STUDY OF VISION BASED HAND GESTURE RECOGNITION USING INDIAN SIGN LANGUAGE

Archana S. Ghotkar¹ and Dr. Gajanan K. Kharate²

¹Pune Institute of Computer Technology, Department of Computer Engineering, University of Pune, Pune, India.
²Matoshri College of Engineering and Research Centre, Department of Electronics and Telecommunication Engineering, University of Pune, Nashik, India.

archana.ghotkar@gmail.com, gkkharate@yahoo.co.in


Abstract- Human Computer Interaction moves forward in the field of sign language interpretation. Indian Sign Language (ISL) Interpretation system is a good way to help the Indian hearing impaired people to interact with normal people with the help of computer. As compared to other sign languages, ISL interpretation has got less attention by the researcher. In this paper, some historical background, need, scope and concern of ISL are given. Vision based hand gesture recognition system have been discussed as hand plays vital communication mode. Considering earlier reported work, various techniques available for hand tracking, segmentation, feature extraction and classification are listed. Vision based system have challenges over traditional hardware based approach; by efficient use of computer vision and pattern recognition, it is possible to work on such system which will be natural and accepted, in general.

Index terms: Indian sign language, vision based hand gesture recognition, hand tracking, segmentation, feature extraction, classification, computer vision, pattern recognition
I. INTRODUCTION

Standard sign languages (SL) are known as Deaf and Dumb languages. SLs are gestural languages which contain symbolic encoded message for communication without speech channel. They are unique in some ways in that they cannot be written like spoken language. Sign language varies from country to country with its own vocabulary and grammar. Even within one country, sign language can vary from region to region like spoken languages. Indian Sign Language (ISL) is a language used by Indian deaf and dumb community [1].

Gestures are powerful means of communication among humans. Among different modality of body, hand gesture is the most simple and natural way of communication mode. Real time, vision based hand gesture recognition is more feasible due to the latest advances in the field of computer vision, image processing and pattern recognition but it has yet, to be fully explored for Human Computer Interaction (HCI) [55-56].

With the wide applications of HCI, now days, it becomes active focus of research [55]. To have an interaction with computer, vision based system is more suitable than traditional data glove based system, as sensors are attached to the data glove and data suit where, user has to wear these cumbersome devices [2]. In this paper, Vision based approach have been discussed for interpreting the Indian sign language using hand modality. A Typical Hand Gesture Recognition system consists of mainly four modules: Gesture acquisition, Tracking and segmentation, Feature extraction and description, Classification and recognition. This paper focuses on a study of sign language interpretation system with reference to vision based hand gesture recognition. An attempt has also been made to explore about the need and motivation for interpreting ISL, which will provide opportunities for hearing impaired peoples in Industry Jobs, IT sector Jobs, and Government Jobs.

The organization of the paper is: Section II gives information about India sign language Section III focuses on a typical sign language interpretation system. Section IV shows human hand skeleton model. Section V describes typical vision based hand gesture recognition system with various methods/techniques available in literature. Section VI focuses on challenges in sign language interpretation and gesture recognition and section VII contains discussion and conclusion.
II SIGN LANGUAGE IN INDIA

In literature, it was found that count of hearing impaired people in India, is more compared to other countries. Not all of them use ISL but, more than one million deaf adults and around half million deaf children use ISL as a mode of communication. Deaf people, who live in villages usually, do not have access to sign language. However, in all large towns and cities across the Indian subcontinent, deaf people use sign language which is not standard sign language. Extensive work and awareness program are being done for implementation of ISL in education systems [3].

In 1970, linguistic work on ISL began and with contribution of a team of researcher from America and Vasishta et al. It was found that ISL is a language in its own right and is indigenous to the Indian subcontinent and resulted in four dictionaries between 1977 and 1982. It was found that 75% signs are same across the region. In 1998, another researcher from Germany (Dr. Ulrike Zeshan) compared signs from many different regions across the Indian subcontinent, including regions such as Orissa, Kerala, Jammu and Kashmir, Bhopal, Chennai, Bangalore and Darjeeling. She also found that on an average about 75% of the signs are similar across different regions [1]. Further work was carried out by Zeshan and Vasishta [3] on developing ISL grammar, ISL teaching courses, ISL teacher training program and teaching material that was approved by the Rehabilitation Council of India in 2002 [1,3]. There are many ISL cells working in India for use and awareness of ISL as well as teaching courses of ISL. Ali Yavar Jung National Institute for Hearing Handicapped, Mumbai released “Basic course in INDIAN SIGN LANGUAGE” [1].

