ISSN:0975-9646

Punamchand Patidar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (3) , 2011, 1305-1308

Sign Gesture Recognition using Combined Features of Sum graph and HMM with Decision tree classifier

Punamchand Patidar, Neetesh Gupta

Department of Information Technology,Technocrats Institute of Technology, Bhopal (M.P.), INDIA

Abstract— **This work introduces an Decision tree-based method to classify and recognize sign gestures from the video. A sign language is used by deaf and dumb for communicating. The language is also used in Human Computer Interactions. We obtain the frames from video and extract the motion area relevant to the occurrence of gesture. From this segmented motion area, we extract the skin part which represents the hand part of a frame. Once skin segment is detected, the image is converted into binary image and is filtered by morphological erosion filter. Furthermore co-occurrence of 0's and 1's are calculated and HMM states are derived from these values. Mean and standard deviation of HMM states and sumgraph are considered as features. The features are classified with the help of decision tree classifier. The work is tested with five distinct signs of numerical one to five as used commonly by us to depict the numbers. Results show that the technique is invariant of the size of the hand or the user but are sensitive to light variations due to dependence of the motion detection on light intensity varying. We have observed that system recognizes the gestures with an efficiency of 83.33%.**

***Keywords*— HMM, Sign Gesture and language , Decision Tree classifier, sum Graph.**

I. INTRODUCTION

A sign language (also signed language) is a language which, instead of acoustically conveyed sound patterns, uses visually transmitted sign patterns (manual communication, body language and lip patterns) to convey meaning—simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions to fluidly express a speaker's thoughts. Sign languages commonly develop in deaf communities, which can include interpreters and friends and families of deaf people as well as people who are deaf or hard of hearing themselves [7]. Generally, each spoken language has a sign language counterpart in as much as each linguistic population will contain Deaf members who will generate a sign language. We recognize five common signs that represent numbers from 1 to 5, which is common to all language. The challenge is to detect the signs in the runtime with hand motion rather than detecting from static images. On the whole, deaf sign languages are independent of oral languages and follow their own paths of development. For example, British Sign Language and American Sign Language are quite different and mutually unintelligible, even though the hearing people of Britain and America share the same oral language.

Similarly, countries which use a single oral language throughout may have two or more sign languages; whereas an area that contains more than one oral language might use only one sign language.

II. RELATED WORKS

Basically four steps are involved to recognize hand gesture or sign in HCI system: motion detection (hand tracking), skin filtering (pre-processing), feature extraction and classification (recognition). Researchers are modifying and improving any steps from that and trying to achieve good recognition efficiency. In order for gestures to be segmented and recognized from video streams, various models have been employed to characterize gestures.

Historically researchers have proposed several methods for both data glove and vision based approach. The thumb objective of gesture recognition is to create a system which can recognize specific human gestures and use them to convey information or for device control. Hand gestures provide a separate complementary modality to speech for expressing ones ideas. Information associated with hand gestures in a conversation is degree, discourse structure, spatial and temporal structure [2]. Pragati Garg *et al.* present a review of Vision based Hand Gesture Recognition techniques for human computer interaction, consolidating the various available approaches, pointing out their general advantages and disadvantages [5].

Vaishali S. Kulkarni et al. developed a system for automatic translation of static gestures of alphabets in American Sign Language. In that system three feature extraction methods (Histogram Technique, Hough, OTSU's segmentation algorithm) and neural network is used to recognize signs. The system deals with images with uniform background [6]. Radu Daniel vatavu *et al.* addresses the problem of visual recognition of several hand postures corresponding to a few operations commonly performed in virtual environments, such as: object selection, translation, rotation and resizing [3]. Salvatore Tabbone et al. proposed a method for both grey level and color object retrieval is presented in this paper. Preliminary tests were performed on color photos [1]. Qutaishat Munib et al., 2007; proposed a system for automatic translation of static gestures of alphabets and signs in American sign language. In that authors used Hough transform and neural networks which is trained to recognize signs. Their system does not rely on using any gloves or visual markings to achieve the recognition task. Instead, it deals with images of bare hands, which allows the user to interact with the system in a natural way. The work presented in their project dealt with static signs of ASL only [4].

