Proceedings of the 2015
IEEE Conference on Robotics and Biomimetics
Zhuhai, China, December 6-9, 2015

Real-Time Bengali and Chinese Numeral Signs Recognition Using Contour Matching

Muhammad Aminur Rahaman[1] Mahmood Jasim[1] Tao Zhang[2] Md. Haider Ali[1] and Md. Hasanuzzaman[1]

***Abstract*— This paper presents a real-time Bengali and Chinese numeral signs recognition system using contour matching. The system converts the captured image into gray scale image. After histogram equalizing and smoothing the gray scale image, the system detects edges using Canny edge detection algorithm. Edge linking is done using morphological operation. The system extracts contours and calculates contour length and area from the detected edges. The system filters unwanted contours based on minimum contour length and area. Then the system encodes the remaining contours into Vector Contours (VC). The system resizes the encoded VC into predefined size. The system generates feature vector based on equalized VC, Auto-Correlation Coefficient (ACC), Normalized ACC and ACC descriptors of equalized VC, which will be used for training and/or testing process. The system recognizes the hand signs based on maximum similarity between test contour and predefined training contours of hand signs using Inter-Correlation Function (ICF). The system is trained using 1000 (10×10×10) contour templates separately for both ten (০ to ৯) Bengali and ten (0 to 9) Chinese numeral signs from 10 signers. The system is tested using another 1000 (10×10×10) contour templates separately for both ten Bengali and Chinese numeral signs achieving recognition accuracy of 95.80% for Bengali numeral signs and 95.90% for Chinese numeral signs with computational cost of 8.023 milliseconds per frame.**

Keywords: Hand sign, Vector Contour (VC), Contour matching (CM), Auto-Correlation Coefficient (ACC), Inter-Correlation Function (ICF).

I. INTRODUCTION

The use of the human hand sign as a simple but robust and natural interface for human–computer interaction (HCI) is increasing rapidly. Vision-based hand sign recognition involves the visual analysis of hand shape, position, and/or movement. Accordingly, the aim of hand sign recognition research is to build a system that can identify and interpret human hand signs automatically. Such a system can be used for manipulation, such as controlling robots or other devices without any physical contact between the human and the machine [1]. It can also be applied as an interpreter for communication through sign languages. Different approaches have been taken for hand sign recognition over the years. Region-based and contour-based approaches are used to segment the hand area from the images. Region-based approaches try to find partitions of the image pixels based on brightness, color and texture of image properties. Contour-based approaches usually start with edge detection, followed by Vector Contour extraction. In the proposed system, contour based approach i.e. Contour Matching is used. Contours are the outer boundary joining all the continuous points along the shape of an object, having same color or intensity. The contour analysis allows describing, storing, comparing and finding the objects presented in the form of the outer contours. Interior points of the object are not accepted for contour analysis and matching. It restricts area of applicability, but it performs in a simple manner using mathematical analysis for image processing with lower computational cost and algorithmic complexity. Contour analysis and matching method is transposition, rotation and scale invariant in image processing [2].

The challenges of hand sign recognition include illumination variation, segmentation of the hand sign from the cluttered background, feature extraction for recognition, noise reduction, signer dependency, high computational cost and proper selection of recognition algorithm, etc.[3]. This paper proposes a signer independent, natural and real-time sign recognition system using contour analysis and matching. Instead of attempting to segment hand sign from the background, the system locates the target hand sign by searching for the best matching contour of hand signs based on training contours. The system recognizes ten (০ to ৯) Bengali and ten (0 to 9) Chinese numeral signs. The test results of the proposed system is compared with the existing reputed sign language recognition systems such as Bengali sign language recognition (BdSLR) system using PCA method by S. Begum et al. [4], Bengali and Chinese numeral signs recognition system using Haar like feature based cascaded classifier and K-nearest neighbor (KNN) Classifier by M. Jasim et al. [5], and BdSLR system using skin-color based segmentation and KNN classifier by M. A. Rahman et al. [3] for the same dataset and environment. This paper is organized as follows. Section II describes the proposed system. Section III presents the

This research is supported and funded by the Information and Communication Technology (ICT) Division, Ministry of Posts, Telecommunications and IT, Government of the People's Republic of Bangladesh.

