ASL Gestures Recognition From Vertical and Horizontal Histograms

Anamika Choudhary[1],
Nitin Rajvanshi[2]

[1] Sr. Assistant Professor, JIET Group Of Institution, Jodhpur, Rajasthan
[2] Sr. Lecturer, Govt. Women's Polytechnic College, Jodhpur, Rajasthan

ABSTRACT

This article present new set of features and method for recognizing hand gestures for American Sign Language (ASL) Finger-spellings. In this work, our aim was to design a system for ASL gestures' recognition which can be used by wearable computer devices. We have used horizontal and vertical histogram of a binary image for recognizing different shapes of hand gestures. Then we are going to apply Hidden Markov
Model (HMM) for classification of incoming gesture's input data. Also, we need to implement these all in real time, that is, on video input.

Index Terms— ASL Gesture Recognition, Finger-spelling Recognition, horizontal-vertical histogram, HMM.

I. INTRODUCTION

THE RESEARCH in the field of Gesture recognition has number of potential applications in HCI (human computer interface), VR(virtual reality), and in machine control in the industrial field. Most of the work related to gesture interface technique has been recognize as *glove based gesture interface methods* and *vision based method.* Gloves based method require user to wear a cumbersome device, and have lots of sensors and wires to connect the device to computer. In the starting phase (60's - 90's) of research related to gesture recognition lots of research has been perform for gloves based method due to unavailability of vision based method that time.

Most of the work related to vision based method start in 90's. Freeman W. [10] has presented an idea to control television through hand gestures. Thomas Huang and Ying Wu [3] present a very good survey on various vision-based gesture recognition techniques. Meaningful gestures can be classified as static hand postures and temporal gestures. Based on this, many techniques like recognition by modeling dynamics, modeling semantics, HMM framework [11, L 4,5,6], neural network and many other techniques has been proposed.

Human-machine interfaces are playing a role of growing importance as information technology continues to evolve quickly. Keyboards have been replaced by handwriting recognition in Palm, and Pocket PC PDAs. Moreover, some companies have developed some new cell phones to help deaf people communicate with another by hand gesture recognition.

In our work, we focus on designing a system for American Sign Language (ASL) finger-spellings hand gesture recognition to be used by wearable computer devices. For this we are using back-view of user's hand. For this, we are using horizontal and vertical histogram of binary image as feature data.

II. THEORY

1) Sign Language:

A deaf person cannot interpret spoken words, at least not by listening. He or she is highly depended on visual information. Sign languages have been emerged or invented to aid this visual communication. Unlike general gestures, sign languages are highly structured so that it provides an appealing test bed for understanding more general principles. However, there are no clear boundaries between individual signs, recognition of sign languages are still very difficult.

Various variant of sign language are available for experimentation like American Sign Language (ASL), Japanese Sign Language (JSL), British Sign Language (BSL), German Sign Language (GSL) etc.. These Sign languages are different on the basis of their movements, number of hand involved, amount of dynamic-temporal information and amount of facial expressions required. The hands have very important role here, since there are countless combinations of finger poses and orientations.

Various algorithms on static hand posture recognition and temporal gesture recognition are surveyed [3, 8]. HMM and its variants can be used in sign language recognition. Due to the complexity of gesture, machine leaning techniques seems promising in this task.

2) Finger-spelling:

Two approaches are there for sign language, some people use '*alphabets of sign language*' to make words, whereas some use direct word gesture available in sign language. For example word "BABY" can be describe by one dynamic gesture, as well as it can also be describe using four finger-

spelling gestures correspond to 'B','A','B','Y'. Second way of representation called *Finger-Spelling*.

In our work we need to investigate the task of recognizing sign language finger-spelling from video. Finger-spelling is used to spell words for which no sign exists e.g. proper names or technical terms, to spell words for signs that the signer does not know, or to clarify a sign unfamiliar to the 'observer' reading the signer.

3) Skin Color Detection:

Color is one of the most discriminative cues in finding objects and is generally input as RGB information via CCD or video camera. The RGB color system is known to be sensitive to lighting conditions and has a high computing cost, since, RGB color system includes the mixed information of color & intensity. There are two alternatives for detection of skin color objects in an image:

- By using template with skin color of a particular size and compare it with blocks of an image (template matching). But this will need processing of pixel by pixel. This method is also sensitive to brightness level of the image.

