Review on Feature Extraction methods of Image based Sign Language Recognition system

Hemina Bhavsar
Assistant Professor
S.S.Agrawal Institute of Computer Science,
Gujarat Technological University
Chandkheda, Ahmedabad, Gujarat
hemina.bhavsar@gmail.com

Dr. Jeegar Trivedi
Assistant Professor
Department of Computer Science & Technology,
Sardar Patel University
Anand, Gujarat
jeegar.trivedi@gmail.com

ABSTRACT

Sign language is the way of communication among the Deaf-Dumb people by performing hand gestures. This paper is present review on Sign language Recognition approaches that aims to provide communication way for Deaf and Dumb Community over Society. Basically There are main two approaches for sign language recognition is Sensor based and Image based. This paper describes review of Image based sign language recognition system. Signs are in the form of hand gestures and these gestures are identified from images as well as videos. Hand gestures are identified and classified according to features of Gesture image. Features are like shape, rotation, angle, pixels, hand movement etc. Features are finding by various Features Extraction methods and classified by Artificial Intelligence methods. The most significance of this paper is to review the key finding of the comparison of feature extraction methods of similar systems used in Image based hand gesture recognition on the base of accuracy rate.

KEYWORDS

Sign Language Recognition, Feature Extraction, Support Vector Machine, Neural Network.

1. INTRODUCTION

Sign languages (SL) are known as Deaf and Dumb languages. SLs are gestural languages which uses sign message for communication by hand without speaking. 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]. There arises the need for sign language interpreters who can interpret sign language to spoken language and vice versa. But, the availability of such interpreters is limited, expensive and does not work throughout the life period of a deaf person. This resulted in the development of automatic sign language recognition systems which could automatically translate the signs into corresponding text or voice without the help of sign language interpreters [3]. Effective Sign Language Recognition system gives the chance to deaf people to express their idea without human translator. Sign Language translator 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]. This paper focuses on a study of sign language interpretation system with reference to vision based hand gesture recognition.
2. SIGN ACQUIRING METHODS

2.1 LEAP MOTION:
Leap Motion controller (figure 1) is a sensor which detects the hand movement and converts that signal into computer commands. It consists of two IR cameras and three infrared LED’s. LED generates IR light signal and camera generates 300 frames per second of reflected data. These signals are sending to the computer through USB cable for further processing[3].

2.2 KINECT SENSOR:
Kinect is Microsoft motion sensor with Xbox 360 gaming console shown in figure 2.it consist of RGB camera, depth sensor and multi-array microphone. It recognizes facial movement and speech [4].

2.3 DATA GLOVE:
This method uses different sensor to detect hand gesture signal. Hand gesture signal is in the form of analog. ADC is used to convert analog signal into digital form. It consists of flex sensor and accelerometer. Flex sensor is used to detect bend signal[5].

2.4 VISION BASED:
In this method web camera used to capture images. After that, image segmentation has done. Feature like palm, finger extracted from input image. Different hand motion that is half closed, fully closed, semi closed was detected. Data is saved in vector and that vector is used for recognition of alphabets [6].

3. METHODOLOGY OF SIGN LANGUAGE RECOGNITION

3.1 IMAGE PROCESSING :
Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispension in which input is image, like video frame or photograph and output may be image or characteristics associated with that image.

3.2 PURPOSE OF IMAGE PROCESSING :
The purpose of image processing is divided into 5 groups. They are:
1. Visualization - Observe the objects that are not visible.
2. Image sharpening and restoration - To create a better image.
3. Image retrieval - Seek for the image of interest.
5. Image Recognition – Distinguish the objects in an image.

In Sign Language Recognition Image processing is used to better extract features from input images. Images are in static image or dynamic image of sign perform by human. In particular, the features that we extract from sign or hand gesture images should be invariant to background data, translation, scale, shape, rotation, angle, coordinates, movements etc.

a. First, images are converted to grayscale from colored.
b. Since the background data is not tested, it should be remove the background from the foreground. By subtracting the background image from an input image.
c. Next step is apply threshold technique to ensure that hand pixels would not be subtracted out, and it will be converted to binary image.
d. Then crop the target part and evaluate various morphological features from image using various feature extraction algorithms.
3.3 SYSTEM ARCHITECTURE:

Collect images of hand gesture including face

Convert images into grayscale image

Apply threshold technique

Crop target part from the image

Apply morphological process to get features

Apply Classification Technique

Identified Word

Generate Database of words with Morphological Features

Fig-1 : Find morphological features from Image and generate Training Database
4. LITERATURE REVIEW

M. V. D. Prasad, P. V. V. Kishore, E. Kiran Kumar, D. Anil Kumar [7] presented methods for Indian Sign Language Recognition. Wavelet based fusion of two weak edge detection models. One is morphological subtraction model and the other is gradient based canny edge operator. Elliptical Fourier descriptors provide shape models used with optimized number of shape descriptors. Principle components determined find the feature vector to a minimum to accommodate all the frames in the video sequence. Classification of the signs is done by Back Propagation Neural Network Algorithm. The recognition rate stands at 92.34%.

