Chapter 1
Introduction

1.1 Introduction

The method of human communication, either spoken or written, consists of the use of words in a structured and conventional way. But the deaf and dumb people cannot hear or speak. So, deaf and dumb people around the world communicate using sign language as distinct from spoken language in their everyday lives. A Sign Language is a visual language that uses a system of manual, facial and body movements as the means of communication. Sign language is not a universal language, and different sign languages are used in different countries [1].

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.

1.2 Motivation of the work

Many people are hearing impaired in our country. Sign languages are used by the hearing impaired people to mainly communicate with other peoples of their kind, and also to some extent with their friends, families as well. Many of these people are intelligent enough to share their opinions and work with rest of the population to move ahead our country [2]. If they get chance to participate in different development organizations in different sectors, they will be able to contribute in country with their intelligence and also can win their disability by their work. They only can communicate with their family members as the members also learn sign language or by observing them while communicating with them. But there is a huge communication gap between the deaf people and the rest of the people in our country.

As deaf people can't listen and also can't talk to ordinary language, they use sign language to communicate among them. On the other hand, people who are not deaf, never try to learn the sign language for interacting with the deaf people. This becomes a cause of isolation of the deaf people. As a result deaf people cannot do anything by his/her own.

Suppose, a deaf patient went in a nearest medical centre as an emergency, he/she is trying to express her/his problem using sign language. But a regular doctor may not be able to understand her problem.

But if the computer can be programmed in such a way that it can translate sign language to text format, the difference between the normal people and the deaf community can be minimized. So, to reduce that gap a system is proposed to translate sign language to ordinary people.

1.3 Challenges of the work

The main challenges in developing such system are the variant of the body structure of different people. Because of the body shape and height, the calculation procedure will be different. And also the light condition of the environment is very crucial here for our hardware device. After detecting the region of interest, motion is another issue to keep in mind. If the gesture is delivered too fast or too slow the system may not detect it. As it is real time project, timing will be challenging too. Similarities between gestures are also a challenge here, for example, in here this project two words gestures almost represent same data. Detecting the correct gesture from the similar ones is mandatory.

1.4 Contribution of the work

The goal of this project is to translate the sign language. For that the first and foremost challenge was to capture the gesture. As this project is based on a run time environment the capture of gesture was very challenging. To solve this problem Kinect Motion Sensor Device is used to capture the gesture. This device can capture the depth data of a moving object. Using those data 8 joint of a moving body is detected. By using the motion data of these joints the gesture has being captured. Also to capture this gesture as far as possible an efficient algorithm is used. Also RGB image of the user is also provided in this project. Moreover a real time environment is created in this project to habituate user with this program easily.

The key objectives of the project have been mentioned in the following:-

- The goal of this project is to translate the sign language.

- To assist the individuals with hearing and speech disability by translating their sign language gestures into its corresponding normal text that is understandable by the general people

- To evaluate the system performance in real environment

The user interface of this program is created in such a way that user should have no problem to use this software. A database is created behind the project to store the sign language along with its corresponding word. Beside those "Adjust Kinect Angle" a new feature is added to the project so that the device can automatically detect the user independent to where the Kinect Motion Sensor Device is placed.

1.5 Organization of the thesis

This thesis paper is organized into six chapters.

Chapter 1 (Introduction) introduces the thesis work and the background of the work.

Chapter 2 (Literature Reviews) provides a brief description of the related theory about the work.

Chapter 3 (System Architecture & Design) introduces the proposed method and describes every step of our work briefly.

Chapter 4 (Implementation) describes the implementation process in details.

Chapter 5 (Experimental Result & Discussion) describes how the proposed model has been deployed and tested.

Chapter 6 (Conclusion and Future Recommendations) concludes our thesis work. This chapter also describes the future works applicable to our work.

.

1.6 Conclusion

In this chapter, we have given the basic overview and the proper introduction of our work. The structure of the whole report is also drawn. In the next chapter, we will see how the theoretical concepts described in this chapter have been integrated into real-time gesture detection module.

Chapter 2
Literature Review

2.1 Introduction:

Sign language translation will play a great role in our communication system. It can decrease the distance between normal people and deaf people. It can make a revolution in our communication system. This can make our communication technique more fast and efficient.

So to do such project it is mandatory to be careful about to choose an efficient algorithm and device & tools used for it.

In this chapter the tools used for this project in foreground and background are discussed in details with their general activities, importance and contribution in this project.

2.2 Sign Language

Sign language is a nonverbal form of communication used especially by people with the inability to speak or hear. It is a separate language with its own grammar and rules [12]. With the advancement of science and technology many techniques have been developed not only to minimize the problem of deaf people but also to implement it in different fields.

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.

.

2.2.1 Sign as a Language

Sign language, any means of communication through bodily movements, especially of the hands and arms, used when spoken communication is impossible or not desirable. The practice is probably older than speech. Sign language may be as coarsely expressed as mere grimaces, shrugs, or pointing; or it may employ a delicately nuanced combination of coded manual signals reinforced by facial expression and perhaps augmented by words spelled out in a manual alphabet. Wherever vocal communication is impossible, as between speakers of mutually unintelligible languages or when one or more would-be communicators is deaf, sign language can be used to bridge the gap.[13]

There are about 70 million deaf hearing-impaired as well as hearing people in the world who use sign language as their first language or mother tongue. Each country has one or sometimes two or more sign languages, although different sign languages can share the same linguistic roots in the same way as spoken languages do. Though it is not an international language, the universal features in sign languages help to make it possible for users of different sign languages to understand one another far more quickly than users of unrelated spoken languages can [1].

