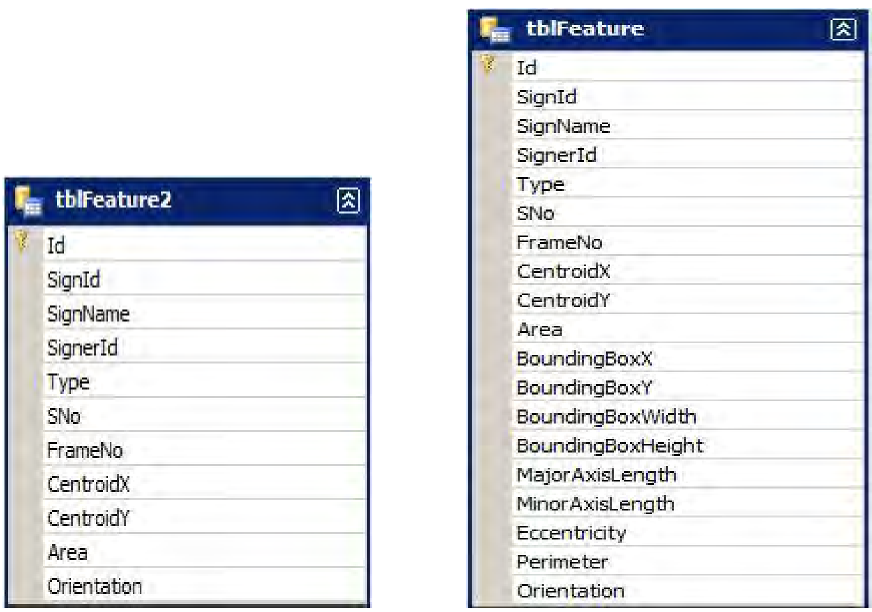


Figure 5.1: Database tables to store feature data

The tables in Figure 5.2 store the cluster matrices (which is found as a result of k-means classification) of signs with 3 and 8 feature sets. The cluster matrices will be stored by sign Id.

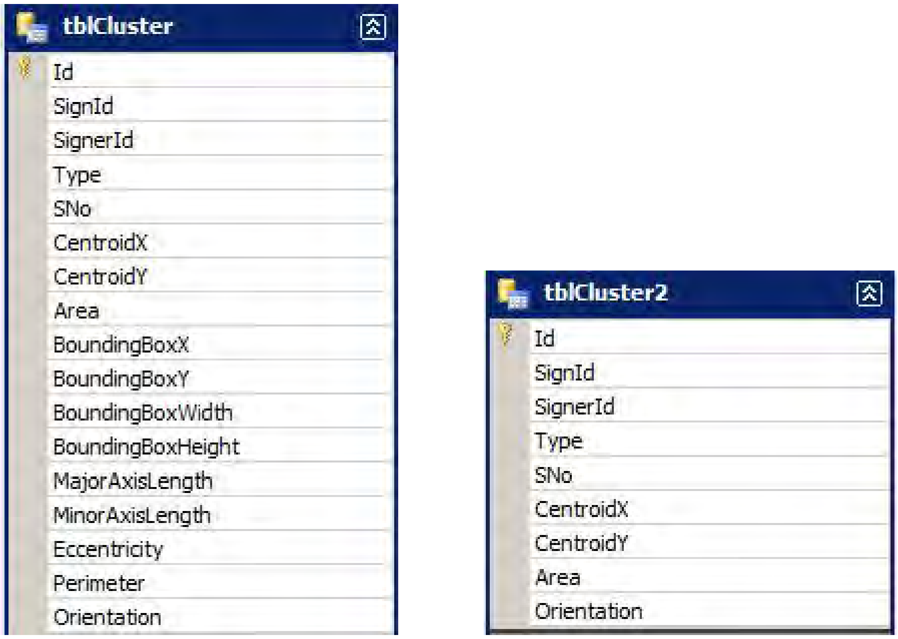


Table 5.2: Database tables to store cluster matrices

After Vector Quantization we will have a sequence of symbols to represent signs. We stored these symbols with their respective Sign Id in tables indicated in Figure 5.3

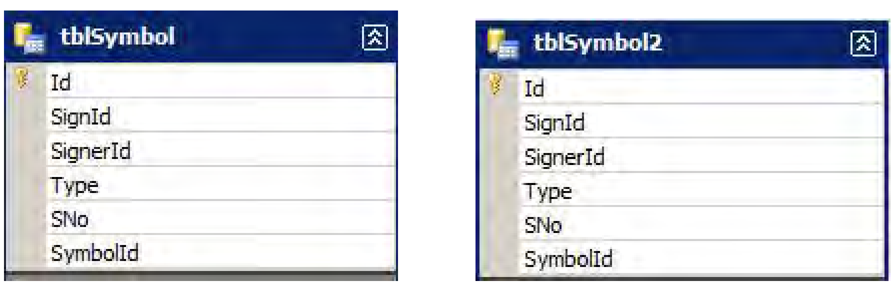


Figure 5.3: Database tables to store representative symbols

The trained HMM for each sign will produce transition and emission probability matrices. These matrices are stored in the tables shown in Figure 5.4. The signs will be identified by Sign Id and the Field Type will store whether the matrix is found by training using 8 features or using only 3 features.