Suriya M., Sathyapriya N., Srinithi M., Yesodha V. [8] this four persons presented system that recognizing sign language of 26 hand gestures in Indian sign language using MATLAB. By using image processing the segmentation can be done. Some of the features are extracted such as Eigen values and Eigen vectors which are used in recognition. The Linear Discriminant Analysis (LDA) algorithm was used for gesture recognition and recognized gesture is converted into text.

Amitkumar Shinde, Ramesh Kagalkar [9] provide a system includes efficient and robust hand segmentation and tracking algorithm to recognize 43 isolated words from the Standard Marathi sign language. In their system they recognize some very basic elements of sign language and to translate them to text and vice versa in Marathi language. The different images were tested and result found that new technique of classification gives 90% accuracy.

Parul Hardeep et al. [10] provide method for recognize sign language. It has three steps: 1. Pre-processing: First input sign image in RGB to convert into Lab colour space where L is lightness a and b are two colour channel. 2. Feature Extraction: It was done using Area, height, Euclidean distance, Average height. 3. Classification: Feed forward back propagation algorithm was used for training and classification. It was provide 85% accuracy.

Druva N. et al. [11] explored the various possible ways of segmentation using different color spaces and models and presents with highest accuracy. In this paper authors compare RGB, Y'CbCr and HSI color spaces. Images obtained from the camera obtained in RGB color spaces. RGB color space is primary color space since it primary colors red, green and blue as its color components. In digital photography is Y’CbCr color space which contain Y’ or luma component which describe brightness, Cb is describe blue difference chroma and Cr is describe red difference chroma component. H or Hue describes primary color, S or Saturation describes as total amount of color and I or Intensity is described total amount of light intensity. Clearly seen
from authors result that HSI model would greatly benefit in order to segment the hand and fingers. Future work include robust video processing algorithm to identify various gesture and keep memory of database minimal.

Andres Jesse Porfiri, Kelly Lais Wiggers, Luiz E. S. Oliveira, Daniel Weingaertner [12], presents a method for recognizing hand configurations of the Brazilian sign language (LIBRAS) using 3D meshes and 2D projections of the hand. Videos were manually segmented to extract one frame with a frontal and one with a lateral view of the hand. For each frame pair, a 3D mesh of the hand was constructed using the Shape from Silhouette method, and the rotation, translation and scale invariant Spherical Harmonics method was used to extract features for classification. A Support Vector Machine (SVM) achieved a correct classification.

Hanning et al. [13] presented hand gesture recognition system based on local orientation histogram feature distribution model. Skin color based segmentation algorithms were used to find a mask for the hand region, where the input RGB image converted into HSI color space. To compact features representation, k-means clustering has been applied. This system was based on static hand gesture and time consuming.

Nasser H. Dardas et al. [14] presented real time system which including detecting and tracking hand in cluttered background using skin detection and contour comparison algorithm after face detection and subtraction and recognition using principle Experiments show that system could achieve satisfactory real time performance as well as classification accuracy above 90% under variable space, orientation, and cluttered background.

Keskin [15] performed the recognition of ASL hand configurations of the 10 digits with videos acquired using Kinect. The method is based on obtaining a 3D skeleton of the hand which, combined with 21 segmented hand parts, form the feature vector. The classifier used in the experiment the SVM had results with an accuracy rate of 99.9%.

El-Bendary et al. [16] developed an Arabic alphabet signs translator with an accuracy of up to 91.3%. Videos are taken of deaf people which convert into text. The features used are rotation, scale and translation invariant. Videos are converted into Frames. For each frame, the distances between three different black pixels are used to construct the feature vector. In the recognition stage, a multilayer Perceptron (MLP) neural network and a minimum distance classifier (MDC) are used.

Quan [17] described hand signals based on spatial and temporal information extracted from video sequences. The database consisted of 30 letters of the Chinese alphabet, with 195 images representing each letter, totaling 5850 images. Color histogram, Hu moments, Gabor filters and Fourier descriptors were used as input features. Support Vector Machine (SVM) used as classifier, and hit rates were 95.55%.

