2014 17th International Conference on Computer and Information Technology (ICCIT)

Real-Time Computer Vision-Based Bengali Sign Language Recognition

Muhammad Aminur Rahaman, Mahmood Jasim, Md. Haider Ali and Md. Hasanuzzaman
Department of Computer Science & Engineering
University of Dhaka
Dhaka, Bangladesh
aminur.wg@gmail.com, mahmood.jasim.0@gmail.com, haider@univdhaka.edu, hzamancsdu@yahoo.com

***Abstract*— This paper presents a real-time computer vision-based Bengali Sign Language (BdSL) recognition system. The system detects the probable hand from the captured image. The system uses Haar-like feature-based cascaded classifiers to detect the hand in each frame. From the detected hand area, the system extracts the hand sign based on Hue and Saturation value corresponding to human skin color. After normalization the system converts the hand sign to binary image. Then the binary images are classified by comparing with pre-trained binary images of hand sign using K-Nearest Neighbors (KNN) Classifier. The system is able to recognize 6 Bengali Vowels and 30 Bengali Consonants. The system is trained using 3600 (36x10x10) training images where each of 10 signers performed 10 signs for each corresponding Bengali alphabet and the system is tested using another 3600 (36x10x10) images of 10 signers. The system is achieved recognition accuracy of 98.17% for Vowels and 94.75% for Consonants.**

Keywords*— *Skin color segmentation, Hand detection, K-Nearest Neighbors (KNN) algorithm, Bengali Sign Language Recognition.

I. INTRODUCTION

Sign language is a nonverbal form of communication used especially by people with the inability to speak or hear. It is a separate language with its own grammar and rules [1]. Bengali Sign Language (BdSL) is structurally different from sign languages from other countries. To perform Bengali Sign Language, two hands are used generally. Many researchers within computer vision communities have attempted to develop sign language recognizers for sign languages used in different countries like American Sign Language (ASL) [2] by Kulkarni et al., British Sign Language (BSL) [3] by Ong et al., Arabic sign language (ArSL) [4] by Naoum et al., etc. Notable work on Bengali Sign Languages include researches done by Pavel et al.[5], S. Begum and Hasanuzzaman [1], M. Jasim et al. [6], etc. The challenges of sign language recognition includes segmentation of the hand area from the image frame, feature extraction for training and recognition, accounting for motion in some of the sign gestures, environmental noise, proper selection of recognition algorithm etc.

To overcome these challenges, this paper proposes a signer independent, natural, low computational complexity and real-time computer vision-based BdSL recognition system using an appearance based approach. The system recognizes 6 Vowels and 30 consonants of BdSL performed by two hands. The rest of the paper is organized as follows. Section II, presents the proposed system description in detail. In section III, the experimental result and discussion is provided. The papers is concluded in section IV.

II. PROPOSED SYSTEM DESCRIPTION

The proposed system detects and segments the hand area from the image frame, extracts hand shape features and trains a model for recognition. After the model generation the system can correctly recognize the Bengali sign language using KNN classifier. Fig.1, presents the architecture of the proposed Bengali Sign Language recognition system.

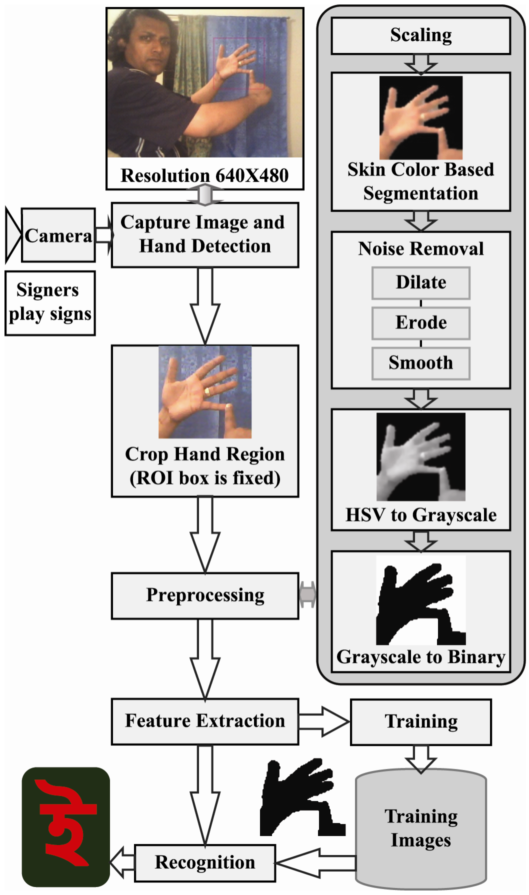

Fig. 1. Architecture of the proposed system.

