Vision-Based Approach for American Sign Language Recognition Using Edge Orientation Histogram

Jayshree R. Pansare
Computer Engineering
Modern Education Society's College of Engineering
Pune, India (Research Scholar, DAVV, Indore, India)
e-mail: jayshree.pansare23@gmail.com

Maya Ingle
School of Computer Science & IT
Devi Ahilya Vishwavidyalaya
Indore, India

***Abstract*—Hand Gesture Recognition System (HGRS) for detection of American Sign Language (ASL) alphabets has become essential tool for specific end users (i.e. hearing and speech impaired) to interact with general users via computer system. ASL has been proved to be a powerful and conventional augmentative communication tool especially for specific users. ASL consists of 26 primary letters, of which 5 are vowels and 21 are consonants. Proposed Real-time static Alphabet American Sign Language Recognizer- (A-ASLR) is designed for the recognition of ASL alphabets into their translated version in text (i.e. A to Z). The architecture of A-ASLR system is fragmented into six consequent phases namely; image capturing, image pre-processing, region extraction, feature extraction, feature matching and pattern recognition. We have used Edge Orientation Histogram (EOH) in A-ASLR system. The system is developed for detection of ASL alphabets based on *Vision-based* approach. It works without using colored gloves or expensive sensory gloves on hand. Our A-ASLR system achieves the recognition rate of 88.26% within recognition time of 0.5 second in complex background with mixed lightning condition.**

Keywords-vision-based approach; Edge Orientation Histogram (EOH); American Sign Language (ASL); static hand gesture; Sim-EOH algorithm

I. INTRODUCTION

Hand Gesture Recognition System (HGRS) for detection of American Sign Language (ASL) has become essential and powerful communication tool for specific users (i.e. hearing and speech impaired) to interact with general users via computer system. Numerous HGRS have been developed for identification of diversified sign languages using effective techniques. There exist two main approaches in the hand gesture analysis namely; vision-based and device-based approach. In vision-based approach the user does not require to wear any extraneous mechanism on hand. Instead the system requires only camera(s), which are used to capture the images of hand gesture symbol for interaction between human and computers. However, conventional device-based approach for HGRS has been designed with external devices such as data gloves, markers, sensory gloves, colored gloves etc. Among the above-mentioned approaches, vision-based approach is the most natural way of constructing HGRS. Moreover, these techniques focus on interpreting the action of hands in different modes of interaction relying upon the application domain. Conversely, applications developed using vision-based approach usually suffer from disadvantages associated with latency, occlusion, or lighting as well [1]. Edge Orientation Histogram (EOH) based hash codes in HGRS is applied on a 3D-address space bounded by hamming distance. System proves 82.1% accuracy against 1000 images comprising of 10 distinct static hand gestures sets [2]. Haar-like and Histogram of Oriented Gradients (HOG) features are applied in HGRS to solve major Haar-like problems with detection rate of 83% [3]. Fuzzy Logic membership approach and contour edge detection technique in HGRS are employed in automatic HGRS of 'Bharatanatyam' dance achieves recognition rate of 85.1% along with time complexity of 2.563 sec. [4]. Real-time HGRS has been developed for recognition of 26 static hand gestures related to A-Z alphabets using centroid of Binary Linked Object (BLOB) method. The system reliably identifies single-hand gestures in real-time and achieves detection rate of 90.19% in complex background [5]. Centroid, DCT, Fourier Transform (FT) and Edge Orientation Histogram (EOH) techniques are used in static HGRS which detects ten DSL digits. With the hand gesture imposed at a distance of 15 inches from a camera using FT technique obtains a success rate of 85% [6]. Static HGRS has been designed to recognize 10 ASL numbers from 0 to 9 using histogram attains success rate of 81.1% with recognition time of 0.5 second [7]. Comprehensive performance study of existing techniques in HGRSs for sign languages have been reviewed for all phases of system architecture [8]. Literature reveals that, various gesture tracking and recognition techniques play a vital role to enhance performance of HGRS systems in real-time progressively. However, there remains an extensive scope to design and develop HGRS for multifold uses. This system will be highly useful to understand the sign postures posed by specific end users to the general users via computer system. On the other hand, this system will act as the learning resource for such young users to learn ASL alphabets also.

With the potential use of technology, it has become easy to design and develop the system as the requirement of a variety of users including specific end users. In this work, we emphasize development of A-ASLR system for recognition of 26 ASL alphabets using Edge Orientation Histogram (EOH) in complex background along with mixed light.

