Bangla Sign Language Interpretation Using Image Processing

Supervisor: Dr. Jia Uddin

Fahmid Nasif Arko	**13101087**
Nujhat Tabassum	**13101027**
Taposhi Rabeya Trisha	**13101163**
Fariha Ahmed	**13101024**

***Department of Computer Science and Engineering*, BRAC University**

Submitted on: 18 April, 2017

DECLARATION

We, hereby declare that this thesis is based on the work results done by ourselves. Other's Materials which we used for help and researcher are mentioned by reference. This thesis, neither in whole nor in part, has been previously submitted for any degree.

Signature of Supervisor

____

Dr. Jia Uddin

Signature of Authors

____

Fahmid Nasif Arko

____

Nujhat Tabassum

____

Taposhi Rabeya Trisha

____

Fariha Ahmed

ACKNOWLEDGEMENTS

We want to thank Almighty ALLAH first, the creator and the owner of this universe, who provided us guidance, strength and abilities to complete this research.

We are very thankful to Dr. Jia Uddin Sir, our thesis supervisor, from the bottom of our heart for his help, guidance and support in completion of this project. We are also thankful to BRAC University Faculty, Staffs of the Computer Science and Engineering, who have helped us a lot and encouraged us in the whole study period at BRAC University.

Finally, we are obviously grateful to our sincere and beloved parents, siblings and friends for their love and help. We are grateful to all of them.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Instant feedback on sign language can greatly improve sign language interpretation. In this project we plan to use efficient methods to detect the hand correctly using skin detection algorithm and removing all noise using MATLAB and classify the image according to the sign gesture performed. In this paper, we propose an image processing based model which will interpret Bangla Sign Language. The purpose of this model is to find Bangla Sign Language recognition accuracy. The model will detect the skin color of every type using YC_BC_R algorithm and use Bag of features for feature extraction and Support Vector Machine (SVM) for training and evaluation. To validate our proposed model, we used our own dataset of Bangla Sign Languages using hand gestures of both male and female. The average accuracy we got from evaluation set is 86%.

Chapter 01

Introduction

1.1 Motivation and Goals

Sign language is a natural language for all human being. By born people try to communicate by expressing signs through making gestures, moving hands, or facial expressions. Sign language is an easygoing medium of communication by which people can communicate, but for some people who cannot speak or hear, this is the only way to communicate. It develops communication and promotes it [1]. People especially who are both unable to hear and unable to speak need to learn sign language for communication. In this case, for communication, sign language is the preferred form. But most of the common people do not understand sign languages. To make an easy communication between the people who use sign language and the people who do not understand sign language, sign language interpreter's service can be arranged. In this way of communication, it is important to know how to make use of the services of the system effectively [2].

Different countries have different languages. Hence, sign language varies from country to country. American Sign Language is considered as International Sign Language among all of the sign languages. In our country, Bengali Sign Language is commonly used by maximum hearing and speaking impaired community. Bengali Sign Language is made from several core one-handed static sign alphabet and number signs. Many researchers have been working on sign language recognition systems for various sign Languages. A system developed by Pavel et al. [3] analyzes video clips of different gestures of sign languages taken as input and gives audio output. Angles of different parts of the hand with body were calculated manually by analyzing captured images from input video frames and stored manually in a database with corresponding audio meanings.

1.2 Project Scopes

This research is focusing to help the hearing and speaking impaired people by developing a system by which sign language can be interpreted to other people both ordinary and hearing and speaking impaired people. This project specifically focuses on Bengali sign language. This project will create an easy communication scope between an ordinary person and who cannot speak or hear, but both have to understand Bengali Language and have to be able to recognize Bengali alphabets and numerals. We hope, our system will reduce the communication gap of the hearing or speaking impaired people with others. Hopefully, it will create more scopes for special children to express themselves. By using our system, hearing or speaking impaired children could ask more questions, thus learn more. Using this system, we hope, negligence towards special children or people who cannot speak or hear will reduce in our society.

1.3 Overview

We have researched different kinds of sign language interpreters or devices for sign language interpretation system. Now we are trying to establish a system that interpret Bangla sign language to common people. And this has been done in two phases. First, from our own dataset it takes the images and apply skin detection algorithm and detects the skin color pixels from it, then makes it binary. From this binary image we will extract features using bag of features method and train the system using SVM classifier.

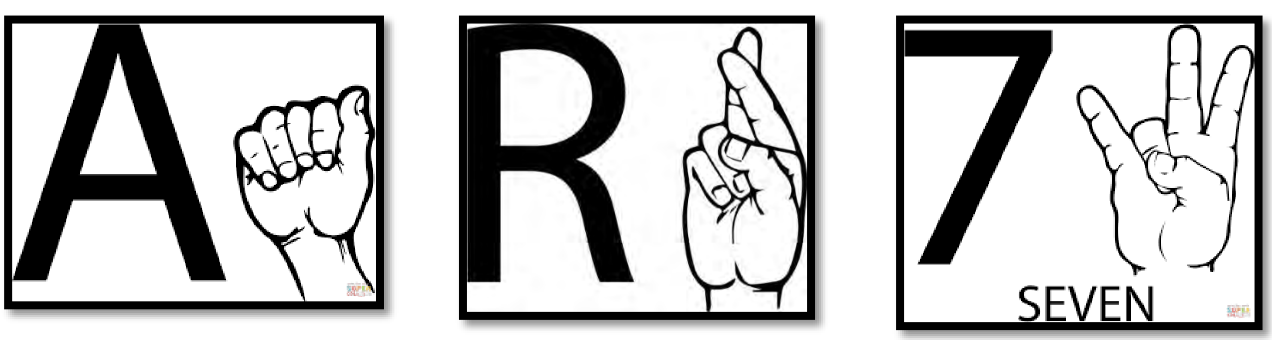
Rest of the paper is organized as follows. In chapter-02 the background study for this system has been described. The proposed model, features, process are included in chapter-03. The Result analysis, statistical analysis with graphical images are shown in Chapter-04 and we concluded the paper in Chapter-05.

Chapter 02

Background Study

2.1 First Sign Language

In the western society sign language started developing in 17th century for a visual language. It is a language with the combination of conventional gesture, hand signs, finger spelling also includes the position of the hand positions to represent a meaningful line. The first American school for the deaf was founded in 1817 by Laurent Clerc and Thomas Hopkins Gallaudet. They are said to be the first founder of American Sign Language [6]. This is actually partly true. Laurent Clerc was from Europe and taught French Sign Language. Thomas Hopkins Gallaudet brought Clerc back to America to start the first American school for the deaf. Like Abbe Charles Michel de L'Epee's school, children from all over the country traveled to attend this school, bringing their home-signs with them. These home-signs, combined with French Sign Language, became American Sign Language. Before 19th century sign languages were confined only to fixed words using finger spelling system [30]. It evolved gradually and now many researches are going on for interpreting real time sign languages to make it understanding easier for everyone. Some inventors acclaim mankind as the originators of the first sign language. This is probably true. Early man, before spoken language, probably used gesture. They most likely figured and created signs for those things they couldn't speak of. In Fig 2.1 there are examples of some American Sign Language (ASL) in (a), (b) and (c) represents accordingly "A", "R" and "7".



(a) (b) (c)

Fig 2.1 Some American Sign Languages: (a) "A", (b) "R", (c) "7"

2.2 Others Sign Language Recognition

There has been a lot of research on sign language interpretation. Most of the works have been done for ASL (American Sign Language). Different types of techniques were used to setup the model. From GES's R.H. Sapat College of Engineering, India there was a paper published on "Intelligent sign language interpretation". Their idea was to make machine understand human language and they also build a HCI (human Computer Interface). The Human Computer interface can understand voice, facial expression and hand gesture of human. They mainly focused on machine learning programming and template matching for better output generation [29].

