Appearance Based Recognition of American Sign Language Using Gesture Segmentation

Author[1] Vaishali S. Kulkarni
M.E. (Digital Systems)
Sinhgad College of Engineering
Dept. of Electronic and Telecommunication
Pune, India

Author[2] Dr. S.D.Lokhande.
H.O.D. (E&TC)
Sinhgad College of Engineering
Dept. of Electronic and Telecommunication
Pune, India

***Abstract*— The work presented in this paper goals to develop a system for automatic translation of static gestures of alphabets in American Sign Language. In doing so three feature extraction methods and neural network is used to recognize signs. The system deals with images of bare hands, which allows the user to interact with the system in a natural way. An image is processed and converted to a feature vector that will be compared with the feature vectors of a training set of signs. The system is rotation, scaling of translation variant of the gesture within the image, which makes the system more flexible.**

The system is implemented and tested using data sets of number of samples of hand images for each signs. Three feature extraction methods are tested and best one is suggested with results obtained from ANN. The system is able to recognize selected ASL signs with the accuracy of 92.33%.

***Index Terms*—ASL, ASL recognition, ASL using ANN**

I. INTRODUCTION

The sign language is the fundamental communication method between the people who suffer from hearing defects. In order for an ordinary person to communicate with deaf people, a translator is usually needed the sign language into natural language and vice versa. International Journal of Language and Communication Disorders, 2005) Sign language can be considered as a collection of gestures, movements, posters, and facial expressions corresponding to letters and words in natural languages.

There are two types of gesture interaction: communicative gestures work as symbolic language (Which is the focus in this project) and manipulative gestures provide multi-dimensional control. Also, gestures can be divided into static gestures (hand postures) and dynamic gestures (Hong et al., 2000). The hand motion conveys as much meaning as their posture does. A static sign is determined by a certain configuration of the hand, while a dynamic gesture is a moving gesture determined by a sequence of hand movements and configurations. Dynamic gestures are sometimes accompanied with body and facial expressions. The aim of sign language alphabets recognition is to provide an easy, efficient and accurate mechanism to transform sign language into text or speech. With the help of computerized digital image processing and neural network the system can interpret ASL alphabets.

A. American Sign language

American Sign Language (ASL) National Institute on Deafness & Other communication Disorders, 2005) is a complete language that employs signs made with the hands and other facial expressions and postures of the body. According to the research by Ted Camp found on the Web site www.silentworldministries.org, ASL is the fourth most used language in the United States only behind English, Spanish and Italian (Comp). ASL is a visual language meaning it is not expressed through sound but rather through combining hand shapes through movement of hands, arms and facial expressions. Facial expressions are extremely important in signing. (www.nidcd.nih.gov(US government)). ASL also has its own grammar that is different from other sign languages such as English and Swedish. ASL consists of approximately 6000 gestures of common words or proper nouns. Finger spelling used to communicate unclear words or proper nouns. Finger spelling uses one hand and 26 gestures to communicate the 26 letters of the alphabet. The 26 alphabets of ASL are shown in Fig.1.