Standard ASL symbol set (comprised of 5 vowels and 21 consonants) is as depicted in Fig. 1 and ASL symbol set used in A-ASLR system is as shown in Fig. 2. Specifically, A-ASLR is exploited for recognition of static hand gestures only. Section 2 entails existing techniques in HGRS for sign languages. We discuss proposed A-ASLR system along with system architecture and experimental results for A-ASLR system in Section 3. In Section 4, performance evaluation of A-ASLR with existing HGRSs is presented. Lastly, we conclude with conclusion in Section 5.

Figure 1. Standard ASL symbols

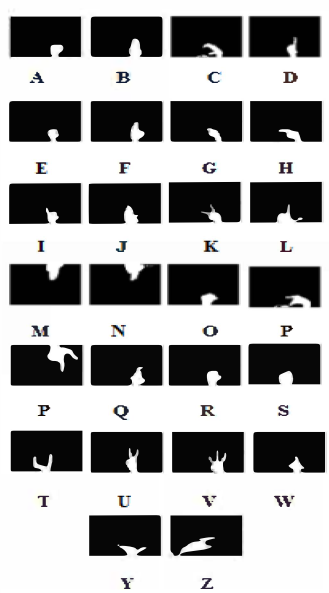

Figure 2. ASL symbol set used in A-ASLR system

II. EXISTING TECNINIQUES IN HGRS FOR SIGN LANGUAGES

Several techniques are applied in HGRS system for recognition of numerous sign languages. The hand gesture recognition exhibits considerable levels of uncertainty casts the processing computationally complex or error prone. There exist various techniques for recognizing hand gesture (s) after some image processing operations that are described as follows:

A. Skin Color Glove Based Technique

Hand detection and tracking method are applied in skin colour filtering for skin segmentation. Detected hand image is converted into black and white image in which white pixels represent the hand portion and remaining part is represented by black pixels. The technique may not differentiate hand from other body objects. The method is fast in execution whereas it is unreliable in complex background with diverse lightening conditions. The input image must contain only hand as an object [9].

B. Feature Based Technique

Haar features and skin colour segmentation are employed for hand gesture recognition. Haar-like features are used to recognize face. Additionally, some special hand features are added to Haar features. The extended features and skin colour segmentation are practiced in hand detection and tracking. The method works in real-time and reliable in complex background. It has been observed that the hand gesture detection results vary in facial background of hand image [10], [11].

C. Machine Learning Based Technique

Support Vector Machine (SVM) classifier is trained by example data points and these classified data points are placed into respective categories. SVM training algorithm is developed to produce a model that predicts about the category input data [12]. The technique is reliable in complex background with different lightening conditions. Thus, the artificial neural networks have become gradually intelligent and execute in real-time only for small set of trained inputs [13].

D. Fingertip Recognition Technique

In fingertip recognition technique the input hand image is filtered by the use of two-dimensional circular filter that smoothen the corners of a binary image. It extracts convex hull points from filtered hand image. For a given non-empty set of points in a plane, the convex hull is the smallest convex polygon which covers all the points in the given set. The contour sets of filtered hand image of an object contain boundary (edge) points [14].

III. PROPOSED ALPHABET AMERICAN SIGN LANGUAGE RECOGNIZER SYSTEM

This recognizer is designed for the recognition of ASL symbols [15]-[21] into their translated version of ASL alphabets. Its system architecture and experimental results are described in this section along with our proposal that is highly concerned with EOH technique used in A-ASLR system subsequently.

A. System Architecture for A-ASLR System

The architecture of A-ASLR system is fragmented into six consequent phases namely; image capturing, image pre-processing, region extraction, feature extraction, feature matching and pattern recognition. System architecture of A-ASLR system is as depicted in Fig. 3.