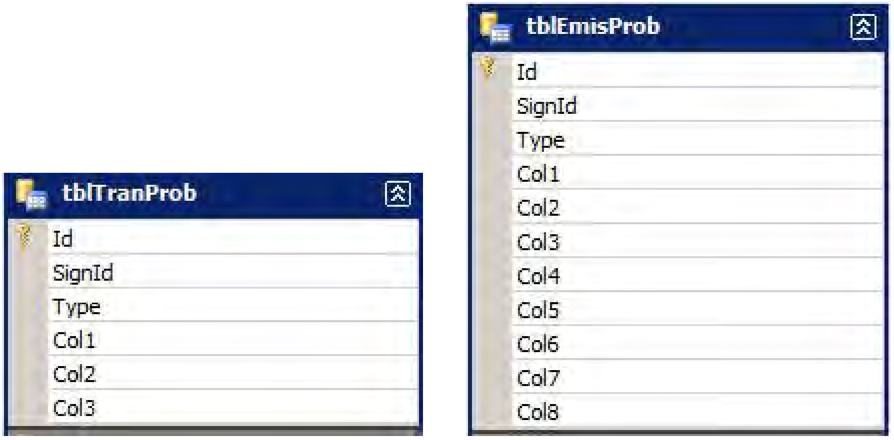


Figure 5.4: Database tables to store transition and emission matrices

5.5 User Interface

The prototype of the system has a user interface which helps to select words from the test set, show the video of the selected word and get the recognized output text. The user interface is shown below.

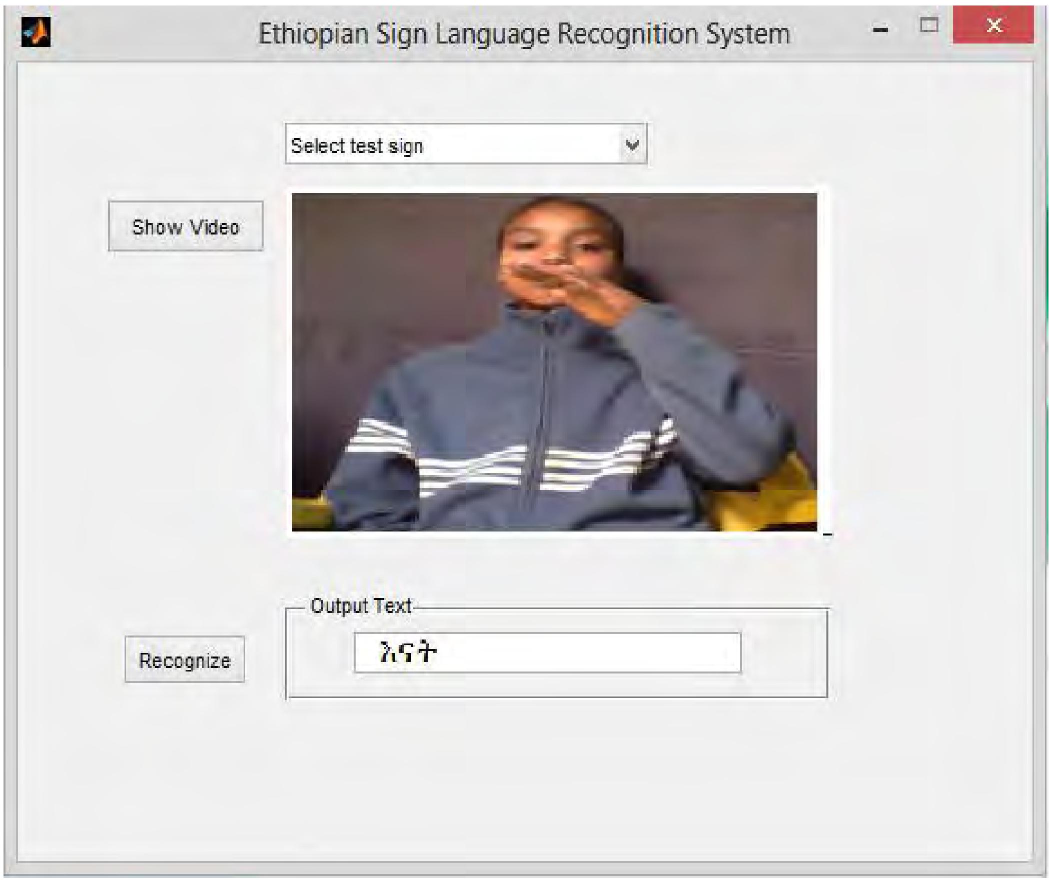


Figure 5.5: User Interface of the prototype

Chapter 6: System Evaluation and Results

In this Chapter we will discuss the evaluation of the system Recognition of Isolated Signs in Ethiopian Sign Language. The evaluation criteria and the results found will be discussed. Further discussion will be made on the results of the evaluation.

6.1 Testing Data

As we discussed in Chapter 5, we took the test data along with the training data. For each sign we have 5 testing videos per signer which were not part of the training process. The testing will be conducted for both reduced feature sets (i.e., the HMM models that were prepared using the position, area and orientation features) and the other trained models which were prepared using the whole 8 feature to check the significance of the other features other than position, area and orientation.

6.2 Evaluation Criteria and Result

The system evaluation is conducted to test the performance of the system. This can be seen from the accuracy of the result the system produces. We divided the evaluation process into two parts. The first part is evaluation based on signers. Since we found differences on the result of the recognition for each signer, we showed the recognition ratio for signers. The result of the recognition for a signer is calculated with the following equation.

$$\text{Recognition ratio of Signer} = \frac{\text{\# of recognized signs of Signer}}{\text{\# of test signs of Signer}} \qquad (6.1)$$

The other part of the evaluation process is evaluation of the overall system performance. The recognition ratio of the system was calculated by mixing the test results for all signers. The following equation is used to calculate the recognition ratio.

$$\text{Recognition ratio} = \frac{\text{\# of recognized signs}}{\text{\# of test signs}} \qquad (6.2)$$

The recognition result calculated using Equations 6.1 and 6.2 are shown in Table 6.1 and Table 6.2. Table 6.1 and Table 6.2 show the results found using the HMM models trained with 8 features.