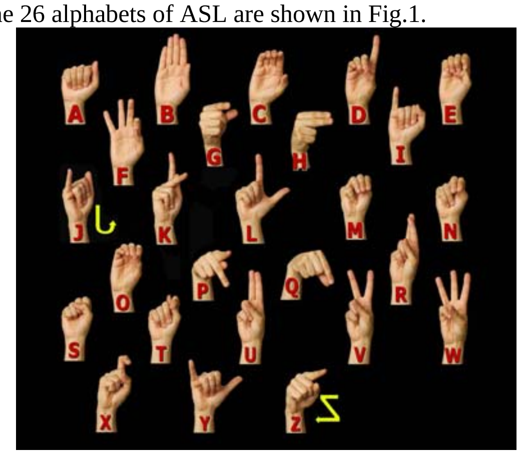

Figure. 1. The American Sign Language finger spelling alphabet

B. Related Work

Attempts to automatically recognize sign language began to appear in the 90s. Research on hand gestures can be classified into two categories: First category relies on electromechanical devices that are used to measure the different gesture parameters such as hand's position, angle, and the location of the fingertips. Systems that use such devices are called glove-based systems. A major problem with such systems is that they force the signer to wear cumbersome and inconvenient devices. Hence the way by which the user interacts with the system will be complicated and less natural. The second category uses machine vision and image processing techniques to create visual based hand gesture recognition systems. Visual based gesture recognition systems are further divided into two categories: The first one relies on using specially designed gloves with visual markers called "visual-based gesture with glove-markers (VBGwGM)" that help in determining hand postures. But using gloves and markers do not provide the naturalness required in human-computer interaction systems. Besides, if colored gloves are used, the processing complexity is increased. The second one that is an alternative to the second kind of visual based gesture recognition systems can be called "pure visual-based gesture (PVBG)" means visual-based gesture without glove-markers. And this type tries to achieve the ultimate convenience naturalness by using images of bare hands to recognize gestures. A number of recognition techniques are available and in some cases can be applied for the two types of vision based solutions (i.e. VBGwGM and PVBG). These recognition techniques can be classified into three broad categories:

1. *Feature extraction, statistics, and models.*
   - *a.* Template matching(e.g. research work Darrell and Pentland, 1993)
   - *b.* Feature extraction and analysis,(e.g. research work of Rubine,1991)
   - *c.* Active shape models "Smart snakes"(e.g. research work of Heap and Samaria,1995)
   - *d.* Principal component analysis(e.g. research work of Birk,Moeslund and Madsen,1997)
   - *e.* Linear fingertip models(Research work of Davis and shah, 1993)
   - *f.* Causal analysis (e.g. research work of Brand and Irfan, 1995).
2. *Learning algorithms.*
   - *a.* Neural network (e.g. research work of Banarse, 1993).
   - *b.* Hidden Markov Models (e.g. research work of Charniak, 1993).
   - *c.* Instance-based learning(research work of Kadous,1995)
3. *Miscellaneous techniques.*
   - *a.* The linguistic approach(e.g. research work of Hand, Sexton, and mullan,1994)
   - *b.* Appearance-based motion analysis (e.g. research work of Davis and Shah, 1993).
   - *c.* Spatio-temporal vector analysis(e.g. research work of Wuek, 1994)

Among many factors, five important factors must be considered for the successful development of a vision-based solution to collecting data for hand posture and gesture recognition

1. The placement and number of cameras used.
2. The visibility of the object (hand) to the camera for simpler extraction of hand data/features.
3. The extraction of features from the stream of streams of raw image data.
4. The ability of recognition algorithms to extracted features.
5. The efficiency and effectiveness of the selected algorithms to provide maximum accuracy and robustness.