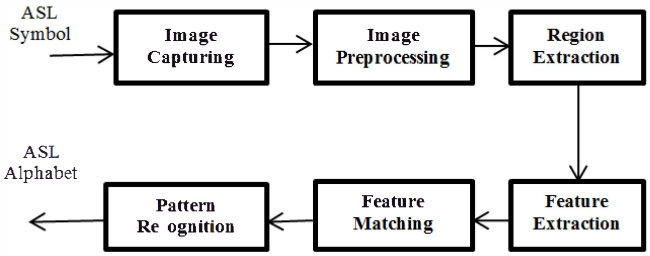

Figure 3. System architecture of A-ASLR system

Image capturing phase deals with capturing ASL symbol image in RGB color space with frame size 160x120 using web camera attached to the laptop. Single frame of image is captured in complex background along with mixed lighting condition and the distance of 1 meter from web camera. Image preprocessing includes preprocessing operations such as color conversion, noise removal and morphological operations. Skin detector is applied for extracting only skin-pixels and ignoring non-skin pixels from input RGB image. Gray threshold is applied on skin pixels with specific probability. Filtering of Gray image is done using Median Filter for preserving edges followed by Gaussian filter is applied on current image for smoothing of image. Morphological operation such as blur, eroaion and dialation are applied on filtered binary image. In region extraction, BLOB of size 80x60 is employed to extract Region of Interest (ROI) of hand in A-ASLR system. EOH technique uses coefficient of EOH to form feature vector. The noise free image obtained from the previous sub-phase (i.e. morphological operations) acts as input in this phase. The input image contain sub-images as regions and these regions are labeled with the use of eight pixel connectivity. We have calculated the area of each labeled region as area set and selected largest area from area set. This selected largest region is set be true value (white pixel) and other remaining regions are set be false value (black pixel). At this stage, image contains only a biggest region, the region as biggest BLOB and this BLOB represents only hand in human image. Feature extraction phase is based on extraction of EOH coefficient and formation of feature vector using EOH coefficients for both running image and training dataset images. EOH technique *is based on EOH descriptor model that measures the similarity between EOH of r*unning image and training dataset images of ASL alphabet. *The CompareHist descriptor is used to evaluate the best suited pattern as ASL alphabet as*

$$C_i = \max_{j \in \alpha} \mathrm{CompareHist}\left(R, X_j\right)$$

In this model, $\mathrm{CompareHist}(R, X_j)$ computes the distance between EOH feature vector of running image and training dataset images, α represents the ASL symbol database for ASL alphabets, R is a running image of ASL alphabet, X_i (j=1, …, N) denotes the training dataset image EOH for ASL alphabets. Feature matching phase matches feature vector of training dataset images and running image using CompareHist (Training Image, Running Image) function using *Sim-EOH algorithm.*. Pattern recognition is comprised of finding similarity between feature vectors of training dataset images and running image thereby using *K-Cluster-EOH-Match algorithm.* Maximum similarity endow with appropriate ASL alphabet (i.e. text)as recognized pattern.

B. *Experimental Results*

Experimental results of our A-ASLR system are as represented in Table I. System executes in complex background along with natural lighting condition and achieves recognition rate of 88.26% with recognition time of 0.5 second.

TABLE I. EXPERIMENTAL RESULTS OF A-ASLR SYSTEM

ASL Symbol	Recognition Rate (%)	ASL Symbol	Recognition Rate (%)
A	95	N	80
B	90	O	80
C	90	P	90
D	90	Q	80
E	90	R	80
F	90	S	90
G	90	T	90
H	90	U	90
I	90	V	90
J	90	W	90
K	90	X	80
L	90	Y	90
M	90	Z	90

IV. PERFORMANCE EVALUTION OF A-ASLR SYSTEM

In this section, the performance of A-ASLR in static background is evaluated for 26 hand gestures. Training dataset comprised of 2600 samples using natural lighting condition in static background. A-ASLR system achieves recognition rate of 88.26% with the recognition time about 0.5 seconds using 10 Mega Pixels web camera.

Various HGRS are developed for recognition of ASL and now present the evaluation of performance of A-ASLR system for detecting 26 ASL symbols in complex background. This analysis has been performed on the basis of some vital factors such as color space, frame size, web camera resolution, and distance from camera, number of training dataset images, background and lighting condition and is depicted in Table II. The following are the important observations associated with the performance evaluation of A-ASLR system for identification of ASL:

A-ASLR system is developed using EOH achieves highest recognition rate i.e. 88.26% with natural lighting condition in complex background. The training dataset comprised of 2600 samples encapsulated in 26 clusters.

- On the other hand, when histogram are used in real-time HGRS attains the recognition rate of 87.82% along with mixed lighting condition in static background.

- Numeric HGRS is designed using histogram and skin color segmentation obtains recognition rate viz. 85% including mixed lighting condition in complex background.
- Statistical HGRS based on EOH achieves recognition rate of 82.1% with mixed lighting condition in complex background.
- In contrast, EOH utilized in DSL Translator received the lowest recognition rate i.e. 52.85 % in cluttered background.
- It is evident from Table II that recognition rate achieved in A-ASLR system shows best outcome as compared to other existing HGRS for identification of sign languages.

V. CONCLUSION

A-ASLR developed for recognition of 26 ASL alphabets using static hand gesture along with natural lightning conditions in complex background. It is observed that the results are more accurate with low-cost 10 mega pixel Frontech E-cam. Fixed position of Web camera attached to laptop captures RGB color space snapshot from distance of 1 meter from camera. Training dataset of 2600 samples is constructed thereby collecting 100 samples for 26 ASL symbols. Low-level feature extraction from extracted region reduces computation and works efficiently while matching feature vectors in real-time. A-ASLR system achieves recognition rate of 88.26% using EOH and displays suitable ASL alphabet along with word.

