# Automatic Recognition of Bangla Sign Language

Thesis submitted in partial fulfillment of the requirement for the degree of

Bachelor of Science
In

## Computer Science and Engineering

Under the Supervision of
**Dr. Mumit Khan**
And
Co-Supervision of
**Md. Zahangir Alom**

By
**Najeefa Nikhat Choudhury**
**ID:09301024**
And
**Golam Kayas**
**ID:09110026**

To

**School of Engineering and Computer Science,**
**Department of Computer Science and Engineering, BRAC University**
**66 Mohakhali C/A, Dhaka-1212**

**December 2012**

# Declaration

This is to certify that the Thesis entitled ”Automatic Recognition of Bangla Sign Language” which is submitted by Najeefa Nikhat Choudhury (ID:09301024) and Golam Kayas (ID: 09110026) in partial fulfillment of the requirement for the award of degree of BSc in Computer Science and Engineering to the Department of Computer Science and Engineering, BRAC University, 66 Mohakhali C/A, Dhaka-1212, comprises only their original work and due acknowledgement has been made in the text to all other material used. The results of this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

**Approved By:**

Supervisor : Dr. Mumit Khan

Co-Supervisor: Md. Zahangir Alom

**Date: 12th December 2012**

# Abstract

Sign Language is the mode of communication among the deaf and dumb. However, integrating them into the main stream is very difficult as the majority of the society is unaware of their language. So, to bridge the communication gap between the hearing and speech impaired and the rest in Bangladesh, we conducted a research to recognize Bangla sign language using a computer-vision based approach. To achieve our goals we used Neural Networks to train individual signs. In the future, this research, besides helping as an interpreter, can also open doors to numerous other applications like sign language tutorials or dictionaries and also help the deaf and dumb to search the web or send mails more conveniently.

# Acknowledgement

Firstly, we would like to thank our supervisor, Dr. Mumit Khan and co-supervisor Jahangir Sir for guiding us through our bachelor thesis. Also, we are extremely grateful to Dr. Khalilur Rahman for providing us with Kinect depth camera for the purpose of our experiments.

We would like to thank the Centre for Disability in Development (CDD) for providing us with reading materials and information on Bangla sign language. We would also like to convey our special thanks to the teachers of Dhaka Bodhir High School for teaching us signs necessary for our experiment and for being highly co-operative and informative.

Special thanks goes to Ahnaf Atef Choudhury (Najeefa's brother) for performing signs required for the training data. Finally, we thank our family, friends and all our teachers for their motivation and support.

# Table of Contents

# List of Figures and Tables

## Chapter 2: Literature Review

### Section 2.1 Sign Language

### Section 2.3 Kinect

### Section 2.5 Artificial Neural Network

## Chapter 3: Sign Language Recognition

### Section 3.1: Tracking

### Section 3.2: Feature Extraction

## Chapter 4: Experimental Results and Evaluation

### Section 4.2: Results Obtained

## APPENDIX

# Chapter 1: Introduction

## 1.1 Motivation and Goals

Sign language is the mode of communication among the deaf and dumb. A common misconception about sign language is that it is universal, which is however not the case. In fact, sign languages, just like spoken languages are unique to a culture and have evolved over time. Moreover, they feature their own grammar and vocabulary and are generally acquired by deaf children as their mother tongue.

Sign languages use manual gestures and body language to convey meaning. Static signs are generally used for alphabets and numbers where hand shapes define each sign. On the other hand, words and sentences are generally expressed through a combination of hand shape, orientation and movement of hands and arms. Additionally, facial expressions exhibit emotions and sometimes head movement, shoulder position, body posture and lip patterns are important parameters in expressing the meaning of a sign. Hundreds of sign languages are in use around the World today, some of which have not yet gained any legal acceptance. In Bangladesh, the Centre for Disability in Development (CDD) has developed a formal sign language for the Bengali deaf and dumb community, which is followed by schools for the speech and hearing impaired countrywide.

Incorporating the deaf and dumb into the mainstream is difficult, mainly due to a lack of knowledge about sign language by the rest of the society. So, as to bridge this communication gap, scientists have been researching on methods to develop automatic sign language recognition systems. This field of research is still far behind and struggling. Moreover, research on recognition of Bangla sign language has not prevailed as it has for some other sign languages. So, our goal is to conduct a research to recognize Bangla sign language.

There are several technologies that can be and has been employed in sign language or gesture recognition. For the purpose of our thesis we are using a computer-vision based approach with the help of Kinect Depth Camera and Neural Networks to recognize signs.

For now, we are limiting recognition of isolated Bangla signs only and are considering only manual features for our experiment.

### 1.2 Report Overview

The following report consists of five chapters. Chapter 1 is the Introduction, which highlights the motivation and goals behind the thesis. Chapter 2, titled Literature Review, outlines all information relevant to the thesis and is divided in five sections. Section 2.1, 2.2, 2.3, 2.4 and 2.5, titled Sign Language, Related Work, Kinect, OpenNI and Processing and Neural Network respectively. Section 2.2 gives a detailed description of research conducted in sign language recognition so far.

Chapter 3 illustrates the details of our thesis experiment spanned over four subsections. Section 3.1, 3.2 and 3.3 explains each phase of the recognition, i.e. Tracking, Feature Extraction and Training in depth. In Chapter 4 we have mentioned the results we obtained and our evaluation of that result. Finally, Chapter 5 is a small conclusion suggesting improvements and featuring future expectations.

An Appendix, which lists all signs we used for our experiment, follows the five sections and at the end there is a Bibliography of reading materials we have referred to during our research.

# Chapter 2: Literature Review

## 2.1 Sign Language

A sign language is a language that instead of using sound to communicate uses manual gestures and body language to convey meaning. The hearing or speech impaired generally uses this mode of communication and today it is extensively studied by linguists.

Many used to believe that sign languages are not real languages, however with the result of studies by linguists this concept is gradually changing. Sign languages contain all necessary components that are present in spoken languages and therefore throughout the world, many nations have started to establish it as a real language. Many also believed that sign languages are dependent on spoken languages, but this is also not the case.

Wherever deaf communities exist, so do sign languages and with increase in the population of sign users, sign languages have been standardized in many countries.

### 2.1.1 Sign as a Language

Despite the common misconception that sign languages are not real language, they are as rich and complex as any spoken language. Sign language is a natural language and is not made up. The acquisition process of sign language is similar to that of spoken languages as they are also adopted by deaf children in school from friends and teachers or at home from parents [4][1]. Also, sign language are closely related to the culture the deaf belongs to and to completely understand a particular sign language it is important to have a clear understanding of the culture it originated in. There is not a universal sign language worldwide, instead there are many national variants and in fact there are also many local dialects in sign even within the same country [4].

Sign languages are not equivalent to pantomime. It is not linked to iconic contents and abstract ideas can as easily be expressed as in any natural language. Also, sign languages are not related to the spoken language of a community. Alphabets in the spoken language are often used to finger spell proper nouns and unknown words, other than that sign languages are completely independent of the spoken language. They come with their own vocabulary and grammar, which is rich in content and feature their own rules. Countries that share a common language, for example, Britain and America, have different sign languages both in terms of grammar and vocabulary. Thus, it is impossible for someone who knows American Sign Language to communicate in British Sign Language unless they learn it [4].

Therefore, sign languages today have been established as natural languages, independent of any other language. Though it has taken a lot of time and debate to make these considerations, the facts eventually led the world to accept these concepts regarding sign languages.