2.2.1 Glove-based gesture recognition

Glove based gesture recognition model was proposed by Christopher Lee and Yangsheng Xu. They were able to recognize 14 letters from hand gesture and the machine could learn new model and update the model for each gesture with 10Hz rate. But after that there were a lot of glove device was designed like Power glove, Sayre Glove and Dexterous Hand master etc. [4]. In 1970 Zimmerman used VPL data glove to determine ASL language. This was the most successful available glove. It is built on original optical fiber sensors along the back of the fingers. Star-ner and Pentland established a glove-environment system that can identify only 40 American Sign Language with a rate of 5Hz.

Lam T. Phi , Hung D. Nguyen and T.T. Quyen Bui of Vietnam Academy of Science and Technology, Hanoi, Vietnam did research on Vietnamese Sign language in 2015. They attached ten flex sensors and one accelerometer. They used the flex sensors for picking up the arch of fingers and the accelerometer was used for detecting movement of hand. Depending on the hand's poses, i.e., vertical, horizontal, and movement, sign language of letters of the Vietnamese alphabets can be divided to category 1, 2, and 3, correspondingly. Firstly, the hand's posture is acknowledged. Next, if the hand's posture fits to whichever category 1 or category 2, a matching algorithm is used to detect a letter. If the posture belongs to category 3, a dynamic time warping algorithm is applied [7].

2.2.2 Real Time Based Bare Hand Gesture Recognition

In 2013 Kashmera Khedkkar Safaya and Prof. (DR.). J.W.Bakal of Mumbai did a research on Real time based bare hand gesture. In this system they used dynamic vision sensor camera for recognizing bare hand gesture. DVS that is Dynamic Sensor Camera is different from conventional camera. The difference is DVS camera only respond with pixels with temporal luminance difference. That reduce the computational cost of comparing consecutive frames to track continuous object. At first they determined the delivery point track number for each frame and checked if number of frame events is less than the given threshold. Then this delivery point is used to detect frame and recognize the hand position. In the extraction phase they calculated the wrist point and estimated the width of hand along with horizontal axis and then update the change in width from right to left with their own algorithm [5].

2.2.3 Bracelet and rings translate Sign Language (Leap Reader)

LEAP READER is a device which helps, the people who do not understand sign language, to understand them. This device is under development. The conceptual device is a combination of a ring and a bracelet [10]. Both bracelets and the set of rings are detachable [11]. One has to wear them together to detect the motion of the fingers. Ring captures the motion of the fingers and send signal to the bracelet and then the bracelet translate the gestures into audible sound or text message through the built in speaker and screen on the bracelet [11]. By this device not only common people but also special people can understand sign languages. This concept was invented by Cao Zu-Wei, Hu Ya-Chun, Huang Ching-Lan, Liao Po-Yang, Tsai Yu-Chi, and Yang Yi-Hsien inspired by Buddhist prayer beads [10]. This project is a 2013 Red Dot Design Award winner [10].

2.2.4 CopyCat: Interactive American Sign Language Game

CopyCat is a game which is designed for two purposes, one is to collect gesture data for American Sign Language (ASL) recognition system and two is to create a scope of practical application to help deaf children acquire language skills when they play the game. The system is consists of a video camera and wrist mounted accelerometers as the primary sensors. The character of the game, Iris the cat, communicates with the user with ASL. CopyCat is designed with a limited, age-appropriate phrase set. While playing the game, if a child cannot provide with the correct sign or the child's sign is poor, Iris the cat looks puzzled and tells the child to try again. If the child is able to give a strong sign, Iris the cat acts accordingly. The user can also choose options (objects shown as icons) to give instruction to Iris. If the child cannot remember the correct phrase to direct Iris, the child can click on the button bearing the picture of any object. The system shows a short video with a teacher demonstrating the correct ASL phrases. Then the child can copy the teacher to communicate with Iris. CopyCat has used a "Wizard of Oz" approach where the interpreter simulates the computer recognizer. This method allows research into the development of an appropriate game interface and also data collection to train Hidden Markov Model (HMM) based ASL recognition system [22].

2.3 Bengali Sign Language

Although in western society this language was familiarized a long time ago but In Bangladesh in 2000 CDD (Center for Disability in Development) took the step to systemize this communication process. About 2.6 million people are deaf in Bangladesh [31]. Bangla Sign language users community is the largest community among the language based minority communities in Bangladesh. Bengali Language is the 5th most widely used writing system language in terms of population. There are 11 vowels which is called a "*sôrôbôrnô* and 36 Consonants which is called *bænjônbôrnô*. For all these alphabets there are separate signs and also for meaningful words. In Fig 2.2 there are pictures of some Bengali Sign Languages (a) Ka ("ক"), (b)Kha ("খ"),(c) Cha ("ছ").

(a) (b) (c)

Fig 2.2 Bengali Sign Languages: (a) Ka ("ক"), (b)Kha ("খ"),(c) Cha ("ছ")

2.4 Bengali Sign Language Recognition

In recent research field Sign Language recognition becoming very popular. Though most of the works are on ASL (American Sign Language). Bangla Sign language recognition research field is still not widely used.

2.4.1 Normalized cross correlation

By Kaushik Deb, Helena Parvin Mony & Sujan Chowdhury of Chittagong University of Engineering and Technology, Chittagong, Bangladesh did a research using Normalized Cross Correlation for two handed sign language recognition. They proposed this model in two steps. Frist a RGB (Red-Green-Blue) color model is implemented to determine heuristically threshold value for adopting applicant regions. After the applicant regions are obtained by applying color segmentation, then actions for refining the candidate region are followed by using two different color wrist band regions and clarifying. Finally, statistically based template matching technique is used for recognition of hand sign regions. Various hand sign images are used to test the proposed method and results are presented to provide its effectiveness [9].

2.4.2 Employing Neural Network Ensemble

By Bikash Chandra Karmokar, Kazi Md. Rokibul Alam and Md. Kibria Siddiquee of Khulna University of Engineering and Technology, Khulna, Bangladesh did a research employing Neural Network Ensemble. In Pre-processing, they took input from webcam and detects skin color of hand. The skin color of the hand has been kept unique in the environment to ensure uniform detection. Then in image processing, captured image has been converted into its threshold value. Then the converted image has been normalized to 30x33 scale pixels by applying normalization process. Then feature extraction method has been applied and NCL [22] algorithm has been used to train these images [23].

2.4.3 Automatic Recognition of Bangla Sign Language

In this project, they interpreted Bangla sign language, using kinect for capturing images and used neural network for training. Their initial goal was to identify isolated signs from movements of hands [24]. They did not include any sign of Bangla alphabet. For pre-processing they used OpenNI framework and then used Artificial Neural Network in MATLAB for training [24].

Chapter 03

System Implementation

3.1 Proposed Model

The dataset used in the proposed model in Fig. 3.1 was collected and created dataset used in the proposed model was collected and created by ourselves so it is a primary dataset. The system works in two phase. First, the system will take the dataset images and apply skin detection algorithm on it and detect the skin color pixels from it. Then it will make the image binary. Then from the binary image it will identify the biggest blob and subtract all other background objects and only keep the hand on it. In the next phase for feature extraction and evaluation we are using Bag of features in category classification. It splits the image grid by grid and takes number of image patches from it. The strongest features are identified and K-means clustering is used for vector quantization. The features or bag of words are stored in the feature vector After that for multi-class SVM (Support Vector Machine) classifier has been used for categorize training and testing set for evaluation.

