



**ADDIS ABABA UNIVERSITY**  
**SCHOOL OF GRADUATE STUDIES**

**ETHIOPIAN FINGER SPELLING CLASSIFICATION: A STUDY TO  
AUTOMATE ETHIOPIAN SIGN LANGUAGE**

**BY**  
**LEGESSE ZERUBABEL**

**A THESIS SUBMITTED TO  
THE SCHOOL OF GRADUATE STUDIES OF ADDIS ABABA UNIVERSITY  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF MASTERS OF SCIENCE  
IN COMPUTER SCIENCE**

September 2008  
Addis Ababa

**Addis Ababa University**  
**School of Graduate Studies**  
**Faculty of Informatics**  
**Department of Computer Science**

**Ethiopian Finger Spelling Classification**

By

Legesse Zerubabel

Members of the Examination Board

1. Sebsibe Hailemariam, Advisor \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_
4. \_\_\_\_\_

## **ACKNOWLEDGMENTS**

This work would not have been possible without the support, encouragement, and patience of several people who deserve recognition for their contributions.

First and foremost, I would like to thank my advisor, Sebsibe Hailemariam, for his support, guidance and patience throughout this work. He has helped me to achieve and learn a great deal of rewarding experience.

I would like to forward my special thanks to the Ethiopian National Association for Deaf for providing reading material and information about Ethiopian Sign Language and Ethiopian Finger spelling. I would like to extend my thanks to Ato Eyasu Hailu and Ato Alemayhu Tefrie for their consultation about Ethiopian Sign language.

I would like to thank all friends and relatives for their valuable comments and suggestions. Additionally, I would like to thank my friends Abera Ababawe, Addisalem Negash, Habtamu Minassie, Mikiyas Teshome, MohamedSani Hassen and Amare Abebe for their comment by reading the thesis document.

Last, but not least, I would like to thank my parents, brother and sister for their moral support and encouragement during my study.

# TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	I
LIST OF TABLES.....	V
LIST OF FIGURES.....	VI
ABSTRACT .....	VII
1. INTRODUCTION.....	1
1.1. MOTIVATION .....	1
1.2. PROBLEM STATEMENT .....	2
1.3. OBJECTIVE.....	3
1.3.1. <i>General Objective</i> .....	3
1.3.2. <i>Specific Objective</i> .....	3
1.4. SCOPE AND LIMITATION.....	4
1.5. METHODOLOGY .....	4
1.6. ORGANIZATION OF THE THESIS .....	5
2. OVERVIEW OF SIGN LANGUAGE.....	7
2.1. DEAF COMMUNICATION .....	9
2.2. SIGN STRUCTURE IN SIGN LANGUAGE.....	9
2.3. FINGER SPELLING.....	10
2.4. SIGNING.....	12
2.4.1. <i>Sign Language</i> .....	13
2.4.2. <i>Sign-Spoken Language</i> .....	14
2.4.3. <i>Signed-Spoken Language</i> .....	15
2.5. ETHIOPIAN SIGN LANGUAGE .....	17
2.6. ETHIOPIAN FINGER SPELLING .....	18
2.7. SUMMARY .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
3. LITERATURE REVIEW AND RELATED WORKS.....	20
3.1. IMAGE REPRESENTATION.....	20
3.2. IMAGE FEATURES .....	22
3.2.1. <i>Principal Component Analysis (PCA) Feature</i> .....	22
3.2.2. <i>Harr-like Feature</i> .....	25

3.3.	PATTERN CLASSIFIER.....	27
3.3.1.	<i>Neural Network</i> .....	28
3.3.2.	<i>Template Matching</i> .....	30
3.3.3.	<i>Boosted Classifier</i> .....	32
3.4.	RELATED WORKS.....	34
3.4.1.	<i>Hand and Shape Detection</i> .....	34
3.4.2.	<i>Sign Classification</i> .....	37
3.5.	THE PROPOSED APPROACH.....	39
4.	<b>ETHIOPIAN FINGER SPELLING CLASSIFICATION SYSTEM ARCHITECTURE AND IMPLEMENTATION.....</b>	<b>40</b>
4.1.	OVERVIEW OF ETHIOPIAN FINGER SPELLING CLASSIFICATION SYSTEM ARCHITECTURE ..	41
4.2.	IMAGE PRE-PROCESSING .....	43
4.3.	FEATURE EXTRACTION .....	44
4.3.1.	<i>Principal Component Analysis Based Feature</i> .....	44
4.3.2.	<i>Harr-like Feature</i> .....	49
4.4.	HAND DETECTION .....	50
4.4.1.	<i>Neural Network Based Hand Detector</i> .....	50
4.4.2.	<i>Boosted Classifier Based Hand Detector</i> .....	53
4.5.	REGION OF INTEREST SEGMENTATION .....	53
4.6.	SIGN CLASSIFICATION.....	54
4.6.1.	<i>Neural Network Based Sign Classifier</i> .....	56
4.6.2.	<i>Template Matching Based Sign Classifier</i> .....	58
4.7.	IMPLEMENTATION OF THE SYSTEM .....	59
4.8.	SUMMARY .....	60
5.	<b>EXPERIMENTAL RESULT AND DISCUSSION.....</b>	<b>61</b>
5.1.	DATA COLLECTION .....	61
5.2.	PERFORMANCE MEASURE.....	63
5.3.	EXPERIMENTAL RESULT .....	63
5.4.	DISCUSSION OF HAND DETECTOR EXPERIMENTAL RESULT .....	68
5.5.	DISCUSSION OF SIGN CLASSIFICATION EXPERIMENTAL RESULT.....	69
5.6.	SUMMARY .....	71

**6. CONCLUSION AND RECOMMENDATION ..... 73**

## LIST OF TABLES

Table 4-1: The result of the neural network based hand detector with different number of node in each hidden .....	51
Table 4-2: The result of the neural network based sign classifier with two hidden layer	57
Table 4-3 : The development tools and the system component relation.....	60
Table 5-1: Collected data distribution.....	62
Table 5-2 : Summary of the experiment type .....	64
Table 5-3 : The result of the neural network based hand detector with PCA driven feature .....	65
Table 5-4 : The result of the neural network based hand detector with Harr-like feature	65
Table 5-5 : The result of boosted classifier based hand detection .....	66
Table 5-6 : The result of neural network based sign classifier with PCA based feature ..	67
Table 5-7 : The result of the neural network based sign classifier with harr-like feature	67
Table 5-8 : The result of the template matching based sign classifier.....	68

## LIST OF FIGURES

Figure 2-1 : The American Sign Language sign for father and mother [ 36] .....	10
Figure 2-2 : Possible hand locations in Signing [9].....	12
Figure 2-3 : The Sign markers of Signed English [8].....	17
Figure 2-4 : Ethiopian Finger Spelling (Source: ENAD) .....	19
Figure 3-1: Rectangle type used in Harr-like feature.....	26
Figure 3-2: Integral Image .....	27
Figure 3-3 : Points used to calculate harr-like feature .....	27
Figure 4-1 : Ethiopian Finger spelling classification system architecture .....	41
Figure 4-2: Reconstructed image with the corresponding number of Principal Component Analysis.....	46
Figure 4-3 : A three stage refinement method to reduce the number of reconstructed image.....	48
Figure 4-4: Neural Network based hand detector topology.....	52
Figure 4-5 : The original image, the image after cropping operation and the image after size adjustment.....	54
Figure 4-6: Amharic alphabet mapping with sequence of zeros and one .....	55
Figure 4-7: The neural network based sign classifier topology .....	56
Figure 5-1 : Some of the collected image of Ethiopian finger spelling sign .....	61
Figure 5-2 : The result of sign classifier graph.....	71

## Abstract

*Ethiopian Finger Spelling is one of the communication means used among Ethiopian deaf. It is used to express names and any concepts that do not have sign in Ethiopian Sign Language. To fill the communication gap that exists among the deaf and between the deaf and the hearing, the Ethiopian Finger Spelling is processed using techniques from image processing and pattern processing.*

*In this thesis work, a new attempt is done to design the architecture and select the appropriate techniques for each component of the Ethiopian Finger spelling classification system. The proposed architecture has components for image capturing, feature extraction, hand detection, region of interest segmentation and sign classification. For the tasks of hand detection and sign classification, experiments are conducted to select the appropriate pattern classifier and feature. In addition, the capability of the principal component analysis (PCA) driven and har-r-like feature with neural network is tested through experiment.*

*According to the experimental result, neural network pattern processing techniques have high detection and classification rate when compared to the cascaded boosted classifier and template matching techniques for the task of hand detection and sign classification respectively. The overall hand detection rate of 99.43 %, 96.59 % and 77.27 % were obtained using neural network with PCA driven feature, neural network with har-r-like feature and boosted cascaded classifier respectively. In addition to this, the overall sign classification rate of 88.08%, 96.22% and 51.44% were obtained using neural network with PCA driven feature, neural network with har-r-like feature and template matching respectively. In particular, neural network that uses har-r-like feature shows better performance for the task of sign classification and neural network with PCA driven feature shows better performance for the task of hand detection.*

**Keywords:** Ethiopian Sign Language, Ethiopian Finger Spelling, Hand detection and Sign Classification

# **1. Introduction**

## **1.1. Motivation**

In general, deaf can communicate with each other and with the hearing by using speech, reading-and-writing, and sign language. Communication through speech is not the best way of communication among deaf. Since the speech communication is one way communication for deaf. In addition to this, the born deaf do not have the capability of producing speech. Communicating through reading and writing requires skill of reading and writing but most of the deaf are not educated because of the shortage of schools that are equipped with the necessary facilities to give education for the deaf. Hence, the reading and writing means of communication is not appropriate for deaf communication. Communication through sign language is the preferable means of communication for deaf to deaf communication when compared to the above specified means of communications. However, it is not the appropriate means of communication between the deaf and the hearing. This is because of the shortage of the hearing societies that have the skill of sign language and the sign language structure is difficult to understand for hearing [8, 11, 35].

Therefore, to minimize the drawback of the sign language over communication between the hearing and the deaf, development of sign language application can be very important. Ideally, translation systems would make communication possible among the deaf people and between the deaf and the hearing.

To develop such system, many researches were conducted on sign languages like American Sign Language [31] and Australian Sign Language [7, 33]. In addition, attempts have been made to develop a system that translates sign language into speech or text. However, for

Ethiopian Sign Language, there had been no such attempt before. Hence, in this work, we have tried to investigate capability of the techniques used in other sign languages over Ethiopian Sign Language.

## **1.2. Problem Statement**

In Ethiopia, there is limited communication between the hearing society and the deaf community. The deaf community uses sign language to communicate among themselves and the hearing society uses spoken language for communication. In addition to this, only few number of the hearing society have the skill of using sign language.

In school, the teaching-learning process is mainly conducted by the discussion and description of the teacher's idea through speech. This prevents the deaf from attending classes at school in the country. To solve this problem either the teachers have to have a sign language skill or sign language interpreters have to be in place. The former is limited due to economic constraints of the country. But there are attempts in the country even though they are very few in number compared to the school required for the deaf in the country. Currently, there are six special schools for the deaf throughout the country. Employing an interpreter also has got many problems like translating subject-specific terms or concepts that are out of the interpreter's knowledge.

Deaf people in the country do not get the information they need at the right time and place. For instance, HIV/AIDS is the main obstacle for developing countries like Ethiopia by killing the youth. To minimize this loss the government launches awareness campaigns using different information dissemination media like radio and television. But these media are not appropriate for the deaf. In addition to this, the medical institutions in the country lack appropriate facilities to give service for the deaf.

Today, the advancement of computer technology is helping people with disabilities in different ways. But, in our country there is no enough research effort to help these people with the availability of computer technology. Therefore, the development of cost effective automated method of processing Ethiopian Sign Language to solve the communication problem between the hearing society and the deaf community is of real importance.

### **1.3. Objective**

#### **1.3.1. General Objective**

The general objective of this work is to design Ethiopian finger spelling classification system and select the best techniques for the designed system from the techniques used in other country's sign language processing systems.

#### **1.3.2. Specific Objective**

In light of this general objective, the research will have the following specific objectives:

- Studying and selecting the appropriate features for hand detection and sign classification.
- Studying and selecting the appropriate pattern classifier for hand detector module of Ethiopian Finger Spelling Classification System.
- Studying and selecting the appropriate pattern classifier for sign classifier module of Ethiopian Finger Spelling Classification System.
- Designing and prototyping the Ethiopian Finger Spelling Classification system.

#### **1.4. Scope and Limitation**

The manual means of deaf communication includes Sign language, Sign-spoken, Signed-spoken and finger spelling [8]. In Ethiopia, the syntactic structure of the manual communication is not studied yet. The only thing we find is the mapping of the finger spelling sign with Gee'z alphabet.

