Automatic Recognition of Dynamic Isolated Sign in Video For Indian Sign Language

A Project Report submitted to Computer Society of India for the MINOR RESEARCH PROJECT (for year 2013-14) by Dr. Anand Singh Jalal

Department of Computer Engineering & Applications
Institute of Engineering & Technology

GLA University
Mathura- 281406, INDIA
April, 2015
Abstract

Sign Language is the only formal way of communication for the mute persons and the hearing impaired. Developing systems that can recognize these signs provide an interface between signers and non-signers by which the meaning of sign can be interpreted. The aim of this project work is to design a user independent framework for automatic recognition of Indian Sign Language which is capable of recognizing various one handed dynamic isolated signs and interpreting their meaning. The proposed approach consists majorly of three steps: preprocessing, feature extraction and recognition. In the first step, i.e. the preprocessing, skin color detection is done followed by face elimination and key frame extraction. A novel method for key frame extraction is proposed which finds out the distinguishing frames from the input video. This algorithm speeds up the system by finding out the most important frames for feature selection. In feature extraction phase, various hand shape features like circularity, extent, convex deficiency and hand orientation are considered. For hand motion, a new feature is proposed called the Motion Direction Code. This feature is efficient in finding out the motion trajectory of the hand while the sign is being performed. Finally, in the recognition phase, Multiclass Support Vector Machine is used to classify the signs and recognize them. Experimentation with vocabulary of 22 signs from ISL is conducted. These signs are one handed dynamic signs which are performed by multiple signers. Results prove that the proposed method for recognition of gestured sign is effective and having high accuracy. Experimental results demonstrate that the proposed system can recognize signs with 90.4% accuracy.
# Contents

Abstract

<table>
<thead>
<tr>
<th>1. Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Motivation and Overview</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Issues and Challenges</td>
<td>8</td>
</tr>
<tr>
<td>1.3 Objectives</td>
<td>9</td>
</tr>
<tr>
<td>1.4 Contribution of the Thesis</td>
<td>9</td>
</tr>
<tr>
<td>1.5 Organization of the Thesis</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. Literature Review</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Introduction</td>
<td>11</td>
</tr>
<tr>
<td>2.2 Previous Work</td>
<td>12</td>
</tr>
<tr>
<td>2.3 Steps in Sign Language Recognition</td>
<td>14</td>
</tr>
<tr>
<td>2.2.1 Hand Detection and Segmentation</td>
<td>15</td>
</tr>
<tr>
<td>2.2.2 Features for Gesture Recognition</td>
<td>15</td>
</tr>
<tr>
<td>2.2.3 Classifier</td>
<td>16</td>
</tr>
<tr>
<td>2.4 Summary</td>
<td>18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. Proposed Methodology</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>19</td>
</tr>
<tr>
<td>3.2 Proposed Framework</td>
<td>20</td>
</tr>
<tr>
<td>3.2.1 Preprocessing</td>
<td>20</td>
</tr>
<tr>
<td>3.2.2 Feature Extraction</td>
<td>24</td>
</tr>
<tr>
<td>3.2.3 Recognition</td>
<td>27</td>
</tr>
<tr>
<td>3.3 Summary</td>
<td>28</td>
</tr>
</tbody>
</table>
4. Experimental Results and Discussions 29
   4.1 Dataset ......................................................... 30
   4.2 Experimental Setup........................................... 31
   4.3 Experimental Results and Discussion...................... 31
   4.4 Summary .......................................................... 36

5. Conclusions and Future Directions 37
   5.1 Summary and Contributions ................................. 37
   5.2 Future Work ..................................................... 38

List of Publications 39

References 40
Chapter 1

Introduction

1.1 Motivation and Overview

The advances in Human Computer Interaction (HCI) in the recent times have led to the growth of man machine communication. With this thrust of development, the technology has contributed significantly to social applications like sign language recognition. Sign language is the prime means of communication between a signer and a non-signer. It is the only formal medium of communication by which mute persons and hearing impaired people express themselves using body movement or facial expressions. Sign language recognition systems make use of mainly hand gestures to communicate information. Hand gesture recognition is another widely growing area which has many applications in various areas such as entertainment industry including interactive virtual tourisms and computer games, controlling industry including 2D and 3D pointing, control of household appliances, objects grasping, command descriptions. Most of the gestures are performed with the hand but also with the face and the body. Hand shape together with its movement and position with respect to the other body parts, forms a hand gesture. Hand gestures are used in every aspect of human communication to accompany speech, or alone for communication in noisy environment and are the most important aspect of the aspect of sign language as most of the information is communicated through hand by the signers.

Developing algorithms and techniques to correctly recognize a sequence of produced signs and understand their meaning is called sign language recognition (SLR). SLR is a hybrid research area involving pattern recognition, natural language processing, computer vision and linguistics [1]. Sign Language recognition systems can be used as an interface between human being and computer systems. Sign languages are complete natural language with their phonology, morphology, syntax
and grammar. A sign language is a visual-gesture language that is developed to facilitate the differently abled persons by creating visual gestures using face, hand, body and arms [2].

SLR and spoken language are drastically different in many respects:

1. The communication medium for spoken language is dependent on sound while, the communication medium for spoken language depends upon the visual channels.

2. Because of the multi-channel of sign language, time is relatively less important for the meaning of a sign as compared to spoken word [3]

3. Phonemes are the smallest constructing unit of words in spoken language. Minimum one phoneme is required to differentiate two words. For sign language, cheremes are used to build up a sign and at least one chereme must be different among different signs [2].

The sign language being used in India is called Indian Sign Language (henceforth called ISL). Similarly the sign language of other parts of the world are American Sign Language (ASL), British Sign Language (BSL), Japanese Sign Language (JSL), Arabic Sign Language (ArSL) and German Sign Language (GSL), etc. There are different dialects of ISL which are present in different parts of India. However, all the dialects possess the same grammatical structure despite having large lexical variation [4]. A sign is a made by cheremes that are can be either manual or non-manual [2]. Cheremes are the sub units of sign which are equivalent to the phonemes of spoken languages. Manual cheremes consists of several parameters like: hand shape, hand orientation, hand location and hand motion and non-manual cheremes are defined by facial expression, head/body posture.

There are three categories of ISL signs. They are: one handed, two handed, and non-manual signs. However, there are certain signs which use only either the manual parameters or non-manual parameters. For example, a single waving hand representing the sign of “Hello” has no non-manual components and sign “Yes” is depicted by nodding the head vertical without any manual component. Figure 1.1 shows the overall Indian sign hierarchy.
One handed signs: These signs make use of only one hand which is called the dominant hand. These signs can be either static or dynamic. Both the static and dynamic signs are divided into manual and non-manual classes of signs. Figure 1.2 represents the examples of one handed static signs having manual and non-manual components.