III. METHODOLOGY

A. *Motion Detection*

The objective of motion detection is to detect the areas in a frame where change in pixel values are significant wrt to previous frames. For our application, rather than considering a compressed video format like MPEG or PAL, we have extracted frames from the video and stored the frames in a specific directory. When images stored in the directory is read sequentially and displayed with proper delay, it symbolizes a video. We assume that each two frames are interdependent on each other, as when pixels from one frame moves or changes, it gives result to a next frame. Motion detection here essentially means separating the background from the image so that only gesture object is considered. In order to do this, we take direct frame difference of two frames, it separates the background from frame2. Thus after background separation, frame will contain only object part. The conceptual block diagram of motion detection shown in figure 1 and 2 respectively.

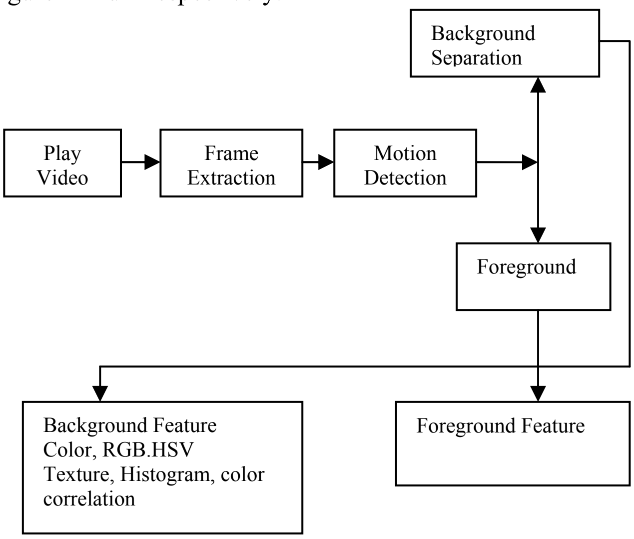


Figure 1: Block Diagram of Motion detection.

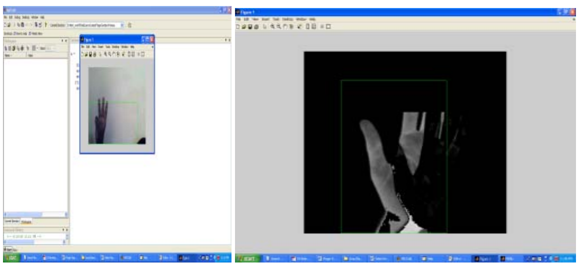


Figure 2: Motion Detection Result

Figure 2 shows the object tracking through motion detection. Once background is separated, the image is left with only the object part.

B. *Skin Filter*

The skin filter is based on the Fleck and Forsyth algorithm [10, 11], which the author proposed as a pre-processing step of finding naked people. Actual color image part is extracted from the current frame over the motion detected area and skin map is extracted from this as shown in figure 3.

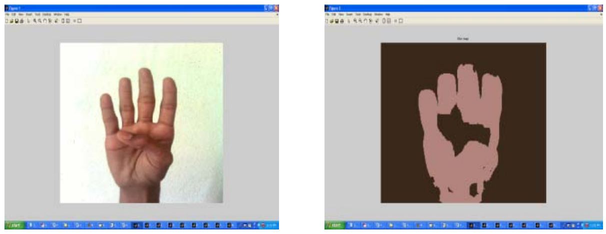


Figure 3: Result of Skin Detection (a) Original RGB image (b) Texture Amplitude Map

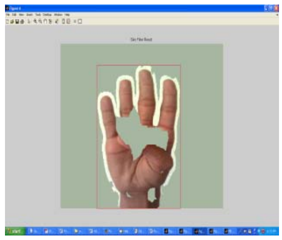


Figure 4: Skin Segmented Video frame.