Therefore, among the manual means of deaf communication only the Ethiopian finger spelling is considered in this thesis. In addition, due to lack of tools that detect motion of the hand, only static signs of the Ethiopian finger spelling are addressed. **Moreover, only spellings signed by right hand are considered in this work.**

#### **1.5. Methodology**

In order to attain the objectives of the research, literatures on contemporary development of image analysis and pattern processing related to sign language classification and object detection were reviewed. From these insight reviews of image analysis techniques and tools that were employed on sign language classification and object detection that were relevant to this work were selected. These image analysis and pattern processing techniques were selected based on the performance they had on the previous related works.

In addition to this, manual means of deaf communications were studied in detail to get a broader understanding of the language and to indicate which part of the language can be automated. To this end, literature reviews of articles, journals and books that are related to sign language specially American and Ethiopian Sign languages are conducted and people who have knowledge about Ethiopian Sign Language have been consulted.

The tools that were used in this research work are described below:

- **Digital Camera:** To capture the finger spelling sign, a SONY digital camera with 7.2 mega pixel resolution was used. In addition, the camera stand is also used to fix the camera in a constant position.
- **Development Environment:** The prototype of this thesis was developed on Window XP operating system with Intel Pentium IV 3.0 GHz processor, 512 MB memory and 80 GB hard disk capacities.
- **Image processing tool:** For pre-processing of image data a Java-based open source image processing software called ImageJ<sup>1</sup> was used.
- **Development tools:** All modules except the image preprocessing part of the proposed system were developed using MATLAB 7.0 Version 14. In addition to this, the GML adaboost matlab toolbox developed by Alexander Vezhnevets [1] was used to develop the system.

## 1.6. Contribution of the work

The result of this work will have application in teleconferencing, videoconferencing and other communication systems. These applications currently play great role by increasing the quality of life in the hearing society. The result of this work helps the deaf community to be qualified for using such kind of applications, as it did for the hearing society.

The main beneficiaries of this work are the deaf community. Moreover, programmers and researchers can benefit from it to develop software that will translate Sign language to Amharic.

---

<sup>1</sup> <http://rsbweb.nih.gov/>

## **1.7. Organization of the Thesis**

The remaining part of the thesis is organized as follows:

- Chapter 2 presents a detailed discussion on the meaning of the word “deaf”, deaf communication, sign language structure, finger spelling, signing, Ethiopian Finger Spelling and Ethiopian Sign Language.
- Chapter 3 provides a review of related works in hand detection and sign classification. In addition, discussions on image representation, image feature extraction and pattern classification techniques like neural network, template matching and boosted cascaded classifiers are presented in detail. Finally, the proposed approaches are presented.
- Chapter 4 presents a detailed discussion on architecture and implementation of Ethiopian finger spelling classification system.
- Chapter 5 presents the experimental results of the hand detector and sign classifier module of the Ethiopian finger spelling system.
- Chapter 6 provides conclusions and recommendations for future research works.

# Chapter One

## 2. Overview of Sign Language

Language is a means of communication used among humans. It can be in spoken or in written format. Languages in spoken form 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. Many spoken languages exist throughout the world. Some of them are used among a small number of communities and others are used internationally in many communities as a mother tongue or as a second language. For example, English language has 510 million speakers all over the world, of which 340 million are native speakers [2].

Communication via speech is appropriate provided that the speaker and the listener share the same medium of communication. Most of the time this communication is real time but sometimes the speaker can dump the speech into a device that can be played back to the target user. There are scenarios in which speech communication cannot be the best choice, in which case written form of communication is another equally important alternative.

To express thoughts in written form, a symbol or collection of symbols is used to represent a letter, word or other form of the language. The art of expressing a thought through symbols is called writing and the nature of writing is called writing system. There are three types of writing systems in the world: namely logographic, syllabic and alphabetic. In logographic writing system each symbol represents a word of the language and in syllabic system the symbol represents the phoneme of the language. Symbols in alphabetic system represent a phoneme just like syllabic but by considering vowel and consonant

independently. Most of the western countries languages use the Latin alphabet and in Ethiopia Gee'z alphabet is used for languages like Amharic and Tigrinya. Both of these alphabets fall under syllabic writing system [3].

Moreover, a written language requires skills of writing & reading. 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. For deaf communication, the third alternative means of communication called visual language is usually used. But it is also used to supplement others in oral communications and in situations where the other means of communication fail to deliver the intended message.

In Ethiopia, the word deaf creates confusion because of the meaning given by some scholars when translated to local languages like Amharic. For example, Dr. Amsalu Aklilu's English-Amharic dictionary put the Amharic word “ደንቆሮ (DENKORO)” for the word deaf. Mostly it created wrong impression among hearing society toward the Deaf. As a result of this, the Amharic meaning doesn't show the exact meaning of the English word deaf.

Currently, this Amharic word is not in use throughout the country. Now, it is substituted by another Amharic phrase “መስማት የተሳነወ (MESMAT YETSANEW)”. This Amharic phrase is officially accepted and in use as a formal way of calling the deaf.

The word “deaf” has two meanings depending on the case of the first letter ‘d’. If the first letter is capital, it represents a society that has its own language and culture. But if it is in small case, it means a person who has partial as well as full hearing impairment [4, 5].

## **2.1. Deaf Communication**

The Deaf have different means of communication to interact with each other and with the hearing society. In general, these communication means can be categorized into two: manual and oral communication. The oral means of communication includes speaking and lip reading. It is a good means of communication among the hearing and the Deaf. The manual means includes different communication media that are expressed using the hand and upper body parts and it involves signing and finger spelling.

Using only oral or manual means is not sufficient because each one of them have advantages and disadvantages. To minimize the disadvantage and to create a better means of communication especially at school, the philosophy of total communication and simultaneous communication are suggested. Total communication is a liberal approach to the use of all the available oral as well as manual means of communication for teaching the deaf. Simultaneous communication is also the philosophy of using both speech and signing at the same time. Among these techniques, total communication is getting wider acceptance and is currently in use in school for teaching-learning purposes [6].

## **2.2. Sign Structure in Sign Language**

Sign Language and other visual language systems are expressed by combining and matching different components of the sign. Signs have the following components [4, 7]:

- i. Hand shape

Different hand shapes are created by bending and stretching the finger and the palm. There are signs created using only one hand and two hands (with the same hand shape or with different hand shapes)

## ii. Location

This is the place where the signing is performed. There are signs that are the same with other components of sign and only different in their location. For example, the sign for Mother and Father in American Sign Language is shown in the figure 2-1.



Figure 2-1 : The American Sign Language sign for father and mother [ 36]

## iii. Movement

There are a wide variety of movements possible with sign language including arcs straight lines and wavy patterns.

## iv. Orientation

There are three possible orientations of the hand: roll, pitch and yaw.

### 2.3. 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. It is a means of communication rather than spelling of any spoken language words. It is used in combination with sign language for proper nouns, names and addresses and for words that

have no sign. The term finger spelling stands for both sign language production and perception [8, 9].

In general, finger spellings are two kinds, those that are represented in one hand and those that are represented in two hands. This means each alphabet of the spoken language is constructed either by using one hand sign or by using two hands sign. Mostly if it is a one-handed alphabet, it is constructed by using the dominant hand (i.e. right hand if the person is right-handed). For example, American and Ethiopian finger spellings are constructed using only one hand and there are countries like Britain that have a two handed finger spelling. Besides, there are some countries like Yugoslavia (Former) and Italy have a combination of one-handed and two-handed finger spelling [8, 9].

Among finger spellings of the world, American finger spelling is the most widely used and studied. In addition to this, World Federation of the Deaf has accepted this system to serve as a base for international manual alphabet and used to design other country finger spelling as a base. For example Ethiopian finger spelling constructed by following the structure American finger spelling as a base. The American finger spelling is produced with 26 different hand shapes except the letters 'J' and 'Z' that incorporate a motion [8, 9].

Some of the rules used when we sign an alphabet through American manual finger spelling include [9]:

- Each letter should be made clearly, distinctly and crisply with a slight pause between words.
- When double letters are formed the hand is opened slightly before repeating the second letter of the series.

e.g. Sunny , haggie, spelling

- To distinguish a word from an abbreviation it is necessary to circle the individual letters clockwise very slightly when they represent names of place or organization.

e.g. U. S. A.

- It should be read in the same way you read the words on this page, in units instead of letter by letter.

The position of the hand when signing and finger spelling is usually in front of the chest in a comfortable manner as shown in the figure 2-1 and the circled parts in the figure are the possible signing positions of a hand.



**Figure 2-2 : Possible hand locations in Signing [9]**

## **2.4. Signing**

Signing is a system or a process of constructing a meaningful sign from manual gesture together with bodily and facial cues by ordering according to their own syntactic rule. It is a major communication means used among the Deaf. It is described in terms of three linguistically broad categories. These are sign language, sign-spoken language, and signed-spoken language. Each of these categories has their own unique feature or characteristic. Most of them are widely used among the Deaf especially, in school, for teaching-learning processes [8]. The following subsections deal with each of these categories of signing.

### **2.4.1. Sign Language**

Sign language is a naturally occurred language used for communication among Deaf and between hearing and the deaf. As a language, it has comparable language structure with spoken language. It uses hands, upper body parts, and facial expressions to construct meaningful sign. It is a visual language that doesn't require speaking and listening to communicate with it.

Currently, many sign languages are used through out the world. Most countries have their own sign language. For example Ethiopian Sign Language, American Sign Language, Britain Sign Language, and so on.

To give you some overview about the difference that exists between sign language and spoken language, let's compare the arrangement of sentences of American Sign Language (ASL) and English. English language arranges sentences in "subject" + "verb" + "object" order but American Sign Language follows "Topic" + "comment" arrangement and to show past or future event it adds "Time" before the topic. A comment is simply a word or a phrase that describes the topic. The following sentences try to show the above specified structures of the two languages [10].

- My car, I wash. (I wash my car)

In the above ASL sentence, the comma separates the topic (“My car”) from the comment (“I wash”). The English equivalent is shown in the brackets. The object of the English sentence serves as a topic for ASL. The space between each word is expressed at the time of signing by a small pause that helps to differentiate a word in a sentence.

- I, from E-T-H-I-O-P-I-A. (I am from Ethiopia)

In ASL, there is no being verb like “am” and there is no sign for the word Ethiopia. As discussed in finger spelling part, if there is no sign for the given word, it is expressed using finger spelling until a sign is assigned for the word. So, the word ETHIOPIA is expressed by using finger spelling of each letter (E T H I O P I A).

- 9:00 MORNING, STORE, I GO (I'm going to the store at 9:00AM.)

In this sentence, the first comma separates the “time” from the “topic” and the second one separates the topic from the comment. In ASL, there is no sign for –ing or –ed form of the verb. So, “going” and “go” are expressed using the same sign without any difference.

#### **2.4.2. Sign-Spoken Language**

The majority of the Deaf belong to hearing families and they are living with the hearing in the same environment. So the Deaf are forced to express their feeling with a dominant spoken language used in that environment. In addition to this, the hearings have a problem of understanding the structure of the sign language, because sign language has different phonology, syntax and semantic structure when compared with spoken language.

To address the above problem, a means of manual representation of spoken language called sign-spoken language was designed. It uses sign from sign language and finger spelling to depict the syntax of spoken language. In schools, this system is used to expose deaf children to the syntax of spoken language and it's also used for teaching other subjects like Geography and Biology.

According to the interview we made with Ato Alemayehu Teferi (former Ethiopian National Association for Deaf (ENAD) president) and Ato Eyasu Hailu (AAU, linguistic department, Sign Language Division coordinator), there is Sign-Amharic language system in Ethiopia that follows the structure of Amharic language grammar and expressed using sign from Ethiopian Sign Language and Amharic finger spelling. Currently, it is in use by Ethiopian Television for its program to present information for Ethiopian Deaf. However, we could not find any written document to certify this and the presence of such systems for other local languages of Ethiopia.

### **2.4.3. Signed-Spoken Language**

It is an artificial system that has been developed to represent precisely the syntax of spoken language with only sign (without finger spelling). The sign can be from sign language or an artificial sign designed by modifying the existing sign or a newly invented sign. Different kinds of such language systems are designed to address particular situations.

The major difference between the Sign Spoken system and this category is that the Sign Spoken system uses the sign of the sign language without modifying it. But in Signed Spoken Language, it is modified or a new one is invented to show the exact syntax of the spoken language. For example, in American Sign Language, there is no sign that shows the past or present continuous form of a given verb. There is only a sign that stands for the

infinitive of each verb. The signs that show past or present or any other tense form are derived either by modifying existing signs for infinitive or creating new signs just to show the exact syntax of the English language.

From American Sign Language different scholars were develop different such kind of model for depicting English language syntax and grammar. Among this, Signing Exact English (SEE) or Signed English is the most widely used model. In this model, new signs can be invented or modified from the existing signs and then create signs for affixes, word endings, plurality, and articles. As shown in figure 2-3, the top two persons show the regular and irregular past forms of the English verb *walk* and *see* respectively. The bottom three indicate the -ing form of the verb *speak* and the form of the adverb *quickly* and adjective *sleepy* of English language. The sign for the word *walk* is created by modifying the existing American Sign Language sign and the rest are newly created signs. Such kinds of model also exist in other language like Spanish and Swedish. To the best of our knowledge, there is no any information to certify the presence of such kinds of models in Ethiopia.



