Abstract—Bridging communication gap between the common man and the hearing impaired people is a big challenge. Thus lot of research is being done in the area of sign language recognition to come-up with the optimum algorithm or solution for cent percent recognition of signs. Till now numbers of techniques are being developed. This article explains wavelet feature based method to recognize the 24 static-image based alphabets A to Z (excluding dynamic alphabets J and Z) of American Sign Language (ASL). This method extracts the wavelet feature vectors of the images. Further neural network is used for the classification of these alphabets. This method is qualified to provide an average recognition rate of 97.47 percent at fourth level decomposition of wavelet features.

Index Terms— American Sign Language, ASL alphabets, Gesture Recognition, Neural Network, Wavelet Transform,

I. INTRODUCTION

Computers are used by many people in their day to day life for all activities. Special input and output devices have been designed over the years with the purpose of easing the communication between computers and humans, the two most known are the keyboard and mouse [1].

The idea is to make computers understand human language and develop a user friendly human computer interfaces (HCI). Making a computer understand speech, facial expressions and human gestures are some steps towards it. Gestures are the non-verbally exchanged information. A person can perform innumerable gestures at a time. Since human gestures are perceived through vision, it is a subject of great interest for computer vision researchers. Coding of these gestures into machine language demands a complex programming algorithm.

Gestures are classified into two distinctive categories: dynamic and static [1]. A dynamic gesture is intended to change over a period of time, whereas a static gesture is observed at an instance of time. Dynamic gestures are considered as consecutive sequences of hand or head or body postures in sequence of time frames. A waving hand means goodbye, is an example of a dynamic gesture and the stop sign is an example of static gesture. To understand a full message, it is necessary to interpret all the static and dynamic gestures over a period of time. This complex process is called gesture recognition. Gesture recognition is the process of recognizing and interpreting a stream continuous sequential gesture from the given set of input data. Dynamic gesture recognition is accomplished using Hidden Markov Models (HMMs), Dynamic Time Warping, Bayesian networks or other pattern recognition methods. Static gesture (or pose gesture) recognition can be accomplished by using template matching, eigen spaces or PCA (Principal Component Analysis), Elastic Graph Matching, neural network or other standard pattern recognition techniques. Template matching techniques are actually the pattern matching approaches. It is possible to find out the most likely hand postures from an image by computing the correlation coefficient or minimum distance metrics with template images.

Gesture recognition has significant application in sign language recognition. Sign languages are the most raw and natural form of languages that could be dated back to as early as the advent of the human civilization, when the first theories of sign languages appeared in history. Sign languages are being used extensively in international signs used by deaf and dumb, in the world of sports, for religious practices and also at work places.

Hearing impaired people have over the years developed a gestural language where all defined gestures have an assigned meaning. The language allows them to communicate with each other and the world they live in. Fig. 1 shows the different gestures in an American Sign Language for the alphabets A to Z.

![Fig. 1 Gestures in American Sign Language](image-url)

II. RELATED WORK

In sign language recognition, it is desirable to use a shape representation technique that will sufficiently describe the shape of the hand while also being capable of fast computations, enabling recognition to be done in real-time. It is also desirable for the technique to be invariant to translation, rotation, and scaling. In addition, a method that will allow for easy matching would be beneficial.

Gesture recognition was first proposed by Myron W. Krueger as a new form of interaction between human and computer in the middle of a seventies [2]. Currently, there are several available techniques that are applicable for hand

Manuscript received June 29, 2015. This work is supported in by Government of India through AICTE’s Grant under Career Award for Young Teacher (CAYT).

Asha Thalange is Assistant Professor at Walchand Institute of Technology, Solapur, India.

Dr. Shantanu Dixit is a Research Guide at Walchand Institute of Technology, Solapur, India.