### 2.1.2 Components and Rules

Hands are the basic means of communicating using sign languages. Hand shapes, hand movement, palm orientation, and hand position are some of the most important components to convey the meaning of a sign [8]. However, signs are not confined to manual features, i.e. hands and arms only. Non-manual features, such as head position, head tilt, body posture, eye movement and lip shapes are also important parameters used in conveying meaning of a sign [1]. In fact, facial features are very important to express emotions. Some of these features are also important to differentiate between questions, negations and affirmations.

Some signs are shown using both hands, while others with only one hand. The right hand is generally called the dominant hand and is used to convey almost all signs unless the signer is left-handed. Signs can be either static or dynamic. In dynamic signs using two hands, both hands can move or one might be static while the other is in motion. In such a case, it is generally the dominant hand that is in motion, while the other is at rest. If both hands are moving simultaneously in a sign, it is important that the hand shapes of both hands are same, but if only one hand is moving at a time in a two-hand sign then the shapes of the two hands can differ [8][1].

One important fact to consider about sign languages is that, same hand shapes or same hand motion can be used to express different signs. For example, in Bangla sign language, window and clean have the same hand movement but the hand shape for the two signs are different. Similarly, picture and table are signed using the same hand shape and palm orientation, but different hand movement. The images of Bangla signs of window and clean and picture and table, illustrating their similarities and differences, are given in figure 2.1 (a) and 2.1 (b) respectively.

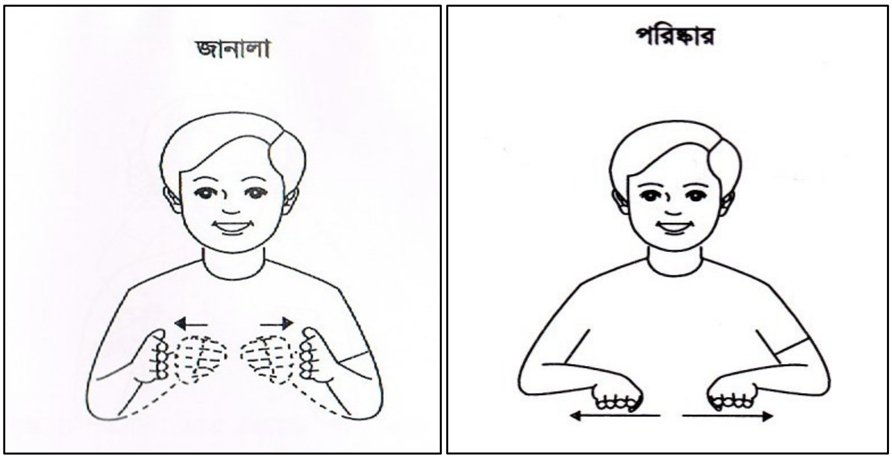


**Figure 2.1 (a) Bangla Sign for Window (left) and clean (right) (figure taken from [8]) Same hand motion (same direction). Different hand shape.**

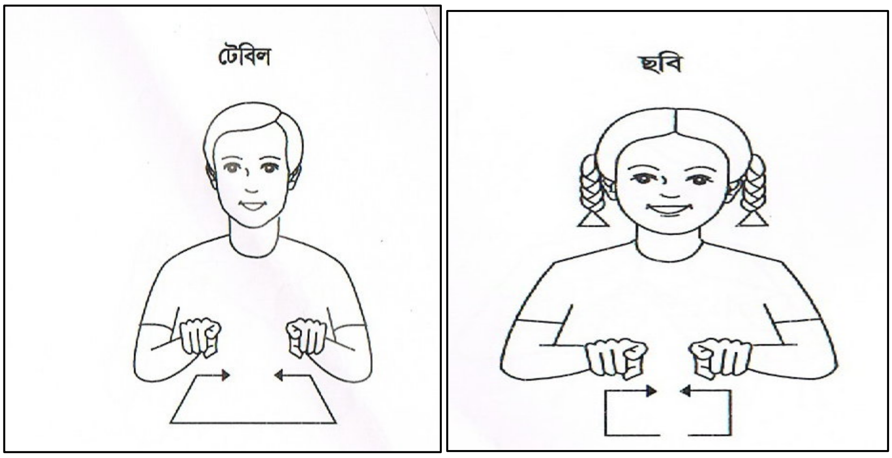


**Figure 2.1 (b) Bangla Sign for Table (left) and Picture (right) (figure taken from [8]) Same hand shape. Different hand motion (different plain).**

Another very important feature to consider is the signing space. All signs are generally produced within this space. The signing space includes the area around the head, the belly and both arms and is shown in figure 2.1 (c). During a sign, if objects and people are present around the signer then they can be referenced by pointing directly at them rather than signing them out.

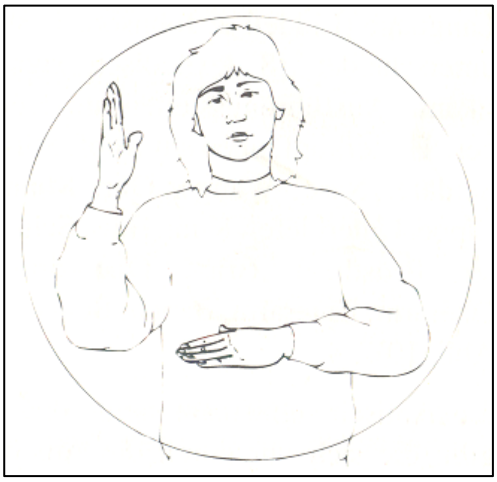


**Figure 2.1 (c) Signing Space (figure taken from [1])**

As stated before, sign language has it's own grammar, which generally does not match the spoken language of that region. Also, just like spoken languages, grammars of different sign languages are different. Signs can also indicate tense and the rules are specific for different sign languages. Moreover, there are many combinations of hand shapes that are not permitted in sign language just like there are illegal alphabetical combinations in spoken languages, for example, the combination "pf" in English, and these rules are again specific to each sign language [1].

To understand a sign completely, all components must be correctly used. Moreover, there are many rules that need to be followed to express a sign accurately. Some of these components and rules might be universal, while others are completely local to a specific culture, society or region.

### 2.1.3 Bangla Sign Language

In Bangladesh a formal sign language has been established only recently. In the year 2000, Center for Disability in Development (CDD) took the initiative to standardize communication with sign languages in this country. Before this step, there were different local variants and no national dialect existed. CDD has published many books rich in vocabulary and grammatical rules and they also provide sign language training in their training center [8]. Other than CDD, there is only one high school for deaf children in Bangladesh, Dhaka Bodhir High School. People in Bangladesh still are ignorant of this mode of communication and thus the deaf children still cannot lead an uncomplicated life here. However, measures are being taken in order to aware the people of Bangla Sign Language and we can hope that eventually life will become much easier for the hearing and speech impaired in the future.

## 2.2 Related Work

A great deal of work has been done in the area of text to sign language conversion. On the other hand, sign language recognition and sign-to-text conversion is relatively less matured. However, there have been recent breakthroughs in the field and the research is only growing.

In 1977 a robotic hand capable of spelling words using alphabets was developed [5]. The initial model was however unable to form letters that required wrist movement. By 1992, a robotic hand was devised that could fluidly produce letters received from text telephone. Finally, in 1994 a fourth generation computer-controlled electromechanical finger spelling robotic hand was developed called Ralph. This robotic hand could accept input from various different sources.

An early approach in sign language recognition in 1991 by Takahashi and Kishino relied on users wearing wired gloves, which only extracted hand shapes and thus was limited to finger spelling or static gestures [1]. In 2002, Ryan Patterson designed a simplistic hand glove that sensed the hand movements of the signed alphabets and then wirelessly transmitted the data to a portable device that displayed the text on the screen. This glove had to be trained for individual signs and was limited to finger spelling [5]. Later, other more complex and efficient systems of sign language recognition using gloves were introduced, some of which are the CyberGlove, VPL Data Glove and the AcceleGlove.