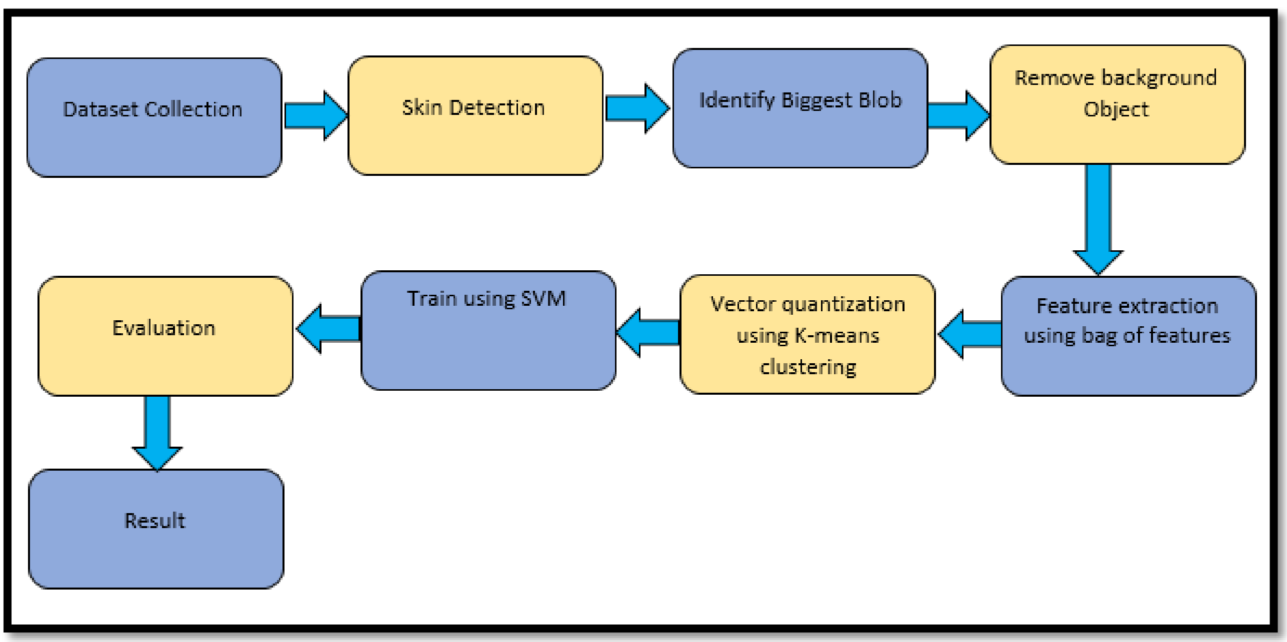


Fig. 3.1 Flowchart of Proposed Model

3.2 Dataset collection

As Bangla sign language recognition is not still a widely researched topic we did not find any dataset on any resources. Therefore, we made our own dataset with the help of our classmates and teachers. We took images of bare hands in different backgrounds. We acquired images for fifteen alphabet of Bengali Language. We took images of thirty people's bare hand in different positions. So in total in the dataset we have eight hundred and ten images. For each alphabet we are getting fifty four images of thirty people in two different posture. We are using thirty percent images for training and other seventy percent for testing and evaluation. Fig.3.2 shows the primary dataset of the system.

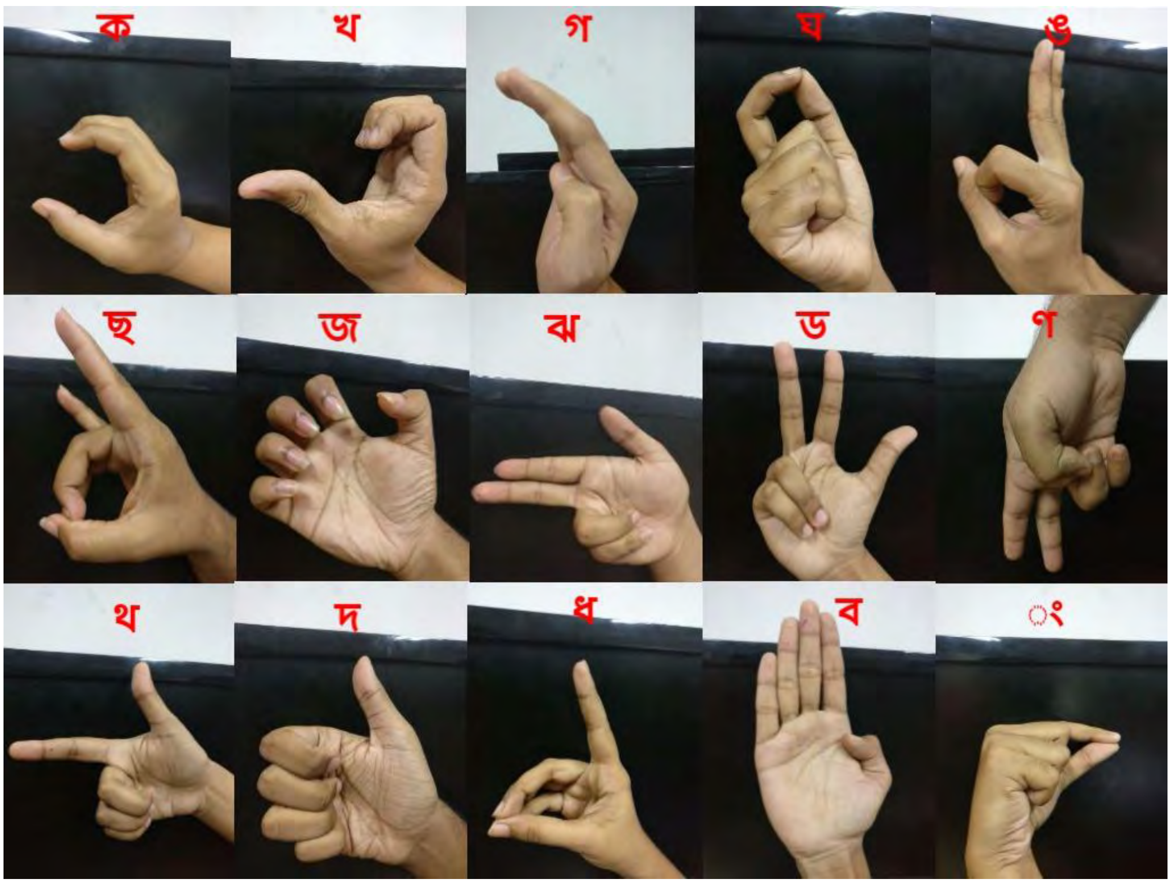


Fig.3.2 Bengali sign language in the Dataset

3.3 Skin Detection

$\mathrm{Y}C_{\mathrm{B}}C_{\mathrm{R}}$ is commonly used in video games, image processing, etc. In this process, luminance information is presented only using Y, and color information is stored as C_{B} and C_{R} which are two color-difference components. C_{B} is the difference between the blue component and a reference value, and C_{R} is the difference between the red components and a reference value [19].

As part of ITU-R BT.601 during the development of a worldwide digital component video standard, The $\mathrm{Y}C_{\mathrm{B}}C_{\mathrm{R}}$ color model was developed. It is a scaled and offset version of the YUV color model. Y is the luma component which has to have 8-bit range of 16-235. As previously stated, C_{B} and $C_{\mathrm{R}}$are the blue-difference and red- difference chroma components respectively, which has to have a nominal range of 16-240 [19].

The equation of the conversion of RGB to Y$C_{\mathrm{B}}C_{\mathrm{R}}$ color space is given below.

$$Y' = 16 + (65.481.R' + 128.553.G' + 24.966.B')$$
$$C_{\mathrm{B}} = 128 + (-37.797.R' - 74.203.G' + 112.0B')$$
$$C_{\mathrm{R}} = 128 + (112.0.R' - 93.786.G' - 18.214.B')$$

Y$C_{\mathrm{B}}C_{\mathrm{R}}$ is luma- independent for which it gives a better performance. The corresponding skin cluster is-

$$Y > 80$$
$$85 < C_B < 135$$
$$135 < C_R < 180,$$
$$Where\ Y, C_B, C_R = [0,255].$$

Chai and Ngan have developed an algorithm that exploits the spatial characteristics of human skin color. A skin color map is derived and used on the chrominance components of the input image to detect pixels that appear to be skin. Chai and Ngan have found the range of C_B and C_R most representatives for the skin-color reference [20] map which are-

$$77 \leq C_B \leq 127$$
$$133 \leq C_R \leq 173$$

Using this Y$C_{\mathrm{B}}C_{\mathrm{R}}$ skin detection method we detected skin colored pixels of the images and then converted them into binary. Then we identified the biggest blob which is the hand from the image and removed background noise. In Fig.3.3 the step by step process of skin detection has been shown. Fig3.3 (a) shows the actual image of a person who is doing a sign. In (b) the image is being converted to Y$C_{\mathrm{B}}C_{\mathrm{R}}$. In (c) the algorithm is detecting the skin and converting to binary image with some noise. Final step is shown in (d) where the noise is getting subtracted and only the detected hand is remaining.