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 [14]. Hundreds of sign languages are being used around the world, which include American Sign Language (ASL), Japanese Sign Language, (JSL), British Sign Language (BSL), Austrian Sign Language (GS), Bangladeshi Sign Language (BdSL) and so on.

Thus, it is impossible for someone who knows American Sign Language to communicate in British Sign Language unless they learn it.

So, we can say, Sign language now a days have been established as a natural language as like independent of other language.

2.2.2 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 [15].

Bangla is the most spoken language in Bangladesh and second in India. About 250 million native and about 300 million total speakers worldwide speaks in Bangla [16]. So a well developed system for Bangla speech to Bangla Sign Language and vice versa is very much needed for communication with deaf people. While discussing further, we will use the term BL for naturally spoken Bangla Language and BdSL for Bangla Sign Language [1].

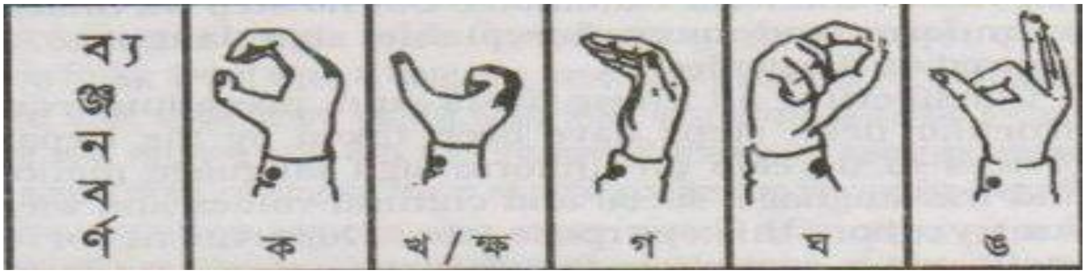


Figure 2.1: Bengali Manual Alphabet Using One Hand [18]

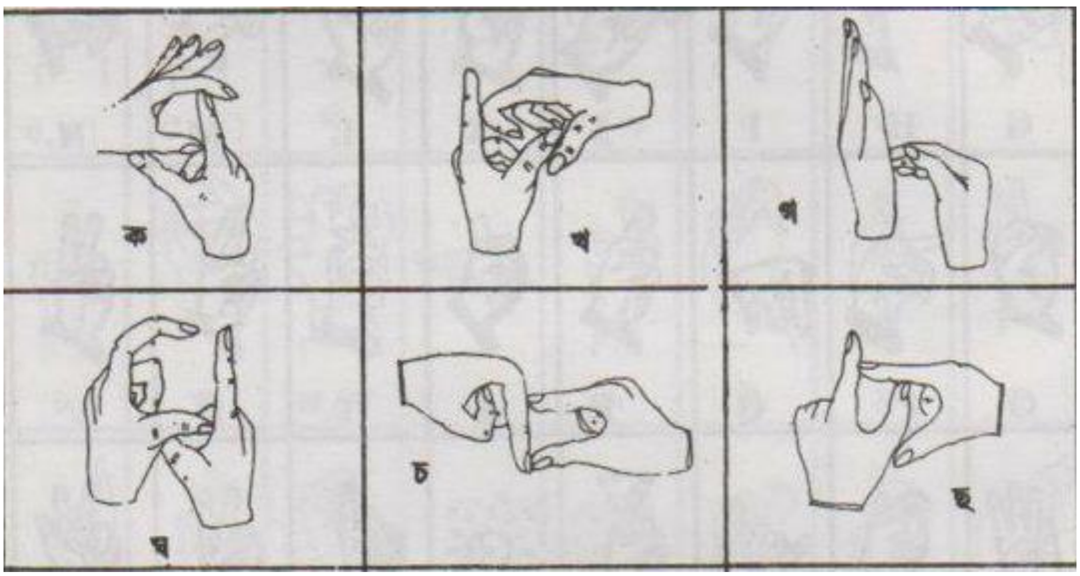


Figure 2.2: Bengali Manual Alphabet Using Two Hands [18]

2.2.2 Components and types

Sign language not only requires movement of hands and arms but also involves specific formation of fingers, facial expressions, and head movements. Most importantly, sign language is not universal. 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 [17] [15].

Here we use the system which is to perform and detect hand gestures that do not involve different alignment of fingertips. The system was able to detect top-to down or left-to-right movement of hands.

Sign language are expressed by delivering gesture or body movement. These gestures depending upon sign language can be divided into two categories. Those are,

1. Static Gesture &
2. Dynamic Gesture.

Single-handed signs can be in a state of motion or can be represented using static or rest position of the hand. In dynamic signs involve the domination of one hand over the other, or both hands share equal priority. The right hand is generally called the dominant hand and is used to convey almost all signs unless the signer is left-handed. 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 [1].

