American Sign Language Recognition System Using Image Processing Method

Amit Kumar Gautam
Department of Electronics and Communication Engineering,
Delhi Technological University Delhi, India
amitgautam.cicdu@gmail.com

Ajay Kaushik
Department of Computer Science and Engineering,
Kurukshetra university, Haryana, India
ajaykaushik777@gmail.com

***Abstract*—The paper aims to propose a novel technique that recognizes finger spelled American Sign Language (ASL) gestures. The external characteristic of hand, i.e. shape based algorithm is being used for recognition. Since almost all of the alphabets have a unique shape, each alphabet is characterized on the basis landmark points marked on the boundary of the hand shown by the signer. A training set is made by training several images of each alphabet and the landmark points of each alphabet which produces a 72 point descriptor are stored in database. The descriptor of test image is then matched with the ones in database. Finally, Euclidean distance classifies the test images to the recognized alphabet. By increasing the number of sampling points there was an increase in accuracy rate. Thus a 180 point descriptor results in better recognition.**

***Index Terms* —** Fingerspelling ASL, external characteristics, landmark, training set, descriptor, Euclidean distance.

I. INTRODUCTION

Apart from facilitating the process of communication for less privileged people, automatic Sign Language Recognition system has various applications in computer vision and entertainment industry also. Existing computer systems can be made more efficient by providing it input in the form of gestures. An average computer user types 33 words per minute when transcribing and a mere 19 when composing. Through extensive training and practice a sign language speaker can speak about 200-220 words per minute. If the average sign speaker can communicate using finger spelling at least ¼ the speed of a regular sign language, they can sign 50 words per minute. Thus if they could sign into a computer, this would be a significant speedup in computer input. Also, there are many gaming technologies that use gesture as an input and this research can be useful for that.

Sign language recognition was developed in the ‘90s. Research related to hand gesture can be classified into two parts. In the first part, electromagnetic gloves and sensors are introduced which consist the hand shape, movements and orientation of the hand. These have limitation such as cost and not suitable for practical use. Second one is computer vision based gesture recognition system, consist of image processing techniques.

Many researchers have been done based on hand gesture recognition using image processing e.g American Sign Language (ASL) recognition system was introduced in [1] in which HSV color model is used to detect hand shape using skin color and edge detection. In [2] HCI system for recognizing hand gesture and faces from a video camera are presented. In this method head position and hand gesture is combined to control equipment. Head position is identified using eyes position, mouth and face center. Automatic gesture area segmentation and orientation normalization of the hand gesture have recognition rate 93.6%. In [3], combination of edge detection algorithm and skin detection algorithm were introduced using MATLAB. For detecting points, the Canny edge detection algorithm is used at which image brightness changes sharply. Gesture identification is done using ANN algorithm for fast computation. Static hand gesture recognition includes three algorithms named K curvature, Convexity defect and Part based hand gesture recognition was presented in [4]. In this microsoft's Kinect camera is used to capture pseudo-3D image, easily segment the input image and track the image in 3D space. It limits with cost of camera. In [5], a static and dynamic hand gesture recognition system was developed in depth data using dynamic time warping. A directional search algorithm allowed for entire hand contour, the K curvature algorithm was employed to locate fingertips over that contour. Identification of Bengali Sign Language for 46 hand gestures was presented in [6]. ANN was trained by feature vectors of the fingertip finder algorithm. A database of 2300 images of Bengali signs was constructed. The experiment showed an accuracy of 88.69%.