Table 6.1: Recognition result by signers

	Total Number of Test Signs	Number of Recognized Signs	Recognition Percentage
Signer 1	125	102	81.6%
Signer 2	125	114	91.2%
Signer 3	125	110	88%

Table 6.2: Result on the overall system recognition

Total Number of Test Signs	Number of Recognized Signs	Recognition Percentage
375	326	86.9%

Table 6.3 and Table 6.4 show the results found using the HMM models trained with 3 features.

Table 6.3: Recognition result by signers using HMM trained with 3 features

	Total Number of Test Signs	Number of Recognized Signs	Recognition Percentage
Signer 1	125	98	78.4%
Signer 2	125	109	87.2%
Signer 3	125	106	84.8%

Table 6.46: Result on the overall system recognition using HMM trained with 3 features

Total Number of Test Signs	Number of Recognized Signs	Recognition Percentage
375	313	83.5%

6.3 Discussion

As shown in Table 6.1 and Table 6.3 there is a result variation between signers. This happened because of the way the signers follow to conduct the sign. Some signers conduct the signing process in such a way that someone can see clearly the components of the sign. In this type of signers we can see the movement direction, the speed variation, the Movements and the Holds of the sign. In addition, these type of signers can show us the shapes of the hand while they are signing. The other type of signers are those who conduct the signing process in a way that someone could not identify the components easily. Usually these type of signers are those who use the language in their everyday activities. The system recognition performance will decline for these type of signers. To tackle these types of problem we need to use more training signs which is conducted by variety of signers.

The other issue that we encounter during experimentation is the wrong recognition result for those signs which have similarity. This problem is encountered due to two reasons.

The first one is the problem of occlusion. To illustrate these problem we use the signs "**መብላት**" (Eat) and "**መጠጣት**" (Drink). Both signs are conducted by moving the hand to the mouth area and moving back to the starting position. Their only difference is the variation of the hand shape while the hand is reached in the mouth area. But we couldn't get the value of these shapes since the hand is occluded with the face when it reaches in the mouth area. These results show us we need to conduct occlusion handling process. Some researchers use colored gloves to detect the shape of the hand during occlusion with the face or with the other hand. The others tried to use frames which were found before the occlusion to estimate the shape of the hand during occlusion.

The other problem is lack of depth information. As we discussed in Chapter 5 we used a webcam in a laptop to gather data. The camera used for recording does not produce depth information. Therefore when we extract features from frames we could only get features which cannot give us information about the depth of the image. For example when the signers conduct the sign "**እኔ**" (Me), they will move their hand to their chest in the z direction in the Cartesian coordinate system. But we can only get information about features in x and y directions only. Due to these reason we have got wrong recognition results.

We used data with different feature sets as indicated in Section 6.1 and previous chapters. This is to show how the addition of new features affect the recognition process in addition to the basic features in sign language (i.e., Centroid, area and orientation). The additional features are important to have more information about the shape of the hands to conduct the sign. As indicated in Tables 6.2 and 6.4 the recognition result when we used all the 8 features is 86.9% which is better than the result we found when we used the 3 features which is 83.5%. The result doesn't show much variation since we used simple signs but the recognition result will be lowered if we use the 3 features for signs which have complex shape information.

Chapter 7: Conclusion and Future Work

In Ethiopia there are very large number of people who are living with the problem of hearing. These people are using Ethiopian Sign Language as their communication medium between themselves. But these people are living with the other part of the community who communicate with spoken language. The deaf people do not have the ability to use spoken language while the hearing people do not have the ability to interpret the sign language. This gap makes the life of the deaf people very difficult. Researches to narrow the gap of these two communities are very important. This study is part of the efforts that were exerted to solve the communication problem of the deaf people with the hearing ones.

7.1 Conclusion

In this study we designed and developed a method which is capable of recognizing Isolated Signs in Ethiopian Sign Language. The system accepts a video of Isolated Signs and returns a text equivalent to the input video. In addition to this we showed that addition of features are important to get more shape information and get a better recognition result.

The process of skin detection applied an algorithm from Mahidra [28] and to get more precise segmentation of the hands and the head we used various filtering algorithms and the 8-connected method. Baum Welch algorithm is used for training our HMM model. We trained a separate model for each sign and Viterbi decoding applied in the recognition process.

In this research 3 signers participated and each signer performed a sign 20 times. Out of these 20 videos we used 15 for training and the rest 5 video for testing purpose.

The system has been tested using the videos which were captured for training purpose. We found overall recognition of 86.9% using the HMMs trained by 8 features whereas we found the overall recognition of 83.5% using the HMMs trained by only 3 features.

7.2 Recommendations and Future Work

In this thesis work we achieved a result in recognizing Isolated Signs in Ethiopian Sign Language. But there are gaps which should be filled in the future researches. Since this research cannot be used as a full translation system of Ethiopian Sign Language, we recommend future works to incorporate the following components.

- When we try to extract necessary information from a sign video one of the challenge is the handling of occlusions. This problem is not covered in this work therefore further efforts should be applied to handle occlusion.
- We cannot incorporate depth information in a sign video since we used a 2D webcam camera for data collection. But this depth information can be found by using cameras which are capable of getting a depth information from a video.
- One of the biggest challenges in this research work was getting available data in Ethiopian Sign Language. Therefore we were obliged to collect all the necessary videos by our own. There should be a database that holds all the signs conducted by different signers. This will greatly assist other similar research works in the area.
- In this research we collected Sign videos which were conducted with a uniform velocity. But these should be enhanced to handling of Signs with variable velocities since velocity in sign language sometimes changes a meaning of the sign.
- We used a static background while we capture videos of the signers. But in real world we cannot find such static background in every corner. Therefore future Ethiopian Sign Language researches should consider dynamic backgrounds.
- Researches should be expanded to the recognition of Continuous Ethiopian Sign Language by using the outputs of the Isolated Sign Language recognitions.