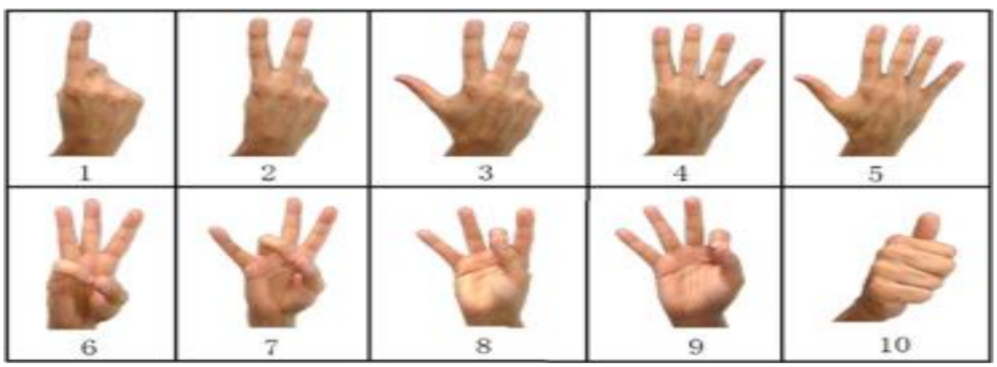


Figure 2.3: Static Gesture[21]

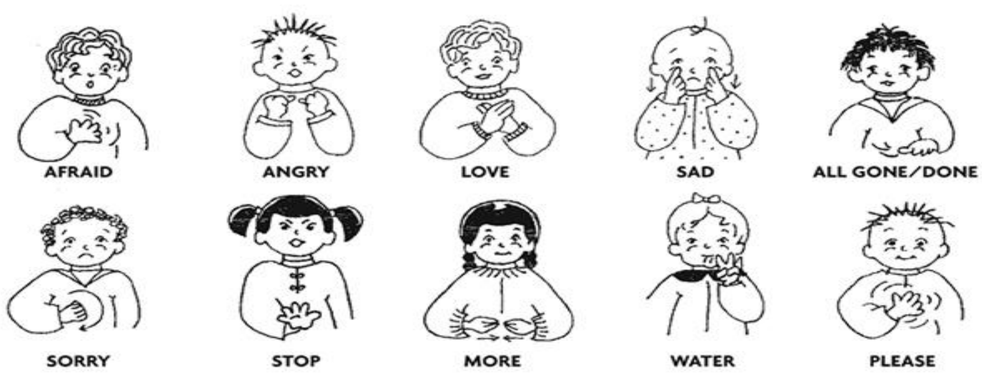


Figure 2.4: Dynamic Gesture [22]

2.3 Kinect Motion Sensor Devices:

Kinect was first announced on June 1, 2009 at E3 2009 under the code name "Project Natal" [18]. Kinect was developed by Microsoft and PrimeSense and was released on November 2010. Kinect is a motion sensing input device by Microsoft for the Xbox 360 video game console and Windows PCs. Based around a webcam-style add-on peripheral for the Xbox 360 console, it enables users to control and interact with the Xbox 360 without the need to touch a game controller, through a natural user interface using gestures and spoken commands. The device features an "RGB camera, depth sensor and multi-array microphone running proprietary software", which provide full-body 3D motion capture, facial recognition and voice recognition capabilities.

The depth sensor consists of an infrared laser projector which captures video data in 3D. It does two things: generate a three-dimensional (moving) image of the objects in its field of view, and recognize (moving) human beings among those objects. The camera transmits invisible near-infrared light and measures its “time of flight” after it reflect off the objects. By doing this it can detect 20 skeletal joint of human body. By calculating the position of each joint at a fixed interval of time it can sense the body movement.

Figure 2.5: Kinect Motion Sensor Device [24]

In this project both depth sensor image and RGB image is taken. Both images are taken in a real time environment.

2.4 Kinect Sensor Components and Specifications

Kinect has 3 main parts, an infrared projector, a RGB camera and a Infrared camera. When Kinect starts to capture a depth image, first power goes to the Infrared projector, this projector shoots an irregular pattern of dots with the wavelength of 700nm to 1mm and hence infrared is invisible to humans. This pattern is projected into the room. Kinect has a diffractive optical element which means when the infrared light is generated at the projector , this light is diffracted to form the irregular pattern of dots. To visualize these dots use a night vision camera and we can see the amazing pattern of dots formed on our subject.

Figure 2.6: RGB Image, Depth Image, Skeleton Joint Point [24]

The skeleton data consists of a set of joints. There are 20 joints of a human body which can be detected by Kinect. These joints are shown in the diagram below.

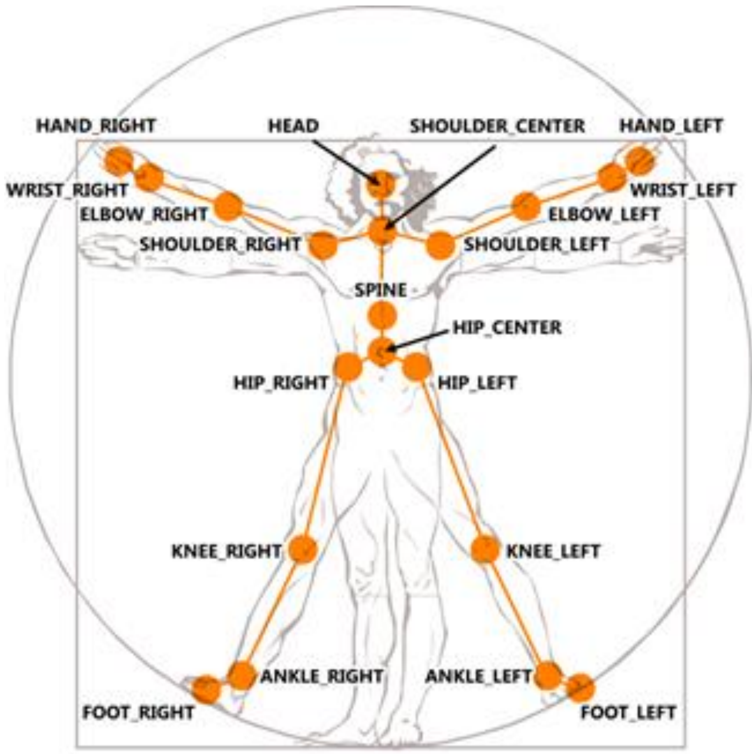


Figure 2.7: Joint Points of Human Body [20]

The depth data is the distance, in millimeters, to the nearest object at that particular(x,y) coordinate in the depth sensor's field of view. The depth image is available in 3 different resolutions: 640x480 (the default), 320x240, and 80x60 as specified using the Depth Image Format Enumeration. The range setting, specified using the Depth Range Enumeration, determines the distance from the sensor for which depth values are received. Each frame, the depth image captured is processed by the Kinect runtime into skeleton data.