Regardless of the approach used (i.e. VBGwGM or PVBG etc.) many researchers have been trying to introduce hand gestures to Human-Computer Interaction field. *Year 1992:* Charayaphan and Marble investigated a way using image processing to understand American Sign Language. Their system can correctly recognize 27 out of 31 ASL symbols. *Year 1993:* Fels and Hinton developed a system using VPL Data glove Mark II with a Polhemus tracker as input device in their system neural network method was employed for classifying hand gestures. *Year 1993:* Another system was developed by Banarse using neural networks. It was vision-based and recognized hand postures using a neocognitron network which is a neural network based on the spatial recognition system of the visual cortex of the brain. *Year 1995:* Heap and Samaria extend active shape models, or "Smart Snakes" technique to recognize postures and gestures using computer vision. In their system, they apply an active shape model and a point distribution model for tracking a Human hand. *Year 1995:* Starner and Pentland used a view-based approach with a single camera to extract two-dimensional features as input to HMMs. The correct rate was 91% in recognizing the sentences comprised 40 signs. *Year 1996:* Kadous demonstrated a system based on power gloves to recognize a set of 95 isolated Aulsan signs with 80% accuracy, with an emphasis on computationally inexpensive methods. *Year 1996*: Grobel and Assan used HMMs to recognize isolated signs with 91.3% accuracy out of 262 sign vocabulary. They extracted the features from video recording of signers wearing colored gloves. *Year 1997:* Vogler and Metaxas used computer vision methods and HMMs to recognize continuous American Sign Language sentences with a vocabulary of 53 signs. They modeled context-dependent HMMs to alleviate the effects of movement epenthesis. An accuracy of 89.9% was observed. *Year 1998:* Yoshinori, Kang-Hyun, Nobutaka, and Yoshiaki used colored gloves and have shown that using solid colored gloves allows faster hand features extraction than simply wearing no gloves at all. *Year 1998:* Liang and Ouhyoung used HMMs for continuous recognition of Taiwan sign language with a vocabulary between 71 and 250 signs with data glove as input device. However their system required that gesture performed by the signer be slow to detect the word boundary. *Year 1999:* Yang and Ahuja investigated dynamic gestures recognition as they utilized skin color detection and affine transforms of the skin regions in motion to detect the motion trajectory of ASL signs. Using a time delayed neural network, they recognized 40 ASL gestures with a success rate around 96%. But their technique potentially has a high computational cost when false skin regions are detected. *Year 2000:* A local feature extraction technique is employed to detect hand shapes in sign language recognition by Imagawa, Matsuo, Taniguchi, Arita, and Igi. They used appearance based Eigen method to detect hand shapes. Using a clustering technique, they generate clusters of hand shapes on an eigenspace. They have achieved accuracy of around 93% recognition of 160 words. *Year 2000:* Symeoinidis used orientation histograms to recognize static hand gestures, specifically, a subset of American Sign Language. A pattern recognition system used a transform that converts an image into feature vector, which will then be compared with the feature vectors of a training set of gestures. The system was implemented with a perceptron network. The main problem with technique is how good differentiation one can achieve. This is mainly dependent upon the images but it comes down to the algorithm as well. It may be enhanced using other image processing technique like edge detection as done in the presenting paper. *Year 2002:* Bowden and Sarhadi developed a non-linear model of shapes and motion for tracking fingersplet American Sign Language. Their approach based on one-state transition of the English Language which are projected into shape space for tracking and model prediction using HMM like approach.

II. SYSTEM DESIGN AND IMPLEMENTATION

The system is designed to visually recognize all static signs of the American Sign Language (ASL), all signs of ASL alphabets using bare hands. The user/signers are not required to wear any gloves or to use any devices to interact with the system. But, since different signers vary their hand shape size, body size, operation habit and so on, which bring more difficulties in recognition. Therefore, it realizes the necessity for signer independent sign language recognition to improve the system robustness and practicability in the future. The system gives the comparison of the three feature extraction methods used for ASL recognition and suggest a method based on recognition rate. It relies on presenting the gesture as a feature vector that is translation, rotation and scale invariant. The combination of the feature extraction method with excellent image processing and neural networks capabilities has led to the successful development of ASL recognition system using MATLAB. The system has two phases: the feature extraction phase and the classification as shown in Fig.2. Images were prepared using portable document format (PDF) form so the system will deal with the images that have a uniform background. The feature extraction applied an image processing technique which involves algorithms to detect and isolate various desired portions of the digitized sign. During this phase, each colored image is resized and then converted from RGB to grayscale one. This is followed by an edge detection technique. The goal of edge detection is to mark the points in an image at which the intensity changes sharply. Sharp changes in image properties usually reflect important invents and changes in world properties. The next important step is the application of proper feature extraction method and the next is the classification stage, a 3-layer, feed-forward back propagation neural network is constructed.