References

[1] Legesse Zerubabel, "Ethiopian Finger Spelling Classification", *Unpublished Masters Thesis,* Addis Ababa University, 2008.

[2] Masresha Tadesse, "Automatic translations of Amharic text to Ethiopian Sign Language", *Unpublished Masters Thesis,* Addis Ababa University, 2010.

[3] Dagnachew Feleke, "Machine Translation System for Amharic Text to Ethiopian Sign Language", *Unpublished Masters Thesis,* Addis Ababa University, 2011.

[4] Eyob Gebretinsae, "Vision Based Finger Selling Recognition For Ethiopian Sign Language", *Unpublished Masters Thesis,* Addis Ababa University, 2012.

[5] Aliaa Youssif, Amal Elsayed, and Heba Hamdy, "Arabic Sign Language (ArSL) Recognition ", *International Journal of Advanced Computer Science and Applications,* Vol. 2, No. 11, 2011.

[6] Oya Aran,"Vision Based Sign Language Recognition*", Unpublished PhD Thesis,* Institute for Graduate Studies in Science and Engineering, Bogazici University, 2008.

[7] Abadi Tsegay and Kumudha Raimond, "Offline Candidate Hand Gesture Selection And Trajectory Determination For Continuous Ethiopian Sign Language", *Journal of Theoretical and Applied Information Technology ,*Vol. 36, No.1, 2012.

[8] Vista Wide,"World Language and Culture", Availble at :*http://www.vistawide.com /languages/top_30_languages.htm, Last accessed on March 16,2013.*

[9] Yonas Admasu and Raimond Kumudha, "Ethiopian Sign Language Recognition using Artificial Neural Network", *10th International Conference on Digital Object Identifier,* pp 995-1000, 2010.

[10] Jim Kyle*," Sign and School: Using sign in Deaf child's development",* Short Run Press Ltd , England,1987.

[11] Ethiopian National Association of the Deaf, *"Ethiopian sign language dictionary",* Ethiopia, 2008.

[12] Susanna Ricco and Carlo Tomasi, "Finger Spelling Recognition through Classification of Letter-to-Letter Transitions", In *Proceedings of ACC,* Vol 3., pp.214-225, 2009.

[13] Kyle Duarte,"The Mechanics of Finger spelling: Analyzing Ethiopian Sign Language", *Sign Language Studies,* Gallaudet University Press, Vol. 11, No.1, pp. 5-21, 2010.

[14] SignGenius.com,"Sign language Software, Sign Language Introduction", Availble at : *http://www.signgenius.com/info-sign-language.shtml, Last accessed on March 18, 2013.*

[15] Start ASL,"Basic Sign Language, Pointers and Explanations for ASL Grammar", Availble at : *http://www.start-american-sign-language.com/basic-sign-language.html#top Last accessed on March 18, 2013.*

[16] American Sign Language University,"American Sign Language: Grammar", Availble at : *http://www.lifeprint.com/asl101/pages-layout/grammar.htm ,page accessed on March 18, 2013.*

[17] Ethiopian National Association for Deaf, *"BERTAT"*, Yearly Magazine, Ethiopia, 1997.

[18] Department of Linguistics and Philology, " Dling" *(Yearly Bulletin), October 2009:* Availble at : *http://www.aau.edu.et/news/ling/Dling%20Bulletin.pdf,Last accessed on March 20, 2013.*

[19] Ulrich Agris, Jorg Zieren, Ulrich Canzler, Britta Bauer and Karl Kraiss, *"Recent developments in visual sign language recognition",* Institute of Man–Machine Interaction, RWTH Aachen University, Springer-Verlag, 2007.

[20] Nathan Naidoo, "South African Sign Language Recognition Using Feature vectors and Hidden Markov Models", *Unpublished Masters Thesis,* University of Western Cape, 2010.

[21] Klimis Symeonidis, "Hand Gesture Recognition Using Neural Networks", *Unpublished Masters Thesis, ,* UniS, 2000.

[22] P. V. V Kishore, Rajesh Kumar, Kiran Kumar, and S. R. C. Kishore, " Video Audio Interface for Recognizing Gestures of Indian Sign Language", *International Journal of Image Processing,* Volume (5) : Issue (4) : 2011.

[23] Rung Huei and Liang Ming, "A Sign Language Recognition System Using Hidden Markov Model and Context Sensitive Search", *Proc. of ACM VRST'96 (ACM International Symposium on Virtual Reality and Software Technology),* pp 59-66, Hong Kong, 1996.

[24] Thad Eugene, "Visual Recognition of American Sign Language Using Hidden Markov Model", *Unpublished Masters Thesis,* Massachusetts Institute of Technology , 1995.

[25] Manar Maraqa, Farid Al-Zboun, Mufleh Dhyabat, and Raed Abu Zitar, "Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks", *Journal of Intelligent Learning Systems and Applications*, Vol 4, pp 41-52, 2012.

[26] Alemayehu Teferi, "Ethiopian Sign Language Manual" , *Unpublished User Guide on Ethiopian Sign Language,* 2007.

[27] Rafael Gonzalez and Richard Woods, "*Digital Image Processing*", Prentice-Hall Inc, Volume 1. Second edition, 2002.

[28] Mahendran, "Skin tone detection Mathlab code", Availble at: *http://mahendranmathan.blogspot.com/2013/01/skin-tone-detection-matlab-code.html,Last accessed on January 12,2013.*

[29] Simon Perreault and Patrick Hebert, "Median Filtering in Constant Time", *IEEE Transactions on Image Processing*, Vol. 16, pp 2389-2394, 2007.