The coordinate system for the skeleton data is a full 3D system with values in meters. It is shown in the following diagram.

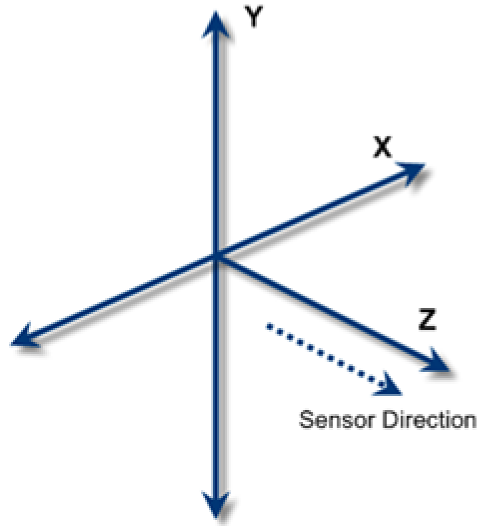


Figure 2.8: Skeleton Detection Parameters [20]

The coordinate system is defined as follows:

The origin (x=0, y=0, z=0) is located at the center of the IR sensor on Kinect

- X grows to the sensor's left
- Y grows up (note that this direction is based on the sensor's tilt)
- Z grows out in the direction the sensor is facing
- 1 unit = 1 meter

In this project RGB image is used to help the user to see his/her movement. That is it will work as a RGB video camera. Though no input is taken from this camera but this only used to habituate user with the program. Depth sensor image is taken as input. By using depth sensor data the coordinate of joint is taken as input. By calculating those position the movement of body is taken and gesture is detected.

2.5 Kinect SDK

Software development kit (SDK) is typically a set of software development tools that allows the creation of applications for a certain software package, software framework, hardware platform, computer system, video game console, operating system, or similar development platform. Here Kinect SDK version 1.8 is used. Kinect for Windows SDK 1.8 includes some key features and samples that the community has been asking for, including new background removal, realistic color capture with Kinect Fusion, improved tracking robustness with Kinect Fusion, HTML interaction sample, multiple-sensor Kinect Fusion sample, adaptive UI sample.

2.6 Kinect Developer Toolkits

Kinect Developer Toolkit enables developers to create applications that support gesture and voice recognition, using Kinect sensor technology on computers. It contains different section. Some of the major sections are:

- **Components:** This section has several components that can help when coding. Each component has their controls or classes that can save you some time to focus on other things. Components include Microsoft.Kinect.Toolkit & Microsoft.Kinect.Toolkit.FaceTracking**.**

- **Documents**: This section contains three things: Release notes & online resources**,** SDK documentation **&** Human Interface Guidelines**.** These help out when code to find your way into the framework or assist you in creating your UI with great UX.

- **Samples:** The toolkit also provide samples in C++/VB.net/C# (WPF) and cover most features like: Audio basics, Color basics, Depth basics, Face Tracking basics, gestures and lots more.

In this project 1.8 version of Kinect Developer Toolkit is used. This version includes updated and new source code samples, Kinect Fusion, Kinect Interactions, Kinect Studio, and other resources to simplify developing Kinect for Windows applications.

2.7 Platforms and Technology

For our Project, We have used Microsoft Visual Studio. It is a powerful IDE that ensures quality code throughout the entire application lifecycle, from design to deployment. Whether to develop applications for SharePoint, the web, Windows, Windows Phone etc. Though many programming language can be compiled through this only C# is used here. Here WPF format is used to build the code.

In this project we have also used the C# programming language. It is class based as well as object oriented programming language. All the frameworks have done by using C#. Also Java and C++ can be used instead of C#. Kinect SDK support C#, C++ and Java. C# has built in library for Kinect SDK. Thus C# has chosen here as programming language here.

Here, MySQL database system is used as a database which holds the word and respective gesture for that word. It can be used in both Web based and desktop based program. Here the gesture and its respective word are stored via database.

2.8 Background and present state of the problem

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.

An MSC Thesis on Sign Language Translator Using Microsoft Kinect XBOX 360 was published by Daniel Martinez Capilla [3]. It was published at University of Tennessee (Knoxville - USA) in 2012. In this thesis, the Microsoft Kinect XBOX 360 is proposed to solve the problem of sign language translation.

The recognition of sign language has begun to appear at the end of 1990. A primary effort was made by using some electrochemical devices to recognize it. The device was used to determine hand gesture parameter like hand’s location, angle, position etc. This approach is known as glove-based system. But this approach compels the signer to wear a cumbersome device. It also encounters problem with accuracy and efficiency of the recognition system [4].

The system named as “Intelligent Assistant” for human machine was developed to communicate the D&M people interfaces [5]. For capturing sound the system had used Microsoft’s Voice Command and Control Engine along with microphone and converts it into text. But it could not perform efficiently in noisy environments. 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 [6].