In [7], Glove named “Velostat” made by conductive material equipped with microcomputer and Bluetooth module is developed for finger-gesture recognition. In [8], a wireless sensor glove is designed for ASL finger spelling gesture recognition consist of five flex sensors on fingers and 3D accelerometer on the back of the hand gives 80 % accuracy. M.S. Sinith et. al. [9] proposed Support Vector Machine (SVM) for HSL recognition with 90 % but limitation occurs to separate the W and O alphabet. In [10], the feature extraction is done by using Principle Component Analysis (PCA) and Fuzzy C-Means (FCM) for classification. In [11] LAB View for recognizing alphabet A-Z is used for feature extraction to calculated centroid of each end of fingers. In [12], a system to translate A-Z sign language is developed. To compute the key point of each posture, SIFT algorithm has been used. Further 3D technique has been used for sign language translation. In [13], 3D shape and motion trajectory of hand from 2D video sequences of ASL words has been recovered. Myoung-Kyu Sohn et.al [14] presented 3D hand motion trajectory technique with depth camera, normalized for translation invariant feature extraction and K-NN classification. Hernandez-Rebollar et.al [15] developed method and apparatus for translating hand gestures though which a sign language can be recognized and hand gestures can be translated to speech or written text. Various sensors are used in their apparatus like hand, arm and shoulder that measures dynamic and static gestures. The da obtained by these sensors is processed by microprocessor to obtain insights and accurate results. A database of gestures is maintained which is used by microprocessor to create output signal. The output signal is then processed to create artificial sound. Accelerometers are mounted on the fingers and thumbs, and on back hand gyroscopes are mounted to detect any hand motion and forearm rotation. This data is transmitted to the microprocessor to search for any orientation and shape information of the arm related to user. In [16], An American sign language translating glove using flex sensor is presented. In this paper flex sensors are used to collect sensory data for the recognition of the American Sign Language. Total number of alphabets in sign language is 26 and gloves are implemented in such a way that it can detect all these different signs. Y. Chuang, L. Chen, and G. Chen [17] worked on Saliency-guided improvement for hand posture detection and recognition. The proposed method is based on machine vision. It is usually a simple camera can dispel system's need and the user has no need to use other Accessories. Recognition rate with appropriate speed can be increased by using this method. A database is maintained which include factors such as variation in scale, place and rotation of objects should be observed. Image size is reduced to increase the speed and accuracy. This stage is named as feature extraction. Q. Munib et. Al [18] presented a method for recognizing 20 signs in American sign language. Image size is reduced to a fix amount and then colored images transformed to grey scale images. canny edge detector is used to detect edges of the image. The resulted image is then processed by Hough transform and features are extracted using this transform. The extracted features are then further used to Train Neural Network (NN). The recognition rate for training images and test images is 99.85% and 86.15% respectively. The total recognition rate is 93% Limitation of this method is that for processing only edges of shapes are used due to which most of the information will disappear.

In [19], H.-D. Yang and S.-W. Lee proposed Robust sign language recognition by combining manual and non-manual features based on conditional random field and support vector machine. In this paper, a system recognizes the 13 hand signs in the American sign language. In this system both Manual as well as non-manual features which are based on CRF and SVM, accuracy of 84.1% is obtained. The features which are extracted from people's face and by use of AMM method are non-manual features. The system presented in this paper is not online. R. Zhou, Y. Junsong, M. Jingjing, and Z. Zhengyou have worked on Robust Part-Based Hand Gesture Recognition Using Kinect Sensor [20]. They have presented a system which can recognize the 13 hand signs with Kinect sensor. For distance measuring in this paper they have proposes a method "FEMD". A separate database is maintained for the images which includes 1000 images. They have not mentioned the number of images which were used at the time of system training. The limitation of the proposed system is that to separate palms' area and forearm user should give a black belt. In [21], A new 2D static hand gesture color image dataset for ASL gestures is introduced. In this paper, a new type of database is used for the recognition of sign language. This dataset includes 36 signs in which 26 of them are based on the alphabets of American sign language and 10 of them include numbers from 0-9. The total number of images in this dataset is 2520 images which include 70 images for each sign and they prepared from 5 different persons to get a better accuracy rate. The recognition rate for training images and test images is 100% and 99.56% respectively. The total recognition rate is 99.8%.