http://dx.doi.org/10.15242/IAE.IAE0815014
gesture recognition. Zimmerman [3] developed a VPL data glove that is linked to the computer to recognize signs. The glove can measure the bending of fingers, the position and orientation of the hand in 3-D space. In vision-based gesture recognition, hand-shape segmentation is one of the toughest problems under a dynamic environment. It can be simplified by using visual markings on the hands. Some researchers have implemented sign language and pointing gesture recognition based on different marking modes [4]. Arpita Ray Sarkar et al. [5] reviews the work carried out in last twenty years and presents brief comparison to analyze the difficulties encountered by these systems, as well as the limitation. The desired characteristics of a robust and efficient hand gesture recognition system have been described. Klimis Symeonidis [6] used an orientation histogram of the image to develop a simple and a fast algorithm to extract features from the static image for comparison and recognize some of the alphabets of a static ASL using neural network. Becky Sue Parton [7] discusses various projects involving sign language and the potential impact these endeavors will have on deaf education and communities at large. He also discusses the use of artificial intelligence in the field robotics, virtual reality, computer vision, neural networks, Virtual Reality Modelling Language (VRML), three-dimensional (3D) animation, natural language processing (NLP), and an intelligent computer aided instruction (ICAI), for sign language manipulation. Noor Adnan Ibraheem et al. [8] illustrates the survey on various recent gesture recognition approaches with particular emphasis on hand gestures. They present the comparative study between different methods with respect to segmentation method, features extracted, number of signs used, classification method used and their recognition rate. Other approaches such as; local linear embedding, Neural Network shape fitting, object based key frame selection, and Haar wavelet representations have been presented in [9], [10]-[12]. S. Nagarajan et al. proposes a static hand gesture recognition system for American Sign Language using Edge Oriented Histogram (EOH) features and multiclass SVM. Rafiqul Zaman Khan [14] discusses key issues of hand gesture recognition system with challenges of gesture system. He also present review of methods of recent postures and gestures recognition system. Summary of research results of hand gesture methods, databases, and comparison between main gesture recognition phases are also given. Advantages and drawbacks of the discussed systems are explained. Nachamai. M [15] attempted to recognize ASLalphabets as part of hand gesture recognition, using the SIFT algorithm. Nicolas Pugeault et al. [16] make use of a Microsoft Kinect device to collect appearance and depth images, and of the OpenNI+NITE framework for hand detection and tracking. A. Karami et al. [17] presents a system for recognizing static gestures of alphabets in Persian sign language (PSL) using Wavelet transform and neural networks (NN) and is able to recognize 32 selected PSL alphabets with an average classification accuracy of 94.06%.

III. SYSTEM DESCRIPTION

The block diagram of the system implemented is as shown in the fig. 2. Here the ASL alphabets A to Z are considered excluding the alphabet J and Z as the sign for these alphabets is not static.

The whole system functioning is divided into four main modules namely:
1) Image capture and pre-processing
2) Image cropping and resize
3) Feature extraction
4) Classification

A. Image Capture And Pre-Processing

The colour image of the sign of an ASL alphabet with a plane black background is captured by a 5 megapixel web-camera concentrating on the palm of the hand. The plane background is used for simplicity of segmentation process. Further, the colour image is converted to gray scale image of size 256 x 420 pixels. Then low pass filtering is applied on the input image. The dataset is created by us. The total of 720 images are captured for training, which consists of 30 images per alphabet sign (A to Z except J and Z), along with a separate set of, a total of 823 images of all alphabets, for testing, by a single signer. The database consists of the images which are obtained with varying scale and contains rotated samples between +45 and -45 degree in order to make system robust.

B. Image Cropping And Resize

In order to make the image scale invariant, the pre-processed image is cropped to segment the background and concentrate only on the palm.

To obtain this, first the gray scale image is converted to binary image with a black background and white hand. This is done by threshold method. This process of segmentation of hand from background is very important and the whole further process of feature extraction and recognition is dependent on the proper segmentation process. Further median filtering and morphological operations are applied to remove noise. Image thinning is applied to some extent to a binary image in order to get clear separation among the edges and enhance the shape. The extent of thinning is limited so that the image is not totally converted to a skeleton. Considering the extreme points of the
hand in all the four directions, the segmented binary image of the hand from the background is obtained. Also by filling the respective white pixels of the segmented image by the gray values from its original gray image cropped gray image is also obtained. This segmented gray image and binary image is further resized to 128x128 pixels. Thus, making the sign image scale independent.

C. Feature Extraction

Wavelet transforms have become one of the most important and powerful tool of signal representation. Nowadays, it has been widely used in image processing, data compression, and signal processing. Unlike conventional Fourier transform, wavelet transforms are based on small waves, called wavelets. The two-dimensional DWT (Discrete Wavelet Transform) is of particular interest for image processing and computer vision applications, and is a relatively straightforward extension of the one-dimensional DWT.

Fig. 3 illustrates the basic, one-level, two-dimensional DWT procedure.

First, the one-dimensional DWT is applied along the rows; second, the one-dimensional DWT is applied along the columns of the first-stage result, generating four sub-band regions in the transformed space: LL, LH, HL and HH. The LL band corresponds roughly to a down-sampled (by a factor of two) version of the original image. The LH band tends to preserve localized horizontal features, while the HL band tends to preserve localized vertical features in the original image. Finally, the HH band tends to isolate localized high-frequency point features in the image. The same procedure is further applied to the LL band which is also known as an approximation coefficients matrix to get further levels of decomposition. In the proposed method the binary image of size 128 x 128 obtained after image cropping and resize is used as the input image for DWT. MATLAB function ‘dwt2()’ using ‘Haar’ wavelets is used to compute 2D DWT. Here 4 levels of decomposition are done and the approximation coefficients matrix of 4th level decomposition is used as feature vector for training. This image is of the size 8 x 8, i.e. of 64 pixels. This 2D data is converted into 1D data of a single vector with 64 elements which itself act as a feature vector obtained from the image of a sign.