A. Hand Area Segmentation

The system detects the hand from the captured images using Haar-like feature-based cascaded classifier [7]. In this system, the Haar-like feature-based cascaded classifier is trained to detect 'Open Hand' and 'Close Hand' from the image frame. Fig.2, presents an example of 'Open Hand' and 'Close Hand' detected by the trained Haar-like feature-based cascaded classifiers.

The process of generating a Haar-like feature-based cascaded classifier is a method of building a boosted rejection cascade [6], which will discard the negative training data to obtain a decision to determine the positive data. W. N. B. W. Ismail used the Haar-like features [7] in their project to find the weak constraints. Each Haar-like feature consists of two or three cascaded "black" and "white" rectangles and the value of a Haar-like feature is the difference between the sum of the gray level values of the pixels within the black and white rectangular regions. Fig.3, shows the example of extended sets of Haar-like features.

To calculate the value of the rectangle Integral Image is used which is a central representation of an image consisting of the total value of gray. Fortunately, this can be done in a single pass over the image using a recurrence equation as shown in (1) [9].

$$ii(x,y) = i(x,y) + ii(x-1,y) + ii(x,y-1) - ii(x-1,y-1) \quad (1)$$

After getting the value of the integral image, the next step is to find a cascaded classifier which is a chain of different stages classifier, where each stage classifier is used to detect whether the image in the sub-window is the desired object (object of interest). The process is presented in Fig.4.

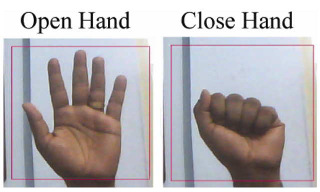

Fig. 2. Example of hand area detection.

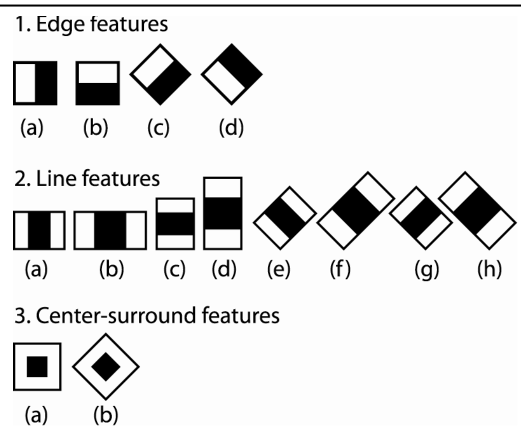

Fig. 3. A set of extended Haar-like features.

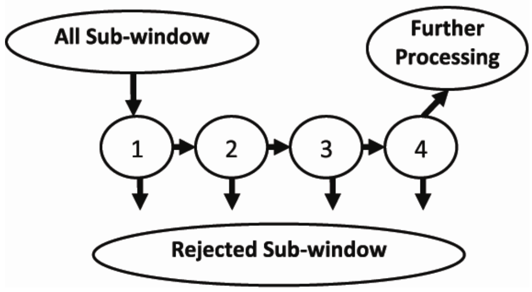

Fig. 4. Object detection using cascaded classifier.

B. Skin Color based Hand Sign Extraction and Preprocessing

After normalization the cropped image containing the hand area is scaled to 0.5. The scaled RGB images are then converted to HSV color system. Then human skin color based hand shape segmentation approach is used. This process filters out the pixels which do not belong to the hands in the current image. This process is based on skin color and background color information. Skin color region is determined by applying threshold values on Hue (H) and Saturation (S) value of HSV color based model. The hand area containing image is thresholded based on the value of (H<=19 and S>=48) to extract the hand shape images. Example of skin color based hand sign segmentation is shown in Fig. 5.

After completing the skin color based hand shape segmentation process to remove the background; dilation, erosion and smoothing technique are used to reduce noise. Gaussian smoothing is used with a filter of 5x5 kernel for better performance as presented in (2).