**Figure 2-3 : The Sign markers of Signed English [8]**

## **2.5. Ethiopian Sign Language**

Ethiopian Sign Language is used among the Ethiopian deaf community. It is derived from American Sign Language because of the first deaf school was established by American missionaries and it served as a medium of instruction in schools. The major difference that exists between American and Ethiopian sign languages is due to the following two reasons:

- Ethiopian Sign Language has borrowed some signs from Nordic countries sign languages especially from Finnish Sign Language.
- There are local signs created and used in specific deaf schools in the country and later included into Ethiopian Sign Language.

Until now, no attempt has been made to find out what grammatical structure and rule the Ethiopian Sign Language has due to lack of experts in the field of sign languages.

However, there are fruitful attempts in the development of modern sign language dictionary in Amharic, Oromiffa and Tigrigna [11].

## **2.6. Ethiopian Finger Spelling**

The Ethiopian Finger Spelling was developed in 1971 by the ENAD and later gained acceptance by the Ministry of Education. Currently, it is used in all deaf schools along with the American finger spelling. The American Finger Spelling is used when we want to express words or concepts in English and any local language that is expressed through Latin script like Oromiffa. However, Ethiopian finger spelling is used when we want to express words or concepts in a language that uses Gee'z alphabet like Amharic and Tigrigna[11].

Ethiopian Finger Spelling has 33 unique hand shapes to represent the first order Amharic alphabet. By adding six different movements on the first order alphabet, the rest six order alphabets for each first order alphabets are constructed. In figure 2-4, the Ethiopian finger spelling with the corresponding Gee'z alphabet is depicted [11].



Figure 2-4 : Ethiopian Finger Spelling (Source: ENAD)

## Chapter Two

### 3. Literature Review and Related Works

Vision based sign language classification is one of the task handled by incorporating techniques from image processing and pattern processing. The image processing techniques prepare the input image for pattern processing to work on it. In pattern processing, the processed images are classified into the corresponding classes. There are different image processing and pattern processing techniques applied for sign language classification application. Among these, techniques that have relation with this work are discussed in this chapter. In addition to this, previous related works on the task of hand detection and sign classification are also presented.

#### 3.1. Image Representation

Image is represented using a continuous two dimensional function  $f(x, y)$  that computes the value or amplitude at spatial coordinates  $x$  and  $y$ . Most of camera's sensors generate an output in terms of continuous voltage waveform. To convert an image representation from continuous voltage waveform to digital form, processes called sampling and quantization are applied on the image. The quantization process digitizes the intensity value and the sampling process digitizes the  $x$  and  $y$  coordinates of the representation function. The resulting images that are represented using discrete and finite values of  $x$ ,  $y$  and intensity are called digital images and each finite element of the image is called pixel [12, 13, 14].

Images can come in binary, grayscale (monochrome), or color forms. In a binary image, each pixel has either zero (Black), or  $2^{n-1}$  (White) value where  $n$  is the sample size. It is the simplest form of image representation which requires less memory and processing time. In

Grayscale image, different color values including black and white are used. For example in an image represented using 8 bit sample size, there are 256 possible color values. Among these, zero represents black and 255 represents white color [13, 14, 15].

In a color image, each pixel value is represented with different dimensionality that hold different information of the basic color of the color model used. Some of the color models are RGB (Red, Green, and Blue), CMY (Cyan, Magenta, and Yellow) and HIS (Hue, Intensity, Saturation). For example, according to the above specified color model each pixel of the color image has three components that correspond to each component of the model. It is possible to transform from one color model to another. For example, the following equations (from equation 1 to 5) are used to transform RGB color space in to CMY and RGB to HSI [12, 16].

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix} \dots\dots\dots (1)$$

And

$$H = \begin{cases} \theta \dots\dots\dots \text{if } B \leq G \\ 360^\circ - \theta \dots \text{if } B > G \end{cases} \dots\dots\dots (2)$$

$$\theta = \cos^{-1} \left( \frac{\frac{1}{2} [(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}} \right) \dots\dots\dots (3)$$

$$S = 1 - \frac{3}{(R+G+B)} [\min(R,G,B)] \dots\dots\dots (4)$$

$$I = \frac{1}{3}(R + G + B) \dots\dots\dots (5)$$

Among the color spaces, the RGB and CMY are not suitable to describe colors that are particular for human interpretation. But they are suitable for hardware implementation when compared with the HIS color space [12].

For an application like sign language recognition, color images contain better information than grayscale and binary images but require high computational cost and special hardware. When we compare grayscale and binary images, grayscale image contains more information but require high computational cost while binary image holds small information but needs less computational cost. Therefore, the grayscale images representation is selected for this work based on the available processing resource.

### **3.2. Image Features**

An image feature is a minimal component of the image that describes the characteristics of the entire image. Using the entire image for classification increases the computational cost of the system. Currently, different types of features are extracted from each image for pattern classification. Among these features, only principal component analysis and Harr-like features are reviewed and presented in this thesis.

#### **3.2.1. Principal Component Analysis (PCA) Feature**

The Principal Component Analysis (PCA) driven feature is a feature extracted by applying PCA technique. PCA is a statistical technique used to reduce the dimensionality of the data. The dimension reduction is performed by projecting the data into new coordinates. The new coordinates are the principal components computed from covariance matrix of the

data. The covariance matrix shows the interdependence exists between each element of the data [14, 37].

To what extent the dimensionality is reduced, is determined by the number of principal components used. If all the existing principal components are used, then the original data can be recovered when the image is reconstructed. On the other hand, if the number of principal components used decreased, the information loss when the image reconstructed is increased. However, the majority of the image information can be recovered using some of the principal components that have relatively higher transformation scaling factor. In PCA, every principal component (eigenvector) has the corresponding value called eigenvalue and It indicates the magnitude of transformation the principal components undertake [14, 37].

The procedure of extracting PCA driven feature is listed below [17]:

- Arrange each image data as a row matrix and collecting all data set into one matrix.
- The covariance of the data matrix is computed to show the interdependence between the data variables.
- From the resulting covariance matrix, the eigenvectors and eigenvalue are computed and then the eigenvectors are sorted in decreasing order of the corresponding eigenvalue.
- Among the existing eigenvectors, the most significant eigenvectors for the particular application domain in consideration are selected.
- The selected eigenvectors are used to transform the original image data into the new coordinate.

Finally, the reduced data is used as a feature for systems like pattern classification and object detection. Using the reduced data as feature will give better accuracy with lesser storage and computational complexity.

Let see how the above stated procedure work with example using three 3 x 3 images and its matrix representation presented below. Each row of the matrix represents one image pixel values and it has nine pixel values.

```

31 32 28 34 31 33 31 29 29
45 56 42 58 41 40 56 44 59
131 126 131 119 134 130 131 129 129
30 27 29 30 31 29 29 27 31

```

To identify the interdependent exist between the variables of the image pixel value, the covariance matrix is calculated from the data matrix and the covariance matrix depicted below.

```

1.0e+003 *
2.3349  2.1829  2.3872  1.9579  2.4159  2.3257  2.2924  2.3399  2.2297
2.1829  2.0816  2.2265  1.8733  2.2433  2.1603  2.1784  2.1926  2.1270
2.3872  2.2265  2.4417  1.9965  2.4718  2.3783  2.3398  2.3915  2.2760
1.9579  1.8733  1.9965  1.6869  2.0096  1.9347  1.9598  1.9672  1.9157
2.4159  2.2433  2.4718  2.0096  2.5056  2.4120  2.3584  2.4192  2.2903
2.3257  2.1603  2.3783  1.9347  2.4120  2.3247  2.2693  2.3293  2.2007
2.2924  2.1784  2.3398  1.9598  2.3584  2.2693  2.2822  2.3014  2.2290
2.3399  2.1926  2.3915  1.9672  2.4192  2.3293  2.3014  2.3456  2.2390
2.2297  2.1270  2.2760  1.9157  2.2903  2.2007  2.2290  2.2390  2.1827

```

After computing the covariance matrix, the next step is calculating the eigenvalue and eigenvector. The computed eigenvalue and eigenvector presented as follow sequentially.

1.0e+004 \*

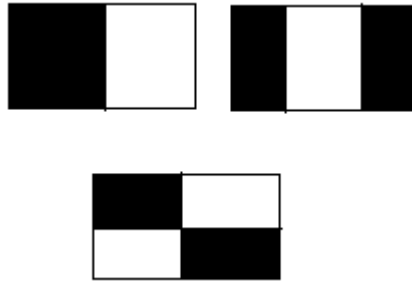
-0.0000  
 -0.0000  
 0.0000  
 0.0000  
 0.0000  
 0.0000  
 0.0008  
 0.0137  
 2.0041

0.4724	0.2218	-0.6089	0.2767	-0.0908	-0.3424	0.0570	-0.1881	0.3410
-0.2526	0.3612	-0.2929	-0.5140	-0.1619	0.0847	-0.4393	0.3556	0.3208
-0.4286	-0.5198	-0.1939	-0.1766	-0.4113	-0.0755	0.3447	-0.2494	0.3484
-0.2523	-0.1651	0.0895	0.7316	-0.1492	0.0281	-0.3047	0.4086	0.2881
-0.1885	0.6170	0.4619	0.1320	-0.1880	0.0730	0.1706	-0.3992	0.3520
-0.1569	-0.2130	0.0852	-0.0610	0.6344	-0.2361	-0.4285	-0.4030	0.3388
0.4718	-0.2019	0.5218	-0.2656	-0.1872	-0.3988	0.0015	0.2939	0.3366
0.4116	-0.2140	-0.0020	-0.0308	0.0102	0.8029	-0.0776	-0.1273	0.3419
-0.1160	0.1015	-0.0704	-0.0137	0.5488	0.0664	0.6111	0.4297	0.3279

As the above matrixes shows, a total of nine principal components exist. However, only the last three principal components contribute more than 90% of the image information. This means, if the data transformed using only the last three principal components, the image can be reconstructed with a minimal loss.

### 3.2.2. Harr-like Feature

Harr-like feature is a rectangular feature first proposed by Viola and Johns [18] in 2002 for object detection purpose. Its capability was tested on face detection system and gave a promising result. Later, it was used for hand and shape detection [19].



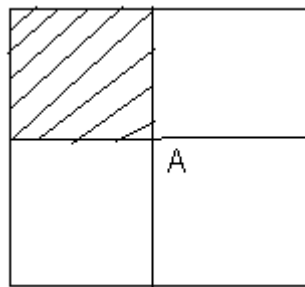
**Figure 3-1: Rectangle type used in Harr-like feature**

The harr-like feature is extracted from three types of rectangles: rectangles with two, three and four compartments. These rectangle types are depicted in figure 3-1 and each compartment is shaded with either black or white. The dimensionality of each compartment can be predetermined. The feature extraction process starts by partitioning the image using each type of the rectangle. After this, the sum of all pixel values fall under white and black compartments are calculated. The feature in that particular image area is the difference between the value in the black compartment and white compartment.

In the case of two-compartment rectangle, the harr-like feature is calculated by subtracting the pixel sum of the black from white compartment. In the case of three-compartment rectangle, the sum of the two black compartment pixels is subtracted from the sum of white compartment pixel. And in case of four-compartment rectangle, the sum of the two diagonal black compartments pixel is subtracted from the sum of the two diagonal white compartment pixel. Finally, the combination of the values computed using each type of rectangle is taken as the harr-like feature for that particular image.

However, extracting the Harr-like feature directly from the pixel value is computationally expensive. To minimize the computational cost, another way of image representation

called Integral Image is used. In Integral Image, every pixel value holds the sum of all pixel values above and to the left of it. For example, in figure 3-2, the pixel value at point ‘A’ is the sum of all pixel values in the shaded region. If we do not use the integral form the shaded region pixel sum is calculated at least three times for each rectangle. However, in integral image, the pixel sum is calculated only ones and as a result the number of computational operation is significantly reduced.



**Figure 3-2: Integral Image**

In this case, the harr-like feature will be computed using the following equation:

$$Harr - like \ feature = \begin{cases} = B - A & \text{if the rectangle has two sub - regions} \\ = C - B + A & \text{if the rectangle has three sub - regions} \\ = 2C + 2B - 3A - D & \text{if the reactangle has four sub - regions} \end{cases}$$

Where, the symbols A, B, C, and D are depicted in figure 3-3.



**Figure 3-3 : Points used to calculate harr-like feature**

### 3.3. Pattern Classifier

Pattern classification is a process of relating features extracted from the raw data with the available target class. It has an application in systems like hand detection, face detection

and sign classification. To accomplish the task, different pattern classification techniques exist and are widely used. Some of the pattern classification techniques are neural network, template matching, boosted classifier, Bayesian classifier and Hidden Markov Model.

In this section, pattern classification techniques like neural network, template matching and boosted classifier are discussed in detail.