[1]Department of Computer Science and Engineering, University of Dhaka, Dhaka-1000, Bangladesh. aminur.wg@gmail.com,jasim@csedu.ac.bd, haider@du.ac.bd, hzaman@cse.univdhaka.edu

[2]Department of Automation, School of Information Science & Engineering, Tsinghua University Beijing 100084, China. taozhang@mail.tsinghua.edu.cn

978-1-4673-9674-5/15/$31.00 © 2015 IEEE

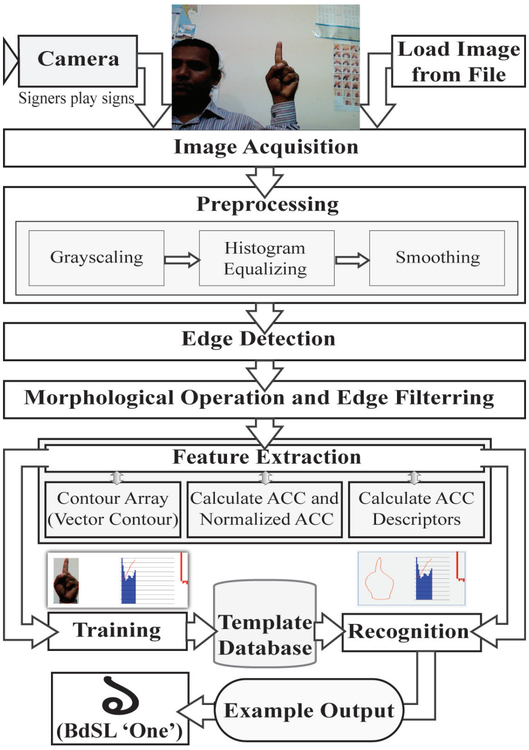


Fig. 1. Block Diagram of the Proposed System.

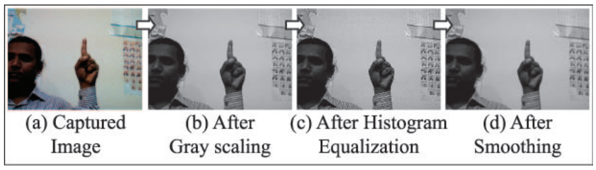


Fig. 2. Example results of the preprocessing.

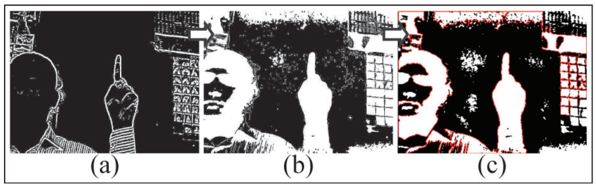


Fig. 3. Example output imges of the (a) Canny edge detection, (b) Morphological operation: dailation and erosion and (c) Unwanted edge filteration.

experimental result with discussion. Finally, the paper is concluded in Section IV.

II. PROPOSED SYSTEM DESCRIPTION

Fig.1 presents the block diagram of the proposed Bengali and Chinese numeral signs recognition system. The system employs necessary preprocessing on captured image. After edges detection from the preprocessed images, the system extracts VC, calculates Auto-Correlation Cofficient (ACC) and normalized ACC, and measures ACC descriptors as features which are used for training and/or testing process. Then the system recognizes the hand signs based on maximum similarity between test contour and predefined training contours of hand signs using ICF. Following subsections briefly describe each process.

A. Image acquisition

The system captures images by using a compatible built-in webcam of ASUS A42F laptop with resolutions 640×480.