Two hand signs: Like one handed sign, similar classification can be applied to two handed signs. However, two handed signs which are dynamic are further classified as:

Type 0: In this type of signs, both the hands are dominant (see Figure 1.3).

Type 1: In this type of signs, one hand is more dominant with respect to the other. The other is termed as non-dominant. The example is shown in Figure 1.3.
The main approaches used in sign language recognition system can be classified as: vision based approach or device based approach [5]. Vision based techniques are used mostly in case of SLR. These techniques make use of image features like color, shape, texture etc. and also require the preprocessing of the input image. The input to the vision based systems can be an image or video which is either recorded from single or multiple cameras or captured in real time through imaging device which is connected to the system. In device based approaches, devices such as motion sensors, data gloves etc. are used. These devices measure the hand shape and motion. These approaches are inconvenient to signer because they require cumbersome devices to be worn by the user and also restrict their movement. The examples of these approaches are shown in Figure 1.4.

The main dimensions of research in SLR can be grouped as: isolated sign recognition and continuous sign recognition. The first type i.e. the isolated sign recognition is concerned with the recognition of single signs without continuation to the other sign. These signs can be either static or dynamic. No other sign is performed before or after the isolated signs in continuation. Thus, this type of sign is not affected by the preceding or succeeding sign [6]. In continuous signing, a complete sentence is
to be recognized consisting of various signs performed one after the other. The aim is to identify the different signs being performed continuously.

Sign language mainly makes use of hand features which could be hand shape, hand orientation, hand location, and hand motion trajectories. There are two methods of extracting the aforementioned features. They are: tracking based or non-tracking based. Hand tracking is a challenging task due to the fact that in a sign language, the movement of hands is very fast and causes motion blur. Hands are deformable objects which change its shape, position and direction while signing. The various methods used for tracking are Kalman filter, sequential Monte Carlo tracking, particle filters and other graphical models. Since hand tracking is very difficult task and not much useful for isolated signs having short duration, some systems like [7], [8], [9] do not use tracking. In the systems which do not require tracking (i.e. the non-tracking based system), hand is segmented from the image through various methods, then feature extraction and recognition is done.

There are three basic movement phases of signs for a sign language. These are: preparation, stroke, and retraction [10]. In the preparation phase, the hand moves towards desired location in the signing space. In the stroke phase, the intended movement is done and hand configuration is changed. In the retraction phase hand is released from final position and signing is finished. The stroke stage is the most important stage of the sign language recognition systems which has the actual information about the sign. But this phase contains many redundancies. These redundancies should be removed in order to speed up the system and contain most important and distinguishing information. Some of the applications of the sign languages are as follows:

**Telecommunications [11]**

Communication through sign language is possible among various remote locations. This is done by using a special-purpose videophone which is meant for the use with sign language or “off-the-shelf” video services designed or an ordinary computer with webcams and broadband. These video devices are developed specially for the sign language communication as they enhance the precision of sign language by providing more frames per second than the “off-the-shelf” services. Since the frame rate is high, different compression techniques can be applied to these videos.
In 1964, “Picturephone”, a videophone developed by AT & T was used for telecommunication aiding the mute and hearing impaired to communicate with each other.

**Sign language interpretation [11]**

In order to make easy communication between hearing people and deaf, sign language interpreters are frequently used. For this interpretation, the signing done requires effort by the interpreter. This effort is due to the fact that sign language vary in their syntax and grammar.

![Figure 1.5](image1.png)

**Figure 1.5** A deaf person communicating with a hearing person using a remote VRS interpreter.

![Figure 1.6](image2.png)

**Figure 1.6** In the Joe Greene jerseyan ASL interpreter appearing at a rally for the Pittsburgh Steelers.

**Home Sign [11]**

Home signs are the signs which are developed within a family by themselves in order to communicate. This happens when a person or more within the family are disable,
then there is a need to communicate with other members of family. Using this self-created signs, the information can be communicated and expressed through various hand gestures, body postures, facial expressions. These signs are also called home signs or kitchen signs as they are understandable within the family or group that has created them.

**Tele presence**

In case of hostile conditions like disasters, system failure, etc. emergency conditions may arise, especially in the remote areas. To tackle these, certain manual operations become necessary. In such situation, physical presence of the human operator near the machine is impossible. Tele presence refers to the technology which makes the person feel that they are present at a place different from their actual location for carrying out a desired task. For E.g. ROBOGEST is one such real time system [12] through which hand gestures are used to control the outdoor vehicles. This system was developed at University of California, San Diego [13].

**Virtual reality**

Virtual reality recreates real environment and gives the feel of the physical presence in that environment and hence providing the reality virtually. This experience is visual and provided through computer screen other devices like stereoscopic displays [14]. Virtual reality is used in various fields like military, education, health care, telecommunication, scientific visualization, etc. Most of the virtual reality based systems make use of hand gestures and body movements for communication. Hence, the sign language recognition methods are helpful in virtual reality as well.

**Gestural Language**

Gestures are the body actions which are used to express the feelings, or to communicate. Gestures are the non verbal way of expressing one’s thoughts without the need of spoken language.

Gestures can be used to convey information at various places. For e.g. in airports, gestured made by ground persons are helpful to communicate and give direction to the pilot. These gestures are predefined and thus understandable and much needed as communication through spoken language is not possible otherwise on
such a distance between pilot and the other person on ground. Similarly, in certain games referees perform gestures that denote a particular meaning. For e.g. the umpire in game of cricket performs different gestures depicting different meaning like four, six, out, no ball, wide ball, etc.

The goal of this chapter is to give the brief introduction about the sign language and also about sign language recognition system. It also discusses the application areas of the sign language recognition system. The remainder of the chapter is structured as follows: Section 1.2 discusses the various issues and challenges of the SLR system. Section 1.3 provides us the objective of proposed work. Section 1.4 states our contribution to the project and Section 1.5 illustrates the organization of the project.

1.2 Issues and Challenges in Indian Sign Language

Developing a system for deaf people to recognize sign language is a very challenging issue due to its cardinal social interest and its intrinsic complexity. The various issues and challenges related to Indian sign language are as follows:

**Issue:** Geographical variation in Indian Sign Language.

**Challenge:** Creating a standard dataset for ISL. Since the sign language within India differs in different regions, there exist no standard dataset for Indian sign language.

**Issue:** Segmentation

**Challenges:** The various challenges which for segmenting the face and hands are occlusion, presence of disturbances and cluttered background. There are possibilities of hand occluding the face which makes segmentation of hand typical. Various disturbances like illumination change, noise, and undesired movement also cause problem in segmentation. It is often difficult to segment the face and hands in cases where the background is cluttered.