### **3.3.1. Neural Network**

Artificial neural network (neural network) is a biologically inspired pattern processing tools. It provides solutions for systems that require pattern processing like character recognition, speech recognition, hand detection and sign classification. To get better performing neural network based pattern classifier, the appropriate network architecture (topology), the training algorithm and the value of different parameters have to be selected carefully.

There are different neural network architectures designed for different application purposes. Among these, the most commonly used network architectures for pattern classification are listed below [20, 21]:

- ***Multilayer Perceptron***: It is a feed forward network that contains a non-linear processing unit called hidden layers between the input and output layers. The number of hidden layers and neurons in each layer varies from application to application. In hidden layers and output layer, a sigmoid activation function is used.

- **Radial Basis function network:** It is also a feed forward network with one hidden layer and a radial basis activation function on the hidden layer nodes. Unlike the multilayer perceptron, it uses a linear activation function on the output layer.
- **Recurrent network:** It distinguishes itself from the feed forward neural network in that it has at least one feedback loop. The presence of the feedback loop has a profound impact on the learning capability of the network and on the performance. Mostly, it is used to process temporal data.

After selecting the appropriate network architecture for the particular application, the next step is training the network. To train the network, there are two approaches: supervised and unsupervised. The supervised approach is used when the input and the expected output are available. The unsupervised approach is used if only when the input feature is available and the expected output is going to be determined automatically by the network. The choice of the learning techniques depends on the application.

Among the training algorithms used in the supervised approach, the back propagation algorithm is the most commonly used algorithm and it is based on error correction learning rule. This algorithm can be applied only on feed forward networks. The algorithm consists of two phases: the forward phase and backward phase. In the forward phase, the network output with the given input is calculated. And in the backward phase, the error of the network at the output layer is computed and then propagated to all contributing neuron in the hidden layer. In addition to this, the connection weight between the consecutive neurons is adjusted based on their error share of each neuron in the backward phase [20].

The two parameters of back propagation algorithm learning rate and momentum have effect on fast generalization of the network and their value should be selected carefully. Mostly,

for supervised learning the learning rate and momentum value are set near zero and one respectively. In addition to this, the number of nodes in each layer should be determined for the particular application. The number of neurons in the input and output layer is determined by the dimensionality of the input feature and the representation of the expected output of the network respectively. However, there is no efficient way to determine the number of nodes in each hidden layers used for the application. Even if it is tedious, trial-and-error is the most commonly used method to determine the number of hidden layer and neuron in each hidden layer [20].

### **3.3.2. Template Matching**

Template matching is one of the techniques used in image processing to classify the object on the image into the available target class. It operates in such a way that by evaluating the similarity of the image with the template by moving it to every position of the image. The region that gives better similarity result when compared with the similarity of the rest of the region will be the possible outcome of the classifier.

The performance of the template matching system depends mainly on the similarity measures used and on the quality of the template selected. Currently different similarity measures exist and are used in different situations. However, there is no single similarity measure that is known to produce the best result in all situations. Among these, the most commonly used similarity measures are listed below [13, 14, 22]:

- ***Sum of absolute difference***: It measures the similarity between the template and the image windows by calculating the absolute value of the difference of the corresponding intensity value of the template and the image windows. The

smaller the calculated value, the more similar the template and the image windows.

- ***Cross-correlation coefficient:*** It is a measure of the linear relationship between the template and the image window. As the template and the image window become more similar, the correlation result becomes larger.
- ***Geometric distance:*** If the template and the image windows contain binary structures like image edge, geometric distance is the preferred metric for similarity measure. To measure the similarity between the template and the image windows, first the template is overlaid on the image windows and then for each image window structure point closeness with the template is determined by the average distance between the corresponding points in the two images. In this measure, the best match is the position where the average distances between the structures of the image window and template is with the minimum value.
- ***Mutual information:*** It is one of the many quantities that measure how much one random variable tells us about another random variable given the knowledge of one another. As geometric distance measure, this metric also begins by overlaying the template on the image window and then the probability that the intensity of a particular pixel in the template lies on the top of intensity of a particular pixel in the image window is computed and becomes their joint probability. If the template and the image windows truly correspond to each other, their intensity will be highly dependent and they will produce high joint probability. However, if the template and the image windows do not correspond to each other, they will produce less joint probability.

- ***Invariant moments***: All of the above similarity measures are affected by rotation of the object in the image. Hence, to make the data rotation invariant the image as well as the template is transformed into moment feature and then similarity will be measured using one of the above specified similarity metrics.

In line with this, the template selection mechanisms used also have an effect on the performance of the system. One of the most commonly used way of selecting templates is selecting one representative image from each class through subjective judgment.

### **3.3.3. Boosted Classifier**

Boosted classifier is a pattern classifier that is constructed from the strong learner and each strong learner is also constructed by combining weak learners using boosting algorithm. This means a boosting algorithm takes a weak learner and returns a strong learner. According to M. P. Carter [23], the weak learner is just a normal learner except, that the error rate of the hypothesis it generates need only be slightly better than that of random guessing.

Boosting algorithm was first proposed by Schapire and Freund [24] and the first form is called Adaboost. Currently, many derivatives of the Adaboost exist and some of them are FrootBoost, LPBoost, TotalBoost, BrawnBoost, ModestBoost and LightBoost. The major variation that exists between the different boosting algorithms is their method of weighting training data points and hypothesis.

To construct the best boosted pattern classifier, the strong classifier is arranged in cascade or tree structure. In cascade structure, the first classifier rejects most of the region and the next classifiers refine the previous classifiers' result. However, in tree structures, like the

cascade structure, the previous classifier rejects more region than the next classifiers refine but the leaf classifier stands for each class in consideration [19].

### **3.4. Related Works**

#### **3.4.1. Hand and Shape Detection**

Hand and shape detection is used to locate the position, size and shape of the hand in a given digital image. Currently, it is an active research area in computer vision. This research can be classified based on the methods used into: glove-based, 2D vision based and 3D vision based [25]. The glove-based method uses a hand glove with a sensor embedded on each joint of the finger. It provides accurate information. However, it is expensive to buy and use besides being heavy to wear in everyday activity. The 3D vision based method gives more reliable information but it is complex to implement and requires more computational time.

The 2D vision based method uses features like skin color, texture, shape, and temporal information as input for the detector. The detector uses techniques like Statistical classifiers and Artificial Neural Network to classify the input pattern. The 2D vision based method currently got considerable improvements in recent years because it requires less number of operations to implement and less time of computation. Besides, it doesn't force the user to wear heavy and expensive.

In this section, we have tried to review papers that are related to hand and hand shape detection. The papers are presented based on their year of publication because the latest papers are extensions of the earlier ones.

One of the researches was conducted by Viola and Jones [18] in 2002. They presented three new methods to the science of image processing. The first contribution is the new way of representing an image called integral image. This representation puts an image point in

terms of the sum of all top and the left pixels of the point. It allows a fast feature evaluation and requires less number of operations per pixel to construct from the original image.

The second contribution is a method of constructing a strong classifier that uses less number of features by modifying existing AdaBoost algorithm. At last the research work contributes a method of constructing cascade of strong classifier in a way that increases the detector speed by focusing on the attention region of the image.

To test and compare the performance of the new method, a 32-layer of cascaded classifier with different numbers of strong classifiers in each layer was constructed to detect the human face. The experiment conducted shows that the face detector constructed with the new method achieves a good result as the number of false positives increase.

Later, this research work was extended by Eng-Jong ong and Richard Bowden [19] for hand and hand shape detection. In this work, boosted cascade of classifier for hand and shape detection are presented. It is based on Viola and Jone object detection method. The constructed system has two layered tree structure and each layer consists of a cascade of strong classifier using AdaBoost and FrootBoost algorithm. The first layer is used for hand detection while the second layer is used for shape detection. To measure the achievement of boosted classifier over hand and shape detection, the experiment was conducted using 5013 images segmented from a video sequence and two independent experiments were conducted to measure the hand detector and hand shape detector performance. Among 5013 image, 2504 were used for testing and the rest were used for training the hand detector. The researches claimed that they achieved 99.8% accuracy for hand detection. For hand shape detection, first all the images were clustered into 300 shapes using k-medoid clustering algorithm and then 900 images were selected from each cluster for testing and the rest of

the images were used for training the shape detector. For hand detection, the researchers stated that they achieved an accuracy of 97.4%.

In addition to the work of [2], Mathias Kolsch and Matthew Turk [26] also extended the Viola and Jones' work for hand posture detection. In addition, they presented a new method that estimates the capability of the detector for the task of hand posture detection without intensive training of a detector with larger data. The estimator uses the value of frequency spectrum analysis of representative images of each hand posture class. This method helps to estimate the detectability of each hand posture prior to training. Those hand posture classes that have less value of frequency have less detectability as compared with those that have high frequency value. This method was tested on cascaded as well as non-cascaded boosted classifiers and showed high detection rate for those hand postures that have high frequency value in contrast with those hand postures that have small value.

In addition to boosted classifier, neural network based classifiers are also applied for object detection. In the work of F. Smach [27], a design of multilayer perceptron neural network for hand detection was presented and in the work of X. Wu [28], a color based self organization neural network was applied for detecting dynamic hand gesture by combining with motion information.

In the above works, the harr-like feature used as input feature and its shows a remarkable performance over the task of hand detection. Moreover, boosted classifier and neural network are used to construct the hand detector. As our knowledge of understanding, there is no previous effort of measuring the performance of the harr-like feature with other feature like PCA driven feature for the task of hand detection and sign classification. To the level of our understanding, the PCA driven feature was not applied for the task of hand

detection in the previous works. Furthermore, the performance of the boosted classifier based hand detector with neural network based hand detector was not compared for the task of hand detection in the previous works. Therefore, all the above stated gaps will address in this work.

### **3.4.2. Sign Classification**

Sign classification is a process of categorizing signs of a sign language based on the feature extracted from image or using an instrument like a glove with a sensor using different techniques. These techniques are applied depending on the nature of the sign: dynamic and static. A dynamic sign is a sign that incorporates motion information with the shape and other information of the hand while static sign does not have motion information.

For classification of static sign, pattern recognition techniques like template matching and neural network are used. The work of Locton and Fitzgibbon[29] used nearest neighbor template matching for recognition of American Sign Language. In this method a simple skin color detection algorithm and a colored wrist band were used to differentiate the hand from other body parts like face. However, these techniques were computationally expensive. So to minimize this limitation, other techniques like clustering of input data and boosted classifier were used. The experiment conducted to test the proposed method was shows an accuracy of 99.7%.

The work of Lamor et. al.[30] presented an approach that used artificial neural network and principal component analysis for Japanese finger spelling recognition. This approach used an image that contains a hand with colored glove. The glove has different colors for the four fingers, thumb and palm to discriminate each finger, the thumb and palm clearly and easily. The features used by the system were extracted by threshold of each color of the

glove in RGB space and from each threshold value, the normalized centroid, eigenvalue ratio and angle of rotation were computed using principal component analysis. This computed value was used as an input for the three layered feed forward neural network that are trained by back-propagation learning algorithm. Another neural network that does not consider palm region information was also used to measure the effect of palm region information on the overall performance of the system. These networks were trained with 1260 images and achieved an accuracy of 89.06% for the case where the palm was disregarded, as opposed to 85.24% with the palm included.

Besides the above work on the recognition of static sign language, a considerable amount of work has been done on the recognition of dynamic sign language using techniques like Hidden Markov Model (HMM), Fuzzy Expert and Artificial neural network. The works of Starner and Pentland [31] and Paul [32] used HMM for recognition of American Sign Language sentences and Auslan finger spelling recognition respectively. In [31], only geometric feature of an image of a person that wears a colored glove are used but in the work of Paul [32] geometric and motion descriptor without a need of wearing colored glove of the signer were used.

In addition to HMM, fuzzy expert system [33] and recurrent neural network [34] are used for recognition of Auslan sign language. The fuzzy expert system uses high level hand posture and motion information as input and the recurrent neural network uses location and orientation information of the hand that are captured through cyber glove as input for the network.

As our knowledge of understanding, there is no such classification method applied for Ethiopian Sign language as well as Ethiopian manual finger spelling. In this work, we will try to select the best classification method for the Ethiopian manual finger spelling.

### **3.5. The Proposed Approach**

These works focus mainly on hand detection and sign classification that will be used as input for Ethiopian finger spelling recognition. Therefore, two pattern classifiers are selected for each task among the pattern classifiers in the previous works and then their performance is compared over static sign of Ethiopian finger spelling.

For hand detection, a boosted classifier as used in E.J. Ong and R. Bowden [19] work and a neural network based hand detector will be constructed and compared. As for the sign classification, the template matching and neural network pattern classifiers that are applied on static sign selected and compared.