B. Preprocessing

The system convertes the captured image into gray scale image. After gray scaling, the system applies histogram equalization using Eq.(1). The example output of histogram equalization is shown in Fig.2(c).

$$I_{eq}(M \times N) = H^{'}(I_{gray}(M \times N)) \tag{1}$$

Where, $I_{eq}(M \times N)$ is the resulted histogram equalized image, $I_{gray}(M \times N)$ is the gray scaled image with $M \times N$ resolutions and $H^{'}$ is the integral of the histogram $H(j)$ of the gray scaled image, $I_{gray}(M \times N)$. The resulted $H^{'}(i)$ is calculated by using Eq.(2).

$$H^{'}(i) = \sum_{0 \leq j < i} H(j) \tag{2}$$

After histogram equalization, the sytem uses Gaussian smoothing operation represented by the convolution of $I_{eq} * G$ with a filter of 5×5 Gaussian kernel G by using the Eq.(3)[6].

$$G(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3}$$

Where, x and y are the horizontal and vertical distances from the origin of the histogram equalized image and σ is the standard deviation of the Gaussian distribution. Fig.2 shows the effects of preprocessing steps on a captured image frame.

C. Edge Detection Using Canny Operator

The system uses canny edge detector which performs beter than other edge detectors [7]. From the smooth image, the system computes the gradient components J_x and J_y and estimates the edge strength, E_s and the orientation of the edge normal, E_o by using Eq.(4) and Eq.(5) [6] respectively. Non-maxima suppression is applied to the strength image and then the system detect edges by applying lower threshold (t_L) and higher threshold (t_H) on all pixels (p,q) of the images. All visited points in the connected contour found is stored for contour analysis. Fig.3(a) shows example output image from the Canny edge detection method.

$$E_s(p,q) = \sqrt{(J_x(p,q)^2 - J_y(p,q)^2)} \tag{4}$$

$$E_o(p,q) = \alpha \; tan J_y/J_x \tag{5}$$

D. Morphological Operation and Edge Filterring

After Canny edge detection the contour is generated with unwanted edges and various noises. Morphological image filtering is used to filter out the noisy pixels. The morphological operation works in the sequence of dilation and erosion by using Eq.(6) and Eq.(7) respectively [3].

$$I_d = I_n \oplus B = \{z|(B'_z) \cap I_n \neq \phi\} \tag{6}$$

$$I_e = I_d \ominus B = \{z|(B)_z \subseteq I_d\} \tag{7}$$

Where, z is the set of all points of the image objects, B is the structuring element with a specific size, I_d and I_e is the dilated and eroded images respectively. The example output of the dilation and erosion is shown in Fig. 3(b).

After completion of morphological operation, the system calculates the contour length and area where, a contour length is the total number of the contour pixels and contour area is the multiplication of maximum height and maximum width of the contour shape. Then the system filters out the unwanted edge by intersection with minimum contour length (k>200) and minimum contour area (k^2<400) from the image, I_e and prepares the image with selected contour as I_c. Fig. 3(c) shows the example output of unwanted edge filtration.

E. Feature Extraction

In this step, the system encodes the selected contour, I_c into Vector Contour (VC). The VC of hand signs is used as main feature for this proposed system to recognize Bengali and Chinese numeral signs. The system resizes the encoded VC as predefined size, calculates ACC and measures normalized ACC and ACC descriptors that are used to form feature vector for the system.

1) Equalized Vector Contour (VC) extraction: The contour is encoded by a sequence consisting of a complex numbers 'a+ib'. The starting point on a contour is fixed. Then, the contour is scanned clockwise, and each vector of offset is represented by a complex number 'a+ib'. Where 'a' is the point offset on x axis, and 'b' is the point offset on y axis. Offset is represented concerning the previous point [2]. Fig.4(a) shows the example process of encoding of a contour of Bengali numeral sign '০'(zero). Each pixel (vector) of a contour is represented by 'Elementary Vector' (EV) and sequence of complex-valued numbers is represented by Vector Contour (VC). Thus, the Vector Contour (VC) of length k can be expressed by Eq.(8).

$$\Gamma = (\Upsilon^0, \Upsilon^1, ..., \Upsilon^{k-1}) \tag{8}$$

Where, Γ represents a VC and Υ represents each EV of the VC.