TABLE II. PERFORMANCE EVALUAION OF A-ASLR SYSTEM WITH EXISTING HGRSS

HGRS for DSL Recognition	Number of Gestures	Recognition Rate	Color Space	Frame Size	Web Camera Resolution (Mega Pixel)	Distance from Camera	Number of Training Set Images	Background	Lighting Condition
Real-time HGRS	46	87.82%	RGB	160*120	8	25 cm	1040	Static	Mixed
HGRS	10	85%	RGB, YIQ	160*120	16, 1.3	0.5 m	20	Complex	Mixed
Statistical HGRS	10	82.1%	RGB	640x320	Variant	Variant	1000	Complex	Mixed
DSL Translator	10	52.85%	RGB	320x240	8	1 meter	1000	Cluttered	Variant
A-ASLR System	**26**	**88.26%**	**RGB**	**160*120**	**10**	**1 meter**	**2600**	**Complex**	**Mixed**

REFERENCES

[1] X. Zabulis, *et al.*, "Vision-based hand gesture recognition for human-computer Interaction," pp. 1–56.

[2] S. U. Ahmad, "Real time rotation invariant static hand gesture recognition using an orientation based hash code," IEEE, 2013.

[3] S. Saha, L., *et al.*, "Fuzzy L membership function based hand gesture recognition for Bharatanatyam dance," 5th Int. Conf. Comput. Intell. Commun. Networks, pp. 331–335, 2013.

[4] S. Ghafouri and H. Seyedarabi, "Hybrid method for hand gesture recognition based on combination of Haar-like and HOG features," IEEE, pp. 1–4, 2013.

[5] J. R. Pansare and M. Ingle, "Real-time static hand gesture recognition for American Sign Language (ASL) in complex background," Journal of Signal and Information Processing, vol. 3, pp. 364-367, 2012.

[6] J. R. Pansare, *et al.*, "Real-time static hand gesture recognition system in complex background that uses number system of Indian Sign Language," Int. J. Adv. Res. Comput. Eng. & Technology, vol. 2, no. 3, pp. 1086–1090, 2013.

[7] J. R. Pansare, *et al.*, "Real-time static Devnagri Sign Language translation using histogram," International Journal of Advanced Research in Computer Engineering & Technology, vol. 2, no. 4, pp. 1-5, 2013.

[8] J. R. Pansare and M. Ingle, "Comprehensive Performance Study of Existing Techniques in Hand Gesture Recognition System for Sign Languages," Int. J. Comput. Sci. Inf. Technol., vol. 7, no. 3, pp. 1343–1347, 2016.

[9] Y. Fang, et al., "A rel-time hand gesture recognition method", in Proceedings International Conference on Multimedia Expo., 2007, pp. 995-998.

[10] L. Yun and Z. Peng, "An automated hand gesture recognition system based on Viola-Jones method and SVMs", in Proceedings 2nd International Workshop on Computer Science and Engineering, 2009, pp. 72-76.

[11] Q. Chen, et al., "Real-time vision-based hand gesture recognition using Haar-like features, 2007", in Proceedings IEEE IMTC, 2007, pp. 1-7.

[12] Y. Ren and C. Gu, "Real-time hand gesture recognition based on vision", in Proceedings Edutainment, 2010, pp. 468-475.

[13] S. Marcel, "Hand posture recognition in a body-face centered space", in Proceedings Conference on Human Factors Computing System (CHI), 1999, pp. 302-303.

[14] S. Marcel, "Hand posture recognition in a body-face centered space", in Proceedings conference on Human Factors Computing System (CHI), 1999, pp. 302-303.

[15] A. Ghotkar, "Vision based multi-feature hand gesture recognition for Indian Sign Language manual," Int. J. Smart Sens. Intell. Syst., vol. 9, no. 1, pp. 124–147, 2016.

[16] G. Tofighi, *et al.*, "Hand pointing detection using live histogram template of forehead skin," in Proceedings 19th Int. Conf. Digit. Signal Process., 2014, pp. 383–388.

[17] S. Shiravandi and M. Rahmati, "Hand gestures recognition using dynamic Bayesian networks," IEEE, pp. 1–6, 2013.

[18] Y. Yao, *et al.*, “Contour model-based hand-gesture recognition using the Kinect sensor,” IEEE, vol. 24, no. 11, pp. 1935–1944, 2014.

[19] H. Hikawa and K. Kaida, “Novel FPGA implementation of hand sign,” IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 1, pp. 153–166, 2015.

[20] A. Kanwar, “An appearance based approach for gait identification using infrared imaging,” IEEE, pp. 719–724, 2014.

[21] Y. Fang, et al., “A rel-time hand gesture recognition method”, in Proceedings International Conference on Multimedia Expo., 2007, pp. 995-998.