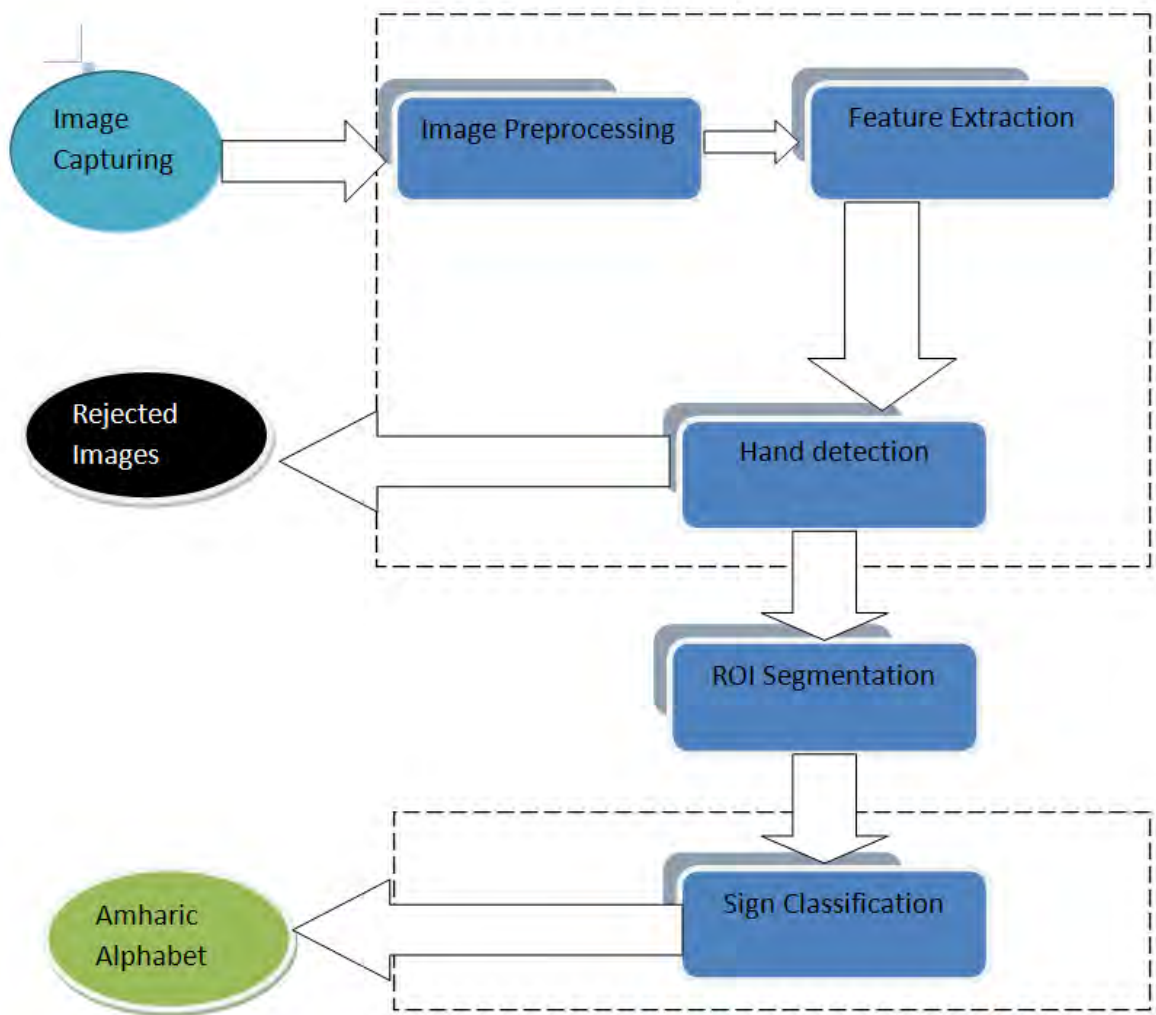
In the next chapter, each of the selected pattern classifiers and the architecture and implementation of sign classification system will be presented in detail.

## **Chapter Three**

### **4. Ethiopian Finger Spelling Classification System Architecture and Implementation**

Ethiopian Finger Spelling is a means of communication used among Ethiopian Deaf to express names of persons and places, and other concepts that don't have appropriate signs in Ethiopian Sign Language by using the syntax of the Amharic Language. As discussed in section 2.6, each Amharic first order alphabet has a unique hand shape sign with the dominant hand. The sign of the rest six order alphabets is constructed from the hand shape of its first order and by incorporating six different hand motions that correspond to each Amharic alphabet order.

In this chapter, the newly proposed Ethiopian Finger Spelling Classification System Architecture and Implementation are discussed in detail.



**Figure 4-1 : Ethiopian Finger spelling classification system architecture**

#### **4.1. Overview of Ethiopian Finger Spelling Classification System Architecture**

The Ethiopian Finger Spelling Classification System (EFSCS) is a system that classifies the hand sign of Ethiopian Finger Spelling into a class that represents each Amharic alphabet. The system accepts the image of the hand sign as an input and returns the associated Amharic alphabet as output. In figure 4.1, each component of the EFSCS is depicted. The rectangular components represent processing components and the circles

represent the output of the system. **The Region of interest segmentation module is not included into the broken rectangle; this is to indicate that the task handled manually.**

Each of the system processing components and their functionalities are described below briefly:

- **Image Acquisition:** The input images of the system are captured from different signers using digital camera.
- **Image preprocessing:** here, the image is preprocessed by applying a sequence of image processing operations to make suitable for the next processing. The image processing operations that are applied on the image are size adjustment, converting the image form into grayscale, contrast adjustment and sharpening.
- **Feature extraction:** From each input data, a representative data is extracted using different feature extraction techniques and later used as input for the hand detection and sign classification module.
- **Hand Detector:** Based on the input from the feature extraction module, here, the image is examined whether it contains a hand object or not. Images that contain a hand object are transferred to the next module for further processing and those images that do not contain a hand object are discarded.
- **Region of interest segmentation:** To increase the performance of the sign classifier by forcing the classifier to focus only on a region that holds target object information. An image processing operation called segmentation is applied on images that pass the hand detector module and then entered into the sign classifier module.

- **Sign Classification:** In this module, the input pattern is classified into the corresponding first order Amharic alphabet output classes.

In the following sections, each component of the system is described in detail. In addition, the implementation of the system is also described.

#### **4.2. Image Pre-Processing**

In image pre-processing, the collected data are prepared for the next processing by applying image processing operations. This image processing operations are applied because the data are susceptible to noise when captured. In addition, processing the collected data before applying image processing operation requires high computational cost. To reduce the noise and computational cost, the following image processing operations are applied on the data:

- ***Size adjustment:*** As stated in section 5.1, the image is captured using a digital camera and it has a dimension of 1632 x 1224 pixels. Processing the image with this dimensionality is computationally heavy with the available processing resource. Therefore, the image size is adjusted to the optimal image size (i.e. 72 x 72 pixels) that is suitable with the available processing resource.
- ***Converting the image into grayscale form:*** The image is captured in RGB form and there is shortage of RGB image processing tools. In addition, processing the RGB image is computationally heavy when compared with binary and grayscale form. Hence, the image is converted into grayscale representation form.
- ***Contrast adjustment:*** In some images after converted into grayscale, differentiating between the background and the foreground object is difficult with

human eye. Adjusting the contrast of the image is reduced the effect of this problem on the performance of the system.

- **Sharpening:** In some images, it is difficult to differentiate how many fingers are standing and the finger is open or close because of blurredness of the image. Therefore, this problem is reduced by applying the sharpening operation.

### **4.3. Feature Extraction**

The Feature extraction component has a paramount importance on pattern processing system by providing appropriate information for the next processing component of the system. To implement the Ethiopian Finger Spelling Classification System feature extraction component, the principal component analysis driven and the harr-like feature are selected among the features used in other countries' finger spelling and sign language classification systems. This selection is performed based on the availability of the algorithm and the tools used to extract the feature.

As described in section 3.4, the principal component analysis was used for Japanese Finger Spelling recognition to extract the features from the information that comes from the glove sensor. The harr-like feature was used for object detection specially for face detection purpose in the previous works. In this work, both features are applied for hand detection and sign classification purpose with some customization.







In the next sub-sections, how each of the features is used for hand detection and sign classification is discussed in detail.

#### **4.3.1. Principal Component Analysis Based Feature**

Principal Component Analysis is a statistical technique used in image processing to reduce the dimension of image data. In this work, the PCA feature is extracted following the procedure specified in section 3.2.1. The number of significant principal components is determined by using subjective judgment. In addition, we have to use a new method that will help us to determine how many images are reconstructed to select the significant number of principal components.

To determine the significant number of principal components used in this work, first one image is selected randomly and the pixel value of the selected image is transformed by using different number of principal components. Among the reconstructed image using the transformed data, the image with the minimal relevant information is selected to determine the significant number of principal components. The selected image and some of the reconstructed images using different principal components are depicted in Figure 4.2.

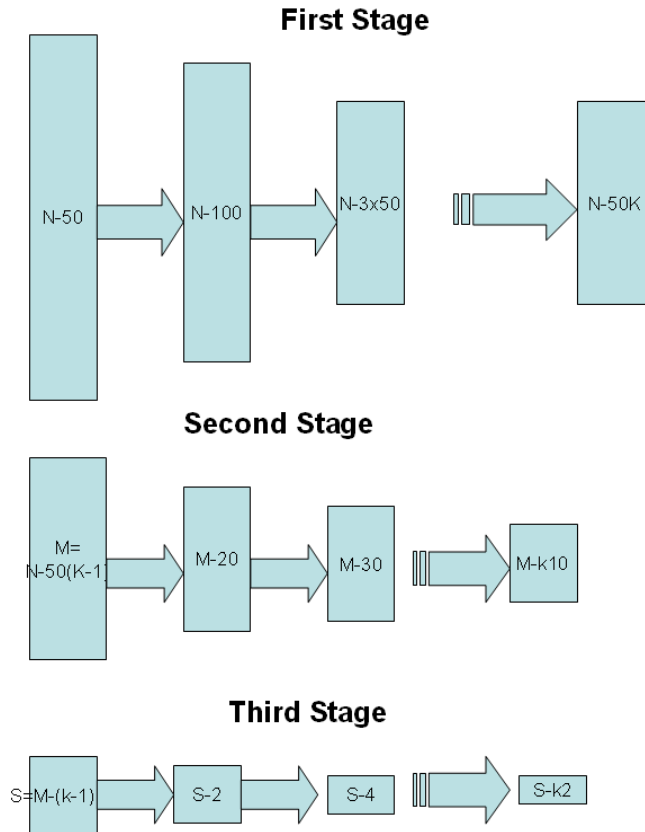
To minimize the number of reconstructed images, a three stage refinement method is used. In the first stage, the image is reconstructed using the top  $N-50$  number of principal components, where  $N$  is the total number of non-zero principal components of the image. This process continues by decreasing the number of principal components used in the previous reconstruction by 50 until all the visible information of the object is lost.

Original Image	
Image with 2680 Principal Components	
Image with 2330 Principal Components	
Image with 1980 Principal Components	
Image with 80 Principal Components	
Image with 40 Principal Components	

**Figure 4-2: Reconstructed image with the corresponding number of Principal Component Analysis**

In the second stage, the image reconstruction is started by using principal component with the highest eigenvalue of the principal components identified before the first stage process stops. Thereafter, the image reconstruction process is performed by reducing the number of principal components in the same fashion with the reduction rate of 10 eigenvectors at each time. This process again continues until all visible information of the object is lost. In the third stage, the image reconstruction starts with the principal components identified before the second process stop. The images reconstructed are performed in the same fashion with the above two stages by using 10 principal component as decreasing factor. Like the previous two stages, this process continues until all visible information of the object is lost. Finally, the image is reconstructed by the significant number of principal components selected among the image reconstructed at the third stage.

In figure 4-3, the three stage image reconstruction refinement method that is described in the above paragraphs is presented diagrammatically. In this diagram,  $N$  stands for total number of principal components and  $K$  stands for the number of maximum image reconstruction in each stage.



**Figure 4-3 : A three stage refinement method to reduce the number of reconstructed image**

To reconstruct the image using different number of principal components, the following procedure is used in this work.

- Calculating the eigenvalue and eigenvector of the data and then the eigenvectors are sorted by descending order of the eigenvalues using the same techniques stated in section 3.2.1.
- The randomly selected image coordinates is projected on the new coordinates that are created with the highest valued principal components or eigenvectors.
- The dimension of the projected data is adjusted to have the same size as the original image (i.e. in our case the original image has 72 X 72 dimension).

- The resulting data is displayed as image and the image is used to decide the significant number of principal components.

As stated in section 5.1, a total of 438 images were collected from different signers and according to the principal component analysis calculation only 2780 of the eigenvectors or principal components exist (i.e. others have zero eigenvalue) among 5184 principal components . In addition, using the above stated procedure to determine the significant numbers of principal components, 80 principal components with highest eigenvalues are selected as significant principal components. Hence, the top 80 principal components are used to transform the training as well as the testing data in this work.

#### **4.3.2. Harr-like Feature**

To extract the Harr-like feature, the image is first divided by 24x24 pixel sized rectangle into sub-regions and from each sub-region three values are computed that correspond to the three types of rectangles used in Harr-like feature as discussed in section 3.2.2. However, before the Harr-like feature is extracted, each image is converted into Integral form. Finally, the feature extracted is normalized (i.e. made between 0 and 1) before usage.