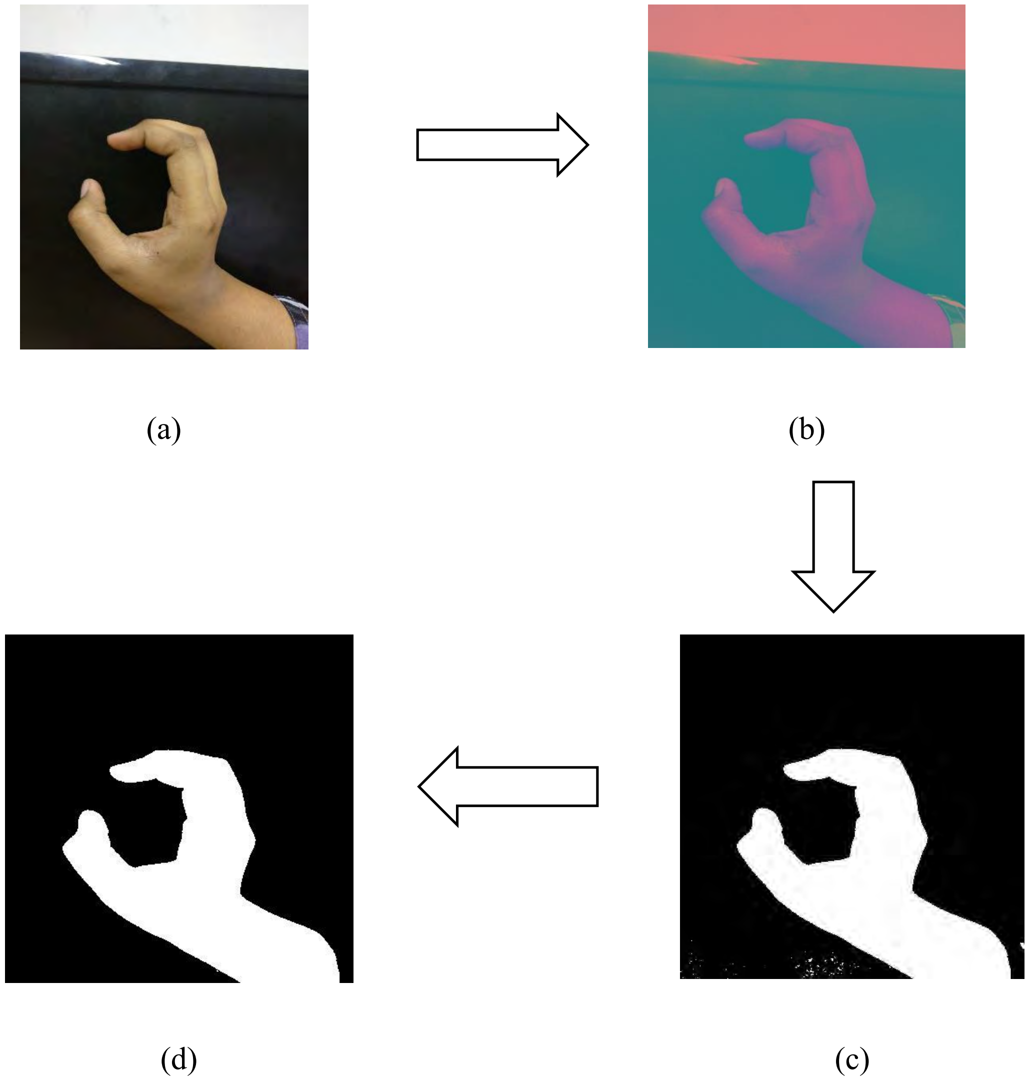


Fig. 3.3 Step by step skin detection and binarization: (a) Actual image, (b) Converted to $\mathrm{Y}C_{\mathrm{B}}C_{\mathrm{R}}$, (c) Detected skin with noise, (d) Final detected hand with biggest blob extraction

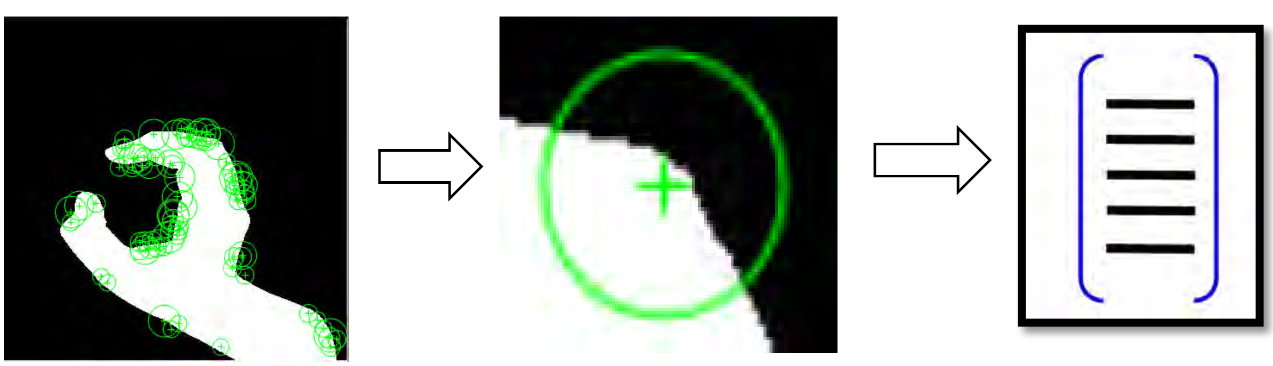
3.4 Feature Extraction

In machine learning, pattern, pattern recognition and image processing, feature extraction starts from a preliminary set of measured data and builds derived value (features) intended to be informative and non-terminated, facility the succeeding learning and generalized steps, and in some case leading to better human interpretation [25]. Feature extraction is very different from feature selection. The former consists in transforming arbitrary data, such as text or images, into numerical features useable for machine learning. The latter is a machine learning technique applied on these features [23].

3.4.1 Bag of Features

For feature extraction phase we used bag of features for category classification. Bag of Features method is one that denotes images as order less collections of local features. The name comes from the Bag of Words representation used in textual information retrieval [16]. The bag-of-words model can be applied in image classification by treating image features as words. In document classification a bag of words is a sparse vector of incident count of words [10]. This process is largely unaffected by position and orientation of object in image. It has fixed length vector irrespective of number of directions. Very successful in classifying images according to the objects they contain. But there are also some disadvantages. There is no explicit use of configuration of visual word positions. This process is poor at localizing objects within an image [11].

The Process works in four phase-extract features using SURF (Speeded up Robust Features) [32], earning visual vocabulary, Quantize features using visual vocabulary and representing images by frequencies of visual words. A visual vocabulary is constructed to represent the dictionary by clustering features extracted from a set of training images. The image features represent local areas of the image. Clustering is required so that a discrete vocabulary can be generated from millions of local features sampled from the training data. Each feature cluster is a visual word. Given a novel image, features are detected and assigned to their nearest matching terms (cluster centers) from the visual vocabulary. The term vector is then simply the normalized histogram of the quantized features detected in the image [16]. In Fig. 3.4 the process of feature extraction in Bag of feature method is shown. First in Fig.3.4 (a) Interest points are identified in the binary image for feature extraction which are represented as green marks. Next SURF descriptors are being calculated which is shown in Fig.3.4 (b).After that the SURF feature descriptors are being stored into feature vector which is shown in Fig.3.4 (c).