Advanced Marathi Sign Language Recognition using Computer Vision For the recognition of the sign language a touch screen based approach is developed in [18]. The author tries to recognize the character generated from the screen sensor and transform to speech signal based on a recognition algorithm. In an approach [19] the author suggests in recognizing the hand gesture based on the finger boundary tracing and fingertip detection.

The author suggested to Identify the American Sign Language based on the hand gesture passed. In [20], Mohandes introduced an automatic recognition of the Arabic sign language letters. Support vector machines were used for classification and moment invariants are used in feature selection. A recognition rate of 87% was achieved.

AlJarrah and Halawani [21] developed a neuro-fuzzy system that deals with images of bare hand signs and achieved recognition rate of 93.55%.

In [22], Al-Rousan and Hussain built an adaptive neurofuzzy interference system for letter recognition. A coloured glove was used to ease the process of segmenting the hands region. The recognition accuracy achieved was 95.5%.

Jason Isaacs and Simon Foo [23] describes system that recognizing 2D hand poses for application in video-based human-computer interfaces. They have developed a two layer feed-forward neural network that recognizes the 24 static letters in the American Sign Language (ASL) alphabet using images. Two wavelet-based decomposition methods have been used. The first produces an 8-element real-valued feature vector and the second a 18-element feature vector. Each set of feature vectors is used to train a feed-forward neural. The system is capable of recognizing instances of static ASL finger spelling with 99.9% accuracy.
<table>
<thead>
<tr>
<th>References</th>
<th>Research based systems</th>
<th>Feature Extraction Techniques</th>
<th>Input Source</th>
<th>Output</th>
<th>Features founded</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.</td>
<td>American Sign language</td>
<td>combines K curvature and convex hull algorithms</td>
<td>Images</td>
<td>Alphabets and Numbers</td>
<td>fingertip finder, eccentricity, elongatedness, pixel segmentation and rotation</td>
<td>94.32%</td>
</tr>
<tr>
<td>25.</td>
<td>Brazilian Sign Language</td>
<td>Lighting Normalization</td>
<td>Kinect sensor</td>
<td>LIBRAS alphabet</td>
<td>Euclidian distance</td>
<td>89%</td>
</tr>
<tr>
<td>26.</td>
<td>Arabic Sign Language</td>
<td>Fisher Linear Discriminant analysis</td>
<td>Microsoft Kinect sensor</td>
<td>20 Arabic language words</td>
<td>Dimension Measures</td>
<td>99.8%</td>
</tr>
<tr>
<td>27.</td>
<td>Thai Sign Language</td>
<td>histograms of oriented gradients</td>
<td>Microsoft Kinect sensor at 0.8 - 1.2 meter distance</td>
<td>16 Hand Gestures</td>
<td>Dimension Measures</td>
<td>83.33%</td>
</tr>
<tr>
<td>28.</td>
<td>Dynamic Gesture Recognition</td>
<td>skeleton 3D trajectory for Key frame detection and Histograms of Cumulative Magnitudes for future extraction</td>
<td>Microsoft Kinect sensor LIBRAS dataset</td>
<td>Global Features like Structural Movements and Local Features like Position, Hand Configuration</td>
<td>Average 94%</td>
<td></td>
</tr>
<tr>
<td>29.</td>
<td>American Sign images</td>
<td>Canny edge detection, seeded region growing</td>
<td>Video Camera</td>
<td>16 Alphabets</td>
<td>rotation, scaling, occlusion and variation</td>
<td>97.13%</td>
</tr>
<tr>
<td>30.</td>