**Issue:** Recognition Accuracy

**Challenges:** Variation in viewpoint. The major difficulties of sign language recognition are feature extraction and the recognition itself. Secondly, nature of information is multi-channel. This means that face, body, hand can be used in parallel to communicate a sign.
**Issue:** Tracking.

**Challenge:** Signing speed varies from person to person. Articulated objects (such as the hand) are more difficult to track than single rigid objects. Because gestures are highly variable from one person to another, and from one example to another within a single person, it is essential to capture their essence-their invariant properties-and use this information to represent them.

**Issue:** Analysis and integration of manual and non-manual signals

**Challenges:** Usage of both static and dynamic hand gestures. The non-manual signs are typical to recognize. In certain cases, the signs use complex hand shape

**Issue:** Gesture Spotting

**Challenge:** The task in gesture spotting is to differentiate the meaningful gestures of the user from the unrelated ones. The garbage movements that come before and after a pure gesture should be removed. Succeeding signs are affected with preceding sign in continuous sign language recognition which is called co articulation problem.

### 1.3 Objective

The objective of this project work is to develop a robust automatic system to recognize signs from Indian Sign Language using vision based approach for one-handed dynamic isolated signs. The proposed system will translate the video of the sign to text. The proposed system’s performance will be tested on large vocabulary of sign from ISL with reduced training samples. The motive of this work is to provide a real time interface so that signers can be able to easily and quickly communicate with non-signers. In India, there is a need of developing an automatic sign language recognition system, which can accomplish the need of hearing impaired people. Unfortunately, till date not much research work has been reported on Indian Sign Language recognition. Moreover, our work has been done on various signers for testing. This had lead the proposed system to be user independent which does not restricts different users to use our proposed system.

### 1.4 Contribution of the Project

The contribution of this project work is to design an automatic system for recognition of sign from ISL. The major contributions of this project are as follows:
• A novel feature known as Motion Direction Code (MDC) is proposed for finding hand motion direction, which distinguishes various isolated dynamic signs of Indian Sign Language. This feature is efficient in finding out the motion trajectory of the hand while the sign is being performed.

• Unlike the majority of previous work, the proposed framework does not require tracking of hand.

• Unlike previous work, the proposed approach does not require any measurement devices such as magnetic trackers or color gloves.

• Developing a framework for automatic recognition of one handed dynamic signs for ISL using multiple features for better recognition rate using Multiclass Support Vector Machine.

• Developing an efficient system for recognition of isolated signs using a novel key frame extraction algorithm to speed up the recognition task.

• Creating a large dataset for varied one-handed dynamic signs of Indian Sign Language.

• No previous work considered the user independent system for dynamic Indian sign language recognition, but the proposed system is able to create a user independent framework. This is ensured by testing the system on various users having different age groups ranging from 5 to 60 years.

1.5 Organization of the Project

This project is organized as follows:

Chapter 2 provides an extended study of literature survey for the main problems of gestural interfaces, namely hand segmentation and detection, tracking and recognition of sign language.

Chapter 3 gives a brief description of the proposed methodology for one handed isolated (static and dynamic) signs recognition. It uses Multi-class Support Vector Machine (MSVM) for recognition of signs using multiple features.

Chapter 4 discusses the conclusion and future scope of the work.
Chapter 2

Literature Review

2.1 Introduction

In this section, the recent work in the area of automatic recognition of sign language is discussed. There are varied techniques available which can be used for recognition of sign language. Different authors have used different techniques according to the nature of sign language and the signs considered. Sign language recognition is mainly consisting of three steps: preprocessing, feature extraction and classification. In preprocessing, a hand is detected from sign image or video. In feature extraction, various features are extracted from the image or video to produce the feature vector of the sign. Finally, in the classification, some samples of the images or videos are used for training the classifier then test sign is identified in image or video. A lot of work has been done on static sign but unfortunately, till date not much research work has been reported for dynamic sign in Indian Sign Language. The proposed work recognizes static and dynamic sign of ISL. Different researchers use the numerous types of approaches in recognizing sign language. We will discuss some of these approaches on the basis of following parameters:

1. Size of the vocabulary, the number of sign classes used for experimental results.
2. The approach used for collecting data for the classification process: instrumented
3. Glove-based or vision based data collection.
4. The types of sign: static sign or dynamic sign.
5. The skin color model used for hand segmentation: YCbCr, HSV, RGB etc.
6. Tracking or non-tracking based methods.
7. The features that they have extracted from the input sign.
8. The classifier used for recognition of sign.
9. The accuracy of existing methods.
10. Research dimension: isolated or continuous sign.

The organization of this chapter is done as follows: Section 2.2 discusses the previous work and steps for sign language recognition are presented in section 2.3. Finally, Section 2.4 contains the summary of the chapter.

2.2 Previous Work

In earlier work on two handed sign recognition, Agrawal et al. [16] proposed a two stage recognition approach for 23 alphabets. Signs are produced by wearing red gloves on both hands for segmentation purpose. The segmented images serve as an input to feature extraction and recognition phase. In stage-I the features that are describing the overall shape of gestures, were calculated and recognition is done through training feature vector without use of any classifier. In stage-II, a recognition criterion was tough and feature vector had binary coefficient. Finally, an output was given whether gesture is correct or not.

In [17], a method for the recognition of 10 two handed Bangla character using normalized cross correlation is proposed by Deb et al. A RGB color model is adopted to select heuristically threshold value for detecting hand regions and template based matching is used for recognition. However, this method does not use any classifier and tested on limited samples.

M.A. Mohandes [18] proposed a method for the recognition of the two handed Arabic signs using the Cyber Glove and support vector machine. Principal Component Analysis (PCA) feature is used for feature extraction. This method is consisting of 20 samples of 100 sign by one signer. 15 samples of each sign were used for training a Support Vector Machine to perform the recognition. The system was tested on the remaining 5 samples of each sign. A recognition rate of 99.6% on the testing data was obtained. When number of signs in the vocabulary increased, the support vector machine algorithm must be parallellized so that signs are recognized on real time. The drawback of this method was to employ 75% images for training and remaining 25% for testing.

Work on two handed signs has been done in Rekha et al. [19]. Here, Principle Curvature Based Region (PCBR) is used as a shape detector, Wavelet Packet Decomposition (WPD-2) is used to find texture and complexity defects algorithms are
used for finding features of finger. The skin color model is used here is YCbCr for segmenting hand region. The classifier used is Multi class non-linear support vector machines (SVM). The accuracy for static signs is 91.3%. However, three dynamic gestures are also considered which uses Dynamic Time Warping (DTW). The feature extracted is the hand motion trajectory forming the feature vector. The accuracy for the same is 86.3%.