An approach in 1995 by Starner and Pentaland featured real time recognition of American Sign Language from video using Hidden Markov Models. In this approach, the signers were required to wear specially colored gloves for hand tracking. In 1998, Vogler and Metaxas used three orthogonally placed webcam to extract body features and then used these 3D data as input to HMMs for continuous signer-dependent sign language recognition. Later in 2007, Dreuw et al. proposed a signer-independent recognition system based on speech recognition techniques, using a single webcam and without the need of wearing any gloves [5][1].

A more recent approach in 2009 by Kelly et al. incorporated non-manual features, namely head movement unlike its predecessors, which mainly focused on manual features. The system relied on a single webcam and the user wearing colored gloves for continuous sign language recognition. The Sign2 Conversion System was designed to convert ASL to written and spoken English. The current system only involves finger spelling in a controlled environment using computer vision; the long-term goal is translation of full sentences in natural environments [10][5]. CopyCat, a game designed for deaf children to develop their working memory and language skills while playing uses a 2D camera and wired gloves for sign recognition [5]. Currently a larger EU-funded project called SignSpeak is ongoing and was introduced by Dreuw et al. It is being built on previous work and focuses on translating continuous sign language to text using a 2D camera and recognizing both manual and non-manual features [9].

Besides all the research mentioned above, students and scientists alike all over the world are conducting many others with a common aim of creating a bridge between the hearing and speech disabled and the not. In Bangladesh, very little has been tried in this field and most of what was started as thesis research, has not been continued after the instigation. Some of the mentionable researches on sign language recognition conducted in Bangladesh are recognizing two-handed Bangla characters using Normalized Cross Correlation [6] and creating 3D models of Bangla signs signed by the deaf community using geometric calculations [7].

### 2.3 Kinect

Kinect was developed by Microsoft and PrimeSense and was released on November 2010. Kinect combines an RGB camera, a depth sensor and a multi array microphone. The best feature of the Kinect camera is its depth sensor, which uses an infrared projector and a CMOS sensor and is capable of tracking users in 3D independent of the lighting condition [3]. Initially developed for the Xbox 360 video game console and windows PC, the Kinect camera is now being used by the computer vision community and many programmers as they realized that this depth sensing technology could be used for many purposes other than gaming [2][3].

This report does not include the details on how Kinect works. However, below is a table for some important hardware specifications [2].

| **Property** | **Values** |
|---|---|
| Angular Field-of-View | 57° horizontal, 43° vertical |
| Frame Rate | Approximately 30 Hz |
| Nominal Depth Range | 0.85 m-0.35 m |
| Nominal Depth Resolution (at 2 m distance) | 1 cm |
| Device Connection Type | USB and External Power Supply |

**Table 2.3 (a) Kinect Hardware Specifications (table data taken from [2])**

Currently, there are three software frameworks available for Kinect, Microsoft SDK, OpenNI and OpenKinect developed by Microsoft, PrimeSense and the hacker community respectively. In 2011, Microsoft and PrimeSense released their versions of frameworks for Kinect. But, before them, the hacker society came up with OpenKinect by reverse engineering the USB stream of data from the Kinect device. These frameworks led to the possibility to develop non-commercial products and thus presenting the computer science community with an excellent and cheap tool to build and research on different computer vision technology [2][1].

Although these frameworks can be used to do skeleton tracking, they do not support tracking of individual fingers and thus hand shape cannot be recognized using these frameworks. For the purpose of our thesis experiments, we have used the OpenNI framework by PrimeSense.

### 2.4 OpenNI and Processing

The OpenNI or Open Natural Interaction driver, released by PrimeSense, has a feature-rich open source framework and can be combined with closed source middleware called NITE for user skeleton tracking and hand gesture recognition [11]. In our thesis we have used the SimpleOpenNI library and it is a wrapper for Processing.

Processing is an open source programming language and environment, which helps create images animations and interactions. The language builds on the Java language, but has a simplified syntax and graphics programming model [12]. Processing also provides a Java wrapper and has been used for our thesis project.

### 2.5 Artificial Neural Networks

Artificial Neural Networks (ANNs) are non-linear mapping structures based on the function of the human brain. They are powerful tools for modeling, especially when the underlying data relationship in unknown. ANNs can identify and learn correlated patterns between input data sets and corresponding target values. After training, ANNs can be used to predict the outcome of new independent input data ANNs imitate the learning process of the human brain and can process problems involving non-linear and complex data even if the data are imprecise and noisy. ANNs have been used for a wide range of applications where statistical methods are traditionally employed. ANNs are being used to solve problems, such as logistic regression, Bayes analysis, multiple regressions etc.

### 2.5.1 Basic Architecture of ANNs

An ANN in basically composed of three types of layers, the input layer, output layer and other hidden layers. Each layer is elaborated below:

1. **Input Layer:** Input nodes take input of the system; according to which the output of the ANN is generated. Usually there is only one input layer.

2. **Output Layer:** There is only one output layer in ANN usually. Output nodes are equal to the number of the outputs of the system.

3. **Hidden Layers:** There is no magic formula to select optimum number of hidden nodes. However some thumb rules are available to calculate the number of hidden nodes. A rough approximation may be geometric pyramid rule proposed by MASTERS(1993). For three layer architecture with n input nodes and m output nodes will have (n*m) hidden nodes. There may be multiple hidden layers as well. Hidden layers provide the ANNs with it's ability to generalize.

Figure2.5.1 (a) provides a picture illustration of the concepts of Artificial Neural Network Layers:

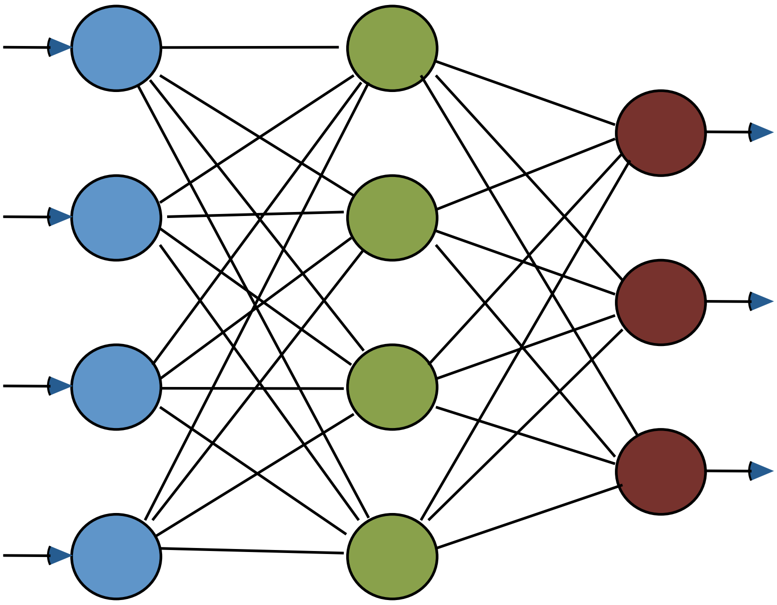


**Figure 2.5.1 (a) ANN Layers: input (blue), hidden (green) and output (brown)**

### 2.5.2 Types of Artificial Neural Networks

The earliest ANNs are the Perceptrons, proposed by the psychologist Frank Rosenblatt (Psychological Review, 1958). Then came the Artron, a statistical switch-based ANN by R. Lee in the 1950s and the Adaline or Adaptive Linear Neuron, by B. Widrow in the 1960s. The Adaline is also known as the ALC (Adaptive Linear Combiner) and it is a single neuron, not a network. In 1988 Widrow introduced a network formulation based on the Adaline and named it the Madaline (Many Adaline). Principles of the above four neurons, especially of the Perceptron, were common building blocks in the development of the most later ANNs.