A report for the degree of Master of Science in Media Arts and Sciences on visual recognition of American Sign Language Using Hidden Markov Models was submitted by Thad Eugene Starner [7]. It was submitted at Massachusetts Institution of Technology in 1995. Here it is told that using Hidden Markov Models (HMM's), a single view camera system is developed that can recognize hand gestures, namely, a subset of American Sign Language (ASL).

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 [8] and creating 3D models of Bangla signs signed by the deaf community using geometric calculations [9].

An intelligent system developed by Rahman et al. [10] is more complicated. Real-time photos of hand wearing a glove containing different dots in each finger were taken as inputs. Analyzing the dots of the graphics in the image is needed to understand what sign has been shown and then pre-recorded audio and written form were shown as output. The technique involved clustering of the dots and mapping the results of this clustering to pre-defined tables. This system was limited to Bengali numerals 1-10 [11].

A sign language translation system developed by Hazari et al. [23] using Kinect motion sensor device. As real-time recognition of a large set of dynamic gestures is considered, some efficient algorithms and models are used. There an efficient algorithm is used to recognize the gesture and translate them.

Most of this project will be accomplished based upon mainly this refrence [23,].

Here Kinect Motion Sensor Device will be used to detect the input by using depth sensor data. For normalizing the input data an efficient algorithm is used. In each project gesture of word is taken as input. Here also gesture of word will be taken as input.

2.9 Conclusion

In this chapter, all theoretical concepts and the basic element of this project has been described. Here Kinect Motion Sensor Device hardware details along with its SDK and Developer Toolkit are discussed in a expand way. The program is a WPF (Windows Presentation Foundation) program.

Chapter 3

Propose System Architecture & Design

3.1 Introduction

Sign Language can be varied in various different ways as mentioned in section 2.2. Out of all techniques tried so far, we have chosen the Computer vision based approach for recognizing signs as opposed to data gloves or other exotic techniques. In this chapter the proposed methodology with necessary diagram are given which will clearly describe the background process of this project. As it is a real time visual project, the system faced many different challenges such as loose clothing, lighting conditions, and different body structure and posture.

3.2 System Methodology

Basically Sign languages are expressed by gesture. Our initial goal is to recognize isolated signs conveyed through movement of hands. Thus, for the purpose of our project we will ignore facial expressions and other parameters involved in signing. For capturing the hand movement from the Kinect device, first we have to plug in the switch of Kinect. When the power supply is available, the camera and depth sensors turn on. The entire sign language recognition process can be broken down into three main phases:

1. Detection joint point
2. Tracking Skeleton
3. Feature Extraction

Initially, the subject has to position him/herself to face the Kinect device and start performing the gesture. The performed gesture is captured by Kinect as a series of frames. When we run the program, Sensor takes some time to start to get the proper power supply. First it turns on the camera to view the object and then it works on the depth sensor. The Kinect then uses an infrared projector and sensor, it doesn't use its RGB camera for depth computation. After finishing the computation of depth map using structured light, the Kinect then detect joint point and the body position. From the detection we get some co ordinate value of joint point which further used to extract the feature of our project. By calculating the value we then train the system to get output. The basic overview of our system architecture is shown in Figure 3.1.

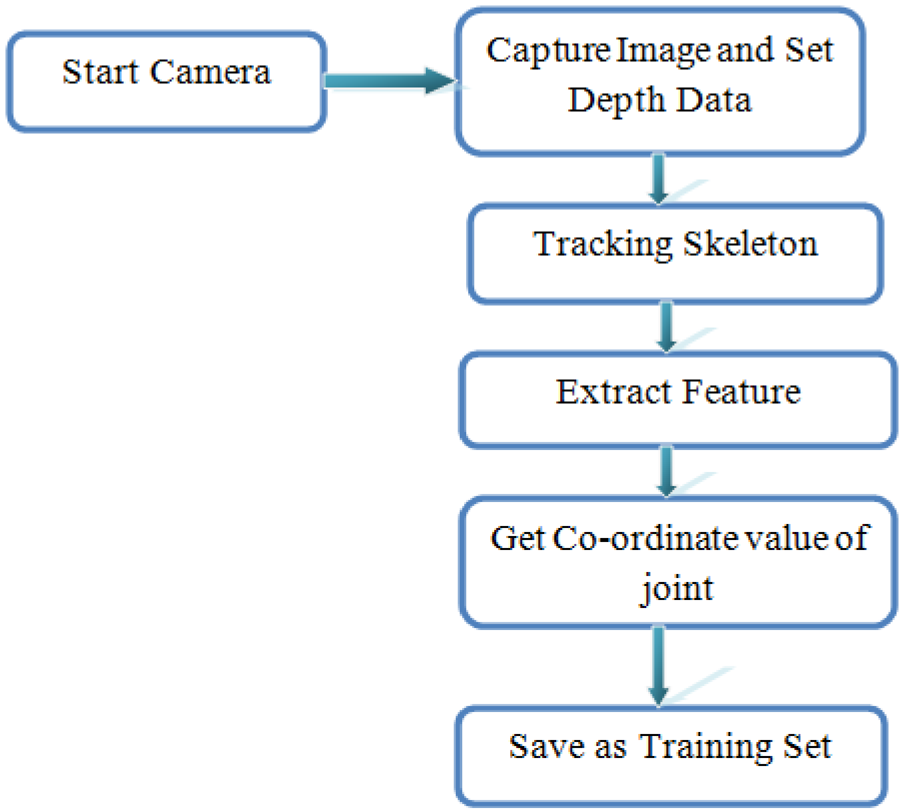


Figure 3.1: System Methodology

3.3 Capture image & Set Depth Data

As stated earlier, we used framework Kinect SDK for windows .It has available parameters for the image capturing engine. *ColorImageFormat* enumeration value chooses the color stream format including the data type resolution, and the frame rate of the data. To get data out from color frame, we used an array and a method represented as *ColorFrame.CopuPixelDataTo* which copies a frame of color data to a pre-allocated array, so it can now be written to an optional writeable Bitmap.