In [5], threshold models have been designed to differentiate between signs and non-sign patterns of American Sign Language in Conditional Random Field (CRF). The recognition accuracy for this system is a 93.5%.

Aran et al. [1] have developed a system called the sign tutor for the automatic sign language recognition. The three stages in this system are: face and hand detector, analysis and classification. User is made to wear colored glove in order to easily detect the hand and remove occlusion problems. For both the hands, kalman filters are used for smoothening hand trajectories and thereafter, features for finding the hand shape are extracted. The classifier used is Hidden Markov Model (HMM). The dataset consisted of 19 signs from ASL. The accuracy for signer-dependent system is 94.2%, while for signer-independent system is 79.61%.

Lekhashri and Pratap [20] developed a system for both static as well as dynamic ISL gesture recognition. The various features extracted are skin tone areas, temporal tracing and spatial filter velocimetry. This obtains the motion print of image sequence. Then pattern matching is used to match the obtained motion prints with the training set which contains the motion print of the trained image sequences. Then, the closest match is produced as the output.

Nandy et al. [9] proposed an approach using direction histogram feature. The classification is done through Euclidean Distance and with also K-nearest neighbor also and both are compared. In the dataset, only isolated hand gestures of 22 ISL signs are taken. The recognition rate is found out to be 90%. The limitation of this approach is the poor performance in case of similar gestures.

Sandjaja and Macros [21] also used color-coded gloves for tracking human hands easily. Multi-color tracking algorithm is used to extract the features. The recognition is done through Hidden Markov Model. The dataset consisted of Filipino sign language numbers. The recognition rate is 85.52%.
Quan [22] proposed extraction of multi-features for the image, where hand is the only object. These features are color histogram, 7 Hu moments, Gabor wavelet, Fourier descriptor, and SIFT features and used support vector machine for classification.

Bauer and Hienz [23] introduced the basic problems and difficulties that arise while performing continuous sign language recognition. In this paper, manual sign parameter information such as hand shape, hand location and orientation are extracted, forming the feature vector. The hand motion is not considered as it is handled separately by HMM topology, and hence not included in feature vector. Gloves of different colors are worn by users to distinguish dominant and non-dominant hand. For each sign, one HMM is modeled. The system was tested on 52 different signs of German Sign language (GSL) and 94% accuracy found for all features. For 97 sign, this accuracy drop to 91.7%.

Alon et al. [24] proposed a unified framework for simultaneously performing temporal segmentation, spatial segmentation and recognition. There are three major contributions of this paper: Firstly, for the detection of hand, multiple candidates are detected in every frame and the hand is selected through a spatiotemporal matching algorithm. Secondly, for easy and reliable rejection of incorrect matches, a pruning framework is used which is based on classification. Thirdly, a sub-gesture reasoning algorithm is used that finds those models of gestures that falsely match to some parts of other large gestures. Skin color combined with motion cues is used for hand detection and segmentation.

Shanableh et al. [25] presented a technique for Arabic Sign Language Recognition. This method works online as well as offline for the isolated gestures. The technique used uses varied spatio-temporal features. The features used for temporal context are forward, backward, and bidirectional predictions. After motion representation, spatial-domain feature is extracted. For classification purpose, Bayesian classifier and k-nearest neighbor are used. The accuracy of this system varies from 97% to 100%.

### 2.3 Steps in Sign Language Recognition

Sign language recognition comprises of three major steps: hand detection and segmentation, feature extraction and classification.
2.3.1 Hand Detection and Segmentation

Hand segmentation is the process of extracting hand means pixels representing the hand are localized in the image and segmented from the background before recognition. In segmentation procedure, a number of restrictions are imposed on background, user and imaging [26]. Restrictions on the background and in imaging are commonly used. A controlled background greatly simplifies the task. It can vary from a simple light background [27] [28] to a dark background [29] [30]. Mostly a uniform background is used. In the case of restriction on user, the user can wear long sleeves [31]. In the case of restriction in imaging, cameras are focused on the hand [27] [32] [33] [34]. Another way to simplify this problem is to adorn the user hand(s) with gloves [16][35]. The chosen color greatly helps the segmentation task.

Skin color segmentation is another approach that can be used to detect hand but drawback of this method that it also finds face, so we have to exclude the face from the image. Skin color can be modeled using simple histogram matching [36], mixtures of Gaussians [37]. The spaces used can be RGB (red, green and blue components) [38], normalized RGB [39], YUV space [40], HSI (hue, saturation and intensity model) [41][42].

2.3.2 Features for Gesture Recognition

Feature extraction is the most important module in sign language recognition system. Since the nature of every sign language and signs considered is different, the reliable features need to be selected. Feature extraction is aimed at finding the appropriate and most distinguishing features for the object. Sign language recognition can be accomplished using manual signs (MS) and non-manual signs (NMS). Manual signals includes features such as hand shape, hand position and hand motion whereas non-manual signals include facial features, head and body motion.

A lot of previous work has been done by extracting appearance based features. This is because these features are simple and have low computational time, and therefore can be used for real time applications. The feature descriptors can be classified as edge, corner, blob or region based descriptors. The various shape based descriptors can be contour based or region based. The region based shape descriptors are further classified as local or global descriptors. These features include region based descriptors (image moments, image eigenvectors, Zernike moments [43], Hu invariants [44], or grid descriptors) and edge based descriptors (contour
representations [45], Fourier descriptors [46]). There are colour, motion and texture based descriptors also.

2.3.3 Classifier

Once features have been computed, recognition of signs can be performed. Sign recognition can be decomposed into two main tasks: the recognition of isolated signs and the recognition of continuous signs. The various recognition techniques include Support Vector Machine (SVM), template matching, neural networks, geometric feature classification, or other standard pattern recognition techniques. For continuous sign recognition, the temporal context needs to be considered. It is a sequence processing problem that can be accomplished by using Finite State Machines (FSM), Dynamic Time Warping (DTW), and Hidden Markov Models (HMM) to cite a few techniques.

In [47], a user independent framework for recognition of isolated Arabic sign language gestures has been proposed. For this, the user is required to wear gloves for the simplification of hand detection and segmentation. K-Nearest Neighbor and polynomial networks are the two classifiers used in this paper and then these two classifiers’ recognition rate are compared. Certain special devices like cyber gloves, or sensors can also be utilized for sign language recognition. These devices find the accurate position and motion of the hand. Though more accurate, these devices are cumbersome and prevent the natural interaction of the signer with the computer.