- Another method [5], take help of color YIQ color model. In YIQ color model, Y-component is responsible for brightness level. Whereas color values are only located in I and Q- component. By studying YIQ-color model, we have find out that skin color pixel can be find out easily by I-component only. Figure (1) shows our assumption. We covert our image from RGB to YIQ by using following transformation matrix:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.595716 & -0.274453 & -0.321263 \\ 0.211456 & -0.522591 & 0.311135 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0.9563 & 0.6210 \\ 1 & -0.2721 & -0.6474 \\ 1 & -1.1070 & 1.7046 \end{bmatrix} * \begin{bmatrix} Y \\ I \\ Q \end{bmatrix}$$

Where, $R, G, B, Y \in [0,1]$ $I \in [-0.5957, 0.5957]$
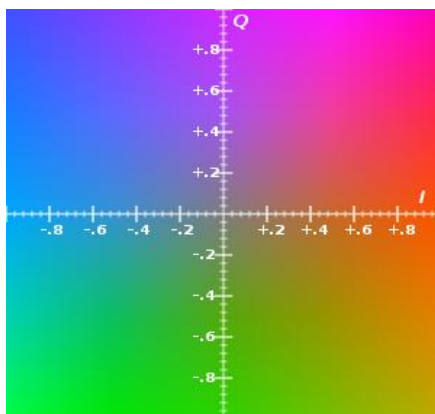and $Q \in [-0.5226, 0.5226]$

Figure 1: YIQ color space at Y=0.5

4) Hidden Markov Model:

HMM model a continuous process as a series of discrete state with state change probabilities. HMM are well suited for the classification of time varying signal that fulfill the Markov property. "*The probability of a state in the future only depends on the current state but not on the past states*".

In describing Hidden Markov Models [7] it is convenient first to consider Markov chains. Markov chains are simply finite-state automata in which each state transition arc has an associated probability value; the probability values of the arcs leaving a single state sum to one. Markov chains impose the restriction on the finite-state automaton that a state can have only one transition arc with a given output; a restriction that makes Markov chains deterministic. A hidden Markov model (HMM) can be considered a generalization of a Markov chain without this Markov-chain restriction. Since HMMs can have more than one arc with the same output symbol, they are nondeterministic, and it is impossible to directly determine the state sequence for a set of inputs simply by looking at the output (hence the "hidden" in "hidden Markov model").

More formally, a HMM is defined as a set of states, one state is the initial state, a set of output symbols, and a set of state transitions. In context of gesture recognition, the state transitions represent the probability that a certain hand position transitions into another; the corresponding output symbol represents a specific posture and sequences of output symbols represent a hand gesture. One then uses a group of HMMs, one for each gesture, and runs a sequence of input data through each HMM.

III. METHODOLOGY ADOPTED

We tried to make a system which is more computational efficient because we are designing our system for wearable computer which have limited resources for computation. In our methodology, we have divided our work in three parts as: image preprocessing, computation of features, and classification of data on the basis of mathematical model.

1) Image Processing Task:

As of now, we are interested in static hand posture analysis, the first task is to extract useful frame from video captured. For this task, we capture information about how much of motion is happening in the video by calculating frame difference and putting certain threshold over it.

In any gesture recognition task first step is always segmentation and tracking of object. Here, our object is hand region. So, in our work, we have taken assumption that only moving *skin color entity* in seen is user's hand and all gesture alphabets need single hand actions only. By these two assumptions, we track out entity i.e. user's hand from seen by applying these following steps:

Step1: Color system conversion from RGB to YIQ.
Step2: Estimation of similarity measure between model and input regions.
Step3: Thresh-holding similarity measures
Step4: Noise removal and dilation.
Step5: Detection of hand candidate region

2) Feature Detection Task:

Once we get the candidate frame and candidate region then we can move on to the task of feature detection. There are many features available for posture shape analysis. Nielson et-al [12] used histogram of Hue, Saturation of skin and Hausdroff distance as features and Visual Memory System (VMS) for matching. Yoon, Yang et al. [5] used combined features of location, angle, velocity for their task. Mark Evergimtom [13] used *HOG* (Histogram of oriented gradient) for detection of British Sign Language Finger-spelling Recognition. Similarly, Gijis Molenaar et. al [14] used HOG for gesture recognition in their project 'SONIC Gesture'. Moment based Size function proposed by Frosini [19] also have been used by researchers [ref] for gesture shape analysis. Also, color glove based technique [Link-2] also seems very attractive. Also, Fourier Descriptor seems good descriptor for static shapes analysis, especially, for shapes having smooth boundaries.

Our method is also doing shape analysis, but, we are using a heuristic approach for it. We are using vertical and horizontal histogram of binary image for shape description. The idea is, first, convert complete image in binary. There will only one skin color object (due to preprocessing task) and some noise

also, but noise are not in much quantity compare to main object (hand). Then we count the number of 1's (presence of hand region) row-wise and column-wise. We call row-wise representation of this as 'horizontal histogram' and column-wise representation as 'vertical histogram'.

Now, we have information about object present in image, contain in two vectors (array). This means, we can represent shape of each gesture alphabet by two vector i.e. vertical vector, horizontal vector. Here size of horizontal array will be equal to number of rows in image; size of vertical array (vector) will be equal to number of column in image.