For working the depth data, it actually represents the distance and the player for every pixel. Here we use *DepthImageFormat* which choose the depth stream format including data type, resolution, and frame rate etc. We used the method *DepthImageFram.MapFromSkeletonPoint*

which maps the depth frame coordinates for a given skeleton point and other method is *DepthImageFrame.MapToColorImagePoint* which maps a depth co ordinate to a color co ordinate. After getting the color stream and depth stream, we also had to set the skeleton stream which works together as synchronized in the event *AllFrameReady*.

3.4 Tracking Skeleton

As the gesture for each word is different, this project has captured the gesture delivered by the user and detects that. The movement of our body joint as we can say, to pass each joint point to a new process is called gesture. To detect skeleton the function named *GetFirstSkeleton* was called and the method *SkeletonFrame* was used which maps the joint point of the subject. In this project total 8 joints are captured. Those are head, center shoulder and shoulder, elbow, wrist & palm(these two create one point named as hand) joint of both hands. But in case of gesture detection only 4 joints are taken into account. Those are elbow & hand joint of both left and right hand. Than the sequence of movement of each joint are stored in the database during training session. During translation session when an user delivers a gesture the sequence of those 4 joint are matched with the database and the word corresponding to that gesture are printed.

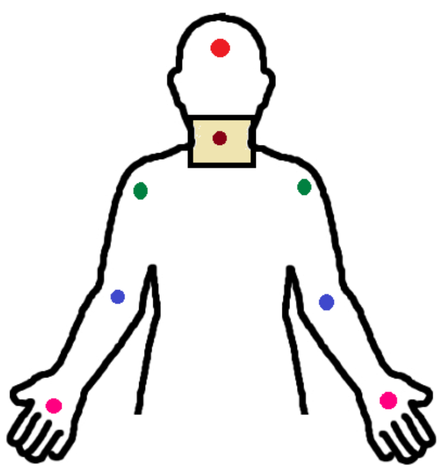


Figure 3.2: Skeleton Tracking

3.5 Extract Feature

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. The step by step process to extract the feature is given below.

3.5.1 Detection sequence of move

As the gesture detection is noisy thus to normalize the data an efficient algorithm is used to detect motion sequence of different joint. A 7*7 square graph created behind the body during gesture detection. In this project when user delivers a gesture the change of position from 1 square graph to another square graph is stored. In this way for each gesture the motion sequence of changing from one graph to another for every 4 joint are stored in the database.

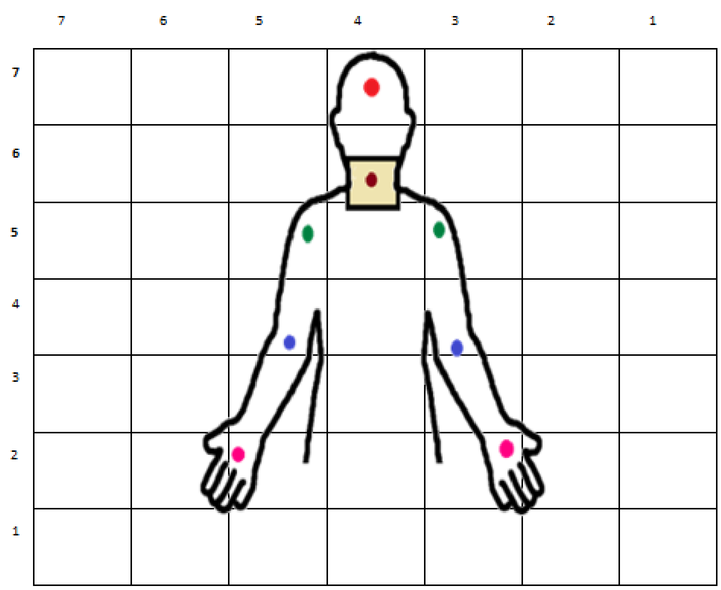


Figure 3.3: A 7*7 Square Graph Created Behind the Body

3.5.2 Calculation of height:

Calculation the height it is an important task for this project because of the variety of the people's height. When the user start the software in training or translation mode than a picture of a person appears. User has to stand like that picture for about 5 seconds. A timer will appear at those 5 seconds. For this timer we use *Dispatcher Timer* class which works by specifying an interval and subscribing to *Tick* event that will occur each time this interval is met .That position standing like picture is called initial position. In meantime when time equals to 1, user's height will be calculated. When a person stands by spreading his two hands as far as possible than the distance between the tip of the middle finger of his/her two hands is his/her height. In this case when the user takes the initial position the vertical distance between his/her shoulder and palm of right hand is calculated. Then the horizontal height between two shoulders is calculated. The user height will be,

$$Height = (shoulderleft.Position.X - shoulderright.Position.X) + (shoulderleft.Position.Y - handleft.Position.Y) *2 \qquad (1)$$

Than depending upon that height the graph is created. So The size of the graph is,

$$Graph = (height\ of\ user)/15*16 \qquad (2)$$

both vertically and horizontally. The size of each square box is,

$$Box= (The\ size\ of\ graph)/15 \qquad (3)$$

both vertically and horizontally.