[30] Aaron, "Image Processing", Available at: *http://angeljohnsy.blogspot.com/2012/05/find-area-perimeter-centroid.html, Last Accessed on June 22,2014.*

[31] Wikipedia, "Feature Extraction", Available at: *http://en.wikipedia.org/wiki/Feature_extraction, Last Accessed on May 15,2014.*

[32] Yang Mingqiang, Kpalma Kidio, and Ronsin Joseph, "A survery of shape feature extraction techniques", *A journal published in Pattern Recognition, Peng-Yeng Yin (Ed.),* pp 43-90, 2008.

[33] Dengsheng Zhang and Guojun Lu, "Review of shape representation and description techniques"*, The Journal of Patter Recognition Society,* Vol 37, pp 1-19, 2004.

[34] The MathWorks*, Inc.,* "Matlab Computer vision toolbox", *1984-2012.*

[35] Christian Vogler, “American Sign Language Recognition: Reducing the Complexity of the task with phoneme based modeling and Parallel Hidden Makov Models”, *Unpublished PhD Thesis,* Department of Computer and Information Science, University of Pennsylvania, 2002.

[36] Achraf Othman and Mohamed Jemni, “Statistical Sign Language Machine Translation: from English written text to American Sign Language Gloss”, *International Journal of Computer Science Issues*, Vol. 8, Issue 5, No 3, 2011.

[37] Thomas Moeslund, Adrian Hilton, Volker Kruger, and Leonid Sigal, “*Visual Analysis of Humans*”, ISBN: 978-0-85729-996-3 (Print) 978-0-85729-997-0 (Online), 2011.

Appendix A: Take video input, Extract and store Features

```
% This code will take a input video, detect skin color,
% segment hands and head and extract features.
% The extracted features will be stored in the database


%% Add the path where training videos are found
addpath('./Data/Signer1/Sign17');

SNo=0;
%Build connection string
z=adodbcnstr('Sql','.','Sign','sa');
%Open connection
cn=adodbcn(z);

% Start loop for the videos

for v=1:20  %20
SNo=SNo+1;
Status=strcat('Processing Video',num2str(v))


VidFileName=strcat('My Movie',num2str(v),'.mp4')

xyloObj = VideoReader(VidFileName);

nFrames = xyloObj.NumberOfFrames;
vidHeight = xyloObj.Height;
vidWidth = xyloObj.Width;

% Preallocate movie structure.
mov(1:nFrames) = ...
    struct('cdata', zeros(vidHeight, vidWidth, 3, 'uint8'),...
           'colormap', []);

IncValue=int32(nFrames/3);
StartNum=1;
EndNum=nFrames;

for Part=1:3



    if(Part==1)
        StartNum=1;
        EndNum=IncValue;
    end
    if(Part==2)
        StartNum=IncValue+1;
        EndNum=IncValue*2;
    end
    if(Part==3)
```

```
        StartNum=(IncValue*2)+1;
        EndNum=nFrames;
    end


% Read one frame at a time.
for k = StartNum:EndNum
    %mov(k).cdata = read(xyloObj, k);
   k
I= read(xyloObj, k);
I=double(I);

%----------------Skin detection Part--------------------------------------
[hue,s,v]=rgb2hsv(I);
cb =  0.148* I(:,:,1) - 0.291* I(:,:,2) + 0.439 * I(:,:,3) + 128;
cr =  0.439 * I(:,:,1) - 0.368 * I(:,:,2) -0.071 * I(:,:,3) + 128;
[w h]=size(I(:,:,1));
 d=zeros(w,h);
for i=1:w
    for j=1:h
        if  140<=cr(i,j) && cr(i,j)<=165 && 140<=cb(i,j) && cb(i,j)<=195 &&
0.01<=hue(i,j) && hue(i,j)<=0.1
            segment(i,j)=1;
            d(i,j)=1;
        else
            segment(i,j)=0;
        end
    end
end
%-----------------------------------Select connected componet to select
%-----------------------------------bigger areas
 cc = bwconncomp(d);
 stats = regionprops(cc, 'Area');
 idx = find([stats.Area] >500);
 d = ismember(labelmatrix(cc), idx);

 d = imfill(d, 'holes');
 d = bwareaopen(d, 500);



%-------------------------Get features and show with image-------------------
-----------
   s  = regionprops(d, 'All');
   centroids = cat(1, s.Centroid);

   %imshow(d)

%---------------------Insert features into database--------------------



   for i=1:size(s,1)

     if(SNo<=15)
```

```
       sql=strcat('INSERT INTO tblFeature (SignId,
SignName,SignerId,Type,SNo,FrameNo,CentroidX,CentroidY,
Area,BoundingBoxX,BoundingBoxY,BoundingBoxWidth,BoundingBoxHeight,MajorAxisLe
ngth,MinorAxisLength,Eccentricity,Perimeter,Orientation)VALUES
(17,''Police'',1,1,',num2str(SNo));%Id,Name,srId,Type,Sno
     else

        sql=strcat('INSERT INTO tblFeature (SignId,
SignName,SignerId,Type,SNo,FrameNo,CentroidX,CentroidY,
Area,BoundingBoxX,BoundingBoxY,BoundingBoxWidth,BoundingBoxHeight,MajorAxisLe
ngth,MinorAxisLength,Eccentricity,Perimeter,Orientation)VALUES
(17,''Police'',1,2,',num2str(SNo));%Id,Name,srId,Type,Sno
     end

sql2=strcat(sql,',',num2str(k),',',num2str(s(i).Centroid(1)),',',num2str(s(i)
.Centroid(2)),',',num2str(s(i).Area),',',num2str(s(i).BoundingBox(1)),',',num
2str(s(i).BoundingBox(2)),',',num2str(s(i).BoundingBox(3)),',',num2str(s(i).B
oundingBox(4)),',',num2str(s(i).MajorAxisLength),',',num2str(s(i).MinorAxisLe
ngth),',',num2str(s(i).Eccentricity),',',num2str(s(i).Perimeter),',',num2str(
s(i).Orientation),')');

       %Run insert and return results

     adodbinsert(cn,sql2);

     if(SNo<=15)
         sql=strcat('INSERT INTO tblFeature2 (SignId,
SignName,SignerId,Type,SNo,FrameNo,CentroidX,CentroidY, Area,Orientation)
VALUES (17,''Police'',1,1,',num2str(SNo));
     else
         sql=strcat('INSERT INTO tblFeature2 (SignId,
SignName,SignerId,Type,SNo,FrameNo,CentroidX,CentroidY, Area,Orientation)
VALUES (17,''Police'',1,2,',num2str(SNo));
     end

sql2=strcat(sql,',',num2str(k),',',num2str(s(i).Centroid(1)),',',num2str(s(i)
.Centroid(2)),',',num2str(s(i).Area),',',num2str(s(i).Orientation),')');

       %Run insert and return results

    adodbinsert(cn,sql2);




  end


end  % end of frame number loop

end % end of part loop
```

```
Status2=strcat('End of Video',num2str(SNo))
end % end of video number loop

%Close connection
invoke(cn,'release')
```