A. Sign language

Sign language is a language which uses gestures instead of sound to convey meaning combining hand-shapes, orientation and movement of the hands, arms or body, facial expressions and lip-patterns. Contrary to popular belief, sign language is not international. As with spoken languages, these vary from region to region. They are not completely based on the spoken language in the country of origin. Sign language is a visual language and consists of 3 major components: finger-spelling: used to spell words letter by letter, word level sign vocabulary: used for the majority of communication, non-manual features: facial expressions and tongue, mouth and body position.

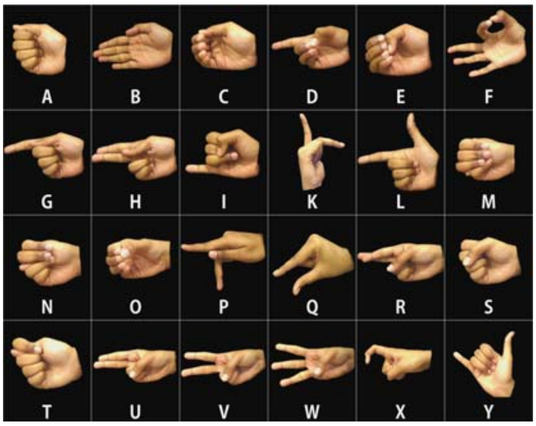


Figure 1. Finger spelling ASL of 24 alphabets.

B. Approach

Our Sign Language Recognition system comprises of 1280x720 resolution HD webcam of laptop to take images of gesture and MATLAB 2009 to process the input and display the output. Several images of each sign were captured except 'J' and 'Z' since both of them were not static signs and required multi frame processing for their identification. Each image was taken with black or dark background for better detection of hand. The external characteristic of the hand i.e. boundary information was used to calculate a unique descriptor for each sign, which was invariant to position and size of hand. The descriptor for each sign constituted the database with which descriptor of each new sign was compared.

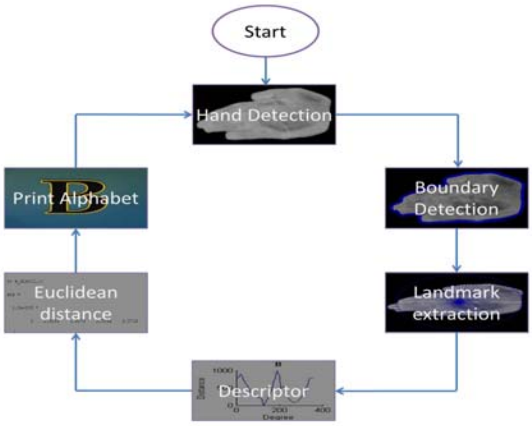


Figure 2. Block diagram of proposed method.

II. METHODOLOGY

A. Hand detection and landmark extraction

From the captured RGB color space domain image, HSV cylindrical coordinates image is obtained. In RGB format, image has three planes, one for each primary color i.e. red, green and blue. In HSV domain, hue (H) specifies color, saturation (S) specifies colorfulness of a stimulus relative to its own brightness and value (V) specifies the intensity.

Skin color was easily detected by specifying thresholds range in hue frame. For distinguishing hand from elbow, a wrist band was worn by the subjects, so that hand can be segregated from other part of body.

Figure 3. Hand with wrist band.

After the hand detection, boundaries were detected and centroid was calculated by considering all the boundary points. From this centroid point, distance is calculated with boundary points at certain angle interval. For creating 72 point descriptor, angle interval is taken as 5 degree and for 180 point descriptor, angle interval as 2 degree.

B. 72 point descriptor

A 72 point vector is constructed which acts as a descriptor for that alphabet. The 72 point vector is obtained by calculating the distance of line joining a point on the boundary of the hand to its centroid. This calculated 72 point vector is then divided by the minimum value in that vector. By doing this, vector now contains the ratios and not the distances in pixels. This made this algorithm translation invariant. The distances are calculated at an interval of 5 degree which justifies the size of the vector i.e. 72 coordinates. For training, this descriptor is made for each alphabet and thus resulting in a 24x72 dimension matrix database i.e. one descriptor for one alphabet. The figure shown below shows a plot of the database.