The system calculates the sum of all EVs of a VC represented by Ω using Eq.(9). If the value of Ω is equal to zero, then the system decides that the detected contour

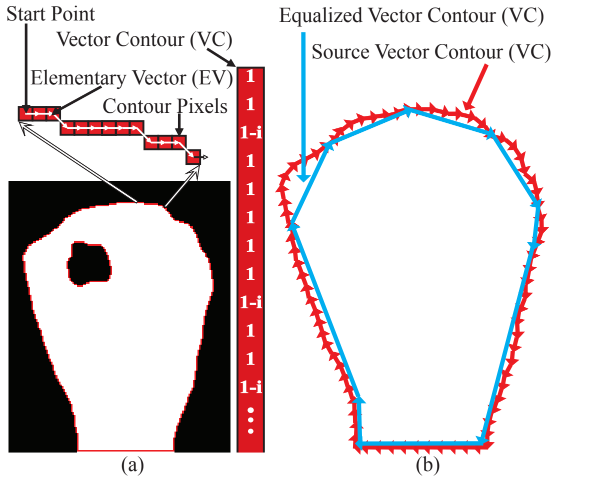


Fig. 4. Example of (a) Complex number encoding process from a contour image and (b) Contour equalization process of Bengali numeral sign '০' (zero).

of a hand sign is a closed contour and extracts this VC for the selected hand sign.

$$\Omega = \sum_{n=0}^{k-1} (\Upsilon^n) \tag{9}$$

Where, k is dimensionality of a VC and Υ^n is the n[th] EV.

The contour analysis method for contour matching adopts identical length of contours. The extracted VC of hand signs have arbitrary length. Therefore, for training and recognition process, the system forces to make all of contours length uniform by resizing the contours that are designated with a predefined length k=50. Fig.4(b) shows the example of contour equalization process of Bengali numeral sign '০' (zero). By this process the system extracts the equalized Vector Contour for each Bengali and Chinese numeral signs.

2) Normalized Auto-Correlation Coefficient: Auto-Correlation Coefficient (ACC) is a scalar product of a vector of contour and itself at various shifts of starting point. After extraction of equalized vectors of contour 'Φ' with length k=50, the system calculates Auto-Correlation Coefficient (ACC) by using Eq.(10).

$$ACC(m) = (\Phi, \Phi^{(m)}) \tag{10}$$

Where, ACC(m) measures the similar length of contours with the value among 0 to 1 and m=0, 1, 2, ..., k-1.

After generation of ACC, the system normalizes the ACC. The normalized ACC is symmetric concerning a central reference k/2. The middle section of Fig.5(b) shows the ACC and normalized ACC represented by dark blue and overlapping red colored graph respectively.

3) ACC Descriptors: Before direct ACC comparison, its descriptors are compared. The descriptors are calculated as Wavelet convolution [8] of ACC using four

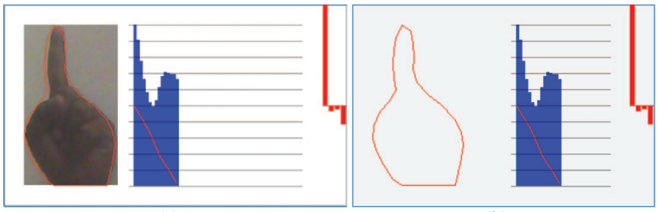


Fig. 5. Example output of feature extraction and feature vector generation (a) Equalized Vector Contour (left part) with ACC and normalized ACC (middle part), and ACC descriptors (right part) and (b) Suitable feature vector of bengali numeral sign 'One' with extracted features.

different filters such as filter1 = {1, 1, 1, 1}; filter2 = {-1, -1, 1, 1}; filter3 = {-1, 1, 1, -1}; and filter4 = {-1, 1, -1, 1}. It increases the performance of ACC comparison. The right section of Fig.5(b) shows the example output of ACC descriptors represented by four red colored bins (right part).