Three later fundamental networks widely used today are:

1. **The Back-Propagation network**: This is a multi-layer Perceptron-based ANN, which gives an elegant solution to hidden-layers learning (Rumelhart , 1986 and others).

2. **The Hopfield Network**: This network was introduced by John Hopfield in 1982. It is different from the earlier four ANNs in many important aspects, especially in its recurrent feature of feedback between neurons and hence, it is to a great extent a seperate ANN-class in itself.

3. **The Counter-Propagation Network:** Proposed by Hecht-Nielsen in 1987, it utilizes Kohonen's Self-Organizing Mapping (SOM) to facilitate unsupervised learning.

For our thesis we used Back-propagation network, so Back propagation learning procedure is described briefly in the next section.

### 2.5.3 Back propagation learning Procedure

Back Propagation network is often considered to be the classic ANN. However, it is less of network and more a training or learning algorithm. The network used is generally of the simple type and are called *Feed-Forward* Networks or occasionally *Multi-Layer Perceptrons (MLPs)*. A Back Propagation network learns by example, i.e. example of what is required from the network is provided to the algorithm, which changes the network's weights so that when training is finished, it will produce the required output, which is known as the Target, for a particular input. Once the network is trained, it will provide the desired output for any of the input patterns. Back Propagation networks are ideal for simple Pattern Recognition and Mapping Tasks. Figure 2.5.4 (a) shows how a Back Propagation Network works.

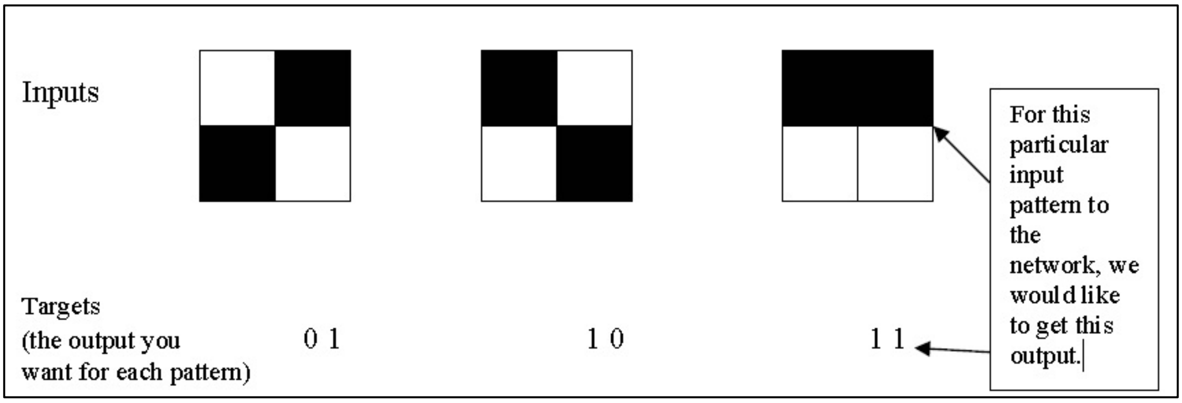

Inputs
Targets
(the output you
want for each pattern)
0 1
1 0
1 1
For this particular input pattern to the network, we would like to get this output.

**Figure 2.5.4(a) Back Propagation Network**

In a Back Propagation Network, the network is first initialized by setting up all its weights to be small random numbers, for example between –1 and +1. Then the *forward pass* is applied, i.e. the input pattern is provided and the output is calculated. As all the weights are random, the output provided initially is completely different from the *Target*. The *Error* of each neuron is then calculated, which is essentially the *Actual Output* subtracted from the *Target*. This error is then used mathematically to change the weights so that the error decreases. Thus, eventually the Actual Output of each neuron will get closer to its *Target* and this part is called the *reverse pass*. The process is repeated again and again until the error is reduced to minimal.

# Chapter 3: Sign Language Recognition

Automatic Sign Language Recognition has been addressed in various different ways as mentioned in section 2.2. Out of all techniques tried so far, we have chosen the vision-based approach for recognizing signs as opposed to data gloves or other exotic techniques. The vision-based approach is a more natural process and is less intrusive to the signer. Our initial goal is to recognize isolated signs conveyed through movement of hands. Thus, for the purpose of our thesis we will ignore facial expressions and other parameters involved in signing.

The entire recognition process cane be broken down into three main phases:

1. Tracking
2. Feature Extraction
3. Training

Before starting our experiment we investigated several techniques and tools. At first we considered hand tracking using a web-cam. However, after extensive training using Open CV's haartraining to track hands, the accuracy was still very low. We figured that due to the variety of skin complexion of Bangladeshi people and the possibility of various hand orientation and shape, this technique would not be efficient. Also handling occlusion of both hands and the face would have become a very cumbersome task in this method and thus we decided to shift to Kinect depth camera, which tracks users independent of any specific conditions and solves the issue of occlusion. Then we came across the OpenNI framework, which provides high quality tracking methods and thus we used the SimpleOpenNI library with Java for the first two phases. For the third phase we chose to use Artificial Neural Networks implemented in MATLAB.

A step-by-step process we followed to conduct the experiment in order to validate our research is given in the form of a flowchart in figure 3 (a).

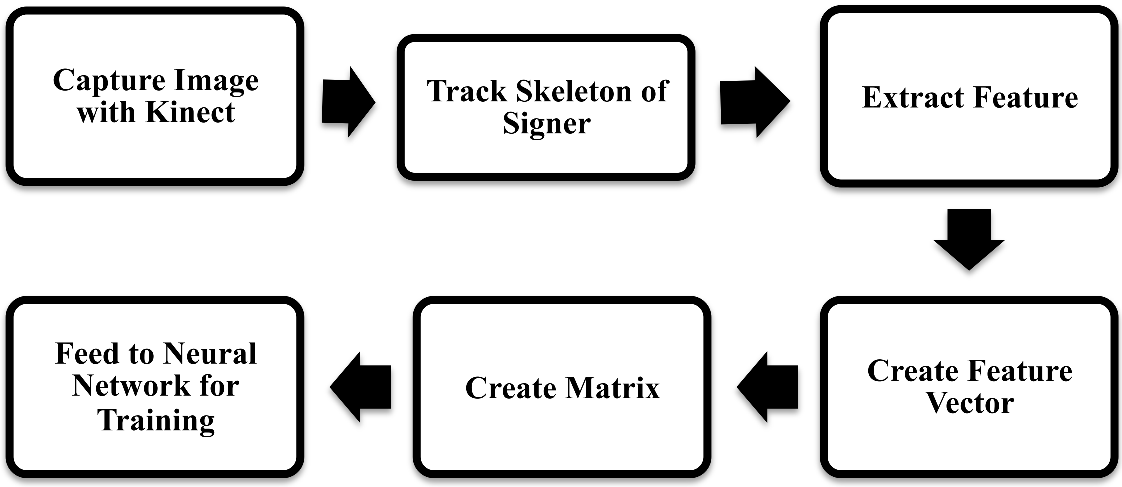


**Figure 3 (a) Step-By-Step Recognition Process**

### 3.1 Tracking

We tracked the signer skeleton, which provides joint information of the entire body. As we focused on manual features only, tracking the hands was the most important task. Besides the hand we also needed to track the head, neck, torso and elbows. These parameters were used only as reference and to help increase the accuracy of recognition.