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

Where, x and y are the horizontal and vertical distance from the origin and σ is the standard deviation of the Gaussian distribution. After removing the noise the HSV color image is converted to grayscale image and then finally is converted to binary image. Example results of the preprocessing steps are shown in Fig.6.

Fig. 5. Example of segmented hand signs based on skin color.

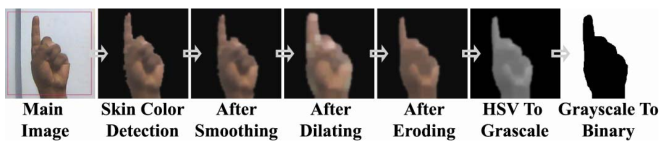

Fig. 6. Example results of the preprocessing steps.

C. Binary Hand Sign Extraction and Model Generation

A set of effective features were extracted to represent hand shapes. Geometrical properties of the hand shapes like finger position, fingertip were examined for real-time processing and recognition. The binary images containing the hand signs are used as extracted features. Fig.7, shows the example of some extracted binary hand signs.

A training module is generated using the extracted binary hand sign using KNN classifier. Separate KNN classifiers are trained for vowels and consonants with K value of 6 and 30 respectively. The system uses 600 (10x10x6) training images for Bengali Vowels and 3000(10x10x30) training images for Bengali Consonants in different illuminations and background. The example of training dataset is presented in Section III. The model is stored for future sign recognitions.

D. Sign Language Recognition

When recognizing a sign language that has been presented to the system, the image frames are captured that construct the sign. The hand postures are segmented from these initial image frames. These frames containing the hand signs are preprocessed and then converts the hand sign to binary image. Extracted binary hand signs are trained or classified by comparing with pre-trained binary images of hand sign using K-Nearest Neighbors (KNN) Classifier. When extracted binary hand sign is matched against the pre-stored training hand signs then the system recognizes the specific sign using the KNN classifier. Examples of Bengali vowel and consonant sign recognition are shown in Fig. 8 and Fig.9 respectively.

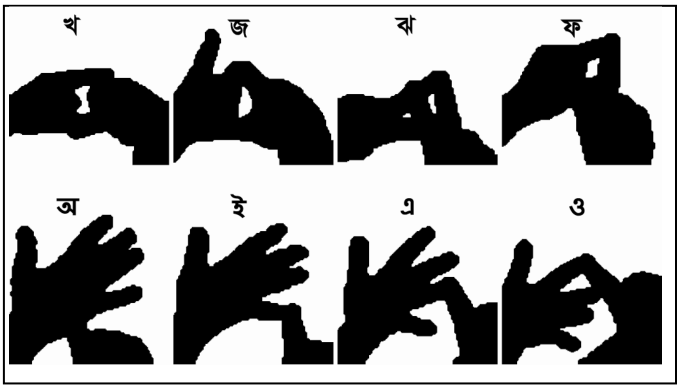

Fig. 7. Example of extracted binary hand signs.

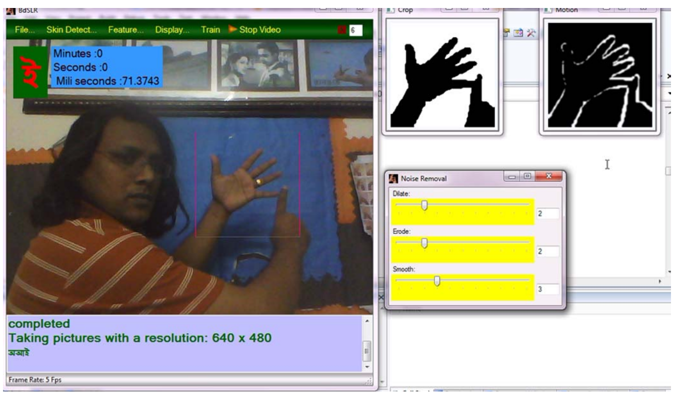

Fig. 8. Example Output for Bengali Vowel Recognition.