<td>Real-time static Alphabet American Sign Language Recognizer</td>
<td>Edge Orientation Histogram (EOH)</td>
<td>Images</td>
<td>Alphabets A to Z</td>
<td>Pixel Regions, Area</td>
<td>88.26%</td>
</tr>
<tr>
<td>31.</td>
<td>Indian Sign Language Translator</td>
<td>2D FFT Fourier Descriptors</td>
<td>Images</td>
<td>10 numbers, 26 alphabets and 10 different phrases, External Boundary Points</td>
<td>Alphabets:8 5.73%, Numbers:9 5.5%, Phrases:97.5%</td>
<td>95.10%</td>
</tr>
<tr>
<td>32.</td>
<td>Sign Language Recognition</td>
<td>Elliptical Fourier Descriptors</td>
<td>Images by 4-Camera Model</td>
<td>Alphabets A to Z, Numbers</td>
<td>Hand shapes</td>
<td>95.10%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>References</th>
<th>Research based systems</th>
<th>Feature Extraction Techniques</th>
<th>Input Source</th>
<th>Output</th>
<th>Features founded</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>33.</td>
<td>Real Time Hand Gesture Recognition System for Android Devices</td>
<td>Color-Based Segmentation, Smoothing</td>
<td>Images from Smart Phone</td>
<td>0-9 Numbers</td>
<td>convex points in contour, point furthest away from each convex vertex</td>
<td>93%</td>
</tr>
<tr>
<td>34.</td>
<td>Bengali Sign Language Recognition</td>
<td>Gaussian Smoothing, Canny Edge Detector</td>
<td>Images</td>
<td>18 Bengali Words</td>
<td>Contours</td>
<td>90.11%</td>
</tr>
<tr>
<td>No.</td>
<td>Application</td>
<td>Feature Extraction Method</td>
<td>Domain</td>
<td>Accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35.</td>
<td>Indian Sign Language Recognition</td>
<td>Fourier Descriptors</td>
<td>Images and Alphabets</td>
<td>96.15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36.</td>
<td>ASL Number Recognition</td>
<td>Image Thinning, Count distance between fingers</td>
<td>Static Images and Numbers</td>
<td>92.09%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37.</td>
<td>Indian Sign Language Recognition using Transform Features</td>
<td>Hybrid Wavelet Transform</td>
<td>Images and Alphabets</td>
<td>91.02%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38.</td>
<td>Indian Sign Language Recognition System under Complex Background</td>
<td>Frame Differencing, Contour Matching algorithm, Calculation of centroids</td>
<td>Video</td>
<td>One Hand: 90.0%, Two hand: 86.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39.</td>
<td>Sign Language Interpretation</td>
<td>Blob Analysis, Noise reduction using smooth Gaussian filter,</td>
<td>Video</td>
<td>92.68%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.</td>
<td>Chinese Sign Language Recognition</td>
<td>9-dimensional feature vector</td>
<td>Kinect</td>
<td>69.32%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41.</td>
<td>Sign Language Recognition</td>
<td>* Contour is achieved through Euler-Largrange equation,</td>
<td>Video</td>
<td>93.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42.</td>
<td>American Sign Language Detection System</td>
<td>* HSV Color Model, * Principal component analysis (PCA) for dimension reduction,</td>
<td>Images and Alphabets</td>
<td>65%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43.</td>
<td>American Sign Language Detection System</td>
<td>* gradient magnitude features, * Random Projection (RP) and Kernel Principal Component Analysis (KPCA)</td>
<td>Static Images</td>
<td>99.8%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Comparison of Feature Extraction Methods
Table 3. Comparison of Feature Extraction Methods