The system generates feature vector based on extracted features equalized VC, ACC, normalized ACC and ACC descriptors for each predefined hand signs of Bengali and Chinese numeral signs that will be used for training and/or testing process. Fig.(5) shows the example output of feature extraction and feature vector generation of Bengali and Chinese numeral sign 'One'.

F. *Training*

After feature vector generation, the system saves maximum suitable contour templates of equalized VC with their properties in two separate binary serialized files for Bengali and Chinese numeral signs as shown in Fig.(5). The system is trained using 1000 (10×10×10) contour templates for BdSL numeral signs and another 1000 (10×10×10) contour templates for Chinese numeral signs from 10 signer where each signer performs 10 signs for each signs.

G. *Recognition*

For testing, the system captures images and generates the feature vector of test hand sign same as the feature vector generation of training contour. Then the system searches through the training database to find a maximum similar contour using ICF, if it is similar enough then the system recognizes the contour of hand sign. ICF of two contour vectors is represented by Eq. (11)[2].

$$ICF(m) = \left(\Phi, \tau^{(m)}\right) \tag{11}$$

Where, $\tau^{(m)}$ represents a vector of contour of test hand sign received from τ by cycle shift by its EV on 'm' of elements, Φ represents a vector of contour of training hand sign. After calculating ICF, the system measures the maximum similarity between two Vector Contours of testing and training hand signs using Eq.(12).

$$ICF_{max} = \left(\frac{ICF(m)}{|\Phi|\,|\tau|}\right) \tag{12}$$

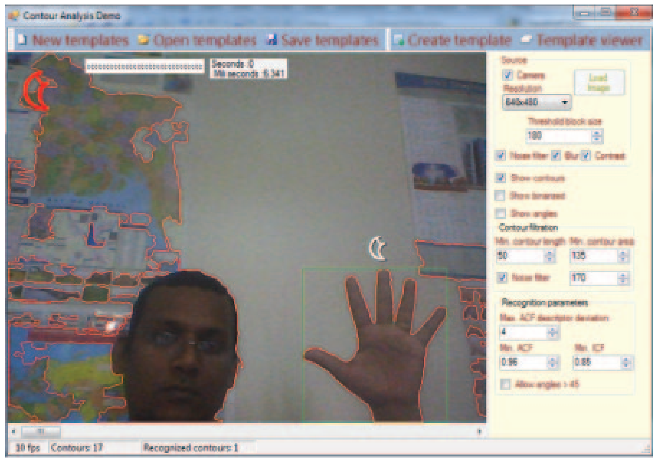


Fig. 6. Example output for Bengali numeral sign '৫' (Five) recognition

Where, ICF_{max} measures the similarity shape of hand signs with the value among 0 to 1. $|\Phi|$ and $|\tau|$ represent the normalized length of contours Φ and τ respectively that are calculated by Eq.(13).

$$|\tau| = \sqrt{\sum_{n=0}^{k-1} |\Upsilon_n|^2} \tag{13}$$

Fig.(6) shows the example of Bengali numeral sign '৫' (Five) recognition.