Detection and tracking of signer skeleton were conducted with the help of Kinect depth camera. Tracking, as stated before could have been achieved by ordinary web-cams as well. However, it requires controlled environment and good lighting conditions. Meanwhile, Kinect is capable of tracking a users full body or hands independent of lighting conditions or other environmental variables. The SimpleOpenNI library provides built-in functions to track the skeleton and also to extract coordinates of the joints.

The result of Kinect and OpenNI tracking a human skeleton using the depth sensor is shown in figure 3.1 (a)

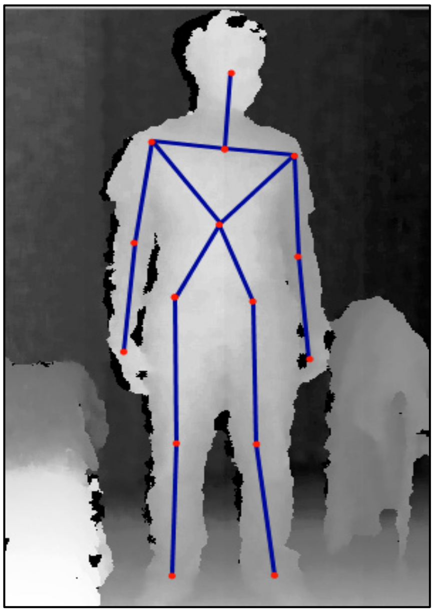


**Figure 3.1 (a) Skeleton Tracking**

## 3.2 Feature Extraction

Each sign is associated with a set of features that need to be extracted in order to distinguish one sign from another. While some features may be different for different signs, others might be similar and so it is necessary to combine several features of a particular sign to maintain higher accuracy. However, it is not required to extract every single feature, in fact, experiments have shown that a few well-chosen features can give a fairly high recognition rate.

Since we are only focusing on manual gestures, we will extract features related to hands only. Our initial aim was to extract the hand shape along with other features. However, as stated before, individual finger recognition is a limitation of the frameworks available for Kinect. Though, some hand shape information could have been obtained by customizing the existing API, it would have taken more time than we have to complete our thesis. So, we decided to proceed without extracting hand shapes for now. We extracted seven features, which were also extracted in the recognition of German Sign Language in [1], however they adopted a different training method than ours and they also considered the neck as the reference point for all calculations, while we took the head as the reference point. The seven features and the calculations are given below.

1. **Two dimensional position of each hand relative to head (xAbs, yAbs)**

$$xAbs = |handX - headX|$$

$$yAbs = |handY - headY|$$

where, (handX, handY) and (headX, headY) are the 2D coordinates of the hands and head respectively.

2. **Distance between both hands (distance)**

$$distance = \sqrt{}((righHandX - leftHandX)^2 + (rightHandY - leftHandY)^2)$$

where, (rightHandX, rightHandY) and (leftHandX, leftHandY) are the 2D coordinates of the hands and head respectively.

3. **Two dimensional movement of each hand or Position of each hand relative to position of hands prior to the last two updates (xrel, yrel)**

$$\mathrm{xrel} = |\mathrm{handXprev} - \mathrm{handX}|$$

$$\mathrm{yrel} = |\mathrm{handYprev} - \mathrm{handY}|$$

where, (handXprev, handYprev) are the 2D coordinates of hands two updates before.

4. **Absolute velocity of each hand (v)**

$$v = \sqrt{}(xrel^2 + yrel^2)$$

5. **Absolute distance of each hand from head (d)**

$$d = \sqrt{}(xAbs^2 + yAbs^2)$$

6. **Two dimensional normalized velocity of each hand (vx, vy)**

$$(vx, vy) = \begin{cases} (0,0) & , xrel = 0 \ and \ yrel = 0 \\ \left(\frac{xrel}{xrel + yrel}, \frac{yrel}{xrel + yrel}\right) & , otherwise \end{cases}$$

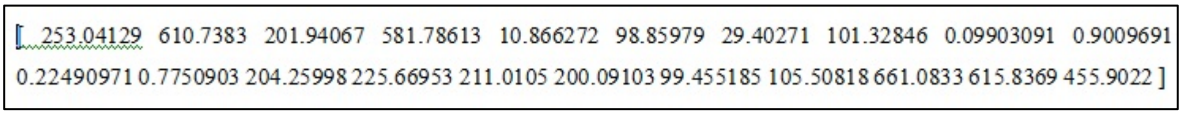
7. **Position of both elbows relative to neck (ex, ey)**

$$(ex, ey) = ((|elbX - neckX|, |elbY - neckY|)$$

The region above the torso was considered as the signing space and the updates were recorded and features were calculated in each frame only if the dominant hand, i.e. the right hand was above the torso, i.e. y-coordinate of right hand was greater than that of the torso. The elbow positions were calculated so that signs where the hand is near the head but differently rotated are easier to understand. All these features were combined into a feature vector (f), which was then used in the training phase. The feature vector is given below:

$$\mathrm{f} = (\mathrm{xAbs}, \mathrm{yAbs}, \mathrm{xrel}, \mathrm{yrel}, \mathrm{vx}, \mathrm{vy}, \mathrm{ex}, \mathrm{ey}, \mathrm{v}, \mathrm{d}, \mathrm{distance})$$

A frame containing all extracted features of both the hands combined together in a particular frame is shown in figure 3.2 (a). It is basically a vector with 21 values.



```
[ 253.04129  610.7383  201.94067  581.78613  10.866272  98.85979  29.40271  101.32846  0.09903091  0.9009691
0.22490971 0.7750903 204.25998 225.66953 211.0105 200.09103 99.455185 105.50818 661.0833 615.8369 455.9022 ]
```

**Figure 3.2 (a): A Feature Vector For A Single Frame**

### 3.3 Training

As stated earlier in section 3.2, prior to training we extracted all features from each frame at run time, provided the dominant hand was inside the signing space. The number of frames required for showing different signs and even one sign twice varies according to the speed of the signer. So, to represent each sign there were an unpredictable and varying number of vectors, which was a problem for creating the training set.

Basically, each frame has 21 values to describe the features we have extracted. A set of vectors will represent a particular sign. If the signer gives the sign quickly then there will be less number of frames in a sign and thus less vectors, whereas if the sign is given slowly number of vectors representing the sign will increase. To avoid this problem we selected 30 random frames including the first and last frame from the vector of the frames to describe a sign. As a result, a sign is described by a 21 X 30 matrix, i.e. 30 frames each containing 21 values. Then we transformed this matrix into a 7 X 10 matrix. To make this transformation we divided the 21 X 30 matrix into seventy 3-by-3 grids. Each grid was replaced by one value, which is the average of every value inside that grid. An example of a conversion of a 6 X 6 matrix into a 2 X 2 matrix using the same technique is given below.

**6X6 Matrix**

| | | | | | |
|---|---|---|---|---|---|
| **2** | **3** | **2** | **4** | **5** | **6** |
| **1** | **2** | **2** | **3** | **7** | **5** |
| **3** | **2** | **1** | **6** | **4** | **5** |
| **1** | **2** | **3** | **6** | **4** | **8** |
| **1** | **3** | **2** | **6** | **7** | **8** |
| **2** | **1** | **3** | **8** | **6** | **7** |

**2X2 Matrix**

| | |
|---|---|
| **2** | **5** |
| **2** | **6** |

After getting a 7 X 10 matrix for each sign we make it a 70 X 1 matrix. For Example the 2 X 2 matrix above will become a 4 X 1 matrix as shown below:

**4X1 Matrix**

| |
|---|
| **2** |
| **2** |
| **5** |
| **6** |

We created this system for five Bangla signs (Brother, Tea, Chair, Door, Notebook) used in daily lives and taught to primary deaf school students. For each sign we had 10 samples and two inexpert signers who learnt these signs specifically for the purpose of this thesis conducted the signs for training. So finally we had ten 70 X 1 matrix and we created one 70 X 10 matrix from these 10 matrices. We created a Back Propagation Artificial Neural Network (ANN) with 3 layers (one input layer with 70 nodes, one hidden layer with 48 nodes, one output layer with 5 nodes). Then each column of the 70 X 10 matrix created earlier was fed as input to our ANN for training the system to recognize the signs..