Appendix B : K-mean Clustering

```
% This code implement the k-mean clustering on the feature data stored in
% the database. By experiment we select the size of k=8
% After applying k- means clustering , the output cluster matrices and
% thier corresponding symbols will be stored in the database.

clear all;
close all;
%Connection for microsoft sql
%Build connection string
s=adodbcnstr('Sql','.','Sign','sa');
%Open connection


cn=adodbcn(s);

sql='select
CentroidX,CentroidY,Area,BoundingBoxX,BoundingBoxY,BoundingBoxWidth,BoundingB
oxHeight,MajorAxisLength,MinorAxisLength,Eccentricity,Perimeter,Orientation
from tblFeature where Type=1 and SignId=17';
sql2='select CentroidX,CentroidY,Area,Orientation from tblFeature where
Type=1 and SignId=17';
x=adodbquery(cn,sql);
x2=adodbquery(cn,sql2);

invoke(cn,'release')

[w h]=size(x);
[w2 h2]=size(x2);

 for  i=1:w
     for  j=1:h
         y(i,j) = str2num(cell2mat(x(i,j)));

     end
 end

 for  i=1:w2
     for  j=1:h2
         y2(i,j) = str2num(cell2mat(x2(i,j)));

     end
 end

clc;
%-----------------------------------------------------
%------------------- Processing K means----------------
% These are the clusters from all signers for the training/Testing Set
 [IDX, C] = kmeans(y,8,'Distance','city','Replicates',5);
[IDX2, C2] = kmeans(y2,8,'Distance','city','Replicates',5);


%-----------------------------------------------------------------
```

```
%-------------------  Saving K means----------------------

%Build connection string
s=adodbcnstr('Sql','.','Sign','sa');
%Open connection
cn=adodbcn(s);

[w h]=size(IDX);
[w2 h2]=size(IDX2);

for i=1:w
 %Sample query to execute
 sql='insert into tblSymbol (SignId,SignerId,Type,SNo,SymbolId) values
(17,1,1,1,';
 sql2=strcat(sql,num2str(IDX(i)),')');
 %Run insert and return results
 adodbinsert(cn,sql2);

end

for i=1:w2
 %Sample query to execute
 sql='insert into tblSymbol2 (SignId,SignerId,Type,SNo,SymbolId) values
(17,1,1,1,';
 sql2=strcat(sql,num2str(IDX2(i)),')');
 %Run insert and return results
 adodbinsert(cn,sql2);

end


%Close connection
invoke(cn,'release')

%----------------------------------------------

% Build connection string
s=adodbcnstr('Sql','.','Sign','sa');
%Open connection
cn=adodbcn(s);

[w h]=size(C);
[w2 h2]=size(C2);
for i=1:w
%  Sample query to execute
 sql='insert into tblCluster (SignId,SignerId,Type,SNo,CentroidX
,CentroidY,Area
,BoundingBoxX,BoundingBoxY,BoundingBoxWidth,BoundingBoxHeight,MajorAxisLength
,MinorAxisLength,Eccentricity,Perimeter,Orientation) values(17,1,1,1,';
 for j=1:h
     if(j==h)
         sql=strcat(sql,num2str(C(i,j)));
     else
         sql=strcat(sql,num2str(C(i,j)),',');
     end
```

```
 end

 sql=strcat(sql,')');
 %Run insert and return results
 adodbinsert(cn,sql);

end

for i=1:w2
 %Sample query to execute
 sql='insert into tblCluster2 (SignId,SignerId,Type,SNo,CentroidX
,CentroidY,Area ,Orientation) values(17,1,1,1,';
 for j=1:h2
     if(j==h2)
         sql=strcat(sql,num2str(C(i,j)));
     else
         sql=strcat(sql,num2str(C(i,j)),',');
     end
 end

 sql=strcat(sql,')');
 %Run insert and return results
 adodbinsert(cn,sql);

end




%Close connection
invoke(cn,'release')

'Complete'
```