(a) (b) (c)

Fig. 3.4 Process for Bag of Features Image Representation, (a) Interest point detection, (b) SURF descriptors being calculated, (c) Descriptors stored in feature vector

The skin detected image is converted to 400×400 pixels for faster computing. The bag of feature uses SURF to compute the local features from the training set. SURF uses a blob detector based on Hessian matrix [33] to find interest points to extract features. To describe the region around the point, a square region is extracted, centered on the interest point and weighted sums in rectangular sub windows are calculated to describe the interest point and the feature descriptor is then put into the feature vector. Then k-means clustering is used to minimize features and calculate the visual words.

3.4.1.1 k-means clustering:

With K-means clustering process we want to minimize sum of squared Euclidean distances between points $x_i$and their nearest cluster centers m_k

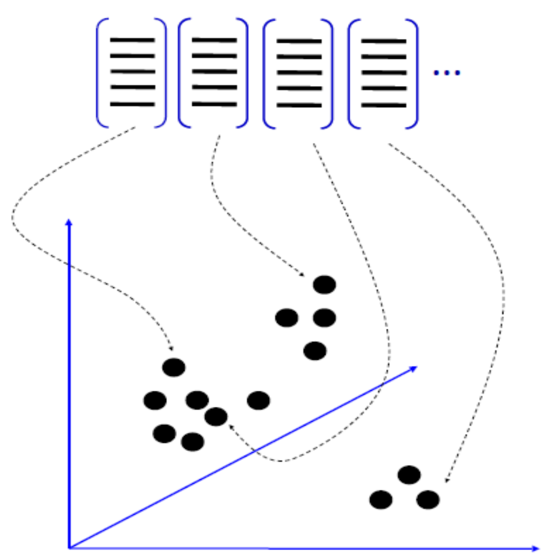
$D(x, m) = \sum_{cluster-k} \sum_{Point\ i\ in\ cluster-k} (x_i - m_k)^2$

Algorithm:

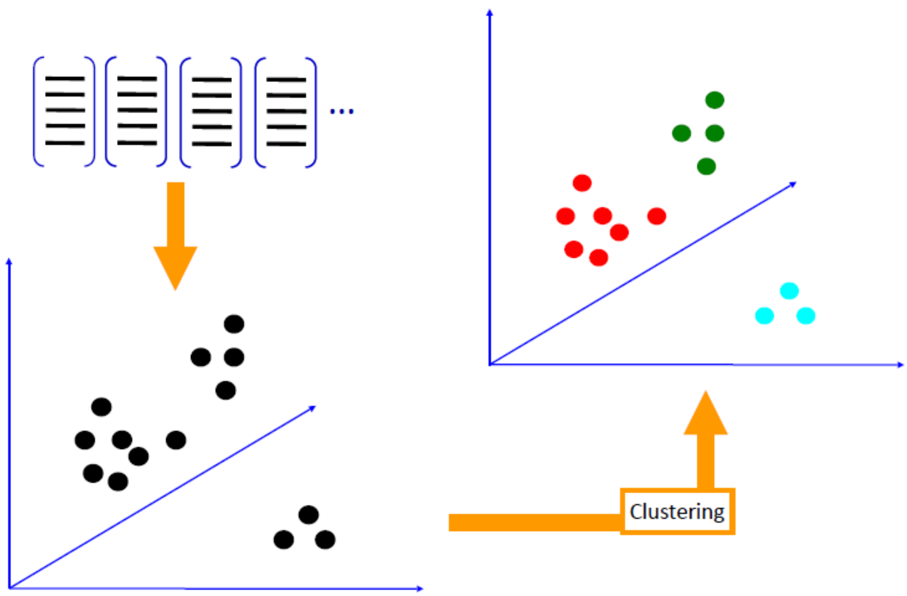
Given k:

1. Select initial centroids at random.
2. Assign each object to the cluster with the nearest centroid.
3. Compute each centroid as the mean of the objects assigned to it.
4. Repeat previous 2 steps until no change.

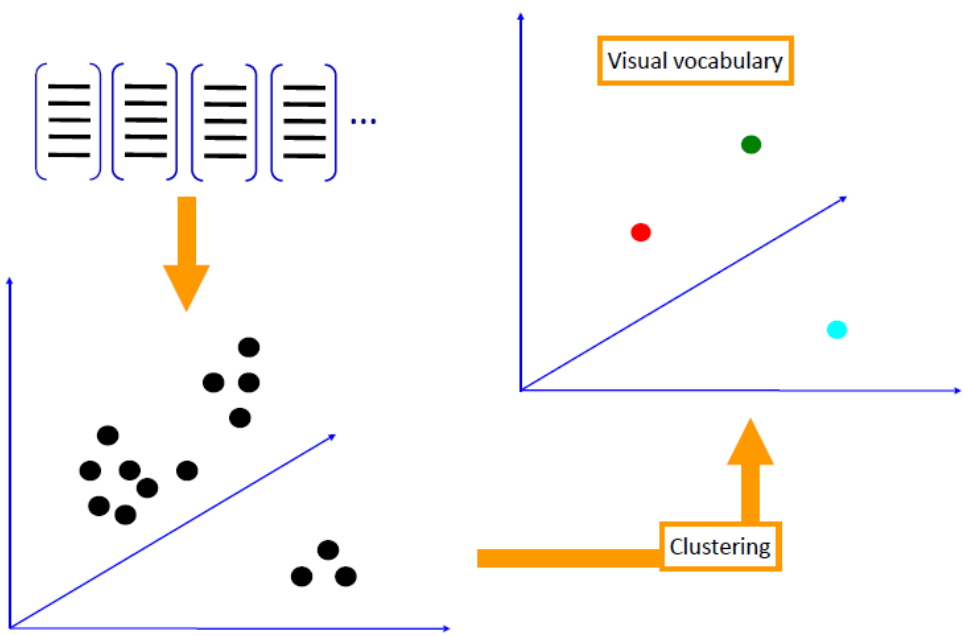
In Fig.3.5 (a), (b) and (c) the step by step process of clustering is shown. Feature vectors are placed and then initial centroid are being selected randomly. After clustering visual vocabulary is getting out.



(a)



(b)



(c)

Fig. 3.5 K-means clustering process: (a) Feature Vector computation, (b) Clustering process, (c) Visual vocabulary identification

From Clustering to Vector Quantization

Clustering is a common method for learning a visual vocabulary or codebook

- Unsupervised learning process
- Each cluster center produced by k-means becomes a code vector
- Codebook can be learned on separate training set
- Provided the training set is sufficiently representative, the codebook will be “universal

The codebook is used for quantizing features

- A vector quantizer takes a feature vector and maps it to the index of the nearest code vector in a codebook
- Codebook = visual vocabulary
- Code vector = visual word

3.5 Training of dataset

For training the dataset we used SVM (Support Vector Machine) classifier. Support vector Machines are a set of supervised learning methods used for classification, regression and outlier’s detection [18]. This is a very popular classifier in gesture recognition. It is very effective in high dimensional spaces. If the number of dimensions is greater than sample number in those cases it is still very operative. It is memory efficient. Different kernel functions can be identified for the decision function. Common kernels are provided but it is also possible to specify custom kernels. But if feature number is greater sample number, the method is likely to give poor performances. This classifier do not openly afford probability estimates, these are calculated using a costly five-fold cross validation [17].

To get *M*-class classifiers, construct set of binary classifierf^1,f^2,… … …, f^M each trained to separate one class from rest. Combining them we get a multi-class classification according to the maximal output before applying the sign function.

Argmax g^j(x) where

$$g^j(\mathrm{x}) = \sum_{i=1}^{m} y_i \propto_i^j k(x, x_i) + b^j$$

Recall: g^j(x) returns a signed real-valued value which can be taken as the distance from the separation (hyper) plane to the point *x*.

Value can also be interpreted as a confidence value. The larger the value the more confident one is that the point x belong to the positive class.