In [48], Back propagation and Kohonen’s self-organizing network has been applied to recognize gestures related to American Sign Language (ASL) for 14 sign vocabulary. The overall accuracy of the system is 86% by using back propagation and reduces to 84% when Kohonen’s network has been applied. However, the low recognition is due to insufficient training data, lack of abduction sensors or over constraining the network.

In [49], Euclidian space and neural networks has been used for recognizing the hand gestures. They have defined some specific gestures and made a test on that. The achieved accuracy is 89%. Since different users perform the signs in different manner, the number of false positives is increased and the recognition rate is also low; the other factor being the occlusion of fingers.
In [50], the authors present a hierarchical structure based on decision trees in order to be able to expand the vocabulary. The aim of this hierarchical structure is to decrease the number of models to be searched, which will enable the expansion of the vocabulary since the computational complexity is relatively low. They used a sensored glove and a magnetic tracker to capture the signs and achieved 83% recognition accuracy, at less than half a second average recognition time per sign, in a vocabulary of 5113 signs.

One of the biggest challenges in sign language recognition arises in the case of continuous sign sentences, which means that a sign is succeeded and preceded by certain other sign, therefore forming a sentence of these signs. This is similar to co-articulation problem in speech. The transition from end of one sign to the start of another sign needs to be identified in order to find the isolated signs within the continuous signing. Such movement is known as movement epenthesis. Among all the techniques in continuous sign recognition, HMM [51] [52] [53] is the most important and widely used technique.

In [33], dynamic hand gestures having both local and global motions have been recognized through Finite State Machine (FSM). In [54], a methodology based on Transition-Movement Models (TMMs) for large-vocabulary continuous sign language recognition is proposed. TMMs are used to handle the transitions between two adjacent signs in continuous signing. The transitions are dynamically clustered and segmented; then these extracted parts are used to train the TMMs. The continuous signing is modeled with a sign model followed by a TMM. The recognition is based on a Viterbi search, with a language model, trained sign models and TMM. The large vocabulary sign data of 5113 signs is collected with a sensored glove and a magnetic tracker with 3000 test samples from 750 different sentences. Their system has an average accuracy of 91.9%.

Agrawal et al. [55] have proposed a user dependent framework for Indian Sign Language Recognition using redundancy removal from the input video frames. The skin color segmentation and face elimination is performed to segment the hand. Various hand shape, motion and orientation features are used to form a feature vector. Finally a MSVM is used to classify the signs with 95.9% accuracy.
2.4 Summary

A lot of work has been done in the case of static isolated signs of Sign Language Recognition systems. Different researchers have used varied methods for the same. Some result in high recognition accuracy at the cost of high computational complexity while some systems are simpler but less accurate. Different datasets pertaining to different corners of the world are created which employ different level of complexity and constraints. Several methods have been proposed to solve the three main problems of vision-based gestural interfaces, namely hand segmentation and detection, tracking and recognition. This chapter discusses the different earlier methods or techniques available and applied in each of the above three phases of the recognition system.
Chapter 3

Proposed Methodology

3.1 Introduction

Sign language recognition systems are being developed in order to provide an interface for the hearing impaired and mute persons. These automatic sign language recognition systems allow the non-signers to interpret the meaning of what the signer wants to convey and therefore facilitating the communication between them. Researches in this direction come under the category of Human Computer Interaction (HCI).

Most of the previous work in Indian Sign Language has focused on static signs and images of signs with constant background and illumination. Some of these have used those images in which only hand is present so that segmentation is not hard, while others have used colored gloves that are needed to be worn by the users while signing in order to detect and segment the hand easily. Moreover, almost all SLR systems developed so far have considered only one signer for training and testing of different signs. These are called the user dependent systems.

The aim of this research work is to design a user independent Automatic Sign language Recognition system which is capable of recognizing various one-handed dynamic signs of Indian Sign Language performed under different background conditions. The proposed system includes key frame extraction and combination of certain new vision based features complemented by multi-class support vector machine (MSVM).

Hence, this chapter discusses the proposed methodology for dynamic isolated sign recognition in video.
3.2 Proposed Framework

The block diagram of the proposed framework is depicted in Figure 3.1. The three main components of this framework are: preprocessing module, feature extraction module and recognition module. In the first step, i.e. the preprocessing, skin color detection is done followed by face elimination and key frame extraction. In feature extraction phase, various hand shape features like circularity, extent, convex deficiency and hand orientation are considered. For hand motion, a new feature is proposed called the Motion Direction Code. Finally, in the recognition phase, Multiclass Support Vector Machine is used to classify the signs and recognize them.

![Figure 3.1. Framework of the proposed system.](Image)

3.2.1 Preprocessing

In the preprocessing step, the video is converted to the skin segmented frames; from which the face is eliminated, so that we’re left with only hand frames. The final and most important step in preprocessing is the key frame extraction in which only the significant frames are selected whose features are calculated in the next step.

Skin Color Segmentation and Noise Removal

The initiation of pre-processing module takes place by converting the video into frames. Thereafter, for every frame, the skin color is detected. For skin color segmentation, the color space used is the YCbCr. In this model, Y represents the
brightness and Cb,Cr are the chrominance values for blue and red light. Here, the YcbCr model is preferred because of the fact that the skin color can be identified with the chrominance component and it separates the brightness value from the chrominance values. The suitable values for Cb and Cr are: $C_b= [77 \ 127]$ and $C_r= [133 \ 173]$ [56]. Figure 3.2(b) shows the results of skin detection algorithm which results in finding the face and hand from the frames of the input video. The advantages of choosing YCbCr model over other models for skin detection are:

1. Since only chrominance component is considered, the algorithm will work in different brightness conditions.
2. It leads to the reduction of feature space from 3D to 2D.
3. The skin tone varies in terms of the luminance component and not on chrominance component. Hence chrominance value Cb and Cr will be almost same for different races.

**Hand Detection by Face Elimination**

After the skin segmentation is done, the noise and face are to be eliminated from all the frames to get the hand as the only segmented object. This is called the hand detection phase. The algorithm for the same is mentioned below in table 3.1. The result of this step is shown in Figure 3.2 (c).

<table>
<thead>
<tr>
<th>Table 3.1: Hand Segmentation Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. For $f_1$, find the largest connected component $L$.</td>
</tr>
<tr>
<td>2. For $i=2$ to $n$ do</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3. Return $h_i$</td>
</tr>
</tbody>
</table>

**Key frame Extraction**

The videos of the dynamic signs consist of large number of frames. All of these frames are not essential in order to determine the meaning of the performed sign, rather, only few important frames from the video are sufficient. These most important and thus
distinguishing frames are known as keyframes. In our system, we have devised an algorithm for finding the keyframes from the video.