III. EXPERIMENTAL RESULT AND DISCUSSION

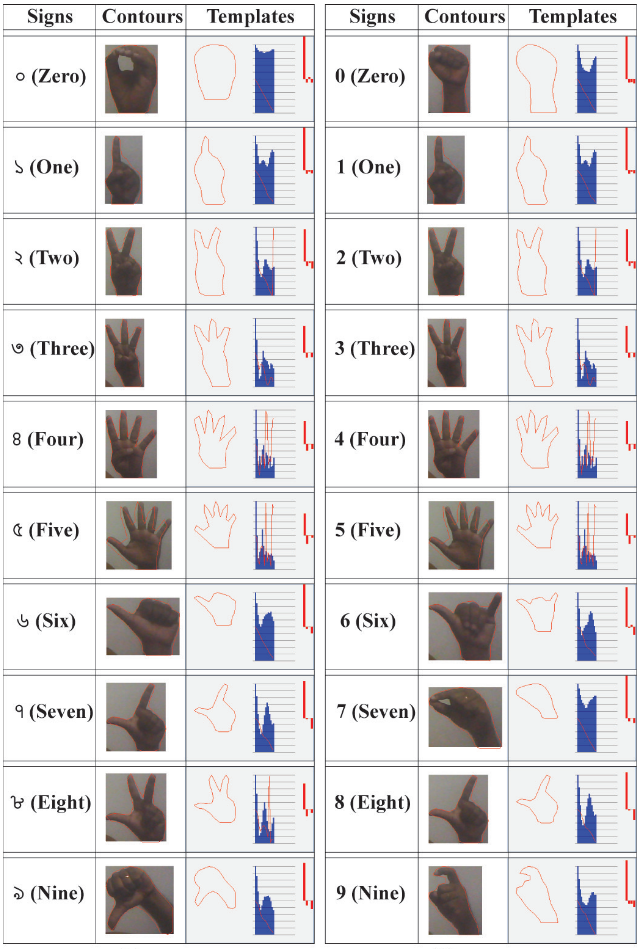
The proposed system uses a built-in webcam of ASUS A42F series laptop for image acquisition. The system uses an ASUS A42F series laptop with Intel Corei3 processor of 2.40 GHz and 2GB RAM. The system uses EmguCV (C# and OpenCV wrapper) [9] in 32-bit operating system of MS Windwos7, as system development platform.

A. *Training Contour Templates Database*

The proposed system is trained for ten (০ to ৯) Bengali and ten (0 to 9) Chinese numeral signs with the equqlized length of VC, k=50. The system asks 10 performers to perform each sign 10 times, where four are female and six are male. This resulted in 1000 (10×10×10) training images for Bengali numeral signs and 1000 (10×10×10) training images for Chinese numeral signs. Fig.7(a) presents the example training dataset of Bengali numeral signs and Fig.7(b) presents the example training dataset of Chinese numeral signs.

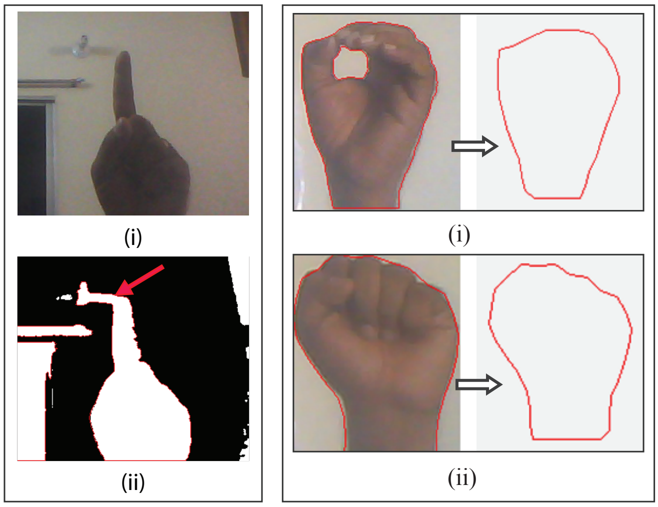
B. *Parameters of Perfomance Analysis*

The system uses two performance parameters such as, Accuracy and Computational cost. Accuracy is calculated using Eq.(14) and the time to capture an image, preprocess, extract features, generate feature vector and match it with the training contour template from the



(a) (b)

Fig. 7. Example training dataset (a) Bengali numeral signs (b) Chinese numeral signs.