# Chapter 4: Experimental Results and Evaluation

To illustrate our experimental results we needed to calculate the accuracy of the recognition of each sign. The accuracy of the system was calculated as described in section 4.1.

## 4.1 Accuracy Measurement

We calculated our accuracy using the following variables:

**Precision:**

In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the search and is given by the equation below.

$$precision = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{retrieved\ documents\}|}$$

Precision takes all retrieved documents into account, but it can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called precision at n or P @ n. For example for a text search on a set of documents precision is the number of correct results divided by the number of all returned results.

Precision is also used with recall, the percent of *all* relevant documents that is returned by the search. The two measures are sometimes used together in the F1 Score (or f-measure) to provide a single measurement for a system.

**Recall:**

Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$recall = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{relevant\ documents\}|}$$

For example, for text search on a set of documents recall is the number of correct results divided by the number of results that should have been returned.In binary classification, recall is called sensitivity. So it can be looked at as the probability that a relevant document is retrieved by the query. It is trivial to achieve recall of 100% by returning all documents in response to any query. Therefore, recall alone is not enough but one needs to measure the number of non-relevant documents also, for example by computing the precision.

Recall and Precision can be defined using the following variables:

**True Positive (tp):** True positive means number of correct results we are looking for.
**True Negative (tn)**: Means correct absence of the irrelevant result in our system's result.
**False Positive (fp)**: Means wrong result in our system's output (Unexpected output).
**False Negative (fn)**: Means missing expected outputs, i.e.the result should be included in system output but is not there.

As recall and precision are defined as follows:

$$precision = \frac{tp}{tp + fp}$$

$$recall = \frac{tp}{tp + fn}$$

Accuracy and True Negative rate are calculated as given below:

$$true\ negative\ rate = \frac{tn}{tn + fp}$$

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn}$$

### 4.2 Results Obtained

We divided the input of our system in three parts:

1. Training data, which we used to train the system given as Input (8 samples for each sign).
2. Testing data created by the same signers who signed the test data for training (2 samples for each sign)
3. Testing data generated by new signer (10 samples for each sign)

The values of tp, tn, fp and fn we obtained for each type of input are given below and the corresponding values of precision, recall, true false rate and accuracy are also calculated.

**First case:**

| TP | TN | FP | FN |
|---|---|---|---|
| **10** | **40** | **0** | **0** |

**Table 4.2.1 (a)**

Precision = 10/(10+0) * 100= 100%

Recall = 10/(10+0)*100 = 100%

True False rate = 40/(40+0)*100 = 100%

Accuracy = (10+40)/(10+40+0+0)*100 = 100%

**Second case:**

| TP | TN | FP | FN |
|---|---|---|---|
| **9** | **39** | **1** | **1** |

**Table 4.2.1 (b)**

Precision = 9/(9+1) *100 = 90%

Recall = 9/(9+1) * 100 = 90%

True False rate = 39/(39+1)*100 = 97.5%

Accuracy = (9+39)/(9+39+1+1)*100  = 96%

**Third case:**

| TP | TN | FP | FN |
|---|---|---|---|
| 7 | 37 | 3 | 3 |

**Table 4.2.1 (c)**

Precision = 7/(7+3) *100 = 70%

Recall = 7/(7+3) * 100 = 70%

True False rate = 37/(37+3)*100 = 92.5%

Accuracy = (7+37)/(7+37+3+3)*100  = 88%

The accuracy we obtained during our thesis is given below in terms of separate type of inputs in figure 4.2 (a) and individual signs in figure 4.2 (c). Figure 4.2 (b) compares the recall, precision and true false rate obtained when the system was provided each type of the three input types.