Appendix C: Training HMM

```
%This part of the code is to change each video to symbol and prepare a
%symbol matrix. After getting a symbol matrix HMM will be trained which
%will be stored as a knowledge base

clear all;
close all;
%Connection for microsoft sql
%Build connection string
s=adodbcnstr('Sql','.','Sign','sa');
%Open connection


cn=adodbcn(s);

sql3='select
CentroidX,CentroidY,Area,BoundingBoxX,BoundingBoxY,BoundingBoxWidth,BoundingB
oxHeight,MajorAxisLength,MinorAxisLength,Eccentricity,Perimeter,Orientation
from tblCluster where SignId=17';
sql4='select CentroidX,CentroidY,Area,Orientation from tblCluster2 where
SignId=17';
c=adodbquery(cn,sql3);
c2=adodbquery(cn,sql4);



[w h]=size(c);
[w2 h2]=size(c2);

 for  i=1:w
     for  j=1:h
         C(i,j) = str2num(cell2mat(c(i,j)));

     end
 end


 for  i=1:w2
     for  j=1:h2
         C2(i,j) = str2num(cell2mat(c2(i,j)));

     end
 end

 TrainCount=0;

for SignerId=1:2

for TrainSize=1:15 %Change to number of training videos from 1 to 15
TrainCount=TrainCount+1;

sql=strcat('select
CentroidX,CentroidY,Area,BoundingBoxX,BoundingBoxY,BoundingBoxWidth,BoundingB
```

```
oxHeight,MajorAxisLength,MinorAxisLength,Eccentricity,Perimeter,Orientation
from tblFeature where Type=1 and SignId=17 and SignerId=',num2str(SignerId),'
and SNo=');
cmdText=strcat(sql,num2str(TrainSize));
sql2=strcat('select CentroidX,CentroidY,Area,Orientation from tblFeature
where Type=1 and SignId=17 and SignerId=',num2str(SignerId),' and SNo=');
cmdText2=strcat(sql2,num2str(TrainSize));

x=adodbquery(cn,cmdText);
x2=adodbquery(cn,cmdText2);



[w h]=size(x);
[w2 h2]=size(x2);

 for  i=1:w
     for  j=1:h
         y(i,j) = str2num(cell2mat(x(i,j)));

     end
 end

 for  i=1:w2
     for  j=1:h2
         y2(i,j) = str2num(cell2mat(x2(i,j)));

     end
 end

[w3 h3]=size(y);
[w4 h4]=size(y2);

for k=1:w3

[s,idx,distance] = euclidean(y(k,:),C);
     Symbol(TrainCount,k) = idx;


end
for l=1:w4

[s,idx,distance] = euclidean(y2(l,:),C2);
     Symbol2(TrainCount,l) = idx;

end



end  % end of training size

end  % end of signer signer
invoke(cn,'release')
```

```
clc

[w5 h5]=size(Symbol);
[w6 h6]=size(Symbol2);
for m=1:w5
    for n=1:h5
        if(Symbol(m,n)==0)
            Symbol(m,n)=1;
        end
    end
end
for m2=1:w6
    for n2=1:h6
        if(Symbol2(m2,n2)==0)
            Symbol2(m2,n2)=1;
        end
    end
end


% Symbol
% Symbol2


TRANS_GUESS = [0.75,0.25,0.25;
     0.3,0.5,0.2;
     0.2,0.2,0.6];
EMIS_GUESS = [1/8,  1/8,  1/8,  1/8,  1/8,  1/8,1/8,1/8;
     1/8,  1/8,  1/8,  1/8,  1/8,  1/8,1/8,1/8;
     1/8,  1/8,  1/8,  1/8,  1/8,  1/8,1/8,1/8;];


[TRANS_EST2, EMIS_EST2] = hmmtrain(Symbol, TRANS_GUESS, EMIS_GUESS);


TRANS_GUESS3 = [0.75,0.25,0.25;
     0.3,0.5,0.2;
     0.2,0.2,0.6];
EMIS_GUESS3 = [1/8,  1/8,  1/8,  1/8,  1/8,  1/8,1/8,1/8;
     1/8,  1/8,  1/8,  1/8,  1/8,  1/8,1/8,1/8;
     1/8,  1/8,  1/8,  1/8,  1/8,  1/8,1/8,1/8;];


[TRANS_EST4, EMIS_EST4] = hmmtrain(Symbol2, TRANS_GUESS3, EMIS_GUESS3);

% Insert the HMM parameters into a database

%-----------------------------------------------

% Build connection string
s=adodbcnstr('Sql','.','Sign','sa');
%Open connection
cn=adodbcn(s);
```

```
[w h]=size(TRANS_EST2);
[w2 h2]=size(EMIS_EST2);
for i=1:w
%  Sample query to execute
 sql='insert into tblTranProb (SignId,Type,Col1,Col2,Col3) values(17,1,';
 for j=1:h
     if(j==h)
         sql=strcat(sql,num2str(TRANS_EST2(i,j)));
     else
         sql=strcat(sql,num2str(TRANS_EST2(i,j)),',');
     end
 end

 sql=strcat(sql,')');
 %Run insert and return results
 adodbinsert(cn,sql);

end

for i=1:w2
 %Sample query to execute
 sql='insert into
tblEmisProb(SignId,Type,Col1,Col2,Col3,Col4,Col5,Col6,Col7,Col8)
values(17,1,';
 for j=1:h2
     if(j==h2)
         sql=strcat(sql,num2str(EMIS_EST2(i,j)));
     else
         sql=strcat(sql,num2str(EMIS_EST2(i,j)),',');
     end
 end

 sql=strcat(sql,')');
 %Run insert and return results
 adodbinsert(cn,sql);

end

[w h]=size(TRANS_EST4);
[w2 h2]=size(EMIS_EST4);
for i=1:w
%  Sample query to execute
 sql='insert into tblTranProb (SignId,Type,Col1,Col2,Col3) values(17,2,';
 for j=1:h
     if(j==h)
         sql=strcat(sql,num2str(TRANS_EST4(i,j)));
     else
         sql=strcat(sql,num2str(TRANS_EST4(i,j)),',');
     end
 end

 sql=strcat(sql,')');
 %Run insert and return results
 adodbinsert(cn,sql);
```

```
end

for i=1:w2
 %Sample query to execute
 sql='insert into
tblEmisProb(SignId,Type,Col1,Col2,Col3,Col4,Col5,Col6,Col7,Col8)
values(17,2,';
 for j=1:h2
     if(j==h2)
         sql=strcat(sql,num2str(EMIS_EST4(i,j)));
     else
         sql=strcat(sql,num2str(EMIS_EST4(i,j)),',');
     end
 end

 sql=strcat(sql,')');
 %Run insert and return results
 adodbinsert(cn,sql);

end




%Close connection
invoke(cn,'release')
% %
% %--------------------------------------------------------------------
'Completed'
```