As discussed in section 3.2, the collected image dimension is adjusted into 72 x 72. In this work, from each 72x72 image data and using the three type of 24x24 rectangle, a total of 27 features are extracted from each image and then the features used for hand detection and sign classification component of the system.

## **4.4. Hand Detection**

Hand detection is one of the research problems addressed by pattern processing techniques like artificial neural network and boosted classifier. It has applications in systems like sign language recognition and human-computer interaction. In this work, artificial neural network and boosted classifier based hand detector are presented. These pattern classifiers are selected based on their performance on the previous works over object detection systems in general and the availability of the algorithm and tools. In the following sub-sections, the two hand detection techniques are discussed in detail

### **4.4.1. Neural Network Based Hand Detector**

As discussed in section 3.3.1, usually three types of neural network architecture are applied for pattern classification purpose. To select the appropriate architectures for our work, we use the following points as selection criteria:

- The availability of the algorithm and implementation of the network architecture and training algorithm.
- Whether the collected data has a temporal relation or not.

Based on the above criteria, the implementation of the network architecture and training algorithm of Radial Basis Function network were not freely available at the time this study was conducted and the recurrent neural network is designed to work on data that have temporal relation but the data collected for this specific work don't have temporal relation. Hence, we chose multilayer perceptron to build the neural network based hand detector.

The other issue considered in neural network based hand detector construction is determining the number of hidden layers and nodes in each layer. According to M. Egmont et. al. [21], two hidden layer can handle any pattern classification problem. Therefore, how many nodes in each hidden layers is can be determined using trial and error method.

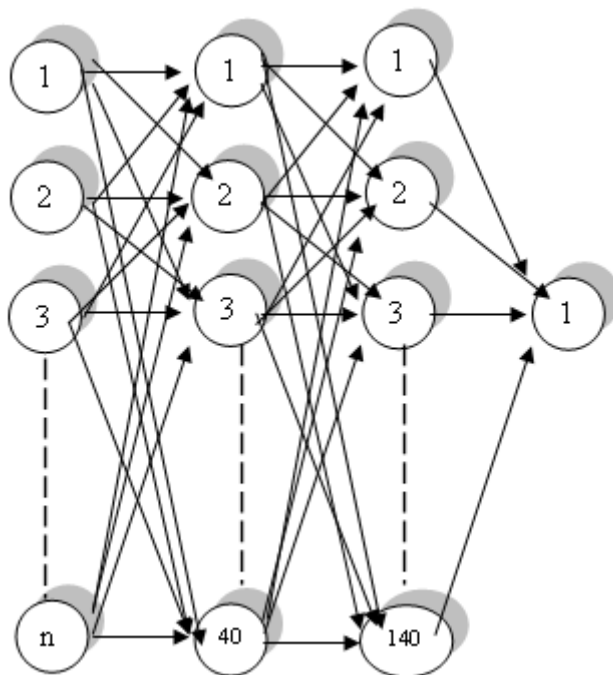
Table 4-1 shows some of the results we found to determine the number of nodes in the two hidden layers of the network using trial and error method. Among the result, the number of node in each hidden layer with the highest detection rate value taken as the right number of node used to construct the number of node in the hidden layers of neural network based hand detector.

**Table 4-1: The result of the neural network based hand detector with different number of node in each hidden layer**

<b>Nodes in 1<sup>st</sup> hidden layer</b>	<b>Nodes in 2<sup>nd</sup> hidden layer</b>	<b>Performance (%)</b>
10	10	96.75
60	60	97.75
90	20	97.86
90	90	97.86
20	20	96.86
80	100	97.86
40	70	98.32
110	70	98.32
70	70	98.43
150	30	98.43
70	40	98.43
<b>40</b>	<b>140</b>	<b>99.43</b>

In addition, the number of nodes in the input and output layers are determined by using the dimensionality of the input feature used and the possible number of target classes or the dimensionality of the output representation.

Hence, the neural network based hand detector of this work contains two hidden layer with 40 and 140 nodes respectively in each hidden layer. The designed network topology is shown in figure 4-4 and it has N number of nodes at the input layer, where N is the dimensionality of the input feature. The value of N is determined by the dimensionality of the input feature used. In addition to this, one node is used in the output layer and if the output of the network is one, this implies that the image holds the hand object otherwise the image doesn't contain the hand object.



**Figure 4-4: Neural Network based hand detector topology**

#### **4.4.2. Boosted Classifier Based Hand Detector**

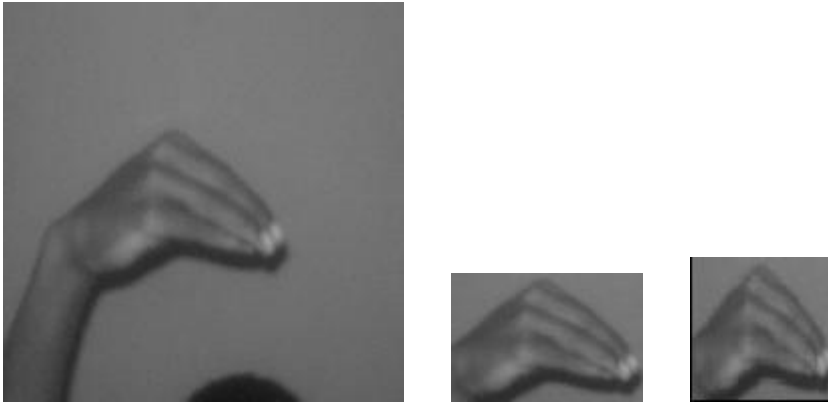
In this work, boosted classifier based hand detector is constructed by following the same procedure as E.J. Ong and R. Bowden [19]. This classifier contains eleven strong classifiers arranged in cascaded manner. Each strong classifier is constructed from weak classifier using Adaboost algorithm. In addition, this detector works only with harr-like feature because we try to construct it using the algorithm stated in [19].

#### **4.5. Region of Interest Segmentation**

Region of Interest Segmentation is the process of finding the minimal rectangular bound of the region that contains the object on the image. This operation increases the performance of the system by removing unwanted information (i.e. Background information) from the image. In addition to this, the size of the image is also reduced. Hence, the computational cost of the system is reduced.

To extract the region of interest for our work, the following simple procedure is applied on the data set before using for the sign classification:

- The coordinate of the object is tagged manually from each image.
- The tagged region of image is cropped.
- The size of the cropped image is adjusted to have the same size.



**Figure 4-5 : The original image, the image after cropping operation and the image after size adjustment**

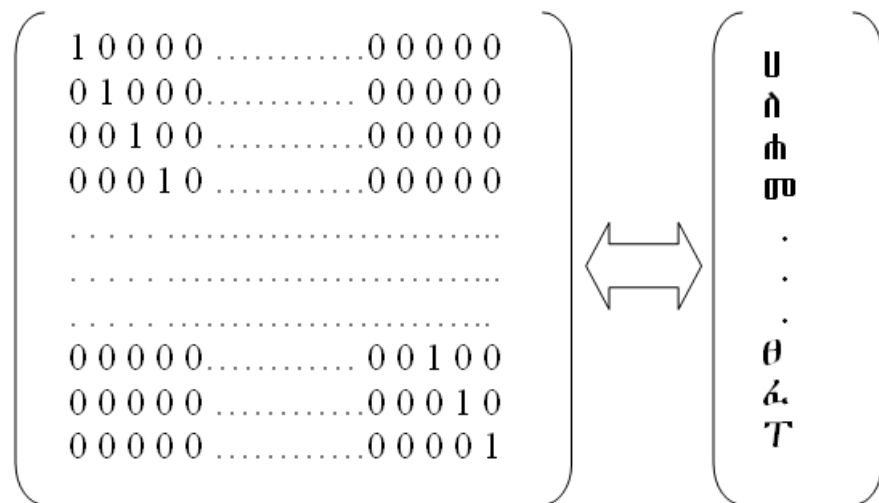
In figure 4-5, the original image before the region of interest segmentation operation applied and the image after the each of region of interest segmentation operations are depicted. As you can see in the figure, the background information and the size of the image is reduced. In addition, unnecessary object is also removed from the image during region of segmentation operation.

#### **4.6. Sign Classification**

In sign classification, the sign from the sign language is associated with the equivalent symbol from spoken language alphabet. In this work, the sign of Ethiopian finger spelling is associated with the alphabet of the Amharic language. As discussed in section 2.5, Ethiopian finger spelling has a one-handed sign using the dominant hand for each alphabet of Amharic. Each sign is constructed using hand shape with or without motion information.

As discussed in section 3.4.2, different pattern classifiers were applied on classification of dynamic and static signs. In this work, two pattern classifiers applied on static sign are selected and compared over Ethiopian finger spelling static signs. The selected pattern classifiers for this task are neural network and template matching.

In the sign classification component, the output of the neural network and template matching based classifier are represented using a sequence of zeros and ones. The total number of zeroes and ones corresponds to the total number of the sign handled by the system. In addition, the position of ones indicates the order of the alphabet in the language. For example, the alphabet HA (**ሀ**) holds the first and the alphabet PE (**ፒ**) holds the last position in Ethiopian finger spelling sign classification system. In figure 4-6, the mapping of the Amharic alphabet with the sequence of zeros and ones is depicted.



**Figure 4-6: Amharic alphabet mapping with sequence of zeros and one**

In the following sections, the neural network and template matching based sign classifiers used for Ethiopian finger spelling are discussed in detail.

#### 4.6.1. Neural Network Based Sign Classifier

The neural network based sign classifier is constructed using multilayer perceptron neural network architecture and this architecture is selected based on the justification stated in section 3.3.1. In addition, the number of nodes in each hidden layer is determined using the same procedure stated in section 3.3.1.

In addition, to avoid the effect of the hand detector error over the sign classifier, the features used for sign classification are extracted from the images that contain the hand object. Those images that contain the hand object are identified manually.

In table 4-2, some of the results of the network with different number of node in each hidden layer are presented. Among the results, the one with the highest overall classification rate and less number of nodes is selected to construct the neural network based sign classifier.

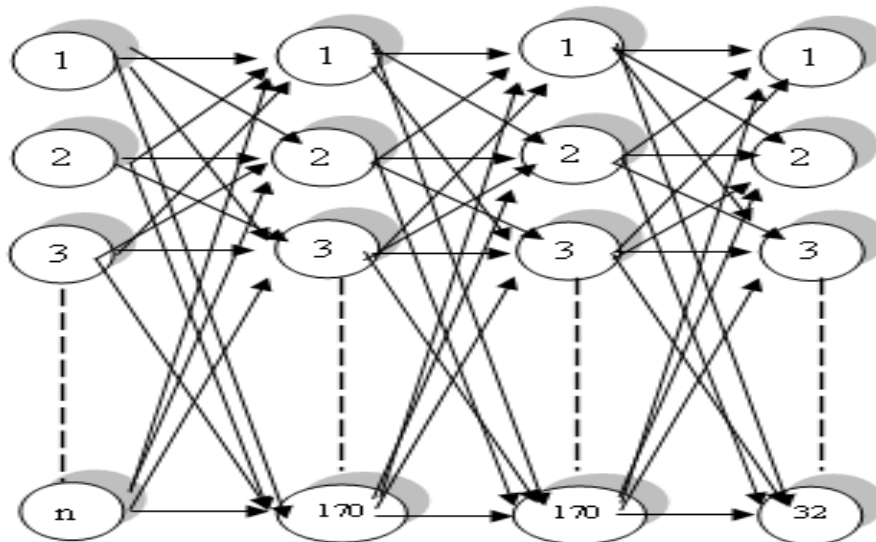


Figure 4-7: The neural network based sign classifier topology

**Table 4-2: The result of the neural network based sign classifier with two hidden layer**

Number		Performance (%)										
1 <sup>st</sup> hidden	2 <sup>nd</sup> hidden	ሀ	መ	ሠ	ረ	ሰ	ሸ	ቦ	ኘ	ኘ	አ	Overall
10	10	38.75	20.00	50.00	20.00	0.00	10.00	0.00	0.00	0.00	0.00	14.51
10	190	67.50	8.89	30.00	0.00	1.25	10.00	26.67	0.00	5.00	7.14	16.10
20	20	43.75	20.00	44.44	10.00	10.00	10.00	7.78	20.00	0.00	4.29	17.32
190	10	40.00	36.67	24.44	12.50	2.50	15.56	27.78	17.14	40.00	22.86	24.15
130	210	67.50	64.44	65.56	78.75	41.25	50.00	40.00	55.71	23.75	61.43	54.76
190	190	48.75	63.33	63.33	70.00	68.75	55.56	47.78	54.29	50.00	55.71	57.80
70	70	60.00	55.56	85.56	66.25	55.00	47.78	46.67	55.71	53.75	51.43	57.93
90	90	57.50	61.11	66.67	80.00	60.00	56.67	61.11	42.86	27.50	67.14	58.29
210	130	50.00	78.89	65.56	60.00	65.00	57.78	50.00	61.43	52.50	57.14	60.00
110	230	77.50	86.67	87.78	83.75	78.75	73.33	95.56	75.71	72.50	82.86	81.71
120	220	73.75	95.56	76.67	95.00	73.75	82.22	80.00	82.86	81.25	85.71	82.68
100	100	75.00	91.11	94.44	90.00	90.00	81.11	70.00	85.71	71.25	88.57	83.66
190	150	73.75	83.33	87.78	90.00	85.00	87.78	84.44	90.00	80.00	78.57	84.15
170	170	70.00	95.56	93.33	96.25	96.25	92.22	83.33	92.86	75.00	91.43	88.66

Hence, the Ethiopian Finger Spelling static sign classifier with neural network is constructed using two hidden layers with 170 and 170 nodes in each hidden layer respectively. As in the case of neural network based hand detector, the node in the input

layer is determined by the dimensionality of the input feature. The number of nodes in the output layer is determined by the dimensionality of the input representation. The diagrammatic representation of the network architecture designed for this system is depicted in figure 4-7.