By observing this plot it can be assured that each descriptor is unique from other and thus this method can help us in distinguishing alphabets on the basis of their shapes. This stored database i.e. matrix is used for matching with the test image.

C. 180 point descriptor

After several tests and trials it was observed that certain alphabets were not easily distinguishable from few others due to their very similar shape. Therefore, rather than finding the best match, the top five matches were calculated and the sign along with these top five matches were sent to another phase of recognition. This new phase was different from previous method by the fact that now instead of 5 degree interval for calculation distance, 2 degree was adopted. This helped in better precision and therefore more chances of accuracy. Thus, now another database was developed which was a matrix of 24 rows and 180 columns (since the each point is 2 degree apart). During runtime, after finding the top five matches, 180 point descriptor is made for the test image and it is matched with the 24x180 dimension training database

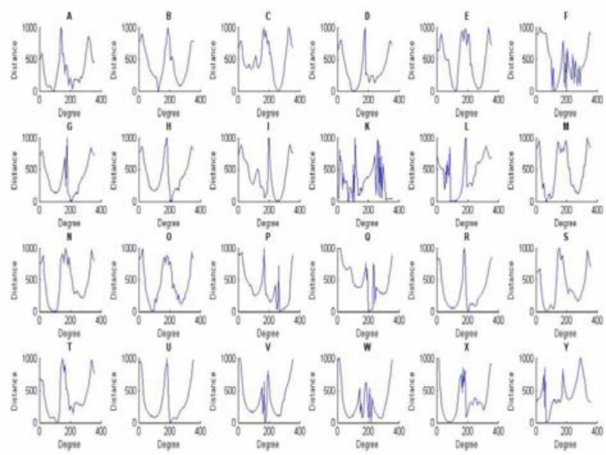
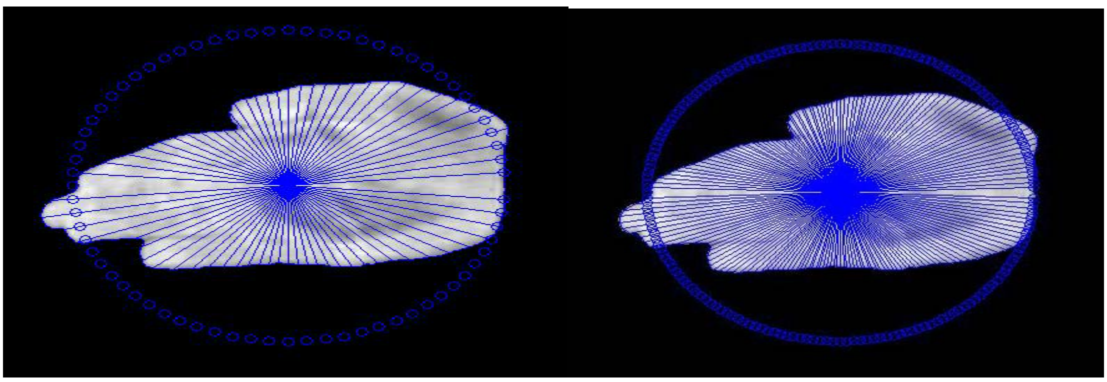


Figure 4. Database histogram.



(a) (b)

Figure 5. Landmark distance from centroid at (a) 5 Degree (b) 2 Degree.

D. Classifier

As the matching descriptors are in array form and are of the same sizes. So, Euclidean distance is used as a classifier. In this classifier, test and training descriptors individual elements are subtracted from their corresponding elements and finally added to get an overall distance. To make this algorithm rotation invariant, 72 point test descriptor is rotated left and right by 6 points and 180 point test descriptor by 15 points on each side. This has provided 30 degree rotation invariance to the developed algorithm. For each rotated test descriptor, Euclidean distance is calculated. The minimum Euclidean distance between descriptor of test image with that of database is considered as best match. By using 72 point descriptor, 5 best matches were selected and 180 point descriptor part was implemented on those best 5 matches. This has provided a two-phase check mechanism for correct recognition. Figure 5 shows the straight lines joined from centroid to the boundary points for 5 degree and 2 degree interval respectively.