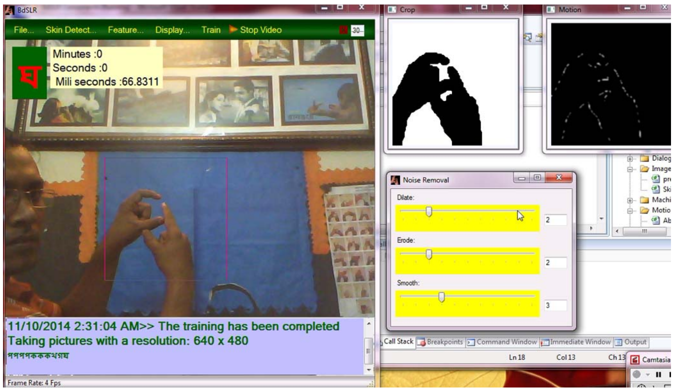

Fig. 9. Example Output for Bengali Consnant Recognition.

III. EXPERIMENTAL RESULT AND DISCUSSION

This system uses a compatible built-in webcam of ASUS A42F laptop for image acquisition. The system uses an ASUS A42F series laptop with Intel Core i3 processor and 2GB RAM. We have used C# OpenCV wrapper (EmguCV) [11] as system development platform.

A. Training Image Database

The system is trained to recognize 36 hand shapes (6 Bengali vowels and 30 Consonants). K value of the variable used is 6 for vowels and 30 for consonants. 10 images of each hand sign are captured from 10 different people for training where four are females and six are males. This resulted in 600 (10x10x6) training images for Bengali Vowels and 3000(10X10X30) for Bengali Consonants. Fig.10 and Fig.11 present the example sets of Bengali vowel and Bengali consonant sign images respectively.

Fig. 10. Example set of training dataset of Bengali Vowel signs.

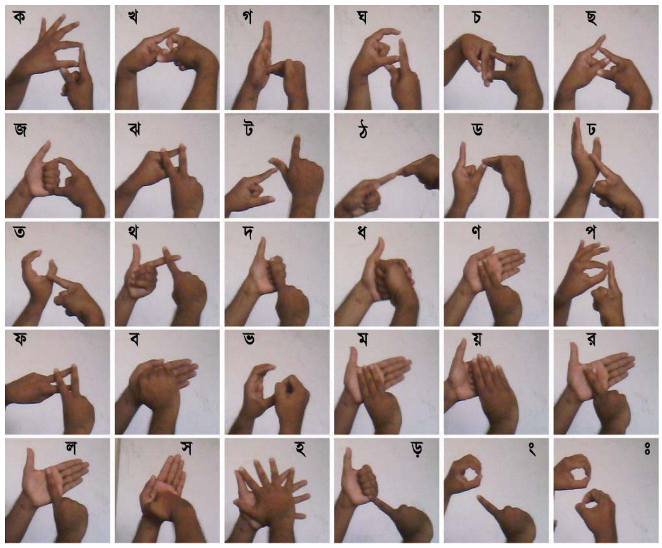

Fig. 11. Example set of training dataset of Bengali Consonant signs.

All of the hand signs are captured in both controlled and cluttered backgrounds. Fig.12, shows example of hand detection in different backgrounds using Haar-like features. It shows the presence of static objects in the background do not reduce system performance. TABLE I, shows the detection rate of ‘Open Hand’ and ‘Close Hand’.

Skin color region segmentation method is used based on HSV to extract the binary hand features containing hand shapes for model training. But skin color detection based on HSV will not work so well under various lighting conditions and backgrounds. In this case, the hand area is segmented using the Haar-like feature-based cascaded classifier, which is mostly illumination and background invariant. We have used a technique in this case that once the ‘Open Hand’ is detected then region of interest (ROI) box is created around the open hand area. By moving the ‘Open Hand’ the ROI box is moved to the non-skin color background position and then the ROI box is set to a fixed suitable position by closing the detected open hand. At the moment, the skin color detection based on HSV starts to process the selected region from the captured images. The confusion related to background and illumination is minimized and our test results show good performance under various illumination and backgrounds.

B. *Sign Language Recognition Result Analysis*

For testing the system, two cases are considered.

- Case-1(C1): Test data generated by the same 10 signers who took part for the generation of the training data (10x10 samples for each sign).
- Case-2(C2): Testing data generated by new signer (10x10 samples for each sign) who didn't take part in generating training data.

The system accuracy is calculated using (3).