After survey, it was found that there are around 405 deaf and dumb schools in India. Most of the schools use their own native sign language as a teaching and learning aid, therefore, for awareness to use of standard ISL as a teaching aid is being done by different ISL cells and NGOs to help Indian deaf and dumb community to bridge the communication gap between them.

There are some common wrong beliefs about sign language which is reported in ISL literature [1]:

i) “Sign language is same all over the world”

ii) “Sign language is not a complete language. It is just a sort of pantomime or gesturing, and it has no grammar”
iii) “Sign language is dependent on spoken language. It is a representation of the spoken language of the hands”
iv) “Sign language is the language of the hands only”
v) “Sign language has been invented by other people to help deaf people”
vi) “Signed Hindi or signed English is better than Indian sign language”

So overcoming these wrong beliefs, there is a need of developing ISL interpretation system to aid Indian hearing impaired people with the help of HCI and making them literate and self-dependent.

Major research work is going on awareness and multilingual Indian sign language dictionary tool [4], so there is a need for Indian sign language interpretation tool. Following may be the major advantages of ISL interpretation.

i) Use and awareness of computer interface through ISL interpretation.
ii) Education and training will be easier through ISL interpretation/visualization for Indian deaf and dumb people.
iii) Serving the mankind by use of technology.
iv) Social aspect like humanity can increase in individual mind by involving physically impaired people in our day to day life.
v) Blind people can also use the same system by extending it for voice interface.

III SIGN LANGUAGE INTERPRETATION SYSTEM

Sign language is not a universal language. Sign language recognition is a multidisciplinary research area involving pattern recognition, computer vision, natural language processing and psychology. Figure 1 shows the typical architecture for sign language interpretation system. It broadly divides into two modules. First module is for converting normal English sentences in to SL (to be understood by deaf people) and another module is for converting SL into English text (to be understood by normal people). For literate hearing impaired people, those who can read English, first module is not required. But for illiterate deaf and dumb people, both modules are essential. In both the module, language processing engine is required which is based on a
particular language rules. Conversion of sign to text includes the area of computer vision, image processing, pattern recognition and language processing with linguistic study.

![Diagram of a typical sign language interpretation system]

**Figure 1. A Typical Sign Language Interpretation System**

The earlier reported work on sign language recognition is shown in Table 1.

<table>
<thead>
<tr>
<th>Sign Language</th>
<th>Modality used</th>
<th>Subset of sign language considered and recognition rate</th>
<th>Reference</th>
</tr>
</thead>
</table>
| Native Indian sign language   | Hand         | Static signs (95.2%)  
Dynamic signs (95.5%)  
Sentences (92.5%)                                                                 | [5]       |
| Indian sign language          | Hand         | Tamil letters(12 vowels and 18 consonants)                                                    | [6]       |
| (south Indian sign language)  | Hand         | Vowels(6), 10 numbers                                                                         | [7]       |
| Bangladeshi sign language     | Hand         | 25 common words with sensor attached on both the hand                                           | [8]       |
| Malay sign language           | Hand         | 23 Arabic words (87%)                                                                         | [9]       |
| Arabic sign language          | Hand         | 15 gestures                                                                                   | [10]      |
| American sign language        | Hand         | Sign database 50                                                                               | [12]      |
| Brazilian sign language       | Hand         | Latin letters                                                                                 | [13]      |
| Chinese sign language         | Hand         | Isolated and continuous sign                                                                  | [14]      |


<table>
<thead>
<tr>
<th>Language</th>
<th>Feature</th>
<th>Vocabulary Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>American sign language</td>
<td>Hand and Face</td>
<td>22 sign vocabulary</td>
<td>[15]</td>
</tr>
<tr>
<td>American sign language</td>
<td>Hand</td>
<td>26 manual alphabets</td>
<td>[16]</td>
</tr>
<tr>
<td>Indian sign language</td>
<td>Hand</td>
<td>ISL alphabets and numbers</td>
<td>[17]</td>
</tr>
</tbody>
</table>

After study and investigation, it was found that there is a relation between human gesture and speech. Speech expression can be replaced by signs going from gesticulation to sign language. ISL is a visual-spatial language. It is having linguistic information in the form of hands, face, arms and head/body posture and movements. Visual channel is active in sign language like speech channel in spoken language. Figure 2 describes the ISL type hierarchy, which could help for design of such system [4]. It is broadly categorized into 1) Manual (hand shape, hand location, orientation and movements) 2) Non-manual (facial expression, eye gaze and head/body posture). In ISL, there are one handed and two handed signs which can be static and dynamic (movement). In two handed sign, some sign contains both hand active (type 0) and some sign contains dominant hand more active than non-dominant hand (type 1). Before begin to design of any sign language interpretation system, it is advisable to go through respective sign language hierarchy. Figure 3 shows ISL manual alphabet set [60].