Hence, assign point x to the class whose confidence value is largest for this point. [21]

In the designed system the SVM (Support Vector Machine) classifier is trained with thirty percent of the dataset from each class which is sixteen images from each class by using the features extracted through bag of features and k-means clustering. Then the trained system is evaluated using the remaining seventy percent images which is thirty eight images for each class.

Chapter 04

Experimental Analysis and Result

4.1 Test and Train sets

In our proposed system for training and testing we have taken fifteen classes of signs. In each class we have fifty-four to fifty-six images. For implementing the system we chose thirty percentage of images for training and the other seventy percent for testing from each class randomly.

Total Class of images = 15

In one class number of images = 54 to 56

Total images = 830

For training the image classifier number of images = 16

For testing and evaluation number of images = 38

4.2 Train set vs. Train set average accuracy

After training the SVM classifier with the train set we evaluated the trained classifier on the train set and got 95% average accuracy. For some class of sign the accuracy we received was 100%. The accuracy for each class of sign is shown in Table 4.1.

Table 4.1 Train set accuracy for each class

Class	Train Set average accuracy for each class	Class	Train Set average accuracy for each class	Class	Train Set average accuracy for each class
ক	100%	ছ	88%	ঢ	94%
খ	94%	জ	94%	দ	100%
গ	94%	ঝ	100%	ণ	100%
ঘ	94%	থ	94%	ব	88%
ঙ	100%	ড	100%	ং	88%

Average accuracy of Train set vs. Train set accuracy = 95%. In Fig. 4.1 we can see the graph for Train set vs. Train set accuracy.

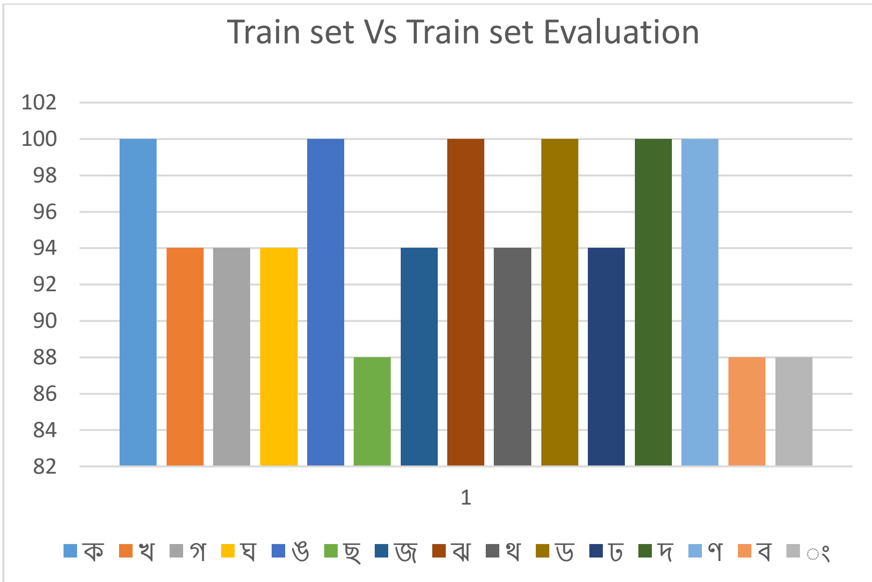


Fig 4.1 Chart for Train Set vs. Train set evaluation

For understanding the data more precisely From Fig. 4.1 another chart was made in Fig. 4.2 where the accuracy rate for train sets are plotted in lines. From Fig. 4.2 we can see that most of the accuracy rate fluctuate between 94% to 100%. Only in two class of signs the accuracy is 88%.

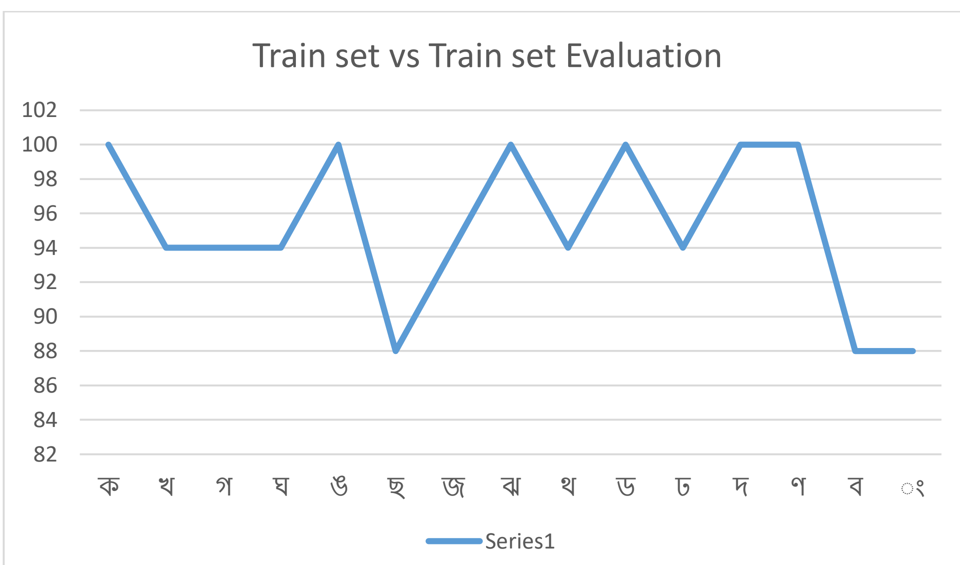


Fig 4.2 Graph for Train Set vs. Train set evaluation

4.3 Train set vs. Test set average accuracy

After training the classifier we tested on the other seventy percentage images that is thirty-eight images from each classes. The average accuracy for the test set we received is eighty-six percentage. Fig. 4.3 shows the accuracy rate of Train set vs. Test set.

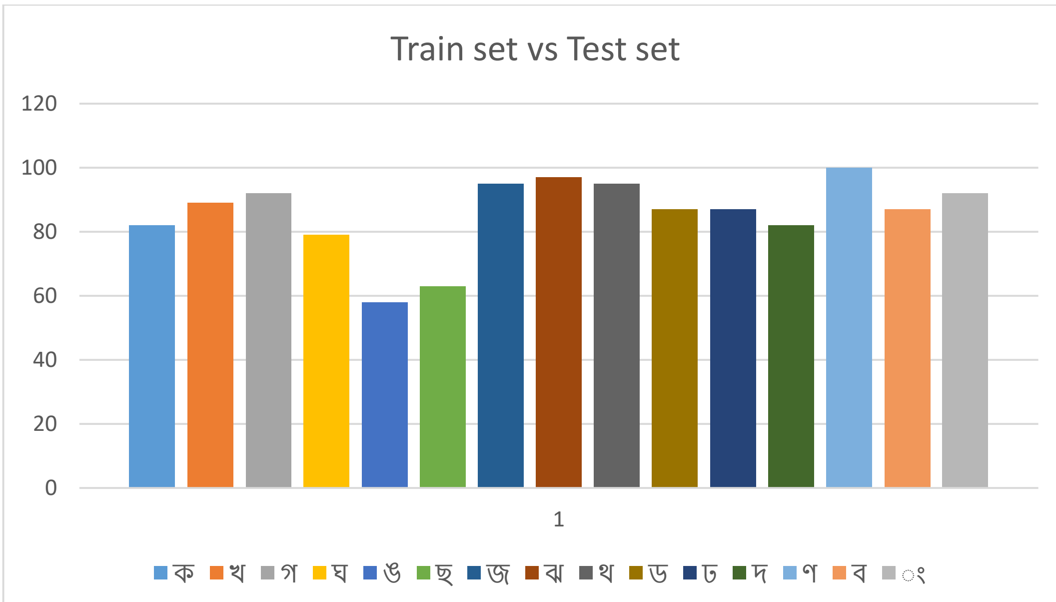


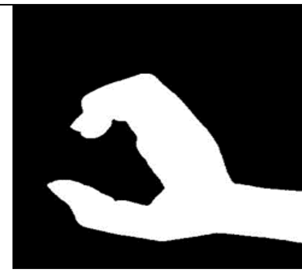
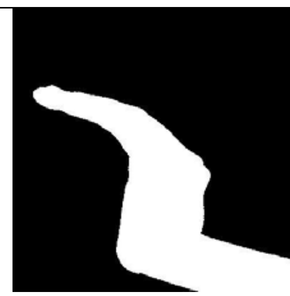
Fig 4.3 Chart for Train Set vs. Test set evaluation