D. Classification:

Multi-layered feed forward back-propagation Neural Network based classification engine is used here. Network with 64 neurons in input layer, single hidden layer with 80 neurons and 24 target neurons in the output layer is created with ‘newff()’ function in MATLAB 2014a. Initially, this network is trained with feature vectors obtained from training set consisting of 30 images of each ASL alphabet obtained from a single user. The function ‘train()’ is used for this purpose which uses default network training function that updates weight and bias values according to Levenberg-Marquardt optimization method. Once the network is trained, the feature vectors obtained from the actual test images is then applied to this network for classification or recognition of an ASL alphabet sign. Further, the detected alphabet is displayed in text form. Here the training and classification is performed by varying the number of neurons in the hidden layer. The maximum recognition rate is obtained at 80 neurons.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

Fig. 4 illustrates the different processing stages of sign of ASL alphabet A. Fig. 4.a shows the captured RGB image. Fig. 4.b shows the RGB to Gray scale image. Fig. 4.c shows the resized 256x420 gray scale image. Fig. 4.d is filtered and morphologically processed binary image. Fig. 4.e shows the cropped binary image of 128 x 128 whose wavelet features are being extracted. Fig. 4.f, 4.g, 4.h shows the LH, HL, HH image of 1st level decomposition. Fig. 4.i, 4.j, 4.k, 4.l shows the approximation coefficients matrix image LL at 1st, 2nd, 3rd and 4th level decomposition.

Table I shows the summary of the recognition of total signs of an ASL alphabets A to Z (excluding J and Z), applied to the system, by considering approximation coefficients matrix at 4th level decomposition of wavelet features of the image, as feature vector. From the result it can be seen that, the system gives higher average recognition rate of 97.47% with wavelet features compared to other methods as mentioned above in related work without using gloves, long sleeves, color marker and considering all the 24 static alphabets of ASL.
Fig. 4: Different processing stages of sign of ASL alphabet A

TABLE I
CLASSIFICATION RESULT OF TEST IMAGES FOR ASL ALPHABETS A TO Z USING WAVELET FEATURES

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>i</th>
<th>k</th>
<th>l</th>
<th>m</th>
<th>Average % of Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Images</td>
<td>31</td>
<td>33</td>
<td>33</td>
<td>34</td>
<td>29</td>
<td>32</td>
<td>29</td>
<td>32</td>
<td>32</td>
<td>40</td>
<td>37</td>
<td>52</td>
<td>97.47</td>
</tr>
<tr>
<td>Detected Correctly</td>
<td>26</td>
<td>33</td>
<td>33</td>
<td>34</td>
<td>28</td>
<td>32</td>
<td>29</td>
<td>31</td>
<td>32</td>
<td>40</td>
<td>37</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>% of recognition</td>
<td>83.87</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>96.55</td>
<td>100</td>
<td>100</td>
<td>96.87</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>96.15</td>
<td></td>
</tr>
</tbody>
</table>

V. CONCLUSION

Compared to glove based or all other complex feature extraction techniques which are applied to limited signs, Wavelet feature extraction technique is a good method to recognize all the alphabets of the ASL with maximum accuracy.

VI. FUTURE SCOPE

Effect of combination of some other efficient features extracted from image along with wavelet features can be used as a feature vector. Also other classification algorithms can be applied in order to increase the recognition rate.

REFERENCES


http://dx.doi.org/10.1109/ijcnn.1999.833533

http://dx.doi.org/10.15242/IAE.IAE0815014
http://dx.doi.org/10.1016/j.jvlc.2005.04.003


http://dx.doi.org/10.1109/icdip.2009.74


http://dx.doi.org/10.1109/icccv.2011.6130290

http://dx.doi.org/10.1016/j.eswa.2010.08.056

Asha Thalange is born in Bangalore, Karnataka, India. She completed her B.E. in Electronics Engineering from Walchand Institute of Technology, Solpaur, Maharashtra, India in 2002 and Masters M.E. in Electronics and Telecommunication Engineering from Sinhgad College of Engineering, Pune, Maharashtra, India in 2009. Currently she is pursuing her Ph.D. at Walchand Institute of Technology, Solpaur.
She has got a total teaching experience of 11 Years. She had published 15 research papers in reputed National and International journals and conference proceedings. Her research area includes signal and image processing.
Ms. Thalange has received a grant from AICTE under the scheme of Career Award for Young Teacher (CAYT) for her research project, received Best Paper Award at ETCC 15, Bhubaneswar, India.

Dr. Shantanu K. Dixit completed his B.E. in Electrical Engineering in 1978, M.E. in Electronics Engineering in 1989 and Ph.D. in Electronics Engineering from Walchand College of Engineering, Sangli, Maharashtra, India, in 2002. He is a member of IEEE, ISTE, IE (FIE).
He has got a total teaching experience of 34 years with 10 years of research experience. He had published nearly 30 research papers in reputed National and International Journals and conferences. His area of interest includes Robotics and Audio / Video Engineering.
Dr. Dixit received a grant from AICTE, India under the scheme of Research Promotion Scheme (RPS).