III. RESULTS

The proposed algorithm shown in Fig. 2 works successfully for all 24 alphabets shown in Fig. 1 that are considered in database. The accuracy rate calculated is shown below:

TABLE 1: ACCURACY RATE

Alphabet	**Training images**	**Using 72 point descriptor**		**Using 180 point descriptor**	
		Correct recognition	Wrong recognition	Correct recognition	Wrong recognition
A	7	7	0	7	0
B	7	6	1	7	0
C	7	5	2	5	2
D	7	7	0	7	0
E	7	7	0	7	0
F	7	4	3	5	2
G	7	3	4	6	1
H	7	7	0	7	0
I	7	4	3	4	3
J	-	-	-	-	-
K	4	2	2	3	1
L	7	7	0	7	0
M	**7**	**3**	**4**	**3**	**4**
N	7	5	2	5	2
O	**7**	**5**	**2**	**6**	**1**
P	7	7	0	7	0
Q	**7**	**5**	**2**	**5**	**2**
R	7	2	5	2	5
S	**7**	**3**	**4**	**3**	**4**
T	7	5	2	5	2
U	**7**	**6**	**1**	**7**	**0**
V	7	5	2	5	2
W	**7**	**7**	**0**	**7**	**0**
X	7	6	1	6	1
Y	**6**	**5**	**1**	**5**	**1**
Z	-	-	-	-	-
Total	**164**	**123**	**41**	**131**	**33**

Table 1 shows the number of training images of each alphabet and the number of sign being correctly recognized. Each alphabet was tested with 72 point and 180 point descriptor. The alphabets A, B, D, E, H, L, P, U and W show 100% accuracy rate i.e. they are correctly recognized each time. 'R' has worst recognition rate of 29%. The overall recognition rate of both methods is summarized in the table 2 below

TABLE 2: SUMMARIZED ACCURACY RATE

Method	Images in Training Set	Correctly recognized	Recognition Rate
72 point Descriptor	164	123	75.0%
180 point Descriptor	164	131	79.9%

IV. CONCLUSION

By increasing the sampling points from 72 to 180, the achieved accuracy rate of the implemented algorithm is 79.9% which can further be increased by adding the internal characteristics of the hand. Implemented algorithm is simple and takes less than a half second for its complete execution. So, for real time application this technique is fast enough to provide quick responses. Also, the developed algorithm is rotation, luminance and translation invariance and also it doesn't need any data glove for the proper recognition.

V. FUTURE SCOPE

The work has further scope of extension. More robust methods should be developed for enhancing the accuracy rate. Since, only the external characteristic of the hand (shape) is being considered till now, therefore internal characteristics such as color or texture can also be considered. Also, two of the excluded alphabets J and Z that require dynamic processing should also be added.

REFERENCES

[1] Sharmila Konwar, Sagarika Borah and Dr. T. Tuithung, "An American sign language detection system using HSV color model and edge detection", International Conference on Communication and Signal Processing, IEEE, April 3-5, 2014, India.

[2] Yo-Jen Tu, Chung-Chieh Kao, Huei-Yung Lin,"Human Computer Interaction Using Face and Gesture Recognition", Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific, IEEE, Kaohsiung.

[3] Shweta. K. Yewale and Pankaj. K. bharne, "Hand gesture recognition system based on artificial neural network", Emerging Trends in Networks and Computer Communications (ETNCC), IEEE, 22-24 April, 2011.

[4] Marek Vanco, Ivan Minarik and Gregor Rozinaj, "Evaluation of static hand gesture recognition", International Conference on Signal and Image Processing (IWSSIP), IEEE, 12-15 May, 2014.