(a) (b)

Fig. 8. (a) Example of unrecognizable hand sign for occlusion (i) Main image (ii) After edge detection and (b) Example of two hand signs with similar outer contour (i) Bengali numeral sign 'Zero' with outer and inner contours (ii) Chinese numeral sign 'Zero' with only outer contour.

database is considered as computational cost in Milliseconds per frame.

$$Accuracy = \frac{R \times 100}{T} \tag{14}$$

Where, R is the number of correctly recognized hand signs and T is the total number of hand signs.

TABLE I and TABLE II present the summarized results of ten (0 to 9) Bengali and Chinese numeral signs recognition by the proposed system (CM) respectively with a comparison with other systems; BdSLR system using PCA [4], Bengali and Chinese numeral signs recognition system using Haar classifier-KNN [5], and BdSLR system using skin color detection-KNN [3] for the same dataset. From the test results, it is evident that Bengali and Chinese numeral signs are recognized with the mean Accuracy of 95.80% and 95.90% by the proposed system respectively which is comparatively better than other systems. From the TABLE I, the mean Accuracy of the proposed system is decreased due to similarity of equalized contours of Bengali numeral signs 'two' with 'seven'; 'three' with 'eight' and 'six' with 'nine' when the angle orientation is greater than 45 degrees as shown in Fig.7(a).

From the TABLE II, the mean Accuracy of the proposed system is decreased due to similarity of equalized contours of Chinese numeral signs 'zero' with 'seven' and 'two' with 'eight' when the angle orientation is greater than 45 degree as shown in Fig.7(b).

TABLE III represents the comparative analysis of the computational costs of the proposed system (CM) with the mentioned system [4], [5], and [3] which represnt that the proposed system is about more than ten times faster than previous systems with keeping the success rate of about 95%. However, the system does not provide proper result incase of occlusion as shown in Fig.8(a). Also the system may not work properly under this condition as like as shown in Fig.8(b).

IV. CONCLUSION

This paper presents a real-time Bengali and Chinese numeral signs recognition system using contour matching. Experimental result shows that the system is capable of recognizing Bengali and Chinese numeral signs (0 to 9) effectively. The system will be capable of recognizing any hand signs, if the system is trained properly. The system generates feature vector based on equalized VC, Auto-Correlation Coefficient (ACC), normalized ACC and ACC descriptors of equalized VC, which will be used for training and/or testing process. The system recognizes the hand signs based on maximum similarity between test contour and predefined training contours of hand signs using ICF. The system achieves recognition accuracy of 95.80% and 95.90% for ten (0 to 9) Bengali and Chinese numeral signs respectively that were trained by 1000 (10×10×10×) signs and tested with another

TABLE I
RESULT OF BENGALI NUMERAL SIGNS RECOGNITION

Bengali Numeral Signs	Total Hand Signs (T)	Correctly Recognized (R)				Accuracy (%)			
		Proposed System (CM)	*PCA [4]*	*Haar Classifier-KNN [5]*	*Skin Detect-KNN [3]*	*Proposed System (CM)*	*PCA [4]*	*Haar Classifier-KNN [5]*	*Skin Detect-KNN [3]*
০(Zero)	100	98	87	93	94	98	87	93	94
১(One)	100	100	89	98	96	100	89	98	96
২(Two)	100	94	84	95	94	94	84	95	94
৩(Three)	100	95	88	94	93	95	88	94	93
৪(Four)	100	95	87	95	94	95	87	95	94
৫(Five)	100	96	90	91	98	96	90	91	98
৬(Six)	100	95	83	90	93	95	83	90	93
৭(Seven)	100	94	85	93	94	94	85	93	94
৮(Eight)	100	95	87	93	94	95	87	93	94
৯(Nine)	100	96	83	90	95	96	83	90	95
					Mean Accuracy	**95.80**	**86.30**	**93.20**	**94.5**