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn} \qquad (3)$$

TABLE I. PERFORMANCE OF HAND DETECTION USING HAAR-LIKE FEATURES

Open Hand (%)	Close Hand (%)
100	98

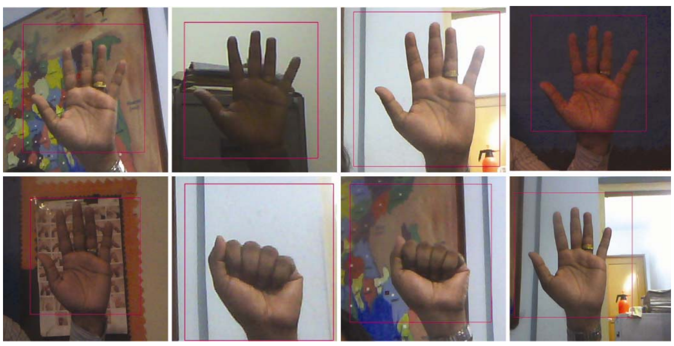

Fig. 12. Example of hand detection in different backgrounds.

Where, tp (True positive) means number of correct output results, tn (True Negative) means number of correct detection of the incorrect input data, fp (False Positive) means number of incorrect result in the system’s output (Unexpected output), and fn (False Negative) means number of missing expected incorrect outputs, i.e. the result should be included in system output but is not there.

TABLE II and TABLE III, present the tp, fp, tn, fn and accuracy of BdSL recognition performed by the proposed system for both Case-1 (C1) and Case-2(C2) for Bengali Vowel and Consonant signs respectively. These values are generated using confusion matrixes for both C1 and C2 which we do not show here in the interest of space.

TABLE IV, presents the overall accuracy and computational cost of the system for Bengali Vowels and Consonants sign recognition by averaging accuracies and computational costs for each individual signs. The efficiency of this real time system is measured based on the time it takes to complete a single matching against the training dataset. The time to capture an image, preprocess, extract features and recognize the Bengali sign using a single test image against the training database is considered for computational cost measurement. The overall accuracy and computational cost of BdSL recognition are 96.46 and 93.55792 milliseconds per frame respectively.

From the test results shown in TABLE II and TABLE IV, it is evident that Bengali Sign Language Vowels are recognized and distinguished with an average accuracy of 98.17% with K value of the classifier set to 6. For both cases C1 and C2, the accuracy of “উ , এ, ও” decreased due to similar shape and training data inconsistency. In particular, to perform the sign “ও” is most difficult. So, the accuracy rate of recognizing the sign “ও” is decreased to 95% that affects on overall accuracy.

From the test results shown in TABLE III and TABLE IV, it is evident that Bengali Sign Language Consonants are recognized and distinguished with an average accuracy of 94.75% with K value of the classifier set to 30. The average accuracy is affected by the accuracy of sign “ফ” which is 63%. For both cases C1 and C2, some consonants such as, “ছ, জ, ঝ, ণ, ফ, য়, র, ল, ড়” obtained less than 95% accuracy. This is caused by several factors which include similarity of the shape cue of “খ” with “জ”; “য়” with “ধ”; “ছ” with “ঝ”; “ড়” with “দ” and “ফ” with “ঝ” and “খ” as shown in Fig.13 and Fig.14. The color images of “ম”, “র” and “ল” are distinguishable but the binary images are very similar to each other as shown in Fig.15.

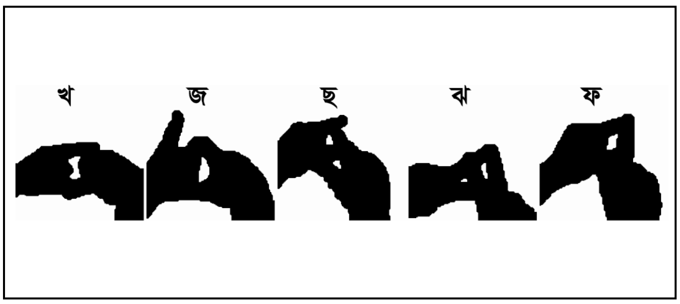

Fig. 13. Similar Binary images of consonants (খ, জ, ছ, ঝ, ফ).