Now, our problem is to comparing two 1-D arrays, (or, we can say two signals), rather than comparing two 2-D images. Also, two signals may be translated or scaled version of each other. By comparing horizontal histogram of input image to the horizontal information of image shape available in database, we can get knowledge about input image's horizontal information. Similarly, by comparing vertical histogram of input image to the vertical information about image shape available in database, we can get knowledge about vertical information of input image. In signal theory, there are some mathematical tools available to do this thing. Most general mathematical tool is *'Correlation'*.

'Correlation' is measure of similarity between two function or signals. More the correlation, more those functions are similar. We have used cross correlation for our work. The general formula for cross correlation is:

$$R_{x,y} (l) = \sum x(n) * y(n-l);$$

$$Correlation\ Coefficient = max(R_{x,y});$$

3) Classification Task:

Classification is an important part of any recognition task. Generally, we have lots of data and we have to classify this data to the class this data is belonging to. For gesture Detection most of the researchers [3, 5, 6] used HMM (hidden Markov model) due to similarity of gesture task to speech recognition [11]. Some of the researchers use variant of HMM some have used PHMM (partly HMM). Other than HMM some researcher applied Neural Network approach and its variant to the system also.

There are two type of gesture to be recognized, one is isolated gestures, and the other is continuous gesture. The presence of silence makes the boundaries of isolated gesture easy to spot. Each sign can be extracted and presented to the trained HMM system individually. In our work, we are concern only about static gesture yet.

We have made database of static gesture images of each alphabet by different (4 person) person to train our system.

IV. EXPERIMENTAL SETUP

There was no ready database available for Back-View of hand gestures. So, we have made our own database for experimentation. We have taken 4 person's hand gesture images for our database and two persons for our testing purpose. For video database we have taken two users video with different gestures.

As every image processing task, we needed a controlled set up for making our database. Also database should be noise free as much as possible. So, we made our database in natural light with controlled background. We have used Webcam Creative™ with average resolution 480x640 size images. Various finger-spelling alphabets shown in figure:

Figure 2: Finger-spelling alphabet A-I and K-Y (sequentially)

From these alphabets we get YIQ color images then by applying threshold on I-component we extract skin color object these will gave us images like as follow:

Figure 3: I-component after threshold

After this, we calculate vertical and horizontal histogram data about YIQ image of the database for each alphabet. Now, we have whole information about each alphabet's horizontal and vertical histogram. By using this information we found correlation of each alphabet with itself and with other alphabet's horizontal and vertical histogram, respectively.

According to theory, correlation of the signal with itself should be higher than correlation of it with other signals. But, when we perform experiment with the available data it also give good (in other words, higher correlation) with other signals also. This may be due to the histogram data (signals) which we have. For example, we have higher correlation with 'i' and 'a' compare to 'a' with itself. For reference, the kind of vertical and histogram data which we have is shown in figure (3) as given below:
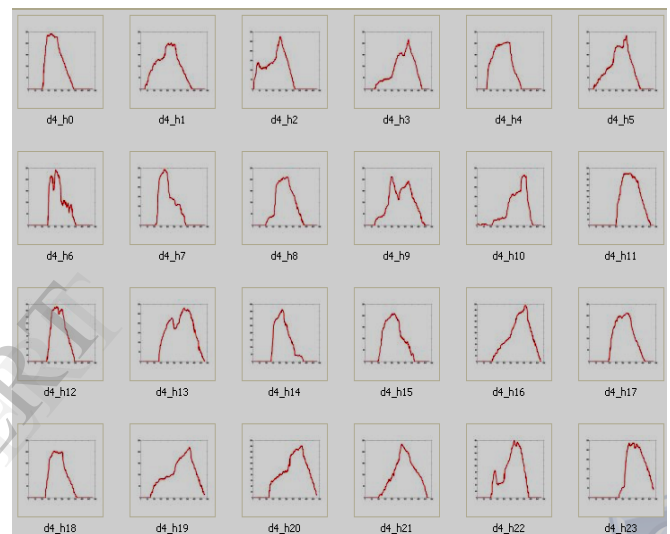

