Real Time Hand Gesture Recognition for Bangla Character using SVM Classifier

Shayla Sharmin
Dept. of CSE, Chittagong University of Engineering & Technology (CUET)

Papia Sultana
Dept. of CSE, Chittagong University of Engineering & Technology (CUET)

Md. Ibrahim Khan
Dept. of CSE, Chittagong University of Engineering & Technology (CUET)

ABSTRACT
Since the very early times of humanity, long before the advent of spoken language, signs and gestures have been in use for communication. Although sign languages are used to bridge the gap wherever vocal communication is impossible, for example to communicate with deaf and mute people as well as with the machines, or difficult if there is a language barrier between the two speakers, there is no optimal work of recognizing sign language in any language as sign languages are vary from country to country. Like any other country Bangladeshi deaf and mute people use their own sign language but unfortunately there are very few works to recognize the sign language which makes the interaction between general and deaf and mute people more tough. In this paper, we proposed an efficient scheme for hand gesture used for Bangla vowel and consonant character recognition in real time adopted by deaf and mute community in Bangladesh. After taking video via webcam as input, hand region is detected by skin color segmentation followed by converting the selected frame into YCbCr. Afterward some pre-processing steps are applied on the image to acquire the region of interest, Hu moment invariants are used to extract features and later on classification and recognizing the character are done by Support Vector Machine (SVM). To use this proposed method as an interpreter for the sign languages of other races no major modifications are required except that the training set should be enriched with desired sign

General Terms
Image Processing, Sign Language

Keywords
Bangla character, real time video, Hu moment invariant, YCbCr, SVM

1. INTRODUCTION
Expression of semantic information through movement of parts of human body like series of hands and arms motion, facial expressions & head/body postures is considered as sign language which are used in different times and in different ways for different reason of different kinds in our day to day communication Nowadays by using hand gesture playing computer games, controlling the computer mouse/or keyboard functions, interacting man and machine interface, controlling mechanical system remotely become more natural and easier. Apart from these application of hand gesture one of the most important application is to communicate with other where vocal conversion is impossible or difficult especially with deaf and mute people.

Normally the deaf people are hear impaired that makes them incapable of talking which compelled them to use sign language to communicate with the people, which is a visual form of communication including the combination of hand shapes, orientation and movement of the hands, arms or the body and facial expressions. It is a comprehensive problem because of the complexity of the visual analysis of hand gestures and the highly structured nature of sign languages. A universal sign language does not exist that is the reason that sign languages like spoken languages are developed differently depending on the community and the region and so various sign languages all over the world are founded, namely American Sign Language (ASL), French Sign Language, British Sign Language (BSL), Japanese Sign Language (JSL) etc.

In Bangladesh, Bangla Sign Language used by the deaf and mute people, differs in the syntax, phonology, morphology and grammar from other country's sign languages. Bangla sign language uses hand gestures, facial expressions, head/body postures, locations of hand with respect to body etc. to represent signs. And it is difficult for general people to learn all the signs used by the deaf and incapable in talking people.

Sign language recognition is a task being solved by many research institutes in the world. Despite of recognizing sign language solved by many research institutes in the world, it lacks a universal parameterization and recognition method that would be widely accepted as a baseline. For image acquisition there are many techniques for example using single camera or multiple camera, glove based acquisition etc. Different types of segmentation techniques such as color segmentation, shape segmentation have been used previously. For feature extraction there are several techniques such as features exclusive for hand (finger count, fingertip, palm center etc.), shape based features (convex hull, moments), edge based features (histogram orientation). Template matching, artificial neural networks, hidden markov model, SVM are the classifiers used in recognition of hand gesture previously. In this paper we recognize hand gesture for Bangla character applying Hu moment invariant as feature extraction and using SVM classifier. The outline of this paper is as follows: Section II describes Bangla sign language and some existing Bangla sign language recognition techniques. Section III presents the proposed system. Experimental studies have been discussed in Section IV. Finally, concluding remarks are explained in Section V.

2. LITERATURE REVIEW

At the end of 1990, the recognition of sign language has begun to appear by some electrochemical devices which were used as primary effort to recognize it [1]. The approach wearing gloves, different types of marker used for determining hand gesture parameter like hand's location, angle, position etc. compel the signer to wear a cumbersome device. Although there are so many works in sign language recognition in the world very few works have been done for Bangla sign language. In this section we will discuss about Bangla sign language and some existing works in this regard.