#### **4.6.2. Template Matching Based Sign Classifier**

The template matching based sign classifier implemented in this work has two components, such as template selector and similarity measure. As discussed in section 5.1, the collected data have a higher variability among the class members and using one template is not enough to represent the class members. Hence, we have tried to select three templates from each class based on the statistical distribution of the data. First, each member image of the class is represented with its mean intensity value. Accordingly, those image data that have mean, minimum and maximum mean intensity value among the class members are selected as templates from each class.

As discussed in section 3.3.2, different similarity measures exist and among them normalized correlation matching and invariant moment are selected based on the characteristics of the data used in this work. The collected image is susceptible to the rotation and transformation because the target object is a movable object. To make the image transformation and rotation invariant the moment feature is calculated and then the similarity of the moment feature of the template and the data are measured using the normalized correlation.

In addition to the correlation based similarity measures, another similarity measure that was used in Data mining and signal processing called dynamic programming is customized to be applied to measure the similarity between the template and data in this work. In this

measure, the similarity is measured by putting the template and image data pixel value into column and row of the result matrix. Each cell of the matrix is populated with the values that indicate similarity between the particular value from the row and column by following the dynamic programming scoring scheme. Finally, the values on the last column are used to determine the degree of similarity between the two parameters.

#### **4.7. Implementation of the System**

To test the effect of the selected features and pattern classifier over Ethiopian finger spelling static signs, the prototype of Ethiopian finger spelling classification system components is implemented using Matlab 7.0, and ImageJ. To capture the input image a SONY digital Camera is used. In addition, the system is implemented on Window XP Professional environment with Intel Pentium IV processor with 3.0 GHZ speed, 512MB Memory capacity, and 80GB Hard disk capacity.

The tools used to develop the prototype are selected based on the following points:

- Free availability of the tool
- It's easiness to use.
- Whether it supports the feature required by this work or not.

In table 4-3, the system component with the corresponding development tools used to implement the component are depicted. Each component is implemented using the tools that have *yes* value on cells of the system component and the tool intersection.

**Table 4-3 : The development tools and the system component relation**

System Component	Development tools			
	Matlab 7.0	ImageJ	GML AdaBoost Matlab toolbox	Digital Camera
Image Acquisition				Yes
Pre-Processing		Yes		
Feature Extraction	Yes			
Region of Interest Segmentation	Yes	Yes		
Hand Detection	Yes		Yes	
Sign Classification	Yes			

#### **4.8. Summary**

In this chapter the architecture of Ethiopian Finger Spelling Classification system were presented. The system has image capturing, image pre-processing, feature extraction, hand detection, region of interest segmentation and sign classification components. Finally, the implementation of the prototype of the system was presented.

In the next chapter, the experiments conducted to measure the performance of pattern processing techniques and features on the two major component of the system (i.e. hand detection and sign classification) with it corresponding results presented in detail.

## Chapter Four

### 5. Experimental Result and Discussion

In order to select the best pattern classifier and feature for Ethiopian finger spelling processing, different experiments are performed. In this chapter, the experiments conducted with its results and the corresponding discussions are presented. In addition, the data collection and the performance measure used in the experiment are also presented.



Figure 5-1 : Some of the collected image of Ethiopian finger spelling sign

#### 5.1. Data Collection

The input image for the Ethiopian finger spelling system is collected from different people using digital camera. Each collected image contains only the hand shape of the selected static sign of Ethiopian finger spelling. The people involved in the data collections are friends and some volunteers and most of them don't have previous knowledge of signing Ethiopian finger spelling. This is because of the problem of getting volunteers as experienced signers. Hence, among the 32 static sign of Ethiopian finger spelling only ten static signs that are easy to sign are selected and collected for this work. The procedures used to collect the data in this work are listed below:

- Finding people that are willing to participate in the data collection.
- Giving 30 minutes training of signing Ethiopian finger spelling for the participants before data collection, if they don't have experience of signing before.
- Selecting signs that are easy to sign for the majority of the participants.
- Collecting the data with digital camera on a constant background and the camera fixed on a photo stand in front of the signer.

Some of the images collected using the above procedures are depicted in Figure 5-1.

In table 5-1, the distributions of the images collected are presented. A total of 438 images were collected and the number of images in each target class is also shown in the table. In addition, the variances existing between each members of the class are also shown and these variances are calculated from the mean intensity value of each image of the class.

**Table 5-1: Collected data distribution**

<b>Alphabet</b>	<b>Number of image collected</b>	<b>Mean variance</b>
<b>ሀ</b>	42	79.84
<b>ሐ</b>	42	791.97
<b>ሐ</b>	43	95.21
<b>ረ</b>	40	66.17
<b>ሰ</b>	40	184.01
<b>ሸ</b>	42	128.38
<b>ቦ</b>	42	94.48
<b>ነ</b>	40	80.91
<b>ኘ</b>	40	86.75
<b>ከ</b>	40	190.12
Negative image <sup>2</sup>	27	

<sup>2</sup> Negative image class contains the images that are only with the background information.

## 5.2. Performance Measure

The performance measure used for the hand detector module begins by classifying the image data based on the result of the detector into three groups of true positive, false negative and false positive. The true positive group includes the system results that are classified into the correct target class. The false positive group includes the system results that are classified the images as the image that holds the hand object but the image doesn't have the hand object. The third group, false negative, includes the images that the system classified as the image that didn't hold hand information but the images have the hand information.

The performance of each group is calculated by dividing the number of images in each group with the total number of test image of that particular target class in consideration. In addition, the overall accuracy of the hand detection module is calculated by dividing the number of images in true positive group with the total number of the test image used in all target class.

However, applying this performance measure for sign classification is not easier as for the case of hand detector because the sign classifier has more than two target classes. Instead, the performance of the sign classifier used considered only the images that fall in the true positive group for each target class. The accuracy of the classifier over each target class is computed by dividing the number of images in each class true positive group with the total number of test images in that particular class. In addition, the overall accuracy of the sign classifier is calculated by dividing the sum of the number of images in each class true positive group with the total number of images used for testing the system.

## 5.3. Experimental Result

The experiment conducted on Ethiopian finger spelling classification system has two major objectives:

- To select the best pattern classifier and the corresponding feature for Ethiopian finger spelling classification system.
- To identify the characteristics of the selected pattern classifiers and the features used.

To meet these objects, six different types of experiments are performed and each experiment is designed by combining the selected pattern classifiers with the selected features. For hand detection, two experiments on neural network based hand detector with harr-like and PCA-driven features are conducted. In addition, another experiment is conducted on boosted classifier based hand detector with harr-like feature.

For sign classification, the first two experiments are conducted on neural network based sign classifier by combining with the harr-like and PCA-driven feature. The third experiment is conducted on template matching based sign classifier. The six experiments described above are summarized in table 5-2.

**Table 5-2 : Summary of the experiment type**

	Experiment	Classifier Type	Feature Type
Hand detection	1	Neural network	PCA driven feature
	2	Neural network	Harr-like feature
	3	Boosted Cascaded	Harr-like feature
Sign Classification	4	Neural network	PCA driven feature
	5	Neural network	Harr-like feature
	6	Template Matching	

The result found on each experiment is presented in detail below:

**Experiment 1:** *Neural network based Hand detector with PCA based feature*

The neural network based hand detector is trained with 80% of PCA based features and the training is performed for about 300 epochs. One epoch means one complete presentation of the whole training data for the network. In addition, the training is performed with learning rate 0.01 and momentum 0.9. Finally, the network performance is tested with 20 % of the PCA based feature. In table 5-3, the number of false positive, false negative and true positive of the neural network based hand detector with PCA based feature are presented.

**Table 5-3 : The result of the neural network based hand detector with PCA driven feature**

<b>Output Group</b>	<b>Number of member</b>
True positive	87/88
False positive	1/88
False negative	0/88
Overall accuracy of the hand detector (%)	98.86

**Experiment 2:** *Neural network based Hand detector with Harr-like feature*

In this experiment, the network training parameters and the partition of the data as a training and testing are the same as the one specified in experiment one. The only difference is the use of the harr-like feature instead of using the PCA based feature. The number of false positives, false negatives and true positives of the neural network based hand detector with harr-like feature are presented in table 5-4.

**Table 5-4 : The result of the neural network based hand detector with Harr-like feature**

<b>Output Group</b>	<b>Number of member</b>
True positive	85/88
False positive	0/88

False negative	3/88
Overall accuracy of the system (%)	96.59

**Experiment 3:** *Boosted classifier based hand detector with Harr-like feature*

The boosted classifier based hand detector that is constructed by following the same procedure as [19] is measured over the Ethiopian finger spelling static sign data. The number of false positives, false negatives and true positives of the boosted classifier based hand detector are presented in table 5-5.

**Table 5-5 : The result of boosted classifier based hand detection**

<b>Output Group</b>	<b>Number of member</b>
True positive	68/88
False positive	12/88
False negative	8/88
Overall accuracy of the system (%)	77.27

**Experiment 4:** *Neural network based sign classifier with PCA feature*

In this experiment, the combination of the PCA based feature with the neural network based sign classifier is tested over Ethiopian finger spelling static sign data. The same network parameters and data partition used in experiment one is also used in this experiment. The accuracy of this sign classifier on each class of the Ethiopian finger spelling static sign is presented in table 5-6.

**Table 5-6 : The result of neural network based sign classifier with PCA based feature**

Target Class	Accuracy (%)
U	71.43
σ	95.24
ω	93.02
ζ	95
η	95
θ	92.5
π	83.33
ι	92.25
γ	75.00
κ	92.25
Overall accuracy of the system	88.08

**Experiment 5: Neural network based sign classifier with Harr-like feature**

In this experiment, the performance of the neural network based sign classifier with Harr-like feature is tested using the same neural network parameter used in experiment one and the result of the experiment achieved on each target class is presented in table 5-7.

**Table 5-7 : The result of the neural network based sign classifier with harr-like feature**

Target Class	Accuracy (%)
U	97.62
σ	95.24
ω	97.67
ζ	100
η	100
θ	100
π	97.62
ι	97.5
γ	95
κ	72.5
Overall accuracy of the system	96.22

### **Experiment 6: Template matching based sign classifier**

In addition to the neural network based sign classifier, template matching based sign classifier performance is also tested over the Ethiopian finger spelling static data. In this test, all images except those selected as templates are used for testing this sign classifier. The result of the template matching based sign classifier is presented in table 5-8.

**Table 5-8 : The result of the template matching based sign classifier**

<b>Target Class</b>	<b>Accuracy (%)</b>
<b>ሀ</b>	30.75
<b>ሐ</b>	28.21
<b>ሠ</b>	32.50
<b>ረ</b>	70.27
<b>ሰ</b>	45.95
<b>ሸ</b>	56.41
<b>ቦ</b>	79.49
<b>ኘ</b>	40.54
<b>ኘ</b>	83.78
<b>ከ</b>	56.76
Overall accuracy of the system	51.44

### **5.4. Discussion of Hand Detector Experimental Result**

The first experiment conducted on the neural network based hand detector with PCA driven feature shows that 87 out of 88 images used for testing are classified correctly and only one image is classified into wrong class. The second experiment is conducted to measure the performance of the neural network based hand detector with harr-like feature and the results show that 85 images out of the 88 images are classified correctly. The rest three images are classified into wrong class. In the third experiment, the performance of boosted classifier based hand detector is measured using the same images that have been used to

test the above two neural network based hand detectors. The experimental results show that 68 images out of the 88 images are classified correctly. But 12 images are detected as containing hand object. However, the images did not actually contain hand object. The rest 8 images are classified as images containing no hand object, but the images did actually contain hand object.

Based on the three experiments, the hand detector with neural network based hand detector shows a better performance when compared with the boosted classifier based hand detector. In addition to this, among the neural network based hand detectors, the hand detector that uses PCA driven feature shows a better performance than the neural network based hand detector that uses harr-like feature. This is because of the PCA driven feature reduces the redundant and noise information from the image that increased the classification rate of the hand detector. However, the harr-like feature uses all the available information without reduction and the noise information which confuse the network.