The basic approach is to select those frames from video in which there is a significant change in either position or shape. For position change, the centroid feature is taken into account. The centroid of every frame is computed and the distance between centroid of every frame (starting from the first) and its successive frame is computed. If the distance is greater than a particular threshold, then the frame is selected, otherwise it is skipped. Here, the threshold used can’t be fixed or kept static as the amount of change in position varies with every sign. So dynamic threshold ($D_{th}$) is used, which is calculated as a function of the average distance between the centroid of every key frame.

Figure 3.2. Illustration of preprocessing steps (a) Sample 12 frames from Video sequences of sign “Only” from our data set (b) Skin color segmentation result (c) Hand detection result
The next factor to be considered for key frame selection is the change in shape. A sign might change in shape without the change in position. So in that context, the feature used is solidity which can distinguish between those frames which have variation in shape. Therefore, for appropriate key frame selection, both of the above mentioned features are selected and a logical operation between them is used to select the key frames. The algorithm for key hand frame extraction is given in table 3.2. The result of this step is shown in Figure 3.3.

![Figure 3.3: Frames extracted by the key frame extraction algorithm](image)

**Figure 3.3.** Frames extracted by the key frame extraction algorithm

**Table 3.2: Key frame Extraction Algorithm**

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Input video $V = h_1, h_2, h_3, \ldots, h_n$, where $n$ is the number of hand frames in the video.</td>
</tr>
<tr>
<td>2.</td>
<td>For $i = 1$ to $n$ do</td>
</tr>
<tr>
<td></td>
<td>a. Compute the centroid $c_1, c_2, c_3, \ldots, c_n$ and solidity $s_1, s_2, \ldots, s_n$ for every frame $h_i$ respectively.</td>
</tr>
<tr>
<td>3.</td>
<td>Compute the threshold values $D_{thr1}$ and $D_{thr2}$.</td>
</tr>
<tr>
<td>4.</td>
<td>Set $j = 1$ and $i = 2$.</td>
</tr>
<tr>
<td>5.</td>
<td>Select $h_j$ as key frame</td>
</tr>
<tr>
<td>6.</td>
<td>While ($i &lt;= n$) do</td>
</tr>
<tr>
<td></td>
<td>a. Compute $d_1 =</td>
</tr>
<tr>
<td></td>
<td>b. If $d_1 &gt; D_{thr1}$ or $d_2 &gt; D_{thr2}$</td>
</tr>
<tr>
<td></td>
<td>i. Select $h_i$ as key frame</td>
</tr>
<tr>
<td></td>
<td>ii. $j = i$</td>
</tr>
<tr>
<td></td>
<td>iii. $i = i + 1$</td>
</tr>
<tr>
<td></td>
<td>c. Else</td>
</tr>
<tr>
<td></td>
<td>i. $i = i + 1$</td>
</tr>
<tr>
<td>7.</td>
<td>Return key frames.</td>
</tr>
</tbody>
</table>

Fig 3.4 shows the graph containing the average number of frames per video of a sign and the average number of key frames extracted by using the proposed key frame extraction algorithm given in table 3.2.
3.2.2 Feature Extraction

For the accurate recognition of the signers, features extracted are of great importance so they must be selected very carefully. For hand gesture recognition, the main parameters to be considered are hand shape, hand motion and hand orientation.

Hand Shape

a) Circularity

It is the measure of similarity of a shape’s circumference with respect to its center. The maximum value for circularity is 1 i.e. for circle. It decreases as shape becomes more elliptical. It can be defined as:

$$\text{Circularity} = \frac{4\pi \text{Area}_{image}}{\text{Perimeter}_{image}}$$  \hspace{1cm} (3.1)

b) Extent

Extent refers to the ratio of pixels in the region to pixels in the total bounding box. It is computed as:

$$\text{Extent} = \frac{\text{Area}_{image}}{\text{Area}_{boundingbox}}$$  \hspace{1cm} (3.2)
Hand Motion

**Motion Direction Code (MDC)**

The chain code gives a unique code when we walk around the boundary of the object starting from a particular point. This proposed code can be used to determine the motion trajectory of the sign being performed. The signer performs the sign with changing the direction of the hand. There could be possibly 8 directions in which the hand can move. (Figure 3.5) if we use a single digit to represent the motion direction, then a unique code can be formed called as the motion direction code which can determine the overall direction of movement of the hand, thus representing the motion trajectory. This ensures the user independence while performing the sign, since the sequence of the motion direction is retained even if it is performed by different users. For n number of total key frames, the length of MDC is n-1.

![Figure 3.5. Motion direction for key frames](image)

Depending upon the direction between two successive frames, a number is assigned to it according to Fig.3. Suppose, between two frames, the hand is moving downwards, then 7 will be assigned. Similarly say afterwards, hand moves upward toward the left then 2 will be assigned. This is repeated for all key frames and the
MDC is the concatenation of all these numbers depicting the code for hand motion trajectory. An example is illustrated in Figure 3.6.

Figure 3.6. Sequence of numbers assigned to motion trajectory

| (a) MDC=22864 | (b) MDC=2864 |

MDC after removing consecutive redundancy (in both cases) = 2864

Figure 3.7. Calculation of MDC, for the sign “sorry”.

Since this feature is applied to the extracted key frames, there could be an ambiguity in chain code when number of key frames extracted in one direction is more for the same sign in different videos. For e.g. For the sign “sorry” the key frames extracted are and give MDC value as 22864 (Figure 3.7.a) while for the same sign the MDC value is 2864(Figure3.7.b) for a different video. To avoid this problem consecutive redundancies are removed from the MDC leaving only 2864 as the resultant code in both the cases. The algorithm for MDC is described in table 3.3.
Hand Orientation

Hand orientation is also an important feature which tells the overall orientation of hand while performing the sign. It refers to the direction of change of hand from one frame to the other. If \((x_c, y_c)\) represent the coordinates of the centroid, then orientation can be calculated as:

\[
\theta = \arctan \left( \frac{y_{c+1} - y_c}{x_{c+1} - x_c} \right)
\]  
(3.3)

3.2.3 Recognition

In the proposed work, multiclass support vector machine has been used to suitably classify the isolated sign among multiple classes.