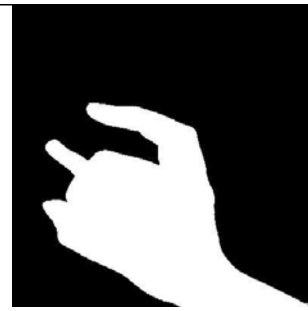
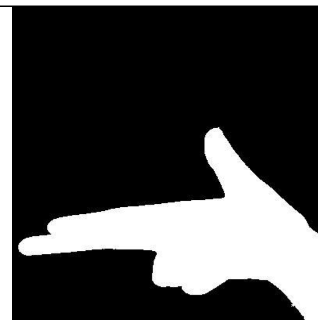
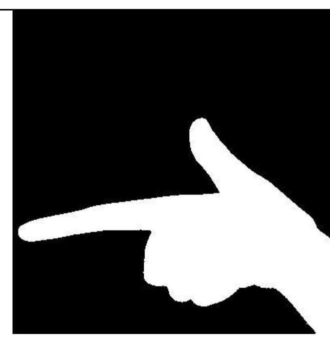
In Fig. 4.4 the graph was plotted again in line for a clear vision. The accuracy rate fluctuates between 82% to 98% but only in two classes fifth and sixth class the accuracy rate goes very low.

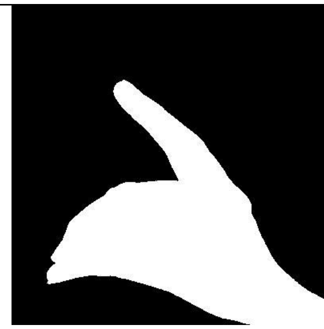
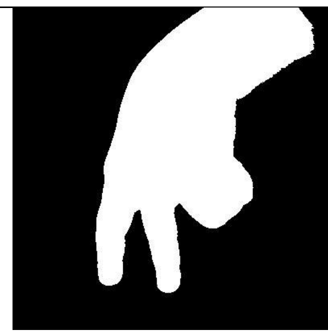
Fig 4.4 Graph for Train Set vs. Test set evaluation

Table 4.2 shows each class with their evaluation accuracy rate and binary noise free image.

Table 4.2 Test set accuracy for each class

Class	Evaluation Set average accuracy for each class	Binary Image of Sign
Ka(ক)	82%	
Kha(খ)	89%	
Ga(গ)	92%	
Gha(ঘ)	79%	
Umo(ঙ)	58%	

Cha(ছ)	63%	
Jo(জ)	95%	
Jha(ঝ)	97%	
Tho(থ)	95%	
Do(ড)	87%	

Dha(ঢ)	87%	
Da(দ)	82%	
Murdho-No(ণ)	100%	
Bo(ব)	87%	
Anussar(ং)	92%	

Evaluation set average accuracy = 86%

4.4 Comparison between Test set accuracy and Train set accuracy

Fig. 4.5 shows the comparison rate between the systems train set and test set accuracy. From this Fig. 4.6 was plotted showing the comparison between two lines test set and train set. From Fig. 4.5 it is visible that the difference between test set and train set accuracy is not so different for all classes.

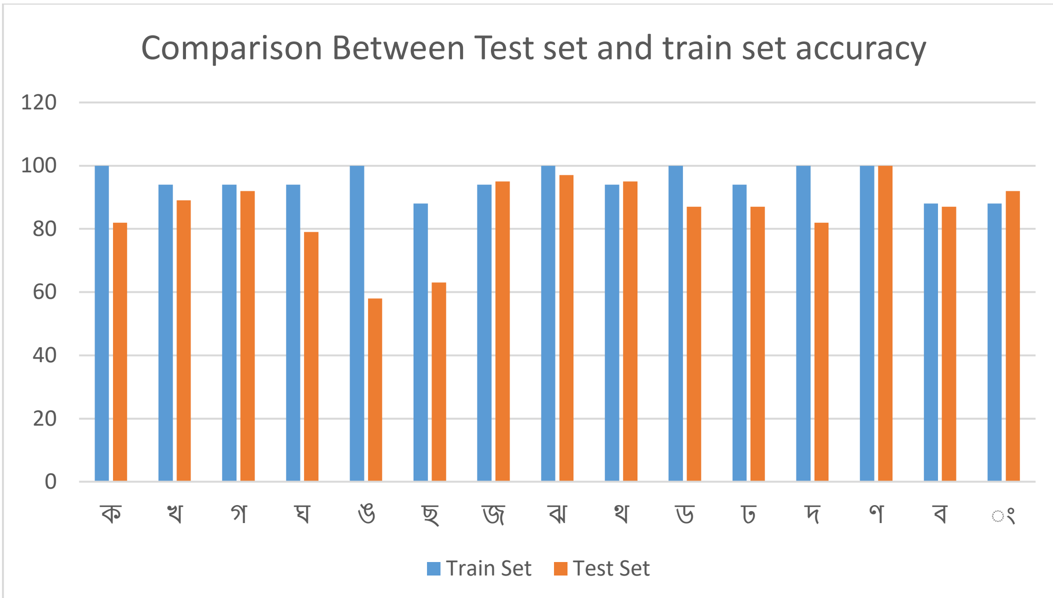


Fig 4.5 Chart showing comparison between test set accuracy and train set accuracy

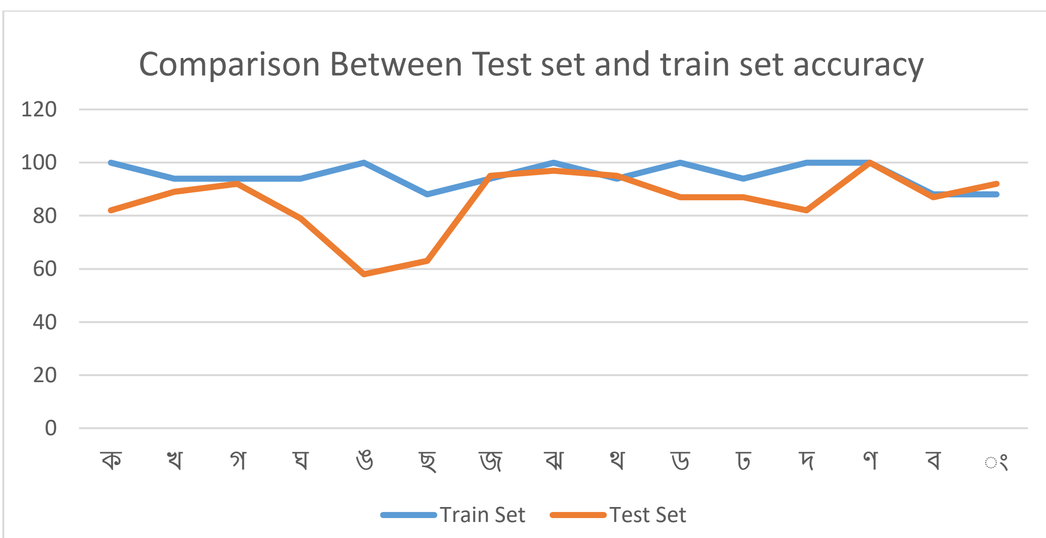


Fig 4.6 Graph showing comparison between test set accuracy and train set accuracy

4.5 Comparative analysis of Bengali Sign Language Detection Techniques

Table 4.3 Comparison between proposed system and other systems.