TABLE II. RESULT OF ACCURACY FOR BENGALI VOWELS RECOGNITION

Bengali Vowels	tp		fp		tn		fn		Accuracy (%)	
	C1	*C2*	*C1*	*C2*	*C1*	*C2*	*C1*	*C2*	*C1*	*C2*
অ	97	95	0	0	3	3	0	2	100	98
আ	97	96	0	0	3	3	0	1	100	99
ই	100	98	0	2	0	0	0	0	100	98
উ	98	96	2	2	0	1	0	1	98	98
এ	99	97	1	2	0	1	0	0	99	98
ও	94	93	4	5	1	2	1	0	95	95
Average Accuracy for C1 and C2									98.67	97.67

TABLE III. RESULT OF ACCURACY FOR BENGALI CONSONANTS RECOGNITION

Bengali Consonants	tp		fp		tn		fn		Accuracy (%)	
	C1	*C2*	*C1*	*C2*	*C1*	*C2*	*C1*	*C2*	*C1*	*C2*
ক	94	92	4	5	2	2	0	1	96	94
খ	95	94	3	5	1	1	1	0	96	95
গ	99	97	1	3	0	0	0	0	99	97
ঘ	95	92	4	6	1	2	0	0	96	94
চ	96	94	4	5	0	0	0	1	96	94
ছ	93	93	5	6	2	1	0	0	95	94
জ	92	92	5	6	2	2	1	0	94	94
ঝ	90	90	5	7	4	3	1	0	94	93
ট	100	99	0	0	0	1	0	0	100	100
ঠ	100	100	0	0	0	0	0	0	100	100
ড	100	98	0	1	0	0	0	1	100	98
ঢ	98	98	2	2	0	0	0	0	98	98
ত	94	92	3	4	3	4	0	0	97	96
থ	98	96	2	2	0	2	0	0	98	98
দ	97	95	3	4	0	1	0	0	97	96
ধ	100	98	0	0	0	2	0	0	100	96
ণ	86	84	11	13	1	2	2	1	87	86
প	97	97	3	3	0	0	0	0	97	97
ফ	61	60	35	35	2	3	2	2	63	63
ব	99	99	1	1	0	0	0	0	99	99
ভ	96	95	4	5	0	0	0	0	96	95
ম	98	96	2	2	0	2	0	0	98	98
য়	93	93	6	6	1	1	0	0	94	94
র	83	80	14	13	3	5	0	2	86	85
ল	90	90	7	7	1	2	2	1	91	92
স	97	95	2	2	1	2	0	1	98	97
হ	100	100	0	0	0	0	0	0	100	100
ড়	95	93	5	5	0	2	0	0	95	95
ং	99	96	0	2	1	2	0	0	100	98
ঃ	90	95	5	2	3	1	2	2	93	96
Average Accuracy for C-1 and C-2									95.10	94.40

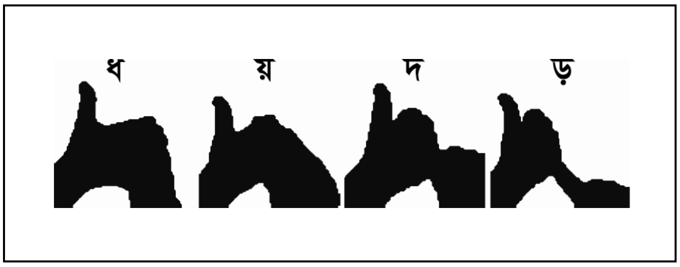

Fig. 14. Similar Binary images of consonants (ধ, য়, দ, ড়).

TABLE IV. RESULT OF OVERALL SYSTEM ACCURACY AND COMPUTATIONAL COST OF BENGALI VOWELS AND CONSONANTS SIGN RECOGNITION

Bengali Sign Language	Average Accuracy (%)	Computational Cost (Milliseconds/frame)
Vowels	98.17	92.37092
Consonants	94.75	94.74492
Average	96.46	93.55792

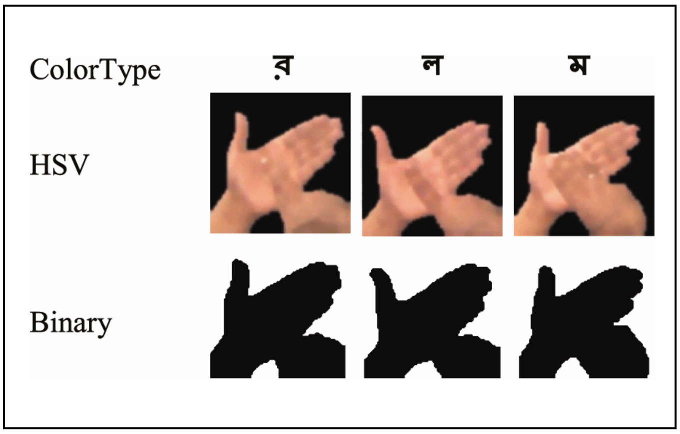

Fig. 15. Indistinguishing problem of binary image.