![Image](https://example.com/image.png)

Figure 2. A hierarchical classification of ISL [4]
Though it is found that hand plays active role in sign language but due to its complex articulated structure consisting of many connected links and joints, hand gesture recognition becomes a very challenging problem. Figure 4 shows skeleton structure and the joints of the human hand, with total 27 degree of freedom (DOF) considering hand wrist. There are widely two terms used in hand gesture recognition system: 1) Hand posture (static hand gesture) and 2) Hand gesture (Dynamic hand gesture). In hand posture, no movements are involved whereas; hand gesture is a sequence of hand posture connected by movement over a period of time [2]. In dynamic hand gesture, again two aspects are considered such as local finger motion without changing hand position or orientation and global hand motion where, position or orientation of hand gets changed. Study of hand skeleton model is very essential for developing any hand gesture recognition system.
Figure 4. Hand Skeleton with DOFs [18]

V VISION BASED HAND GESTURE RECOGNITION

In any typical hand gesture recognition system, a good and strong set of features, description and representation are required. In the current state of the art, due to the limitation of data glove/sensor based approaches, vision based (appearance based) and 3-D hand model based approaches are being used [2]. One of the major tasks in hand gesture recognition is the description of the gesture. Various methodologies are found in the literature such as, statistical and synthetic based approaches. In statistical representation, one can represent it, in the form of feature vector and then apply classification and recognition algorithm; whereas synthetic gesture recognition gesture can be represented in the form of tree, string or graph and decision rule such as graph matching, decision tree and string matching.

Now days in the field of Human Computer Interaction, Hand Gesture recognition [HGR] is an active research topic. In this section, various approaches and techniques have been explored related to hand gesture recognition.
Recognizing gesture is a complex task which involves many aspects such as object detection, object description, motion modeling, motion analysis, pattern recognition and machine learning even psycholinguistic studies also required [4].

![Figure 5. The Typical Hand Gesture Recognition System](image)

Figure 5 shows typical architecture of HGR system. Hand tracking and segmentation are to be done on captured video and feature extraction is to be done on segmented hand image which is further given to classification and recognition phase. Output is to be printed or executed, depending on the application. In this paper, survey has been carried out on various methods adopted by various researchers on defined steps.

a. Hand Tracking and Segmentation
After capturing and separating frames from videos, the elementary and important task is detection and segmentation of hands. There are various approaches and techniques available in literature [19] but the results vary, images to images due to the limitation of vision based approach such as variable lightning condition, variation of skin color, detection of hand in complex background. Pixel and region based segmentation techniques are available [20, 21]. In India, there is deviation in human skin color tones. So, finding out adaptive color model is a big challenge for skin color detection [22]. It has been observed that HSV and YCbCr color model gave better result for skin color detection [23] than other models due to the separation property of luminance and chrominance component. Some researcher used additional marker or color gloves for hand segmentation using color thresholding, but for natural interface bare hand interaction is always preferred. Supervised as well as Unsupervised Learning Model such as Bayesian classifier [24] can be used for skin color segmentation [25-26]. Unsupervised learning such as, K-mean clustering is also a good option for skin color segmentation [27]. 2D Tracking algorithm [28] gives the position information of hand such as color tracking, motion tracking, template matching, blob tracking, Multiple cues integrating methods are available. It has been noticed that
tracking algorithm such as meanshift [29], camshaft [30], viola jones [31] with appropriate color space gave better segmentation result in complex background.

b. Feature Extraction
After studying hand skeleton model it has been noticed that shape is the important visual feature of the hand. Zhang and Lu [37] gave classification of shape representation and description techniques based on contour and region. In contour based method, shape features are extracted from the shape boundary whereas, in region based method features are extracted from the whole shape.

i) Contour-based shape representation and description methods are chain Code, Polygon, B-spline, Perimeter, Compactness, Eccentricity, Shape Signature, Hausdoff Distance, Fourier Descriptor, Wavelet Descriptor, Scale Space, Autoregressive, Elastic matching.

ii) Region-based shape representation and description methods are Convex Hull, Media Axis, Area, Euler Number, Eccentricity, Geometric Moments, Zernike Moments, Pseudo-Zernike Moments, Legendre Moments [37].