Appendix D : Recognition

```
% This code implements the recognition component of the system. The trained
% HMMs and the testing signs will be used.

clc
clear all;
close all;


s=adodbcnstr('Sql','.','Sign','sa');
%Open connection


cn=adodbcn(s);

'SignerId     SignId     TrainSignId     MinIndex'
for SignerId=1:2
```

```
for SignId=1:17

% for TrainSignId=1:5
for TrainSignId=16:20
% SignId=10;
% TrainSignId=1;
% SignerId=2;



sql3=strcat('select
CentroidX,CentroidY,Area,BoundingBoxX,BoundingBoxY,BoundingBoxWidth,BoundingB
oxHeight,MajorAxisLength,MinorAxisLength,Eccentricity,Perimeter,Orientation
from tblCluster where SignId=',num2str(SignId));
sql4=strcat('select CentroidX,CentroidY,Area,Orientation from tblCluster2
where SignId=',num2str(SignId));
c=adodbquery(cn,sql3);
c2=adodbquery(cn,sql4);



[w h]=size(c);
[w2 h2]=size(c2);

 for  i=1:w
     for  j=1:h
         C(i,j) = str2num(cell2mat(c(i,j)));

     end
 end

 for  i=1:w2
     for  j=1:h2
         C2(i,j) = str2num(cell2mat(c2(i,j)));

     end
 end

%----------------------------------------------------------------------------
---------------------------
%------------ Change the matrix to symbol


sql=strcat('select
CentroidX,CentroidY,Area,BoundingBoxX,BoundingBoxY,BoundingBoxWidth,BoundingB
oxHeight,MajorAxisLength,MinorAxisLength,Eccentricity,Perimeter,Orientation
from tblFeature where Type=2 and SignerId=',num2str(SignerId),' and
SNo=',num2str(TrainSignId),' and SignId=',num2str(SignId));
```

```
sql2=strcat('select CentroidX,CentroidY,Area,Orientation from tblFeature2
where Type=2 and SignerId=',num2str(SignerId),' and
SNo=',num2str(TrainSignId),' and SignId=',num2str(SignId));
x=adodbquery(cn,sql);
x2=adodbquery(cn,sql2);
[w h]=size(x);
[w2 h2]=size(x2);

 for  i=1:w
     for  j=1:h
         y(i,j) = str2num(cell2mat(x(i,j)));

     end
 end
 for  i=1:w2
     for  j=1:h2
         y2(i,j) = str2num(cell2mat(x2(i,j)));

     end
 end


 [w3 h3]=size(y);
 [w4 h4]=size(y);

for k=1:w3

[s,idx,distance] = euclidean(y(k,:),C);
     Symbol(k) = idx;


end

for k=1:w4

[s,idx,distance] = euclidean(y2(k,:),C2);
     Symbol2(k) = idx;


end


%-------------------------------------------------------------------
%------------------------ Check the log probability
MinIndex=0;
Min=0;
for ModelCount=1:17

sql=strcat('select Col1,Col2,Col3 from tblTranProb where
SignId=',num2str(ModelCount),' and Type=1');
sql2=strcat('select Col1,Col2,Col3 from tblTranProb where
SignId=',num2str(ModelCount),' and Type=2');
```

```
x=adodbquery(cn,sql);
x2=adodbquery(cn,sql2);
[w h]=size(x);
[w2 h2]=size(x2);

 for  i=1:w
     for  j=1:h
         Tran(i,j) = str2num(cell2mat(x(i,j)));

     end
 end

 for  i=1:w2
     for  j=1:h2
         Tran2(i,j) = str2num(cell2mat(x2(i,j)));

     end
 end

sql=strcat('select Col1,Col2,Col3,Col4,Col5,Col6,Col7,Col8 from tblEmisProb
where SignId=',num2str(ModelCount),' and Type=1');
sql2=strcat('select Col1,Col2,Col3,Col4,Col5,Col6,Col7,Col8 from tblEmisProb
where SignId=',num2str(ModelCount),' and Type=2');

x=adodbquery(cn,sql);
x2=adodbquery(cn,sql2);

[w h]=size(x);
[w2 h2]=size(x2);

 for  i=1:w
     for  j=1:h
         Emis(i,j) = str2num(cell2mat(x(i,j)));

     end
 end

 for  i=1:w2
     for  j=1:h2
         Emis2(i,j) = str2num(cell2mat(x2(i,j)));

     end
 end

Pi= rand(1,8);
Pi=Pi/sum(Pi);



%Prob=pr_hmm(Symbol,Tran,Emis,Pi);
Prob=pr_hmm(Symbol2,Tran2,Emis2,Pi);

if(ModelCount==1)
    Min=abs(Prob);
```

```
    MinIndex=1;
else
    if(abs(Prob)<Min)
        Min=abs(Prob);
        MinIndex=ModelCount;
    end
end

end

% Min

aa=strcat(num2str(SignerId),'-----',num2str(SignId),'-----
',num2str(TrainSignId),'-----',num2str(MinIndex))


end % end of TrainSignId
end %end of signid
end %end of signer id
invoke(cn,'release')
```

Declaration

I, the undersigned, declare that this thesis is my original work and has not been presented for a degree in any other university, and that all source of materials used for the thesis have been duly acknowledged.

____

Tefera Gimbi

This thesis has been submitted for examination with my approval as an advisor.

____

Yaregal Assabie (PhD)

Addis Ababa, Ethiopia

June 2014