However, our test results show better performance than existng reputed BdSL recognition systems such as [1], [5], [6] and [10] that were developped using Principle Component Analisys (PCA) Method, 3D modelling, KNN Clasifier for static one-handed Bengali Number recognition module and Artificial Neural Network respectively.

IV. CONCLUSION

This paper presents a signer independent and real-time computer vision-based Bengali Sign Language recognition system. Haar-like feature-based cascaded classifier is used to detect the probable hand area from the captured image frames. Skin color-based segmentation method is used to extract hand sign and then extracted hand images are converted to binary images those are used as either training or testing images. Finally, K Nearest Neighbor classifier is used to recognize both Bengali Sign Language Vowel and Consonant signs. The system is trained using 3600 (36x10x10) training images where each of 10 signers perform 10 signs for each corresponding

Bengali alphabet and also the system is tested using another 3600 (36x10x10) images of 10 signers. The system achieved the overall computational cost and accuracy of BdSL recognition are 93.55792 milliseconds per frame and 96.46 respectively. However the system has few limitations. The system cannot properly segments hand area if some objects rather than hand has skin like colors. The system cannot properly distinguish some signs such as “র” with “ল” and “ফ” with “ঝ” because of their similarity among their binary images. This limitation may solve using edge detection method. The performer faces difficulties to perform few alphabets such as “ও, থ, জ, ঠ, র” due to camera position. The problem may be overcome by using over head camera or multiple camera. The system is applicable for human machine communication or sign and non-sign human communication using Bengali sign.

REFERENCES

[1] S. Begum and M. Hasanuzzaman, “Computer vision-based Bangladeshi Sign Language recognition system”, MS thesis, Department of CSE, Dhaka University, Bangladesh, and ICCIT- 2009, 2009, p. 414.

[2] Vaishali S. Kulkarni and S.D. Lokhande, “Appearance Based Recognition of American Sign”, (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 03, 2010, p 560-565.

[3] E. Ong, H. Cooper, N. Pugeault and R. Bowden, “Sign language recognition using sequential pattern trees”, In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012.

[4] R. Naoum, Hussein, H. Owaied, S. Joudeh, “Development of a new Arabic Sign Language recognition using K-Nearest Neighbor algorithm”, Journal of Emerging Trends in Computing and Information Sciences, VOL. 3, NO.8, ISSN 2079-8407, August 2012.

[5] D. S. H. Pavel, T. Mustafiz, A.I. Sarkar and M. Rokonuzzaman, “Modeling of Bengali Sign Language expression as dynamic 3D polygons for developing a vision based intelligent system for dumb people”, National Conference on Computer Processing of Bangla, Independent University, Bangladesh, 2004.

[6] M. Jasim, T. Zhang and M. Hasanuzzaman, “A real-time computer vision-based static and dynamic hand gesture recognition system”, International Journal on Image and Graphics, 2014.

[7] W. N. W. Ismail, “Object detection system using Haar-Classifier”, Faculty of Electrical & Electronics Engineering, Universiti Malaysia Pahang, May 2009.

[8] G. Bradski, "The OpenCV library", Dr. Dobbs Journal of Software Tools, 2000.

[9] http://www.codeproject.com/Articles/441226/Haar-feature-Object-Detection-in-sharp. Accessed at 08/04/2014.

[10] M. Atiqur Rahman, Ahsan-Ul-Ambia, M. I. Abdullah and S. K. Mondal, “Recognition of static hand gestures of alphabet in Bangla Sign Language”, IOSR Journal of Computer Engineering (IOSRJCE), ISSN: 2278-0661, ISBN: 2278-8727 Volume 8, Issue 1, Nov-Dec. 2012, PP 07-13.

[11] http://sourceforge.jp/projects/sfnet_emgucv/downloads/emgucv/2.0.1.0/Emgu.CV.x86_2.0.1.0_windows_installer.zip/ Accessed at 08/04/2013.