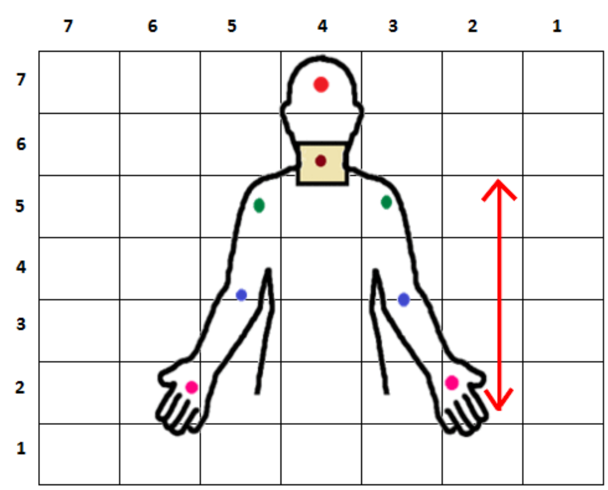


Figure 3.4: Calculation of Height

3.5.3 Variation of height:

As different person have different body structured, so in our project we can say that height will be vary for each person. Thus the sequence of movement of joint for different person of different height will be different on the same graph. Thus the graph size varies for different person depending upon their height. At the time 1, a few things will be calculated as well as the height of the person before the training or translation and the graph created depending upon that height. The system can translate from one person at a time. Thus the height also detected for one person at a time.

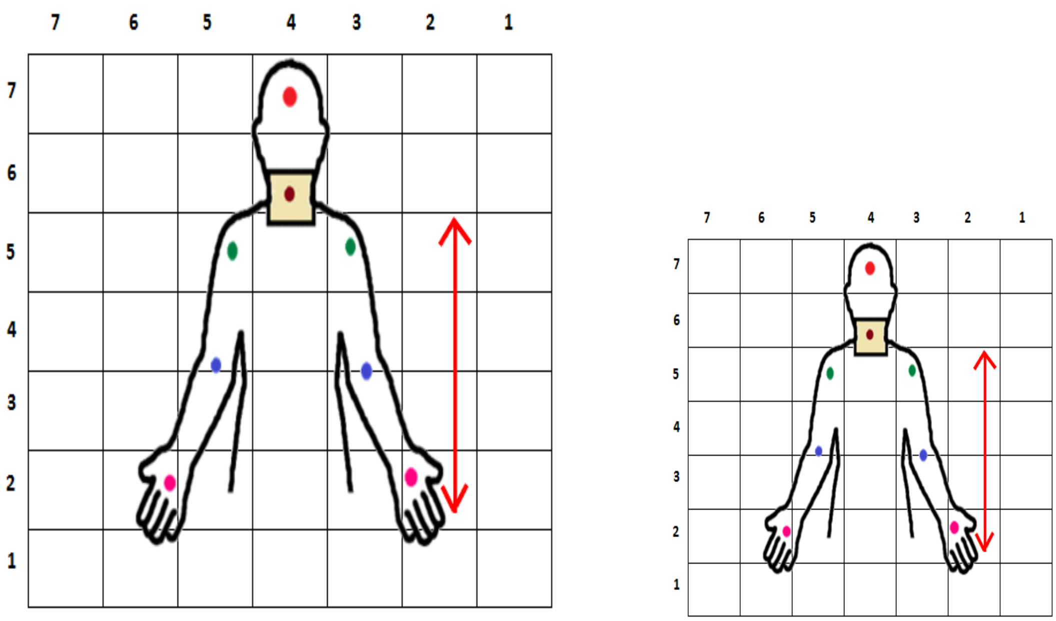


Figure 3.5: Variation of Height

3.5.4 Calculation of Sequence

When a gesture is delivered the sequence of two joint points that is the palm of both hands are calculated separately. That is the sequence or position change of these two joints is stored and the combination of these two data is the expression of a gesture. During training session sequence for each two joints are stored and during translation session the motion sequence of these two joints are calculated and matched with database. In Figure 3.6 the sequence of right palm for a specific gesture are shown.

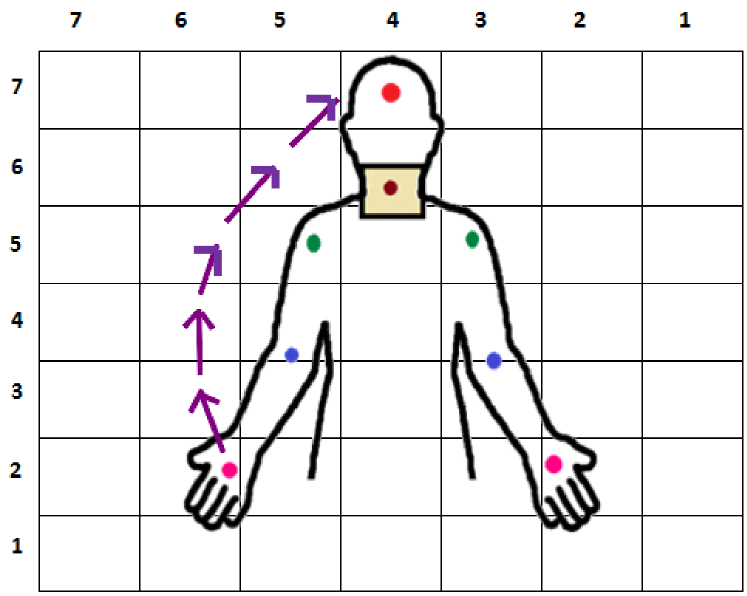


Figure 3.6: Sequence of Right Hand Movement

3.5.5 Save as training set

When a joint point remains inside a box the vertical coordinate and horizontal coordinate is combined and stored in database. The vertical coordinate is multiplied by 100 and the horizontal coordinate is added to it.