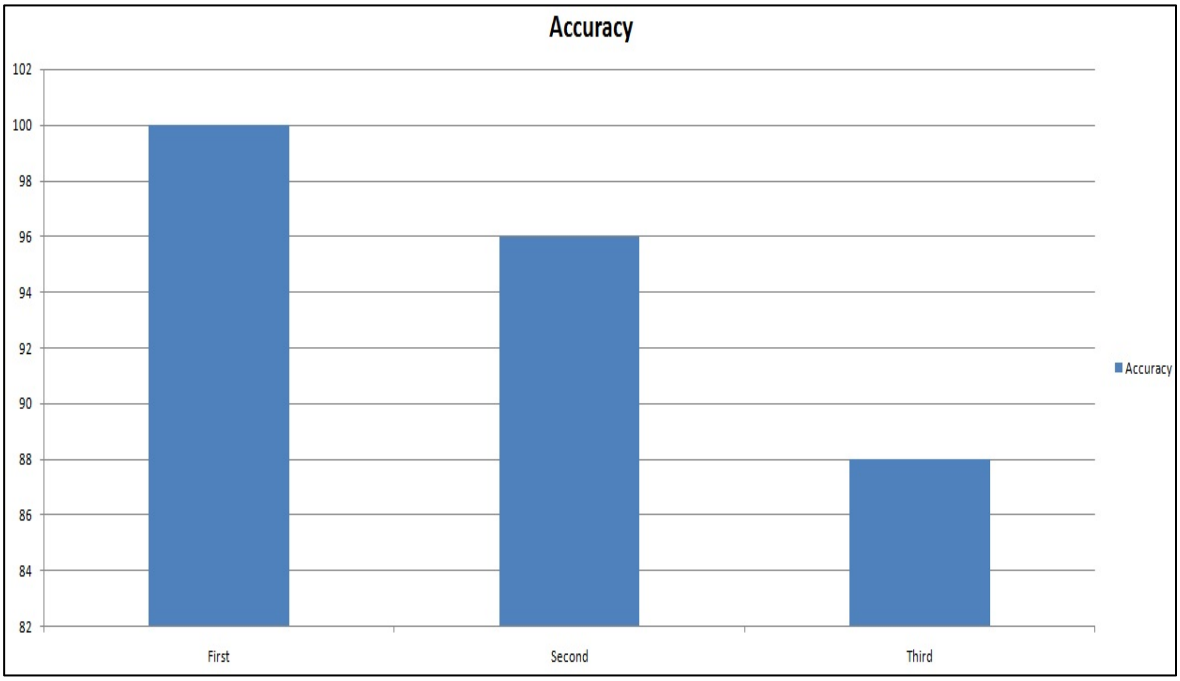


**Figure 4.2 (a): Accuracy comparison of three types of input**

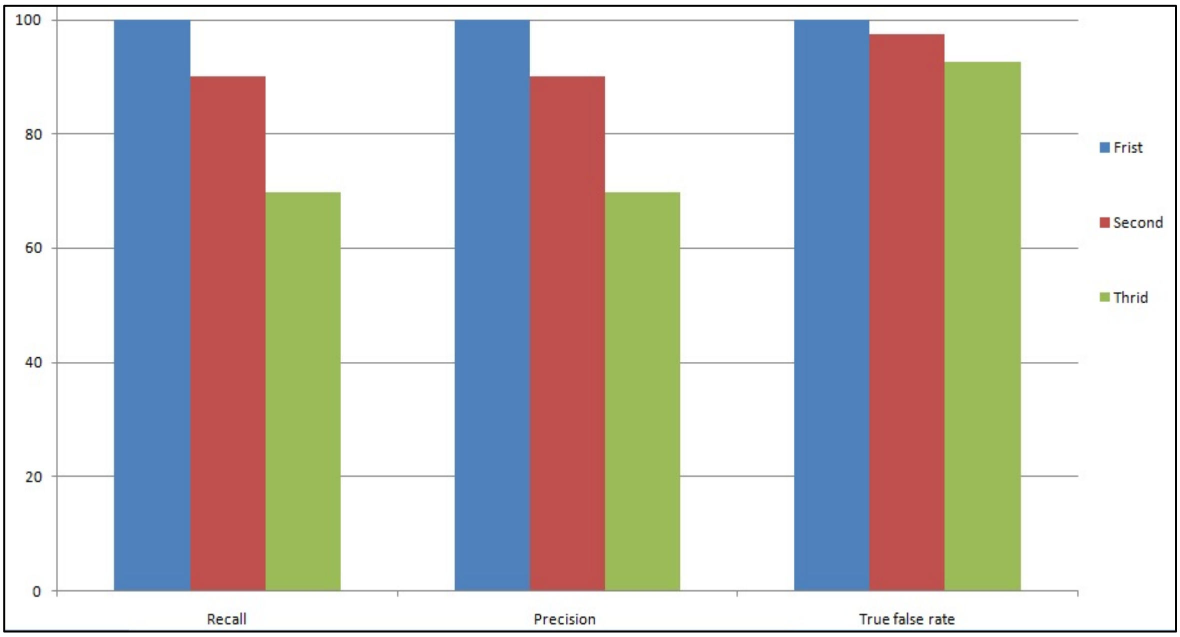


**Figure 4.2 (b): Recall, Precision and True False Rate comparison of three types of input**

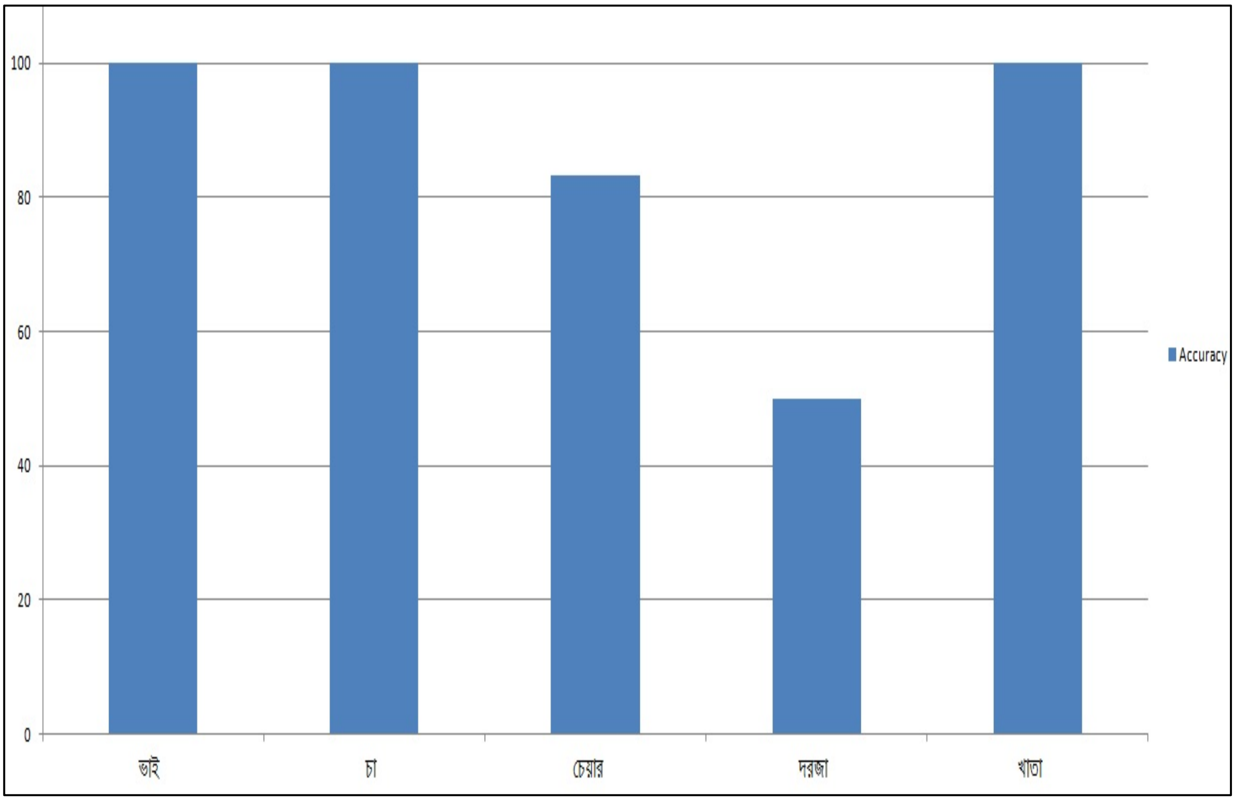


**Figure 4.2 (c) Accuracy for individual signs**

## 4.3 Evaluation of Results

We can observe in figure 4.2 (a) that when given the training data as input the recognition rate is 100 percent, meaning that the training was successful. Then when a separate test data was given to the neural network, the accuracy rate was 96 percent. The same signers who signed for the training sample provided this data. When the third input is given to the network the accuracy rate decreases to 88 percent. Thus, it cane be seen that the system performance becomes worse when the input data is provided by signers the system does not know. Also from figure 4.2 (b) we can see that the recall, precision and true false rate is 100 percent and above 90 percent for the first and second type of input respectively. For the third input type recall and precision is just about 70 percent.

In figure 4.2 (c) we can see the accuracy rate of individual signs. The accuracy rate of brother (“bhai”, 1st from left), tea (“cha”, 2nd from left) and notebook (“khata”, last from left) are 100 percent each. However, chair (“chair” , 3rd from left) and door (“dorja” , 4th from left) have very medium (85 percent) and low (50 percent) accuracy rate respectively.

# Chapter 5: Conclusion and Future Work

Our thesis spans over a very small subset of automatic sign language recognition. There is a lot more that needs to be and can be done in this field to achieve the final goal of reducing the communication gap between the hearing impaired and the rest.

We need to improve the work we have done in our thesis to make recognition of isolated signs more accurate. Firstly, features related to hand shapes have to be extracted, otherwise signs with similar hand movements and position will be difficult to separate using the existing features that has been extracted. The next step will be to increase the number of training samples and to try training methods other than neural network to analyze which one produces the best result for Bangla Sign Language recognition. Also, the numbers of trainers need to be increased and this will improve the rate of recognition.

After improving on the current work we have done so far, our next aim will be to research on extensions to the project. At first, recognition of both static and dynamic signs need to be incorporated within the same system, which has not been achieved to greater accuracy yet, especially for Bangla Sign Language. Then continuous sentences need to be recognized and mapped to the corresponding spoken grammar. Finally, facial expressions and other sign parameters mentioned in section 2.1.2 need to be accounted for to understand beyond the literal meaning of the sign and to reduce errors in recognition. Also, text-to-sign and sign-to-text systems need to be combined together in one system to make the communication a two way process.

In the future, continued research in this area, besides helping as an interpreter, can also open doors to numerous other applications like sign language tutorials or dictionaries and also help the deaf and dumb to search the web or send mails more conveniently.