Author reference	**Technique Used**	**Sign Language**	**Average Accuracy**
Adnan Eshaque, Tarek Hamid, Shamima Rahman, and M.Rokonuzzaman [26]	Matching signs with previously used knowledge based	Bengali	82%
M. Atiqur Rahman, Ahsan-Ul-Ambia[27]	Multillayer Feedforward ANN	Bengali	80.9%
Manar Maraqa, Farid Al-Zboun, Mufleh Dhyabat and Raed Abu Zitar[28]	Multillayer Feedforward ANN	American	79.33%
Proposed Model	Bag of features and SVM classifier	Bengali	86%

In [27] for Bengali Sign Language recognition they used Artificial Neural Network and the accuracy they got 80.902%. Their dataset was not enough for getting a better output from ANN recognition method. Using of other statistical recognition method like Hidden Markov Model (HMM) with more features could improve their accuracy rate. Also there was a limitation that for learning Neural Network (NN) the feature vector should have been integer values.

With the use of Artificial Neural Network another research was made [28] where they studied the American Sign Language. The accuracy they got was 79.33%. If they used any other recognition method the accuracy may have gotten a little higher.

Chapter 05
CONCLUSIONS AND FUTURE WORKS

5.1 Conclusion

Without sign language, deaf and dumb people would have totally been ignored worldwide. As our country is a populous country, the percentage of these kind of people is more than average. Therefore, for the development of our country involvement of these special people is very much necessary.
Our country is also a developing country. Still many of our people are not well educated and some of them are illiterate. Huge percentage of people of our country do not know English efficiently. As Bangla is our mother tongue, it is easier for everyone to interconnect in Bangla in our country. Hence, Bangla sign language is necessary for special people to communicate with the people of all sphere of life in our country. While developing the system, we have focused on the special people of all classes of our country. Through our system, those hearing and listening impaired people can communicate with all kind of people. This will make their life as like as normal people. And this will play an important role in the development of our country.

5.2 Future Work

In the future we will try to make our system more efficient. We only tested for fifteen Bengali alphabets as we did not have enough resources for our dataset. We had to create all the datasets by ourselves which took a lot of time. In future we will measure the accuracy for all the Bengali alphabets and even for numerical numbers too. We proposed this model based on still images but up next we will try to make it in real-time.

In sign language, there are also individual signs. There is individual signs for chair, table, book, eat, and so many more. If we try to include all, it will be a huge dataset. We will try to include them too.

Reference

[1] Hsiao, Ko-Jen, Tse-Wei Chen, and Shao-Yi Chien. "Fast fingertip positioning by combining particle filtering with particle random diffusion." *Multimedia and Expo, 2008 IEEE International Conference on*. IEEE, 2008.

[2] Darroch, Kathy, and Liza Marshall. "Interpreting. NETAC Teacher Tipsheet."

[3] D. Shahriar Hossain Pavel, Tanvir Mustafiz, Asif Iqbal Sarkar, M. Rokonuzzaman, "Geometrical Model Based Hand Gesture Recognition for Interpreting Bengali Sign Language Using Computer Vision", ICCIT, 2003.

[4] Lee, Christopher, and Yangsheng Xu. "Online, interactive learning of gestures for human/robot interfaces." *Robotics and Automation, 1996. Proceedings, 1996 IEEE International Conference on*. Vol. 4. IEEE, 1996.

[5] Safaya, Kashmera Khedkkar, and Asst Prof Rekha Lathi. "Bare Hand Gesture Recognition-A study." (2012).

[6] Bellugi, Ursula. *The signs of language*. Harvard University Press, 1979.

[7] Phi, Lam T., et al. "A glove-based gesture recognition system for Vietnamese sign language." *Control, Automation and Systems (ICCAS), 2015 15th International Conference on*. IEEE, 2015.

[8] Ahmed, Syed Tauhid, and M. A. H. Akhand. "Bangladeshi Sign Language Recognition using fingertip position." *Medical Engineering, Health Informatics and Technology (MediTec), 2016 International Conference on*. IEEE, 2016.

[9] Kaushik Deb, Dr, et al. "Two-handed sign language recognition for bangla character using normalized cross correlation." *Global Journal of Computer Science and Technology* 12.3 (2012).

[10] Nowak, Eric, Frédéric Jurie, and Bill Triggs. "Sampling strategies for bag-of-features image classification." *Computer Vision–ECCV 2006* (2006): 490-503.

[11] "INRIA Summer School 2010 Cordelia Search Large | Information Retrieval". *Scribd*. Web. 2017.

[12] Bouchard, Guillaume,and Bill Triggs. "Hierarchical part-based visual object categorization." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005.

[13] Alexander, Voskresensky, and Ivanova Marina. "A Methodical and Didactical Complex "Sign Language"".

[14] Darroch, Kathy, and Liza Marshall. "Interpreting. NETAC Teacher Tipsheet." (1998).

[15] Pavel, Dewan Shahriar Hossain, et al. "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*. 2004.

[16] O'Hara, Stephen, and Bruce A. Draper. "Introduction to the bag of features paradigm for image classification and retrieval." *arXiv preprint arXiv:1101.3354* (2011).

[17] Suykens, Johan AK, and Joos Vandewalle. "Least squares support vector machine classifiers." *Neural processing letters* 9.3 (1999): 293-300.

[18] Zisserman, A. "Lecture 2: The SVM classifier." *C19 Machine Learning* (2011): 12.

[19] Mahmoud, Tarek M. "A new fast skin color detection technique." *World Academy of Science, Engineering and Technology* 43 (2008): 501-505.

[20] Chai, Douglas, and King N. Ngan. "Face segmentation using skin-color map in videophone applications." *IEEE Transactions on circuits and systems for video technology* 9.4 (1999): 551-564.

[21] Cheong, Sungmoon, Sang Hoon Oh, and Soo-Young Lee. "Support vector machines with binary tree architecture for multi-class classification." *Neural Information Processing-Letters and Reviews* 2.3 (2004): 47-51.

[22] Lee, Seungyon, et al. "A gesture-based american sign language game for deaf children." *CHI'05 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2005.

[23] Guyon, Isabelle, et al., eds. *Feature extraction: foundations and applications*. Vol. 207. Springer, 2008.

[24] Choudhury, Najeefa Nikhat, and Golam Kayas. *Automatic recognition of Bangla sign language*. Diss. BRAC University, 2012.

[25]Guyon, Isabelle, and André Elisseeff. "An introduction to feature extraction." *Feature extraction*. Springer Berlin Heidelberg, 2006. 1-25.

[26] Eshaque, Adnan, et al. "A novel concept of 3d animation based'intelligent assistant'for deaf people: for understanding bengali expressions." *International Conference on Computer and Information Technology*. 2002.

[27] Rahman, Md, Md Abdullah, and S. Mondal. "Recognition of Static Hand Gesture of Alphabet in Bangla Sign Language‖." *Conference (IOSRJCE)*. Vol. 8. No. 1. 2012.

[28] Maraqa, Manar, and Raed Abu-Zaiter. "Recognition of Arabic Sign Language (ArSL) using recurrent neural networks." *Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference on the*. IEEE, 2008.

[29] Pramada, Sawant, et al. "Intelligent sign language recognition using image processing." *IOSR Journal of Engineering (IOSRJEN)* 3.2 (2013): 45-51.

[30] Baker-Shenk, Charlotte Lee, and Dennis Cokely. *American Sign Language: A teacher's resource text on grammar and culture*. Gallaudet University Press, 1991.
[31] Alauddin, Mohammad, and Abul Hasnat Joarder. "Deafness in Bangladesh." *Hearing Impairment*. Springer Japan, 2004. 64-69.
[32] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." *Computer vision–ECCV 2006* (2006): 404-417.
[33] Hiriart-Urruty, Jean-Baptiste, Jean-Jacques Strodiot, and V. Hien Nguyen. "Generalized Hessian matrix and second-order optimality conditions for problems withC 1, 1 data." *Applied Mathematics & Optimization* 11.1 (1984): 43-56.