A. Feature Extraction Phase

Images of signs were resized to 80 by 64, by default “imresize” uses nearest neighbor interpolation to determine the values of pixels in the output image but other interpolation methods can be specified. Here ‘bicubic’ method is used because if the specified output size is smaller than the size of the input image, “imresize” applies a low pass filter before interpolation to reduce aliasing. Therefore we get default filter size 11-by11. To alleviate the problem of different lighting conditions of signs taken and the HSV “(Hue, Saturation, Brightness)” non-linearity by eliminating the HSV information while retaining the luminance. The RGB color space (Red, Green and Blue which considered the primary colors of the visible light spectrum) is converted through grayscale image to a binary image. Binary images are images whose pixels have only two possible intensity values. They are normally displayed as black and white. Numerically, the two values are often 0 for black and either 1 or 255 for white. Binary images are often produced by thresholding a grayscale or color image from the background. This conversion resulted in sharp and clear details for the image. It is seen that the RGB color space conversion to HSV color space then to a binary image produced images that lack many features of the sign. So edge detection is used to identify the parameters of a curve that best fir a set of given edge points. Edges are significant local changes of intensity in an image. Edges typically occur on the boundary between two different regions in an image. Various physical events cause intensity changes. Goal of edge detection is to produce a line drawing of a scene from an image of that scene. Also important features can be extracted from the edges. And these features can be used for recognition. Here canny edge detection technique is used because it provides the optimal edge detection solution. Canny edge detector results in a better edge detection compared to Sobel edge detector. The output of the edge detector defines ‘where’ features are in the image. Canny method is better, but in some cases it provides extra details more than needed. To solve this problem a threshold of 0.25 is decided after testing different threshold values and observing results on the overall recognition system.

1. Feature Extraction Methods Used.

a. Histogram Technique
b. Hough
c. OTSU’s segmentation algorithm
d. Segmentation and Extraction with edge detection

B. Classification Phase

The classification neural network is shown (see figure 3). It has 256 instances as its input vector, and 214 output neurons in the output layer.

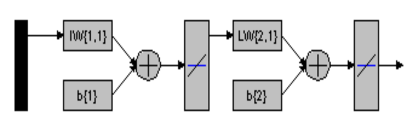

Fig. 3: Classification network.

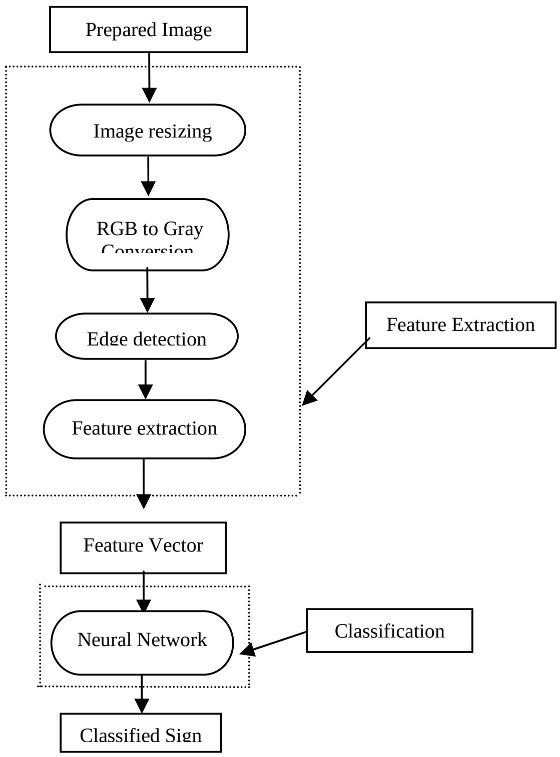

Figure. 2. System Overview.

classification phase includes network architecture, creating network and training the network. Network of feed forward back propagation with supervised learning is used.

III. EXPERIMENTAL RESULTS AND ANALYSIS

The performance of the recognition system is evaluated by testing its ability to classify signs for both training and testing set of data. The effect of the number of inputs to the neural network is considered.

A. Data Set

The data set used for training and testing the recognition system consists of grayscale images for all the ASL signs used in the experiments are shown see fig. 4. Also 8 samples for each sign will be taken from 8 different volunteers. For each sign 5 out of 8 samples will be used for training purpose while remaining five signs were used for testing. The samples will be taken from different distances by WEB camera, and with different orientations. In this way a data set will be obtained with cases that have different sizes and orientations and hence can examine the capabilities of the feature extraction scheme.