Multi-class Support Vector Machine (MSVM)

The Support Vector Machine (SVM) is a machine learning algorithm which is used for binary classification. SVM uses supervised learning technique and was proposed by Vapnik and his colleagues in 1979 [57]. The main concept of SVM is to find a hyperplane that distinguishes the positive and negative class with the objective to maximize the width of the hyperplane between the classes. The orientation of hyperplane and the distance from the origin of the hyperplane are the parameters which need to be considered. SVM as a classifier is gaining popularity because of its various

<table>
<thead>
<tr>
<th>Table 3.3: Algorithm to Compute MDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Set MDC=0</td>
</tr>
<tr>
<td>2. For i =1 to n-1 do</td>
</tr>
<tr>
<td>MDC=MDC*10+m,</td>
</tr>
<tr>
<td>// where, m is the motion direction between frames (k_i) and (k_{i+1}) (Figure 3.5)</td>
</tr>
<tr>
<td>3. Remove the consecutive redundancy in the MDC to get the overall motion direction.</td>
</tr>
<tr>
<td>4. Return MDC</td>
</tr>
</tbody>
</table>

Dept. of CEA, GLAU, Mathura
features like the ability to accurately separate the classes even for a large dataset with small training samples. SVMs have applications in pattern recognition like face detection, handwritten digit and character recognition, information and image retrieval, gender classification, prediction, text detection and categorization, and many more. The equation of the hyperplane can be represented as the function which classifies the samples as:

\[ f(x) = sgn(wx + b) \]  

(3.4)

Where, \( w \) is the unit vector perpendicular to the hyperplane and \( b \) is the bias term.

SVM was originally developed for binary classification. However, it needs to be extended to solve the classification problems involving multiple classes. The popular methods for multiclass classification using SVM are: “one-against-all,” “one-against-one,” and directed acyclic graph SVM (DAGSVM). For our work, we have utilized one-against-one approach [58]. In this approach, the \( i_{th} \) binary classifier uses the pattern of class \( i \) (+1) as positive examples and the patterns of class \( j \) (-1) as negative examples, and thus forming a feature vector for every class. To find the result, the distance between the vectors of the test sample is calculated with all the trained samples. The class with minimum distance is produced as the output. For \( n \) number of classes, the total binary classes formed are \( n(n-1)/2 \) i.e. \( O(n^2) \). Let the \( i_{th} \) decision function that separates class \( i \) from the other classes with the maximum margin be-

\[ D_i(x) = w_i^T(x) + b_i \]  

(3.5)

Where \( w_i \) is the \( l \)-dimensional vector, \( (x) \) is the mapping function that maps \( x \) into the \( l \)-dimensional feature space, and \( b_i \) is the bias term.

### 3.3 Summary

In this chapter, we have proposed a vision based method for recognition of isolated signs of Indian Sign Language using skin color segmentation, key frames extraction, using new features for hand description and multiclass SVM classifier. The key frame extraction algorithm used speeds up the system by finding out the important frames for feature selection. The novel features used for hand description have proved to provide accurate results for the recognition.
Chapter 4

Experimental Results and Discussion

This chapter covers the experiments performed on the proposed framework, the dataset used, the environmental conditions, setup, constraints imposed. Furthermore, the results of different experiments are calculated and are also compared with previous works in Indian Sign Language Recognition.

4.1 Dataset

No standard dataset is available for Indian Sign Language. Therefore, a new dataset is created comprising of 22 one handed dynamic isolated signs of ISL. In this dataset, various signers perform the different one handed dynamic signs. This is done to ensure that multiple users can use the proposed system and it is not restricted to be used by single user. In addition to it, different background conditions having uniform or non-uniform background are taken into account. The dataset is created by copying the signs from the videos available online at [http://www.indiansignlanguage.org]. 15 signers have emulated those signs with little variations. These signers have different ages ranging from 5 to 60 years. The complete database of signs comprises of the training and testing samples. The number of videos used for training is 4; the remaining videos of each signs are then used for testing. The users have repeated each of these signs multiple number of times with variation in speed, movement, etc. Signer dependent systems are those which require training and testing to be done by the same person. But the proposed system is user-independent; i.e. the signs trained by one signer can be recognized even if different user performs that sign.
4.2 Experimental Setup

The videos of signs of ISL were shot by digital camera, Sony Cybershot 14 megapixel, placed at 85cm from the subject. The user is made to wear full sleeves clothing to extract the skin region easily. The proposed system is trained by one or two users; and tested on multiple signers on various videos of the one handed isolated dynamic signs. Fig 4.2 shows the snapshot of the GUI of the proposed system. The video is inputted and tested. After processing, the result is displayed.

4.3 Results and Discussion

We evaluate the performance of the entire Sign Language Recognition System in terms of accuracy. The accuracy of proposed system is calculated as follows:

\[
Accuracy = \left( \frac{\text{correctly classified gestures}}{\text{total number of gestures}} \right)
\]  

(4.1)
Experiment 1:

In this section, the experimental results performed on the proposed system are discussed. For this experiment, the dataset consists of the dataset consisted of multiple users and thus the system is made user independent i.e. the training and testing is done by different users. For testing, multiple users are considered. To test the adaptability of the proposed system, users of different age group are made to perform signs. This includes users from with age ranging from 5 to 60 years. We have tested the system on 15 users. The recognition rate is lower than that of signer dependent system. The number of training samples considered is 4. Thereafter samples of multiple users are then tested. The multiple users of the system are shown in figure 4.1. The signers used for testing are from different age groups in order to test the effectiveness of the proposed system i.e. whether it is capable of handling the user independence in terms of different types of users with different hand movement, speed, way of performing sign, etc.

The results of various samples of each of the 22 signs are shown in table 4.1. The graph depicting the number of samples per sign and the number of correctly classified samples is shown in Figure 4.4

Figure 4.2 shows that the same sign performed by a child and adult is recognized by the proposed system. This is due to the use of all user independent features like Motion Detection Code, etc.

![Snapshot of the GUI depicting the result of sign “down” performed by multiple signers](image-url)
Figure 4.3 shows the snapshot of the GUI in which multiple users are performing the same sign. The results show that same sign is recognized correctly even when performed by different users.

**Figure 4.3.** Snapshot of the GUI depicting the result of sign “after” performed by multiple signers
Table 4.1 Classification results of the proposed approach