2.1 Bangla Sign Language

The deaf community in Bangladesh use sign language which is called Bangla Sign Language Anthology (BSLA). This BSLA is controlled by Centre for Disability Development (CDD) [2]. The SDSL is a Disable People Organization (DPO) in Bangladesh that is run by Sign Language Users. From the very common experience of life style, history and culture, deaf people and sign language users become united and formed SDSL in 2008 [2]. The aim of the organization is to create an inclusive society where sign language user's emancipation will be secured with their dignity and equity and to stand beside the individuals with hearing and speech difficulties in Bangladesh to preserve, protect and promote their human, civil, culture and linguistic rights. Various form of Bangla sign language of two handed [3] are shown in figure 1

Fig 1: Two handed Bangla Sign language [3]

According to the manual of BdSL [4], there are approximately 5000 sets of gestures for alphabets, numbers and common words. In this paper, only the Bangla sign language for Bangla alphabet has been considered for recognition purpose. The Sign Language Alphabet is a set of alphabetic finger signs. It is a rich and complex visual-spatial language. It has a vocabulary and syntax of its own. It is different from other sign languages and from the spoken language. To express a full meaning only hand gesture is not enough. Bangla sign language includes hand shapes and movements, facial expression and also body movement.

In Bangla language we have 50 characters, 40 alphabets and 10 numbers. There are 11 vowels (Shoroborno/ স্বরবর্ণ) and 39 consonants (Benjonborno/ ব্যাঞ্জনবর্ণ) in Bangla language. Among them some characters pronounced similar and also some gestures of Bangla sign language are really hard to recognize as few of them are very similar to look which make the recognition task more difficult. Bangla sign language can be one handed or two handed. In this paper, two handed Bangla characters have been recognized as in case of communication use of two hands can produce more gesture than one handed and this procedure is much easier and efficient to express anything.

2.2 Previous work

In [1], they worked with single hand Bangla character. And they have tested 47 images. The inputs of Bangladeshi sign language have been taken by webcam and later on recognized by efficient Neural Network Ensemble (NNE).[2] Works with Bangla characters and they used two wrist bands to detect the hand region. They recognize only 11 letters and use template matching as classifier.

In [3] the system only requires the images of the bare hand for the recognition. But the only limitation of this system is that, for learning NN, the feature vector should have integer values only. In [5] single handed Bangla sign have been recognized. And it has been observed that single handed signs are really difficult to understand. And for this reason the main concern which is to communicate with deaf people become difficult.

The main emphasise of this paper is to develop a system for two handed Bangla sign language without using any special marker and using a fast classifier. Here 34 different hand signs have been collected by different people who are used to test the effectiveness of this system

3. PROPOSED FRAMEWORK

At this section, the overview of the system is going to be introduced. The creation of a database with all the images is the starting point of the project which is used for training and testing. Various BSL databases on the Internet, photographs and videos have been taken with a digital camera. Sample of trained and tested images are shown in figure 2

(a) training image **(b) test image**

Fig 2: Sample images of database

The training dataset is developed having 34 sets for 34 gestures where each set contains 10 images of each gesture. The training set has images with black background to reduce noise that makes the training set steadier. But in case of test images containing different backgrounds are considered which makes the system more effective. After acquiring the image using webcam, some operations which are known as

preprocessing is applied that helps us to get smoother region of interest. After that features will be extracted using Hu moment invariants and use SVM to recognize the letter. The block diagram is showing the system overview of the proposed method.