# APPENDIX

As mentioned before we have used five signs to measure the accuracy of our method during the research. Information on how to perform these signs is given below. All signs shown here belong to the Bangla Sign Language vocabulary only.

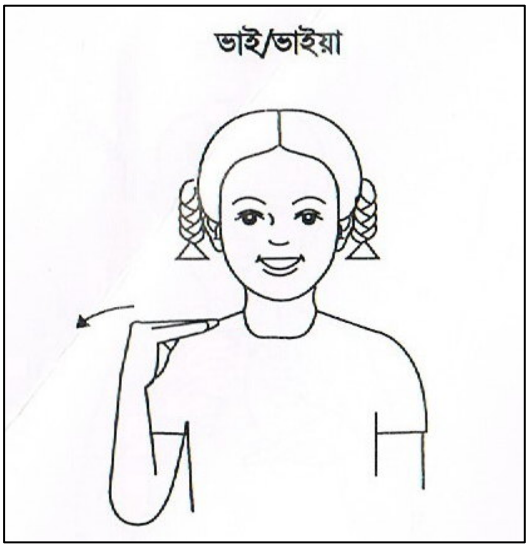


**Figure 0.1 Brother (figure taken from [8])**

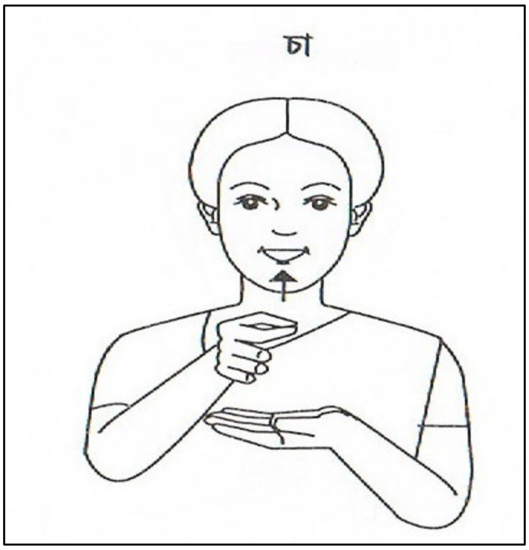


**Figure 0.2 Tea (figure taken from [8])**

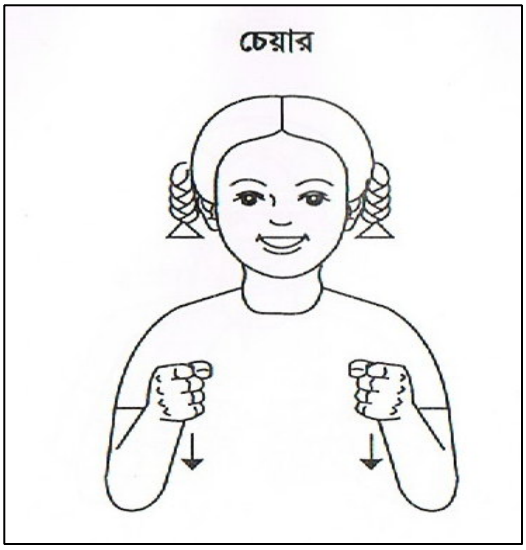


**Figure 0.3 Chair (figure taken from [8])**

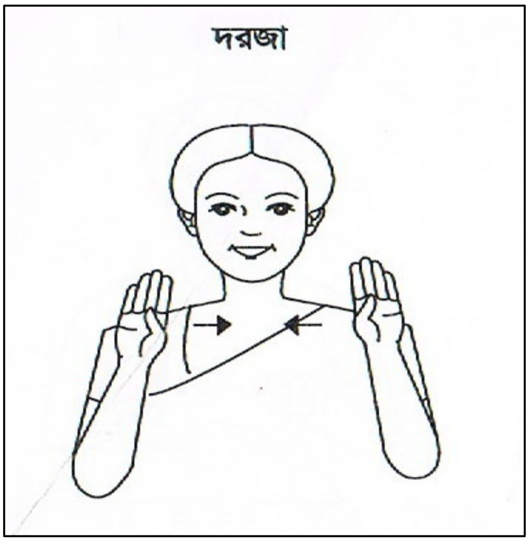


**Figure 0.4 Door (figure taken from [8])**

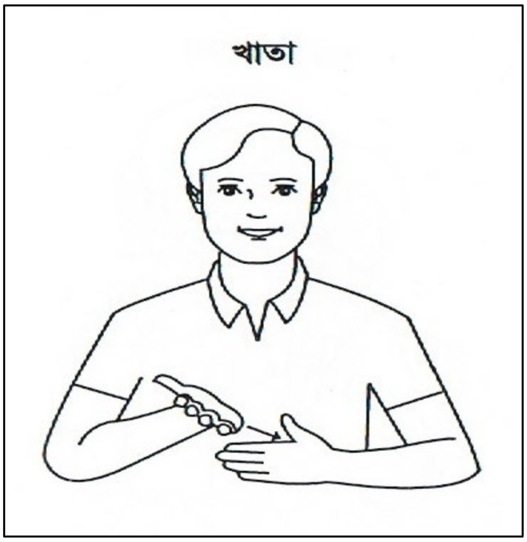


**Figure 0.5 Notebook (figure taken from [8])**

# BIBLIOGRAPHY

[1] Lang, Simon. *Sign Language Recognition With Kinect.* Institut for Informatik, Freie Universitat Berlin, Berlin: Freie Universitat Berlin, 2011.

[2] M.R. Andersen, T. Jensen, P. Lisouski, A.K. Mortensen, M.K. Hansen, T. Gregersen and P. Ahrendt. *Kinect Depth Sensor Evaluation For Computer Vision Applications.* Department of Engineering , Aarhus University, Aarhus University, 2012.

[3] Kinect Fact Sheet, Microsoft News Center. June 2010. http://www.microsoft.com/presspass/presskits/xbox/docs/KinectFS.docx.

[4] *Sign Language.* http://en.wikipedia.org/wiki/Sign_language.

[5] Parton, Becky Sue. "Sign Language Recognition and Translation: A Multidisciplined Approach From the Field of Artificial Intelligence." *Journal of Deaf Studies and Deaf Education* (Oxford Journals) 11, no. 1 (2005).

[6] Kaushik Deb, Helena Parvin Mony & Sujan Chowdhury. "Two-Handed Sign Language Recognition for Bangla Character Using Normalized Cross Correlation." (Global Journals Inc. (USA)) 12, no. 3 (February 2012).

[7] Dewan Shahriar Hossain Pavel, Tanvir Mustafiz , Asif Iqbal Sarkar, 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.* 2004.

[8] Nahid Sultana Juthy, Broj Gopal Shaha,Md. Sharafat Ali Shojol. *Ishara Bhashay Jogajog (Communicating through Sign Language).* Dhaka: Center for Disability in Development, 2005.

[9] "Scientific Understanding and Vision Based Technological Development for Continuous Sign Language Recognition and Translation." *SignSpeak Project*, Annual Public Report. http://www.signspeak.eu/

[10] Sarella, Kanthi. *Formulation of an Image Processing Technique for Improving Sign2 Performance.* Final Report.

[11] *OpenNI.* http://openni.org/

[12] *Processing. http://processing.org/*

[13] Precision and recall: http://en.wikipedia.org/wiki/Precision_and_recall

[14] Danial Graupe. *Principles of Artificial Neural Networks,* 2nd Edition. Advanced Series on Circuits and Systems, vol. 6.

[15] Girish Kumar Jha. *Artificial Neural Networks.* Indian Agricultural Research institute.