Figure 3: Horizontal histograms for each alphabet in ASL (A-I and K-Y)
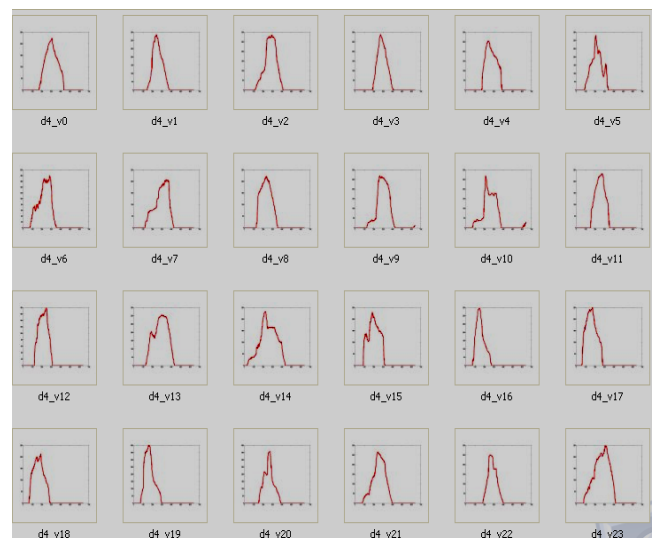

Figure 4: Vertical histograms for each alphabet in ASL (A-I and K-Y)

So, we found that two signals having highest correlation may not be the same signals. Due to this, we have to study the nature of data we have in the database, that is, horizontal and vertical histogram. For reference, table-1 and table-2 shows the correlation data of each database image with other images available for the same person.

V. FUTURE WORK

As, simple correlation is not giving the exact solution to our problem, we need to study the nature of incoming image with each alphabet's image available in database. Also, we can find tangent information of horizontal and vertical histogram.

The part of correcting the finger-spelling using Hidden Markov Model is yet to be done.

VI. BIBLIOGRAPHY

[1] A. Mcandrew, "Introduction to Digital Image Processing with MATLAB", Cengage Learning Publication, 2004.

[2] Gonzalaz, Woods, Eddins, "Digital Image Processing", Pearson Education Press, 2004.

[3] Ying Wu and Thomas S. Huang, "Vision-Based Gesture Recognition: A Review", Gesture-Based Communication in Human-Computer Interaction, Volume 1739 of Springer Lecture Notes in Computer Science, pages 103-115, 1999, ISBN 978-3-540-66935-7

[4] D.S. Zhang, G.J. Lu, "A comparative study on shape retrieval using Fourier descriptors with different shape signatures", Proceedings of the International Conference on Multimedia and Distance Education, Fargo, ND, USA, pp. 1–9, June 2001.

[5] Yoon, Soh, Yang, "Hand gesture recognition using combine feature of location, angel and velocity", Pattern Recognition Papers, page 1491-1501, 2001.

[6] F. Chen, C. L. Huang, Chih-Ming Fu, "hand gesture recognition using a real time tracking Method & hidden markov model", journal of Image and vision computing, page 745-758,volume 21(8), 2003.

[7] A. Starner and T. Pentland, "Visual recognition of America Sign Language using Hidden Markov Models", Int. Conf. on Automatic Face and Gesture Recognition, page 189–194, 1995.

[8] M. A. Moni, A B M Shawkat Ali, "HMM based Hand Gesture Recognition: A Review on Techniques and Approaches", Int. Conf. on Computer Science & Information Tech (ICCSIT) 2009, pp. 433 – 437, 2009

[9] P. Dreuw, D. Keysers, T. Deselaers, and H. Ney. "Gesture Recognition Using Image Comparison Methods". In International Workshop on Gesture in Human-Computer Interaction and Simulation (GW), Lecture Notes in Computer Science, volume 3881, pages 124-128, Ile-de-Berder, France, May 2005.

[10] Weissman C Freeman W. "Television control by hand gestures". International Workshop on Automatic Face and Gesture Recognition. 26-28:179-183, 1995, Switzerland.

[11] L.R. Rabiner, "A toturial on hidden Markov models and selected. applications in speech recognition," Proc. of IEEE, vol.77, no. 2, pp. 257-286, Feb.1989.

[12] Elena S-Nielsen, Luis A-Canalís and, Mario H-Tejera,: Hand Gesture Recognition for Human-Machine Interaction. Journal of WSCG, Vol.12, No.1-3, Plzen, Czech Republic, 2003.

[13] S. Liwicki and M. Everingham. "Automatic recognition of fingerspelled words in British sign language". Proceedings IInd IEEE Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB'09), CVPR2009, pages 50-57, June 2009.

[14] Gijs Molenaar, "Sonic Gesture", master's thesis, University of Amsterdam, October 2010.

[15] P. Frosini. "Measuring Shapes by Size Function". In Proc. of SPIE on Intelligent Robotics and Computer Vision, Boston, Mass., volume 1607, pages 3–26, 1991.

Links:

a. http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/COHEN/gesture_overview.html

b. http://www.physorg.com/news193566126.html

c. http://www.tsi.enst.fr/tsi/enseignement/ressources/mti/descript_fourier

d. http://www.comp.leeds.ac.uk/roger/HiddenMarkovModels/html_dev/main.html

e. http://www.autonlab.org/tutorials/hmm.html

f. http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/