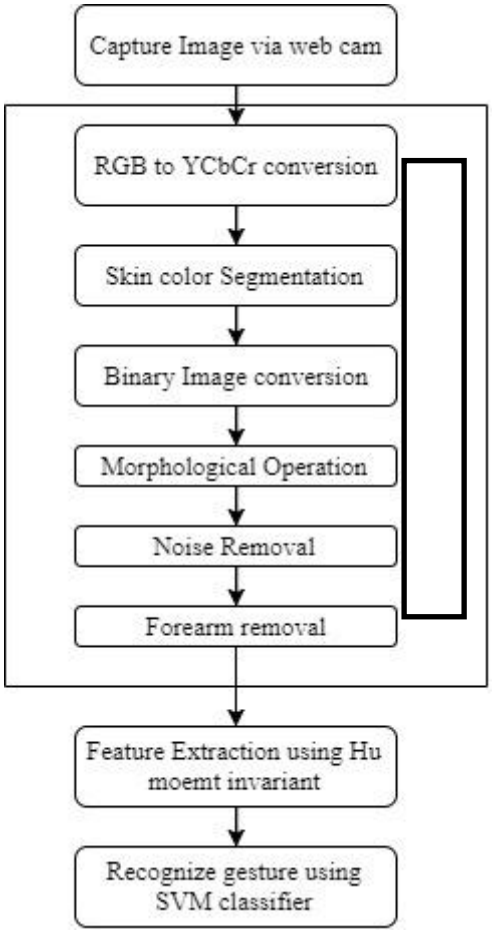


Fig 3: Block diagram of the proposed system

3.1 Image Acquisition

In term of webcam, the camera caaptured frame in one second per one image. In this case one should hold the position for 3 second to select the desire image. If the system find similar images for cosecutive three seconds it will take the middle one.

Fig 4. Select RGB image as input via web cam

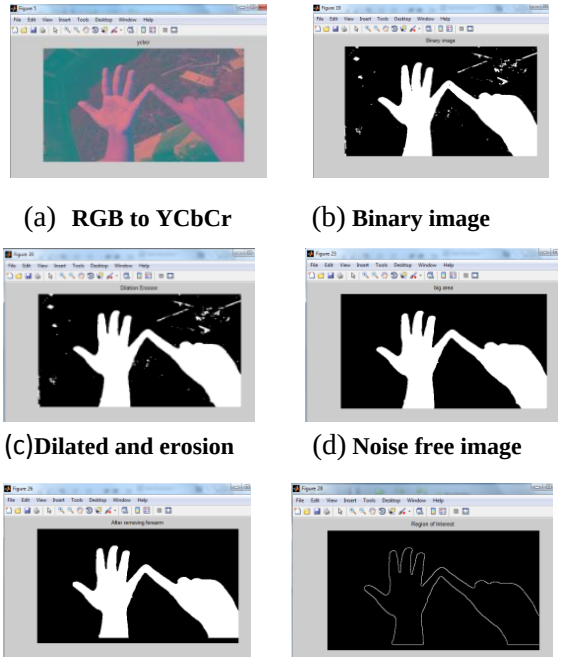
3.2 Preprocessing

3.2.1 YCbCr conversion

The image taken from the video is a RGB image. In this project segmentation of the hand region is done by skin color segmentation. Segmentation is a process to simplify the image representation into something which is easier to further analysis. First the RGB image is converted into YCbCr form shown in figure 4 (a). YCbCr color model is chosen because in YCbCr color space the luminance and chromatic information are separated. After convert the RGB image into YCbCr color space region of interest (ROI) is defined as it is the hand portion that has skin color. After taking images from the database, the mean and standard deviation of Cr and Cb channel is calculated. And the ranges of skin color where blobs that are segmented as skin blob are

$Cb \geq 106$ & $Cb \leq 128$; $Cr \geq 134$ & $Cr \leq 157$

Y component is not considered in this method because it distributes uniformly across (0, 255) and is too sensitive to the alternation of luminance. Figure 4(b) shows the binary image of the hand region from the input image but it consist some noises as some of the segment in the image can have also color similar as skin color.



(a) **RGB to YCbCr** (b) **Binary image**

(c)**Dilated and erosion** (d) **Noise free image**

(e) **Removing forearm** (f) **Region of interest**

Fig 4: Steps of image pre-processing

3.2.2 Morphological Operation

Morphological operation like erosion, dilation is done (figure 4(c)) on the segmented binary image to improve the binary threshold image quality and removes holes from the image. At first, dilation operation is performed and then erosion operation which is known as opening operation. Dilation followed by erosion is called closing operation. By applying opening operation stands for erosion followed by dilation, tiny objects and smooth the boundary of objects with large size are erased. After morphological operation the ROI contains has some noises which will be reduced in the next section to get better result.

3.2.3 Noise Reduction

Salt and pepper noise is a common form of noise. Binary threshold results this type of noise. Because of this noise errors are introduced during the imaging process. So, removal of the salt and pepper noise is necessary. In the image frame there may come some unwanted component which should be removed from the image frame. In the image frame there may come some unwanted component which should be removed from the image frame. Based on the assumption of hand

region as the big connected component, other small connected component with respect to hand region will be considered as noise. After morphological preprocessing of the binary image, Flood fill algorithm label the connected component. Consider the two big connected components and remove all the others as noise. Then if the posture contains two hands without touch with each other, the two larger components will be the two hands. But if the hands are in touch with each other then remove the second component if it contains less binary data than the half of largest component. After this process the image frame contain only the hand region and a noiseless image frame is achieved. Hand region is assumed as the big connected component whereas other small connected components with respect to hand region will be considered as noise. After calculating the area of each and every connected component of a binary image and the bigger area are considered as the ROI. After this process the image frame contain only the hand region and a noiseless image frame is achieved in Figure 4(d).