<table>
<thead>
<tr>
<th>S.No</th>
<th>Sign</th>
<th>Total no. of Signs</th>
<th>No. of Signs correctly recognized by MSVM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Abbreviation</td>
<td>20</td>
<td>19</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>After</td>
<td>28</td>
<td>25</td>
<td>89.2</td>
</tr>
<tr>
<td>3</td>
<td>Bad</td>
<td>16</td>
<td>15</td>
<td>93.7</td>
</tr>
<tr>
<td>4</td>
<td>Down</td>
<td>14</td>
<td>13</td>
<td>92.8</td>
</tr>
<tr>
<td>5</td>
<td>Fan</td>
<td>15</td>
<td>14</td>
<td>93.3</td>
</tr>
<tr>
<td>6</td>
<td>Fish</td>
<td>16</td>
<td>15</td>
<td>93.7</td>
</tr>
<tr>
<td>7</td>
<td>Goodbye</td>
<td>15</td>
<td>13</td>
<td>86.6</td>
</tr>
<tr>
<td>8</td>
<td>Happy</td>
<td>16</td>
<td>14</td>
<td>87.5</td>
</tr>
<tr>
<td>9</td>
<td>Leader</td>
<td>14</td>
<td>13</td>
<td>92.8</td>
</tr>
<tr>
<td>10</td>
<td>Morning</td>
<td>18</td>
<td>17</td>
<td>94.4</td>
</tr>
<tr>
<td>11</td>
<td>Night</td>
<td>16</td>
<td>14</td>
<td>87.5</td>
</tr>
<tr>
<td>12</td>
<td>Office</td>
<td>20</td>
<td>18</td>
<td>90</td>
</tr>
<tr>
<td>13</td>
<td>Only</td>
<td>14</td>
<td>12</td>
<td>85.7</td>
</tr>
<tr>
<td>14</td>
<td>Paint</td>
<td>16</td>
<td>14</td>
<td>87.5</td>
</tr>
<tr>
<td>15</td>
<td>Quick</td>
<td>14</td>
<td>13</td>
<td>92.8</td>
</tr>
<tr>
<td>16</td>
<td>Right</td>
<td>18</td>
<td>17</td>
<td>94.4</td>
</tr>
<tr>
<td>17</td>
<td>Sad</td>
<td>15</td>
<td>13</td>
<td>86.6</td>
</tr>
<tr>
<td>18</td>
<td>Sometimes</td>
<td>14</td>
<td>13</td>
<td>92.8</td>
</tr>
<tr>
<td>19</td>
<td>Sun</td>
<td>14</td>
<td>13</td>
<td>92.8</td>
</tr>
<tr>
<td>20</td>
<td>Up</td>
<td>20</td>
<td>17</td>
<td>85</td>
</tr>
<tr>
<td>21</td>
<td>Word</td>
<td>14</td>
<td>13</td>
<td>92.8</td>
</tr>
<tr>
<td>22</td>
<td>Wrong</td>
<td>20</td>
<td>17</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>367</td>
<td>332</td>
<td>90.46</td>
</tr>
</tbody>
</table>
Figure 4.4. Graph showing the total number of samples and correctly classified samples

Experiment 2:

The proposed method in [55] was tested on our dataset consisting of varied repetitions of 22 dynamic isolated signs of Indian Sign Language. It was found that our proposed system achieves higher recognition rate than [55].

<table>
<thead>
<tr>
<th>Method</th>
<th>Agrawal et al [55]</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>82.28%</td>
<td>90.4%</td>
</tr>
</tbody>
</table>

Table 4.2 Comparison of accuracy in [55] and proposed work

Table 4.3 shows the comparison of proposed work with other approaches with various parameters such as nature of the dataset, vocabulary size, segmentation, classification methods, and feature extraction techniques etc.
Table 4.3: Comparative study of various approaches for ISL Recognition

<table>
<thead>
<tr>
<th>Comparison Parameter</th>
<th>Works in Indian Sign Language or Other Sign Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature of Dataset</td>
<td>ISL alphabets</td>
</tr>
<tr>
<td>Size of Vocabulary</td>
<td>23</td>
</tr>
<tr>
<td>Approach</td>
<td>Red colored gloves</td>
</tr>
<tr>
<td>Type of Gestures</td>
<td>Static</td>
</tr>
<tr>
<td>Segmentation and Tracking</td>
<td>RGB model</td>
</tr>
<tr>
<td>Features Extracted</td>
<td>Shape Descriptors</td>
</tr>
<tr>
<td>Classification Methods</td>
<td>Threshold based</td>
</tr>
</tbody>
</table>

Experiment 3:
In the proposed framework, a novel feature called the Motion Detection Code (MDC), has been proposed to find out the hand motion. In this experiment, an existing feature for hand motion used in [59] has been incorporated in place of MDC to find the effectiveness of the proposed MDC feature. The results show the fall in accuracy when existing feature is used instead of the proposed MDC.

Hand motion is one of the parameter considered for hand gesture recognition in [59]. In this, the motion of the hand is described as a temporal sequence of points with respect to the centroid of the hand. It can be represented as:

\[ p_t = (x_t, y_t) \quad t = (1, 2, 3, \ldots, T) \]  

(4.2)
Where, $T$ is the total number of hand frames and $p_i$ represents the centroids of the gesture path (an ordered set of points) which depicts the hand motion.

<table>
<thead>
<tr>
<th>Method</th>
<th>Wang et al [59]</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>84.46%</td>
<td>90.4%</td>
</tr>
</tbody>
</table>

**Table 4.4** Comparison of accuracy in [56] and proposed work

### 4.4 Summary

In this chapter, the various experiments conducted on the proposed system have been discussed. The results show that the system is able to recognize varied one handed dynamic signs of Indian sign language efficiently. We have also compared the performance of our system with the other approaches. The system is tested on multiple users having different age groups. The proposed system is capable of handling user independence with reliable recognition rate.
Chapter 4
Conclusions and Future Directions

4.1 Summary and Contributions

In this project, we have addressed the problem of automatically recognizing one-handed isolated dynamic signs in video for Indian Sign Language.

Chapter 3 proposed a vision based Indian sign language recognition method using skin color segmentation, key frames extraction, multi-features extraction and multiclass SVM classifier. Finding out the most important and distinguishing frames from the video leads to the removal of similar or non-essential frames from video and hence speeds up the system. Multiple features such as hand shape, hand orientation and hand motion improve the performance of the system. A new feature, Motion Detection Code, for finding out the overall motion direction of the hand is also proposed which has proved to be effective in different signs.

There is no standard dataset available for ISL signs; therefore, we have created our own dataset. The signer copied the signs from examples available online at [http://www.indiansignlanguage.org] that were played on a monitor. Gesture videos were recorded by a digital camera, Sony Cybershot 14 megapixel, placed at 85cm from the subject. A vocabulary of 22 different signs is used. The gesture database is divided into training and testing sets. In classification, we have used 4 samples of video of each sign for training and the remaining used for testing. System is trained and tested on multiple signers. The database is composed of varying number of repetitions for each of 22 sign classes which are performed by multiple users. These signers vary in their age ranging from 5-60 years. Total number of videos used is 367. This framework is user-independent; i.e. the signs trained by one signer can be recognized if different user performs the sign.
4.2 Future Work

We proposed the user independent method for recognition of dynamic isolated sign recognition using single handed sign and not considered non-manual parameters. Our future work will focus on following issues:

(a) Adding non-manual parameters for the complete sign language recognition;
(b) Sign dataset will also include two handed dynamic signs;
(c) Fusion of classifiers for better recognition rate.
List of Publications

References