[5]] Guillaume Plouffe and Ana-Maria Cretu, "Static and dynamic hand gesture recognition system in depth data using dynamic time warping" IEEE Transactions on Instrumentation and Measurement (Volume: 65, Issue: 2, Feb. 2016)

[6] Angur M. Jarman, Samiul Arshad, Nashid Alam and Mohammed J.Islam, "An automated Bengali sign language recognition based on finger tip finder Algorithm", International journal of Electronics & Informatics, 2015.

[7] Eunseok Jeong, Jaehong Lee, and DaeEun Kim, "Fingergesture Recognition Glove using Velostat", 11th International Conference on Control, Automation and System, Oct. 26-29, 2011.

[8] Netchanok Tanyawiwat, and Surapa Thiemjarus, "Design of an Assistive Communication Glove using Combined Sensory Channels", Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference, 9-12 May 2012.

[9] M.S. Sinith, Soorej G Kamal, Nisha B.,Nayana S, Kiran Surendran, and Jith P S, "Sign Language Recognition Using Support Vector Machine", 2012 International Conference on Advances in Computing and Communications, 9-11 Aug.,2012.

[10] M. Ashraful Amin, and Hong Yan, "Sign Language Finger Alphabet Recognition From Gabor-PCA Representation of Hand Gestures", Proceeding of the 6th International Conference on Machine Learning and Cybernetics, 19-22 Aug. 2007.

[11] Andreas Domingo, Rini Akmeliawati, and Kuang Ye Chow, "Pattern Matching for Automatic Sign Language Translation System using LabVIEW", International Conference on Interlligent and Advanced Systems 2007, 25-28 Nov. 2007.

[12] Rudy Hartanto, Adhi Susanto, and P . Insap Santosa, "Preliminary Design of Static Indonesian Sign Language Recognition System", Information Technology and Electrical Engineering (ICITEE), 2013 International Conference, 7-8 Oct. 2013.

[13] Liya Ding and Aleix M. Martinez, "Three-Dimensional Shape and Motion Reconstruction for the Analysis of American Sign Language", Computer Vision and Pattern Recognition Workshop, 17-22 June 2006.

[14] Myoung-Kyu Sohn, Sang-Heon Lee, Dong-Ju Kim, Byungmin Kim, Hyunduk Kim, "3D Hand Gesture Recognition from One Example", 2013 IEEE International Conference on Consumer Electronis (ICCE), 11-14 Jan 2013.

[15] L. Hemandez-Rebollar, "Method and apparatus for translating hand gestures," US Patent 7.565.295, lui 2009.

[16] M. G. Kumar, M. K. Gurjar, and M. S. Singh, "American sign language translating glove using flex sensor," Imperial Journal of Interdisciplinary Research, vol. 2, no. 6, 2016.

[17] Y. Chuang, L. Chen, and G. Chen, "Saliency-guided improvement for hand posture detection and recognition," Neurocomputing, vol. 133, pp. 404-415, 6/10/ 2014.

[18] Q. Munib, M. Habeeb, B. Takruri, and H. A. Al-Malik, "American sign language (ASL) recognition based on Hough transform and neural networks," Expert Systems with Applications, vol. 32, pp. 24-37, 1// 2007.

[19] H.-D. Yang and S.-W. Lee, "Robust sign language recognition by combining manual and non-manual features based on conditional random field and support vector machine," Pattern Recognition Letters, vol. 34, pp. 2051-2056, 12/1/ 2013

[20] R. Zhou, Y. Junsong, M. Jingjing, and Z. Zhengyou, "Robust Part-Based Hand Gesture Recognition Using Kinect Sensor," Multimedia, IEEE Transactions on, vol. 15, pp. 1110-1120,2013.

[21] A. Barczak, N. Reyes, M. Abastillas, A. Piccio, and T.Susnjak, "A new 2D static hand gesture colour image dataset for asl gestures," 2011.