<table>
<thead>
<tr>
<th>References</th>
<th>Research based systems</th>
<th>Feature Extraction Techniques</th>
<th>Input Source</th>
<th>Output</th>
<th>Features founded</th>
<th>Accuracy</th>
</tr>
</thead>
</table>
| 44.        | Real-Time Computer Vision-Based Bengali Sign Language Recognition | * Haar-like feature-based cascaded classifiers  
* Hue and Saturation value | Images | 10 Bengali alphabet | Geometrical properties of the hand shapes | 96% |
| 45.        | Sign Language Recognition | Principal Component Analysis(Matching Frames of Test image and Data Base image) | Video | 10 Alphabets | Pixel Segmentation, Hand Shape | 90% |
| 46.        | American Sign Language Recognition | 3D Geometric Invariant Feature | Image | Alphabets | All feasible triangle area patches constructed from 3D coordinates triplet. | 95% |
| 47.        | AUTOMATICSIGN LANGUAGEIDENTIFICATION | Hu-moments(for shapes),discrete grids (for locations), XORs (for movements) | Images | 19 signers for British and Greek sign languages | Hand-shape, orientation, location and movement. | 95% |
| 48.        | Static Indonesian Sign Language Recognition System | contours | Images | Alphabets | Contours, rectangles, center points | 62.6% |
| 49.        | Persian Sign Language Recognition | * mean-shift (CAMSHIFT) algorithm for hand tracking,  
* radon transform and discrete cosine transform (DCT) | Video | 20 dynamic signs | Angles | 95.56% |
| 50.        | LIBRAS Sign Language Hand Configuration Recognition | * 3D meshes,  
* Silhouette method for shape, scale invariant Spherical Harmonics method for rotation, translation and scale | Video | 61 Hand Configuration | shape (scale, rotational and translational invariance) | 96% |
| 51.        | Recognizing Words in the Sign System for Indonesian Language | * Skeleton data features indicate the angle between human joints and Cartesian axes,  
* Color images are transformed to gray-scale and their features are extracted by using Discrete Cosine Transform (DCT) with Cross Correlation (CC) operation  
* depth features are extracted by running MATLAB region props function to get its region properties | Kinect | SIBI words | Angle, Shape, Depth | 96.67% |
<table>
<thead>
<tr>
<th>References</th>
<th>Feature Extraction Techniques</th>
<th>Input Source</th>
<th>Output</th>
<th>Features founded</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>52.</td>
<td>Recognizing and Classifying Indonesian Sign Language</td>
<td>Discrete Cosine Transform (DCT), Cross Correlation Function (With 5 frames)</td>
<td>Kinect camera</td>
<td>alphabet (A to Z) and numbers 1 to 10</td>
<td>Cross Correlation values between same rows or columns</td>
</tr>
<tr>
<td>53.</td>
<td>Recognition of Indian Sign Language</td>
<td>shape descriptors, HOG descriptors (Histogram of Oriented Gradient) and SIFT (Scale Invariant Feature Transform)</td>
<td>Images</td>
<td>36 different signs</td>
<td>Eccentricity of an ellipse, Aspect Ratio, Compactness, Solidity, Orientation, Spreadness, Roundness, local intensity gradients by HOG, A Number of key points/interest points using SIFT</td>
</tr>
<tr>
<td>54.</td>
<td>LDCRFs-Based Hand Gesture Recognition</td>
<td>Polar systems</td>
<td>stereo color image sequences from video</td>
<td>alphabet characters (A - Z) and numbers (0 - 9)</td>
<td>location, orientation and velocity</td>
</tr>
<tr>
<td>55.</td>
<td>Video Gestures Identification And Recognition Of Indian Sign Language</td>
<td>Fourier Descriptor</td>
<td>video</td>
<td>4 to 5 sentences, some words</td>
<td>shape signatures :centroid distance and complex coordinates (position function)</td>
</tr>
<tr>
<td>56.</td>
<td>Indian Sign Language Recognition</td>
<td>Hu Moments and motion trajectory</td>
<td>video by Kinect</td>
<td>4 sign gestures of “A,” “B,” “C,” and “Hello.”</td>
<td>Hu Moments and motion trajectory</td>
</tr>
<tr>
<td>57.</td>
<td>Sign Language Recognition</td>
<td>Sparse Observation (SO) description</td>
<td>RGB-D data captured by Kinect</td>
<td>20 categories of gesture</td>
<td>Motion Trajectory Feature</td>
</tr>
<tr>
<td>58.</td>
<td>Static Hand Gesture Recognition of Persian Sign Numbers</td>
<td>Thinning Method</td>
<td>Images</td>
<td>Persian Sign Numbers</td>
<td>length of thin segments , end points of image</td>
</tr>
<tr>
<td>59.</td>
<td>A Static Hand Gesture Recognition Algorithm</td>
<td>localized contour sequence (LCS) based feature</td>
<td>Images</td>
<td>Alphabets</td>
<td>Shape, Dimension, Edges</td>
</tr>
<tr>
<td>60.</td>
<td>Real-time Ukrainian sign language recognition system</td>
<td>fingertips location and pseudo 2-dimensional image deformation model</td>
<td>videos</td>
<td>85 signs</td>
<td>Hand Shape</td>
</tr>
</tbody>
</table>
In this review paper, different techniques of sign language recognition are reviewed on the basis of feature extraction. For sign acquiring methods, vision based future extraction methods are more reliable. We can easily find different features of sign like hand shape, rotation, angles, movements, coordinates, pixel intensity etc from images. It can be possible to make real time system which continuously capture sign generate sentences. According to this paper more research has been done on words, alphabets and numbers. In future it will go more in dimension on continues sentences of signs. Further review can be possible for image processing technique on the base of camera type, Image pixel resolution, frame compression, image background, Image Distance etc.

6. REFERENCES

[27] Chana Chansri, Jakkree Srinonchat,” Reliability and Accuracy of Thai Sign Language Recognition with Kinect Sensor”, 978-1-4673-9749-0/2016 IEEE.


[47] Binyam Gebrekidan Gebre, Peter Wittenburg, Tom Heskes ”, AUTOMATICSIGNLanguAGeIDENTIFICATION”, 978-1-4799-2341-0/2013 IEEE.