In hand recognition problem, shape contour is important than whole region so, contour based methods are mostly used. But for complex sign, sometimes region based methods are more suitable because it contains all the available information [37]. In case of the new signer for performing gesture, there may be chances for angle deviation, shifting of signer space (translation) can occur. Hand size (scaling) of the signer can also vary. So, while choosing feature extraction method, care must be taken that it should be invariant to translation, rotation and scale. SLs contain large set of vocabulary, use of one of the feature extraction techniques is not sufficient. Practically combination of feature vector and motion vector is the better choice to get accuracy. Table 2 shows the earlier reported work on hand gesture recognition on various segmentation and feature extraction techniques.

c. Classification and Recognition
An efficient classifier and recognition method plays very important role in any gesture recognition system. This step goes forward with the pattern recognition and machine learning field. Any pattern recognition problem is classified into two methods i) supervised and ii)
unsupervised classification. Though an intensive research is being carried out for the last 60-65 years in the field of pattern recognition, but the complex pattern with variant to translation rotation and scale is still unresolved [44]. Various supervised classification methods are available such as nearest neighborhood classification with Euclidean distance [45] and other similarity measures [46], Bay’s classifier [47], Neural network [48], Hybrid Recognizer [49], linear regression model and unsupervised classification methods such as clustering methods: K-mean, Fuzzy k-mean, Minimum spanning tree, Single link, Mutual neighborhood, Single-link, Complete link, Mixer decomposition. In sign language interpretation, as the previous classes are known, supervised classification is the good choice. For construction of sentence followed by sequence of signs, Hidden morkov model [50-51] is useful. Vision-based hand gesture recognition system also needs to meet the requirements including real-time performances, accuracy and robustness, so use of correct classifier is the need of the any machine learning system. Training and testing the system is the very important aspect of any research work. There are many error estimation methods available such as redistribution methods, Holdout method, Leave-one Out method, Rotation method, n-fold cross validation and bootstrap method. Depending on the availability of sample data and required performance one can choose the error estimation method for analysis of results. Some researcher worked on hybrid classifier or cascaded classifier to get best performance [49].

Table 2: Survey on Different Segmentation and Feature extraction Techniques

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Various available techniques/methods available in literature</th>
<th>Background/illumination</th>
<th>Accuracy/ Remark</th>
</tr>
</thead>
</table>
| Segmentation and Tracking techniques | YCbCr color space, K-means embedded particle filter for two hand tracking [33]                  | Simple as well as cluttered background                               | Accuracy: 83%
  a) worked better than mean shift algorithm,
  b) tracking fail for rapid movement of hand.                                                |
<p>|                                  | Tower method for hand tracking [34]                                                                | Simple background                                                   | Faster than camshift                                                                                  |
|                                  | Two hand segmentation with Haar-Like feature and adaptive skin color model [35]                  | Complex background                                                  | Accuracy: 89% to 98% for four movement                                                               |</p>
<table>
<thead>
<tr>
<th>Feature extraction techniques</th>
<th>Kalman filtering and a collapsing method [36]</th>
<th>Complex background</th>
<th>Satisfactory results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viola Jones method for tracking [31]</td>
<td>White background and different lightning condition</td>
<td>Fast and most accurate learning-based method for object detection</td>
<td></td>
</tr>
<tr>
<td>Color based segmentation using HSV, L<em>a</em>b color spaces and camshift method for tracking [30]</td>
<td>Simple and complex background</td>
<td>Camshift tracking with HSV color model gives better result in complex background, different lighting condition and skin color</td>
<td></td>
</tr>
<tr>
<td>Angle and distance from endpoint [38]</td>
<td>Complex background</td>
<td>Accuracy: 92.13% No. of gesture used : 10</td>
<td></td>
</tr>
<tr>
<td>Haar wavelet, Code word scheme [39]</td>
<td>uniform</td>
<td>Accuracy: 94.89% No. of gesture used : 15</td>
<td></td>
</tr>
<tr>
<td>Location, angle, velocity and motion pattern P2-DHMMS [40]</td>
<td>Complex background</td>
<td>Accuracy: up to 98% No. of gesture used : 36</td>
<td></td>
</tr>
<tr>
<td>Orientation Histogram, Neural network [41]</td>
<td>Complex background</td>
<td>Accuracy: up to 90 No. of gesture used : 33</td>
<td></td>
</tr>
<tr>
<td>Co-occurrence Matrix, local and global features [42]</td>
<td>Complex background</td>
<td>Accuracy: 93.094% No. of gesture used : 30</td>
<td></td>
</tr>
<tr>
<td>Key trajectory point selection, trajectory length selection, location feature extraction, orientation feature extraction, velocity and acceleration [43]</td>
<td>Uniform</td>
<td>Accuracy: Static-92.81% Dynamic:87.64% No. of gesture used : 26</td>
<td></td>
</tr>
</tbody>
</table>