### **5.5. Discussion of Sign Classification Experimental Result**

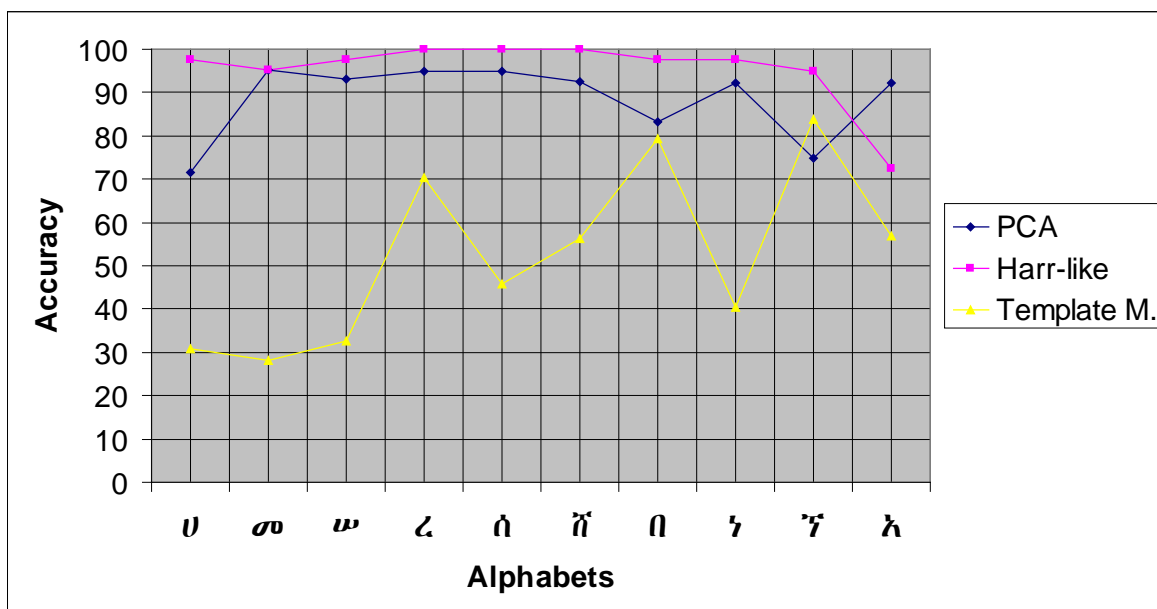
According to the experiment conducted on the neural network based sign classifier with PCA driven feature, the ሀ and ኘ signs have relatively a smaller classification rate. However, the signs of the Amharic alphabet መ, ረ and ሰ relatively have a better classification rate. In general, the classifier classified 78 images into the correct target class out of the 88 images used for testing.

In the experiment conducted on the neural network based sign classifier with harr-like feature, all the signs of ረ, ሰ and ኘ alphabets are classified correctly. However, relatively the sign of the alphabet አ scored less classification rate when compared with the classification rates of the other classes.

The experiment on the template matching based sign classifier shows that the signs of **U**, **o**, and **w** classes relatively have the least classification rate and the signs of **7** and **l** have relatively the higher classification rate than other classes.

Hence, the neural network based sign classifier has higher classification rate when compared with template matching based sign classifier. In addition to this, when we compare the neural network based sign classifier that uses PCA driven feature with the one that uses harr-like feature, the neural network based sign classifier with harr-like feature gives a better result. This is because of some of the information of the image is lost due reduction in the case of PCA driven feature but in the case of harr-like feature all the available information are used. However, the network needs more information like the position of the fingers and orientation of the hand to classify the sign than for the information need for hand detection.

In addition to this, as shown in figure 5-2, the harr-like feature has better classification rate over signs that have up right position like **L**, **n** and **ñ** but have small classification rate for signs that require the hand to rotate horizontal and forwarded direction (i.e. sign of alphabet **o** and **k**). Therefore, the overall performance of the neural network based hand detector with harr-like feature is better than the neural network based hand detector with PCA-driven feature because of 8 of the selected sign out of 10 have the up right position.



**Figure 5-2 : The result of sign classifier graph**

In general, if the number of significant principal components increases and also testing and training data include more signs that have different hand position other than up right position, the performance of the neural network based sign classifier with PCA feature shows better classification rate than the neural network based sign classifier with harr-like feature.

## 5.6. Summary

In this thesis work, a total of six experiments were conducted to measure the effect of different pattern processing techniques with features like harr-like and PCA driven feature over the task of hand detection and sign classification. The experimental result indicates that the neural network has higher detection and classification capability over the collected Ethiopian finger spelling signs when compared with boosted classifier and template matching techniques for hand detection and sign classification task respectively.

In addition to this, the capability of the harr-like and PCA driven features on neural network are also tested for hand detection and sign classification purposes. For hand detection, the PCA driven feature shows a better result over the hand sign of Ethiopian Finger spelling data collected for this work. However, the harr-like feature shows a better classification result over the task of sign classification when compared with the PCA driven feature.

In the next chapter, the conclusion made based on the experimental result and the indication future further research presented.

## Chapter Five

### 6. Conclusion and Recommendation

In this thesis work the issue of classifying the Ethiopian Finger spelling sign is addressed. The classification system architecture proposed in this work includes components like data collection, feature extraction, hand detection, region of interest segmentation and sign classification. Among these components, experiments are conducted to identify the better techniques for feature extraction, hand detection and sign classification modules among the techniques used in other countries' Sign Languages.

For the features extraction module, the PCA-driven and Harr-like features are selected from the features used in previous related works. In this work, the significant number of principal components of the PCA driven feature is determined through subjective judgment and the Harr-like feature is extracted by using 24 x 24 pixel sized rectangles to create the three type of rectangles used in Harr-like feature extraction. In addition to this, the extracted feature data is divided into training and testing set.

In the hand detection module, the neural network and cascaded boosted classifier based hand detector are constructed. For the neural network based hand detector, the optimal number of nodes on each of the hidden layers is determined through trial-and-error method. The cascaded boosted classifier based hand detector is constructed by following the same procedures specified in the work of E.J. Ong and R. Bowden [19].

In this work, the neural network and template matching based sign classifiers are constructed. By using the same procedure used to determine the optimal number of nodes of the neural network based hand detector, the optimal number of nodes in the hidden

layers of the neural network based sign classifier is determined. The template matching based sign classifier that uses three templates from each target classes is constructed and to measure the similarity that exists between the image window and the template, a dynamic programming and normalized correlation with rotation invariant moment is used.

Finally, six experiments are conducted to measure and compare the performance of the hand detector and the sign classifier with the extracted feature. The result of the experiment shows that the neural network pattern classifiers have a better performance with other techniques: template matching and boosted classifier. In addition, the neural network with PCA driven feature performs better on hand detection task and with the harr-like feature performs better for the task of sign classification. However, if the number of significant principal components increased and the data include more signs that have hand position other than up right, the PCA feature shows a better performance than harr-feature for the task of sign classification.

To make Ethiopian deaf beneficiary from the computerized system, many additional researches must be conducted in collaboration with expertise in the field of sign language linguists. Therefore, the following recommendations are made for further research and improvement:

- Enhancing the Ethiopian Finger spelling classification system by including components that detect and classify the motion of the hand.
- Enhancing the Ethiopian Finger spelling classification system by including components that handle left handed finger spelling sign.

- Testing the performance of the Ethiopian finger spelling classification system by applying other techniques
- Studying the syntax of Ethiopian Finger Spelling and Ethiopian Sign Language.
- Enhancing the Ethiopian Finger Spelling classification by incorporating the language associated grammar and syntax.

## References:

- [1]. Alexander Vezhnevets :*GML AdaBoost matlab toolbox*, 2005.
- [2]. World Language and Culture, [http://www.vistawide.com/languages/top\\_30\\_languages.htm](http://www.vistawide.com/languages/top_30_languages.htm), Last accessed on 25 December 2007
- [3]. Baye Yimam.: *Ethiopian Writing System*, <http://www.ethiopians.com/bayeyima.html>, Addis Ababa University , Department of Linguistic, 1992
- [4]. Daniel Assefa: *Amharic Speech Training for the Deaf*, Masters Thesis, Addis Ababa University, Computer Science Department, 2006
- [5]. Sandler, W. : *An Overview of Sign Language Linguistics*. In Keith Brown, (ed.), *Encyclopedia of Language and Linguistics*. 2003.
- [6]. Kyle, J : *Sign and School: Using sign in Deaf child's development*. 1987
- [7]. Kadous, M. W.:*Recognition of Australian sign language using instrumented gloves*, Honours Thesis, School of Computer Science and Engineering, University of New South Wales, 1995
- [8]. Evans, L.: *Total Communication: Structure and Strategy*. Florida: Gallaudet College press, 1982
- [9]. Riekehof, L.L.: *The Joy of Signing: The Illustrated Guide for Mastering Sign Language and the Manual Alphabet* (2<sup>nd</sup> edition). Gospel Pub House, 1987
- [10] American Sign Language: Grammar, <http://www.lifeprint.com/asl101/pages-layout/grammar.htm> , Last accessed on 15 February, 2008
- [11] Ethiopian National Association for Deaf: BERTAT (Yearly Magazine). Addis Ababa: Ethiopia, 1997
- [12] Gonzalez, R. C. & Wood, R. E.:*Digital image processing* (2<sup>ed</sup>), Pearson Education,

2003

- [13] Awcock, G. W. & Thomas R.: *Applied Image Processing*, Mc Graw Hill, 1995
- [14] Acharya, T. & Ray, A. K.. *Image processing principles and Application*, A JOHN WILEY & SONS MC, 2004
- [15] [http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL\\_COPIES/OWENS/LECT2/node5.html](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT2/node5.html), Last Accessed on March 30,2008
- [16] Color Model, <http://academic.mu.edu/>, Last Accessed on March 29,2008
- [17] A tutorial on Principal Component Analysis,  
[http://www.cs.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf) ,  
Last Accessed on may 03, 2008.
- [18] Viola, P. & Jones, M.: *Robust Real-time Object Detection*. Int. Journal of Computer Vision, 2002.
- [19] Ong, E.J. & Bowden, R.: A boosted classifier tree for hand shape detection. *In Face and Gesture Recognition Sixth IEEE International Conference*, pages 889–894, 2004.
- [20] Haykin, S.: *Neural network A Comprehensive Foundation*. Prentice-Hall Inc., 1994
- [21] M. Egmont et. Al.: *Image Processing with neural networks ---a review*, The Journal of the pattern recognition society, 2002
- [22] L. Ding et. Al. : *Volume Image registration by template matching*, Image and Vision Computing, 2001
- [23] Carter, M. P.: *Boosting a Simple Weak Learner for Classifying Handwritten Digits*, Computer Science Technical Report, Department of Computer Science: Dartmouth College, 1998
- [24] Freund, Y. & Schapire, R. E. : *A decision-theoretic generalization of on-line learning and an application to boosting*. Journal of Computer and System

Sciences, 55(1):119—139, 1997

- [25] Xiaojuan Wu, Liqun Xu; Boyang Zhang, Qingguo Ge.: Hand detection based on self-organizing map and motion information. *In Proceedings of the 2003 International Conference on Neural Networks and Signal Processing Volume 1*, Page(s):253 – 256, 2003
- [26] K'olsch, M. & Turk M. :Robust Hand Detection. *In Proc. IEEE Intl. Conference on Automatic Face and Gesture Recognition*, 2004.
- [27] Smach, F.et. al.: *Design of a Neural network Classifier for face detection*. Journal of computer science, 2006
- [28] Wu, X.:Hand detection based on self-organizing map and motion information, *IEEE Int. Conf. on Neural network and Signal processing*, 2003
- [29] Lockton, R., & Fitzgibbon, A. W. :Real-time gesture recognition using deterministic boosting. *In British Machine Vision Conference '02*, 2002
- [30] Lamar, M. V., Bhuiyan, M. S., and Iwata, A.:Hand alphabet recognition using morphological PCA and neural networks. *In Proceedings of IEEE SMC 99*, 1999
- [31] Starner, T. & Pentland, A. :Real-time American sign language recognition from video using hidden markov models. *In SCV95, p. 5B Systems and Applications*, 1995.
- [32] Goh P. :*Automatic Recognition of Auslan Finger-spelling using Hidden Markov Model*. Masters Thesis, University of Western Australian, 2005
- [33] Holden, E.-J.,*Visual Recognition of Hand Motion*, A PHD. Thesis, University of Western Australia, 1997
- [34] Vamplew, P., and Adams, A.:Recognition and anticipation of hand motions using a recurrent neural network. *In IEEE International Conference on Neural Networks*, 1995

- [35] <http://clerccenter.gallaudet.edu/supportservices/series/4010.html> , Last Accessed on August 12,2008
- [36] [www.lifeprint.com/asl101/pages-signs/m/momdad.htm](http://www.lifeprint.com/asl101/pages-signs/m/momdad.htm), Last Accessed on August 25, 2008
- [37] I.T Jolliffe : Principal Components Analysis 2<sup>nd</sup> edition , Springer series in Statistics , 2002.