3.2.4 Removing Forearm

The Forearm part is useless for the feature extraction. Forearm part is removed from an image based on human hand anatomy Figure 4(e). Forearm always follow a non-increasing radius shape up to the wrist part of the hand. This feature can be written as $d_i > d_{i+1,}$ Where d_i=diameter of the forearm up to the wrist Based on this phenomenon the forearm part is clipped from the image and a normalized image is found for further processing.

Then by canny edge detection algorithm we get our region of interest the hand portion for the further analysis in Figure 4 (f).

3.3 Feature Vector Extraction

It is very important to choose appropriate features. But there is no universal and optimal method to characterized image of hand gesture. As in the matter of feature, moment invariants are considered in this project adapted idea from Hu [6]. But some of the gestures of Bangla characters are difficult to distinguish from each other. Hu moment invariant [6] has been selected for recognizing gesture of Bangla Characters. The moment can be applied in the image processing as the binary image can be considered as a bivariate distribution function. The moments help to provide a generic representation of any object to extract easily from an image. Hu proposed the concept of moment invariants first time. He proposed a set of moment invariants with characteristics of translation invariance, scale invariance and rotation invariance. These invariances were obtained by a nonlinear combination of moments. This proposed method gives almost same moment invariants for an image with different rotation, scaling and transformation. Hu used algebraic invariants to derive these expressions. And these are applied to the moment generating function under a rotation transformation. Such statistical moments work directly with regions of pixels. The moments most commonly used are the seven invariant moments of Hu of order 2 and 3. The advantage of using Hu invariant moment is that it can be used for disjoint shapes. In particular, Hu invariant moment set consists of seven values computed by normalizing central moments through order three[7].The 7 moment invariants proposed by Hu can be expressed as

$$\emptyset_1 = \eta_{20} + \eta_{02} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

$$\emptyset_2 = (\eta_{20} - \eta_{02})^2 + (2\eta_{11})^2 \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

$$\emptyset_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad \dots\dots\dots\dots\dots (3)$$

$$\emptyset_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad \dots\dots\dots\dots\dots\dots (4)$$

$$\emptyset_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - (3\eta_{21} - \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (5)$$

$$\emptyset_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad \dots\dots\dots\dots\dots\dots\dots (6)$$

$$\emptyset_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \dots (7)$$

Where the normalized central moments are defined as $\boldsymbol{\eta_{pq}=\frac{\mu_{pq}}{\mu_{00}^{\gamma}}}$, **where** $\boldsymbol{\gamma = \frac{p+q}{2} + 1}$ **for** $\mathbf{p + q = 2, 3}$

The 2-D moment of order (p+q) of a digital image *f*(x, y) of size M × N is defined as

$$M_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x,y) dx\, dy \; p, q = 0,1,2 \dots\dots (8)$$

The central moments are defined as

$$\mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \bar{x})^p \, (y - \bar{y})^q f(x,y) \; p, q = 0,1,2 \dots . (9)$$

Where, $\bar{x} = \frac{M_{10}}{M_{00}}$ $\bar{y} = \frac{M_{01}}{M_{00}}$

If an instance has translation, scaling or rotation the 7 invariants for that instance in three positions are not that much differ. So it is easy to recognize any instance. The 7 Hu moment invariants have different amount of calculation. The most useful information of image is usually stored in less computational moment with low order. And the high order moment invariants are of tremendous amount of calculation and contain detail information which is prone to be affected by noise which makes it difficult to distinguish the differences using moments with high order objects. So only first six moment of invariant are used in this work.

3.4 Classification

There are some reputed method for classification for recognizing gesture such as template matching, neural network, hidden marcov model, support vector machine (SVM) etc. SVM has chosen as the classifier in this project. SVM is a very powerful and very widely used both within industry and in Academia and compared to both the logistic regression and neural networks, the support vector machine or the SVM sometimes gives a cleaner and sometimes more powerful way of learning complex nonlinear functions.