TABLE II
RESULT OF CHINESE NUMERAL SIGNS RECOGNITION

Chinese Numeral Signs	Total Hand Signs (T)	Correctly Recognized (R)				Accuracy (%)			
		Proposed System (CM)	*PCA [4]*	*Haar Classifier-KNN [5]*	*Skin Detect-KNN [3]*	*Proposed System (CM)*	*PCA [4]*	*Haar Classifier-KNN [5]*	*Skin Detect-KNN [3]*
0 (Zero)	100	95	85	92	91	95	85	92	91
1 (One)	100	100	89	98	96	100	89	98	96
2 (Two)	100	95	84	94	93	95	84	94	93
3 (Three)	100	100	87	95	93	100	87	95	93
4 (Four)	100	96	89	95	94	96	89	95	94
5 (Five)	100	95	90	91	99	95	90	91	99
6 (Six)	100	98	83	90	93	98	83	90	93
7 (Seven)	100	89	83	91	92	89	83	91	92
8 (Eight)	100	94	89	92	94	94	89	92	94
9 (Nine)	100	97	83	91	96	97	83	91	96
					Mean Accuracy	**95.90**	**86.20**	**92.90**	**94.10**

TABLE III
COMPARATIVE ANALYSIS OF THE COMPUTATIONAL COST AMONG DIFFERENT METHODS

Method	Computational Cost (Milliseconds per frame)
Proposed System (CM)	8.023
PCA[4]	121.110
Haar Classifier-KNN[5]	78.080
Skin Detect-KNN [3]	88.09

1000 ($10\times10\times10\times$) signs for both cases with the mean computational cost of 8.023 milliseconds per frame. The proposed system is attractive for the simplicity and low computational cost that permit to run the system in real-time. The proposed system can be applied as an interpreter for communication with sign and non-sign people as well as it will assist for human-robot interaction based on sign language recognition.

References

[1] M.K. Bhuyan, K. F. MacDorman, M. K. Kar, D.R. Neog, B. C. Lovell, P. Gadde, "Hand pose recognition from monocular images by geometrical and texture analysis." In: Journal of Visual Languages and Computing, Vol. 28, (2015), p.39–55

[2] R. Kolar, A. Thakar, M. Shabad, "Image segmentation for text recognition based on boundary analysis", International Journal of Emerging Technology and Advanced Engineering, Vol. 4(2), (2014), p. 294-298, ISSN 2250-2459.

[3] M. A. Rahaman, M. Jasim, M. H. Ali, M. Hasanuzzaman, "Real-Time Computer Vision-Based Bengali Sign Language Recognition." In: Procedings of the 17th Int'l Conf. on Computer and Information Technology (ICCIT2014), pp. 192-197, IEEE, Dhaka, Bangladesh (22-23 December, 2014), doi: 10.1109/ICCITechn.2014.7073150

[4] S. Begum, M. Hasanuzzaman, "Computer vision-based Bangladeshi Sign Language recognition system." In: MS thesis, Department of CSE, University of Dhaka, Bangladesh, and 12th ICCIT- 2009 Dhaka, Bangladesh, (2009), pp. 414-419, doi: 10.1109/ICCIT.2009.5407274.

[5] M. Jasim, T. Zhang, M. Hasanuzzaman, "A real-time computer vision-based static and dynamic hand gesture recognition system" In: International Journal on Image and Graphics, vol. 14(01n02) (2014), doi: 10.1142/S0219467814500065

[6] R. C. Gonzalez, R. E. Woods, "Digital Image Processing." (3rd Edition). Prentice-Hall, Inc., Upper Saddle River, NJ, USA (2006).

[7] J.Canny, "A Computational approach to edge detection." In: IEEE Transaction Pattern Analysis Machine Intelligence, vol. 8(6), (Nov. 1986), pp. 679-698.

[8] F. P. R. Antonio, R. Robles, "The convolution theorem for the continuous wavelet transform", In: Elsevier B.V., Signal processing, Vol. 84(1), (January 2004), P.55–67.

[9] http://sourceforge.net/projects/emgucv/files/latest/download Accessed at 08/01/2015.