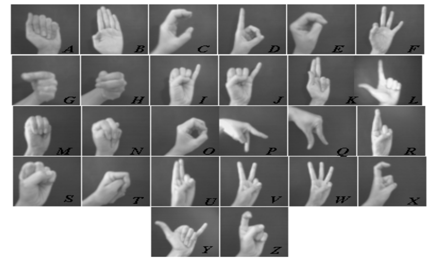

Fig. 3. ASL signs used in the system. (Actual)

B. Recognition Rate

The system performance can be evaluated based on its ability to correctly classify samples to their corresponding classes. The recognition rate can be defined as the ratio of the number of correctly classified samples to the total number of samples and can be given as

$$\text{Recognition rate} = \frac{\text{number of correctly classified signs}}{\text{Total number of signs}} * 100\%$$

C. Experimental Results

The network is trained on 8 samples of each sign. samples of same size and other features like distance rotation and lighting effect and with uniform background are taken into consideration while discarding the others .

Table 1 Results of training 8 samples for each sign with (0.25) Canny threshold

Sign	Recognized samples	Misclassified samples	Recognition rate (%)
A	7	1	66.66
B	7	1	66.66
C	7	1	66.66
D	8	0	100
E	8	0	100
F	8	0	100
G	7	1	66.66
H	7	1	66.66
I	8	0	100
J	8	0	100
K	7	1	66.66
L	7	1	66.66
M	8	0	100
N	7	1	66.66
O	7	1	66.66
P	8	0	100
Q	8	0	100
R	7	1	66.66
S	7	1	66.66
T	8	0	100
U	8	0	100
V	8	0	100
W	8	0	100
X	6	2	33.33
Y	8	0	66.66
Z	6	2	33.33
Total	**193**	**15**	**92.78**

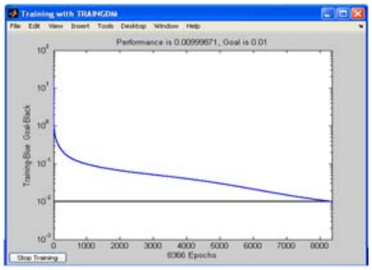

Figure. 4 Training chart for a network trained on 8 samples for each sign, (0.25) Canny threshold

D. GUI Simulating Results (sign to text)

For testing the unknown signs we have created an GUI as shown in Fig. 5 which provides the user an easy way to select any sigh He/She wants to test and then after clicking on the Apply pushbutton it will display the meaning of the selected sign.

Figure. 5 Example on GUI simulating sign 'D'

Fig.6 GUI of original 'F' sign Fig.7 identification of Sign 'F' with rotation

E. *GUI Simulating Results (Sign To Text)*

A text to sign interpreter means if the user types any sign or sentence corresponding signs are shown so that normal person to deaf people communication can be done. Examples of showing the sign to text converters are shown (See figure 6). Fig. 6 shows when the user type the name 'BOB' in the text box its corresponding signs are appear on the screen one by one above the text or spelling.

Figure. 7 Example on GUI Showing Real Time Result for 'BOB' alphabets

IV. Hardware and Software

The system is implemented in MATLAB version 6.5. The recognition training and tests were run on a modern standard PC (1.5 GHz AMD processor, 128 MB of RAM running under windows 2000.) WEB-CAM-1.3 is used for image capturing.

Conclusion

The system is proved robust against changes in gesture. Using Histogram technique we get the misclassified results. Hence Histogram technique is applicable to only small set of ASL alphabets or gestures which are completely different from each other. It does not work well for the large or all 26 number of set of ASL signs. For more set of sign gestures segmentation method is suggested. The main problem with this technique is how good differentiation one can achieve. This is mainly dependent upon the images but it comes down to the algorithm as well. It may be enhanced using other image processing technique like edge detection as done in the presenting paper. We used the well known edge detector like Canny, Sobel and Prewitt operators to detect the edges with different threshold. We get good results with Canny with 0.25 threshold value. Using edge detection along with segmentation method recognition rate of 92.33% is achieved. Also the system is made background independent. As we have implemented sign to text interpreter reverse also implemented that is text to sign interpreter.

Future Scope / Challenges

The work presented in this project recognizes ASL static signs only. The work can be extended to be able to recognize dynamic signs of ASL. The system deals with images with uniform background, but it can be made background independent. It is overcome and it is made background independent. The network can be trained to the other types of images. It is important to consider increasing data size, so that it can have more accurate and highly performance system.