C. *Feature Extraction*

- Sum Graph or projection histogram is column wise sum of the binary image. The graph gives an approximation of the peaks which are most important in detection of number of fingers extended for gesture (figure 5).
  Let us consider that our Input image be I(i,j);Then the sum graph represents the column wise summation of the binary color values.
  $S(j)= \Sigma jI(i,j) \qquad J=1:N \qquad (1)$

  Where N=number of columns

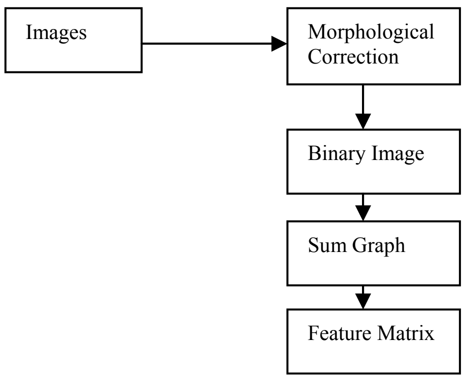


Figure 5: Block Diagram of Sum Graph Feature Extraction

- Calculated co-occurrence of 0's and 1's we determine the probability matrix of appearance of 1 w.r.t 0 and 1 and probability of occurrence of 0 w.r.t 1 and 0.This matrix is use to generate the states of hidden markov model(HMM). Mean and Standard deviation of this state are extracted and are used as feature vector for decision tree classifier (figure 6).

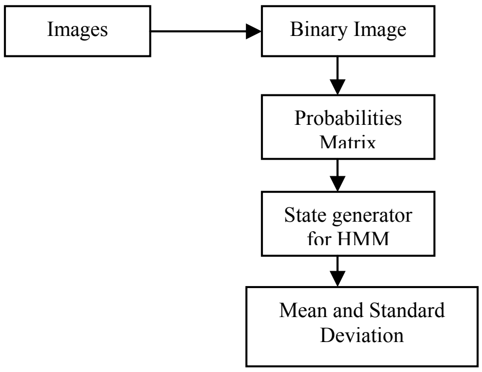


Figure 6: Feature Vector using HMM

D. Classification

The decision tree classifier is one of the possible approaches to multistage decision making; table look-up rules, decision table conversion to optimal decision trees, and sequential approaches are others. The basic idea involved in any multistage approach is to break up a complex decision into a union of several simpler decisions, hoping the final solution obtained this way would resemble the intended desired solution. A complete review of *multistage* recognition schemes is given by Dattatreya and Kanal [9]. Some of the differences between the different multistage schemes are addressed in Kulkarni & Kanal [8].

We have combined feature vectors like sum graph with HMM model which gives a very good feature set. We have used decision tree classifier which gives a very good recognized result.

The basic steps of methodology adopted here is as follows.

1. Background separation and area tracking: First we obtained continues frames from video and stored in to specific directory and we converted all RGB frames (images) to Gray scale with size of 240*320 and pixels have value less than 30(black) we considered it 255(white) and after that we converted some boundaries areas in to white and subtracted current frame with respect to others.

   Once background separated we convert image to binary and used imdilate function in MATLAB to fill the gaps in an image .We calculated top and bottom point as diagonally now we got the rectangle where moment occurred.
2. Select the motion detected hand part image. N numbers of such images are pre-categorized into five groups for training.
3. The binary converted image is used for extracting the sum graph and mean and standard deviation of HMM states are considered as features.
4. Decision tree classifies the sum graph and mean and standard deviation of HMM state combined features. Those that are classified are selected as correct detection and rest are rejected.

IV. RESULTS

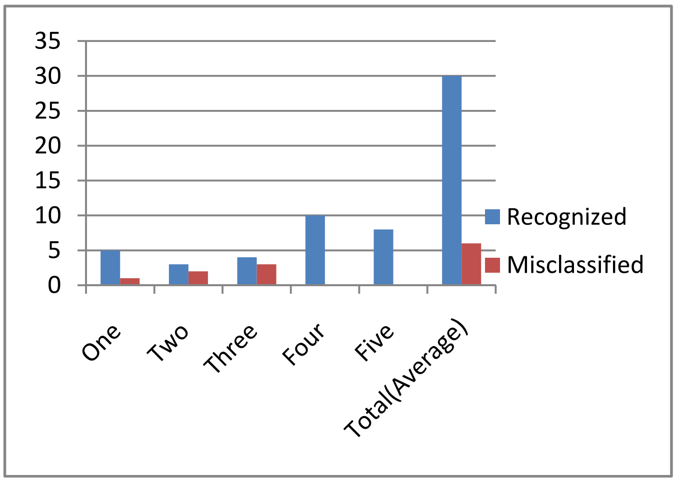
Following Table 1 shows that system recognizes the gestures with an overall accuracy of 83.33%.