Considering a data set with positive and negative examples, this data is linearly separable and this means that there exists a straight line, although there is many a different straight lines, they can separate the positive and negative examples perfectly. An SVM is a linear classification algorithm that maximizes the distance between the decision line (discriminator) and the closest example to it in the training set. Since the system needs to classify between several gestures are needed to classify several binary SVMs in a

Classifier Tournament structure have been combined. For every pair of classes, we train a binary classifier that classifies between them. After new image features are computed in the preprocessing stage, the features are entered into all the binary classifiers. Each binary classifier returns a class number representing a gesture.

In between two method of multiclass SVM the system approached with the one versus rest method. A training set with 34 classes has been created. The training set of 34 separates set is then turns into two class classification problems. For each class i[th] we train a logistic regression classifier $h^i\theta(x)$ to predict the probability that y=i. To make a prediction, on a new input x, pick the class i that maximizes Max $h_\theta^i\ (x)$……. (10), where $h_\theta^i\ (x)$= p(y=1|x;θ)
The gesture that has the highest probability of classifiers wins and is the result of the classifier.

3.5 RECOGNIZING LETTER

After classification the proposed system show the corresponding image of the given hand gesture. Here only one gesture of a video has been recognized. The output here in figure 5 is a vowel of Bangla character "ই"/'Rossho E' pronounced as 'E'

Fig 5: Output Image

4. EXPERIMENTAL RESULT AND ANALYSIS

For proving the effectiveness and accuracy of the proposed system, a number of experiments have been carried out. The hand gesture recognition for Bangla characters module is implemented in the MATLAB environment. An Intel Pentium Dual Core 2.20 GHz machine with a 32-bit operating system and 1 GB RAM was used for testing. In the experiments, images sized 256×256 were used. A sign image dataset has been created, which contains 1020 images. The dataset is classified as a training set that has 340 images, 10 images for each sign with black background and a testing set that has 680 images, 20 images for each sign with different background of 20 different people. [1] Provides the equation for accuracy of the work which is Output the correct recognition accuracy= (Correctly classified/Total input)*100. From 340 Training images successfully classified images are 334 so the accuracy is (334/340) % =98.24%. From 680 tested image successfully classified images are 597 so the accuracy is (597/680) % =87.79%.

5. CONCLUSION

Gesture classification is a challenging work. Strong classification system is requires for accurate classification. In this proposed classification system, any sign posture can be classified. If segmentation of a test image gives a better result this system gives a better result. One of the limitations of this project is that the system fails to detect hand region in full skin-colored background. Another one is if the noise is larger than the hand region after segmentation then the hand portion is considered as noise according to this proposed algorithm. In this case the system will fail.

The initial limitation of this project is that is dependent much on color based segmentation. Though color is the initial detector of any object, in future a shape based rule can be defined for hand ROI detection which will solve the problem of having skin color background as well as having skin color portion larger than the hand region.

6. REFERENCES

[1] Bikash Chandra Karmokar, Kazi Md. Rokibul Alam and Md. Kibria Siddiquee , "Bangladeshi Sign Language Recognition employing Neural Network Ensemble", International Journal of Computer Applications (0975 – 8887) Volume 58– No.16, November 2012 43

[2] Dr. Kaushik deb, Helena Parveen Mony & Sujan Chowdhury "Two Handed Sign Language Recognition for Bangla Sign Character using Cross Correlation" Global journal of Computer Science and Technology, Volume 12, Issue 3, February 2012

[3] Md. Atiqur Rahman, Dr. Ahsan-Ul-Ambia, Md. Ibrahim Abdullah and Sujit Kumar 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,pp. 07-13, 2012.

[4] Centre for Disability in Development (CDT), "Manual on Sign Supported BangIa," in Computer Vision and Image Understanding, 1-50, 2002.

[5] Foez M. Rahim, Tamnun E Mursalin, Nasrin Sultana, "Intelligent Sign Language Verification System – Using Image Processing, Clustering and Neural Network Concepts"

[6] M.K. Hu "Visual pattern recognition by moment invariants", IRE Trans information Theory, vol.8, no.2, pp.179-187, Feb. 1962

[7] Neha S. Chourasia, Kanchan Dhote, Supratim Saha, "Analysis on Hand Gesture Spotting using Sign Language through Computer Interfacing ", International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 3, Issue 3, May 2014

[8] Muthukrishnan.R and M.Radha, "Edge Detection Techniques for Gesture Recognition", International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 6, Dec 2011.