Acknowledgment

The authors wish to thank to guide Dr. S.D. Lokhande for his valuable guidance for this work. and the volunteers who has given gesture images in the required format no of times whenever required till to the results.

References

[1] American Sign Language recognition using multidimensional Hidden Markov Models. Journal of information science and engineering 22, 1109-1123(2006) .HONGGANG WANG1, MING C. LEU2 AND CEMIL OZ3 Department of Industrial Engineering Purdue University.
[2] Development of a new "Sign Writer" program Daniel Thomas Ulrich Noelpp-d.noelpp@gmx.ch http:/www.signwriter.org.
[3] CULSHAW MURRY (1983) "It Will Soon Be Dark. The situation of disabled in India. Delhi lithouse publications
[4] Deshmukh Dilip (1994) the status of sign language in deaf education in India. Signpost. Newsletter of International Sign Language Association 7(1) 49-52.

[5] Starner, T. (1995). Visual recognition of American Sign Language using hidden Markov models. Master's thesis, Massachusetts Institutes of Technology.

[6] Duda, R. O., & Hart, P. E. (1972). Use of Hough transformation to detect lines and curves in pictures. Communications of the ACM, 15, 11-15.

[7] Kourosh Khoshelham. Extending the use of Hough Transform to detect 3D objects in laser range data. ISPRS Workshop on Laser Scanning 2007 and SilviLaser2007, Espoo, September 12-14,2007,Finland

[8] Ong, S., and Ranganath, S. (2005) Automatic sign language analysis: a survey and the future beyond lexical meaning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(6).

[9] Symeoinidis, k. (2000). "Hand gesture recognition-using neural networks." Master's thesis, University of Surrey.

[10] Vamplew,p.(1996). Recognition of sign language using neural networks PhD Thesis, Department of computer Science, and University of Tasmania.

[11] Watson, R. (1993). A survey of gesture recognition techniques. Technical report TCD-CS-93-11, Department of computer Science, Trinty College Dublin.

[12] C.W. ong and surendra and ranganath Automatic sign language analysis: A survey and future beyond lexial meaning sylvie IEEE transactions on pattern analysis and machine intelligence, Vol27,No.6,June 2006.

[13] Thad sarner,Joshua weaver and alex pentcana Real time American sign language recognition using desk and wearable computer based video. M.I.T. Media laboratory perceptual computing section technical report No.466 to appear IEEE PAMI'98'

[14] Thuy Tuong nguyen,Xuan Dai Pham,Jae work Jeon "An improvement of the standard Hough Transform to detect line segments" member IEEE-2008 978-1-4244-1706-3/0/$25

[15] "Pattern classification by Assembling small Neural networks" IEEE Proceedings of International Joint conference on Neural networks,Montreal,Canada,july 31-August4,2005.

[16] C.K.Lee and S.P. Wong (Science Direct) A mathematical morphological approach for segmenting heavily noise-corrupted images

[17] "ASL recognition based on Hough Transform and neural Network" Expert system with application 32(2007) 24_37 ScienceDirect

[18] Jung-Bae Kim,Kwang-Hyun Park,Won_Chul Bang and Z.Zenn Bien Div.Of EE,Dept of EECS,KAIST,Daejeon,Republic of Korea.Continuous Korean sign language recognition using gesture segmentation and HMM. IEEE

[19] Joe Naoum-Sawaya,Mawaya,Mazen Slim,Sami Khawam and Mohamad Adnan Al-Alaoui Dynamic system design for American Sign Language Recognition by Electrical and Computer Engineering Department American University of Beirut,Lebanon. IEEE

[20] Real-Time American Sign Language Recognition from Video Using hidden Markov Models by Thad Starner and Alex Pentland AAAI Technical Report FS-96-05.

[21] Knoxville, Sitapa Rujikiet gumjorn July 31,2008.Segmentation Methods for Multiple Body parts. Project in Lieu of thesis. Presented for the Master of Science Degree. The university of Tennessee,