Gesture/Sign	**Recognized Samples**	**Misclassified Samples**	**Recognition rate (%)**
One	5	1	83.33%
Two	3	2	60%
Three	4	3	57.14%
Four	10	0	100%
Five	8	0	100%
Total(Average)	30	06	**83.33%**

Table 1: Recognition Rate of Proposed System

$$\text{Recognition Rate} = \frac{No.of\ Correctly\ Classified\ Samples}{Total\ No.of\ Samples} \times 100\ \%$$

Following graph shows description of tabular forms.



V. CONCLUSION AND FUTURE WORKS

In this work we recognized sign language from hand gestures from video frames. The work is modelled in such a way that it is able to recognize number of fingers that a user is showing. Therefore in the work, we have considered mainly 5 gestures. We have combined feature vectors like sum graph with HMM which gives a very good feature set. Further we have classified with decision tree classifier which produces a very good recognition result. We have observed that system recognizes the gestures with an overall accuracy of 83.33%. The misdetection mainly is caused by the improper segmentation due to light variance in motion frames. The work can be further improved by Zernike moment and Thin plate Spline shape descriptor as features which will minimize the misdetection rate.

REFERENCES

[1] Salvatore Tabbone, Laurent Wendling. "Color and grey level object retrieval using a 3D representation of force histogram", *Image and Vision Computing*, pp. 483–495, 2003.

[2] Harshith.C, Karthik.R.Shastry, Manoj Ravindran, M.V.V.N.S Srikanth, Naveen Lakshmikhanth "Survey on various gesture recognition techniques for interfacing machines based on ambient intelligence" International Journal of Computer Science & Engineering Survey (IJCSES) Vol.1, No.2, November 2010.

[3] Radu Daniel Vatavu, Christophe Chaillou. "Visual recognition of hand postures for interacting with virtual Environments"."Stefan cel Mare" University of Suceava.

[4] Qutaishat Munib, Moussa Habeeb, Bayan Takruri, Hiba Abed Al-Malik. "American Sign Language (ASL) recognition based on Hough transform and neural networks" Expert Systems with Applications 32 (2007) 24–37.

[5] Pragati Garg, Naveen Aggarwal and Sanjeev Sofat."Vision Based Hand Gesture Recognition" World Academy of Science, Engineering and Technology 49 2009

[6] Vaishali S. Kulkarni , Dr. S.D.Lokhande. "Appearance Based Recognition of American Sign Language Using Gesture Segmentation" International Journal on Computer Science and Engineering Vol. 02, No. 03, 2010, 560-565.

[7] Keith Brown and Keith Allan *Oxford Handbook of the History of Linguistics (090621).*

[8] S. Rasoul Safavian and David Landgrebe. "A Survey of Decision Tree Classifier Methodology". School of Electrical Engineering Purdue University, West Lafayette, IN 47907.

[9] Laveen N. Kanal. "On pattern, categories, and alternate realities". *11th International Conference on Pattern Recognition,* August 31 - September 3, 1992.

[10] Fleck, M., Forsyth, D., and Bregler, C., "Finding Naked People". *In Proc. of the ECCV, Vol. 2, 592-602, 1996.*

[11] David Forsyth and Margaret Fleck (1999) "Automatic Detection of Human Nudes". *International Journal of Computer Vision 32/1, pp. 63-77.*

[12] Oi Mean Foong, Tan Jung Low, and Satrio Wibowo "Hand Gesture Recognition: Sign to Voice System(S2V)" *World Academy of Science, Engineering and Technology 42 2008.*