**VI CHALLENGES IN SIGN LANGUAGE INTERPRETATION**

Literature shows that due to the challenges of vision based system, most of the researchers till date have limited their work, to small subset of a full sign language [5]. To work on full sign language interpretation, close collaboration with SL interpreter and deaf people is required [52].
Annelies Braffort [52], mentioned worry about work going on sign language interpretation by hearing researcher. In this context, Annelies Braffort [52] suggested to hearing researcher to ask themselves these questions:

i) “Is there really an interest in my research for the deaf community?”

ii) “Is what I call sign Language in my papers really Sign Language?”

Sign language function and spoken language functions are totally different. SL is fundamentally based on spatial properties and iconicity properties. Hand parameters such as shape, movement, orientation and location as well as facial expression, mouth movements are considered to understand the sign. These parameters occur simultaneously and are articulated in space. Building of syntactic and semantic based rule system is required because one sentence in a spoken language can be represented by a single sign in SL. Britta and Karl-Friedrich [53] reported some difficulties in terms of sign language:

i) Occlusions problem while performing sign

ii) Signer position may vary in front of camera while performing sign

iii) Working on 2D camera give loss of depth information

iv) As each sign varies in time and space, so there may be a change in position and speed with same person or person to person

v) Co-articulation problem (link between preceding and subsequent sign)

Spoken language and sign languages are totally different in their linguistic structure. Design of rule based system is also a challenging task. In ISL, proper noun is not pronounced as in spoken, but it has been used like pointing to the identity of the particular person with their gender (male/female). English language is used the structure of sentence: SUBJECT, VERB, OBJECT, but In ISL, always this structure does not use. In some cases it is used as SUBJECT, OBJECT, VERB. These are the some important linguistic properties which should be studied before design of rule based system (language processing engine).

VII DISCUSSION AND CONCLUSIONS

The major objective of this paper was to give significance of ISL as an interpretation language and focus on various methods/techniques available for vision based hand gesture recognition. Researchers are facing major problem of availability of standard database. Major work is going on for ISL multilingual multimedia dictionary tool. Most of the researchers are working on their
own created dataset. In this context as per the Annelies Braffort [52], if we follow some linguistic ethics and go forward with the help of SL trainer and deaf people, we can really give justice for implementation of SL interpretation system. The deaf assistive system can bridge the communication gap between hearing impaired and normal people without isolating them in the society. Table 3 shows some required additional information, which is not covered in the previous text.

Table 3: Summary of additional parameters required for vision based hand gesture recognition.

<table>
<thead>
<tr>
<th>Attribute/Parameter</th>
<th>Solution</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capturing Devices</td>
<td>2D web camera, 3D camera</td>
<td>For cost effective solution, normal 2D web camera is being used for HCI application but now day’s 3-D camera such as Microsoft kinect sensor is also available in the market which provides the depth as well as skeleton information. Using this camera major of the image preprocessing work gets reduced.</td>
</tr>
<tr>
<td>Application of Gesture Recognition</td>
<td><strong>Multidirectional control</strong>- 3D design, advertising, gaming, computer interface[56 ], robotic interface and behavior [57], mobile application, virtual reality, home care and security[ 58] <strong>Symbolic language</strong>: Sign languages</td>
<td>Research work is going on variety of application, which will replace traditional hardware devices by hand but anticipated limited signs are used in multidirectional control but robustness and speed are the essential performance parameters. Whereas, in sign languages vocabulary set is predefined and it is huge so it is challenging task for researchers.</td>
</tr>
<tr>
<td>ISL training courses, dictionary</td>
<td>Training courses at level (A,B,C), course material, ISL Dictionary</td>
<td>[59-63]</td>
</tr>
<tr>
<td>Implementation softwares</td>
<td>VC++, JAVA, PYTHON using OpenCV, MATLAB, C#, .NET</td>
<td>Use of the OpenCV is advisable for real time application because of free cost and execution speed.</td>
</tr>
</tbody>
</table>
In this paper, authors tried to focus on challenges for vision based and sign language interpretation system with an objective, to give an overall glimpse of SL interpretation need, existing image processing and pattern recognition techniques available in the literature. It will be a great contribution to the Indian hearing impaired through working on Indian sign language, so that they are enabled to become self-respecting citizens and despite their deafness and muteness can play a useful role in the society.

Every God creature has an importance in the society, remembering this fact, let us try to include hearing impaired people in our day to day life and live together.

REFERENCES
[63] http://rehabcouncil.nic.in