For example if a joint point coordinate be (2,3) than its combine reading will be as 3*100 + 2=302. In this way the sequence are stored in the database.

3.6 Architecture of proposed method

As spoken language is different from every country to country, sign language also varies. In our project Bangladeshi sign language (BdSL) is used. In BdSL, each gesture contains different word not sentence. By combining those words or gesture a full meaningful sentence is found.

In this project the software translates each gesture only in Bangla word not the full Bangla sentence. In this project any user can also train the software to learn each gesture for new word which is not available in the database. Then that word will be included in database and user can use it later. In other way, user also only can translate the gesture which is available in the database.

That means in this project there are two features:

1. Translation &
2. Testing.

The step by step process of our project is given below:

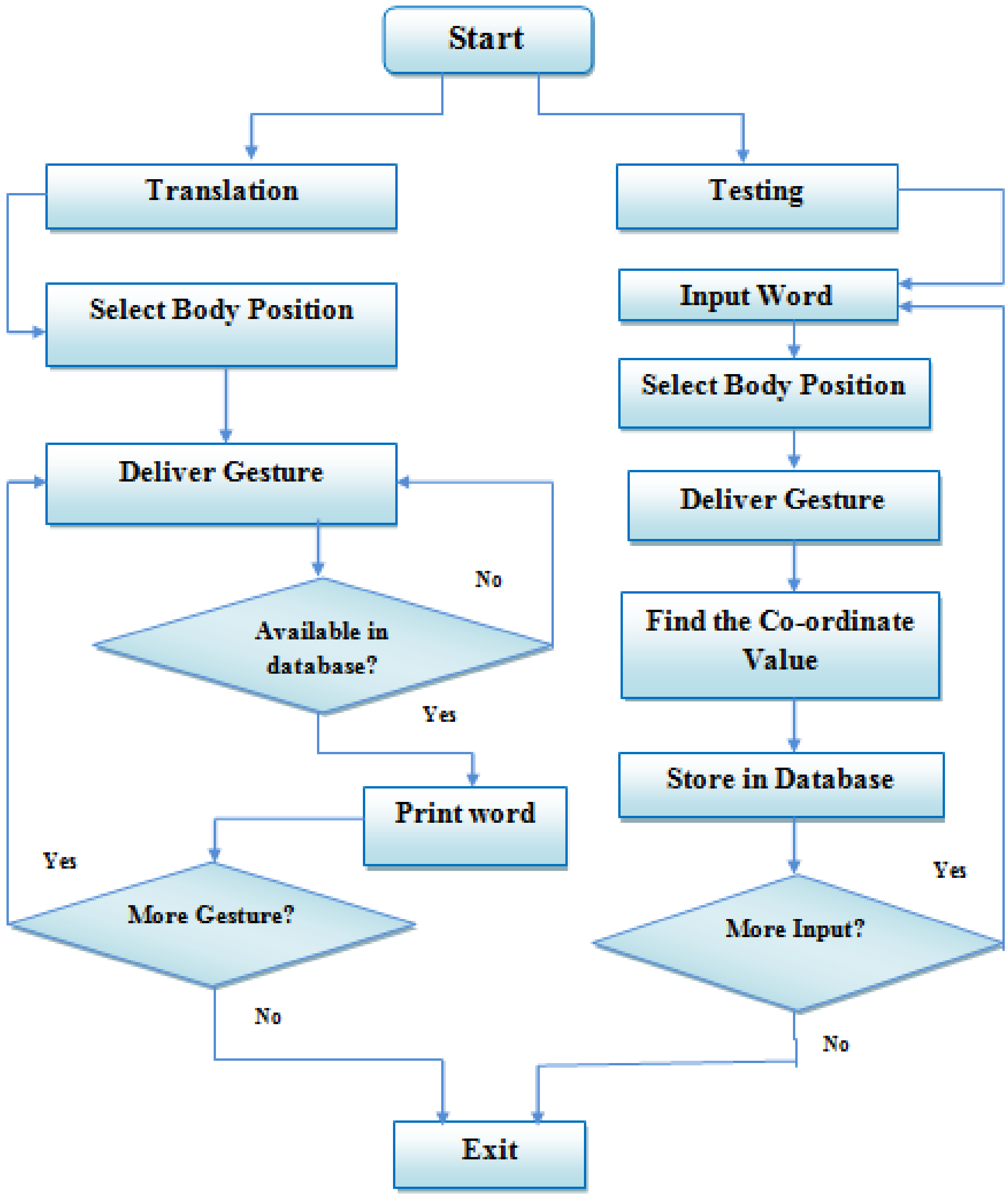


Figure 3.8 : System Architecture

3.6.1 Start

First user enters into the software. Then two buttons will appear: Training & Translation. User has to press any of the button.

3.6.2 Testing

If any user selects the testing button, user can train the system by putting the word along with the gesture by his own. On other words, if any sign language is not available user can input it by him/her self. So if user has to input a new word and its corresponding gesture he has to choose.

3.6.2.1 Input word

When user clicks the testing button, a textbox will appear where user has to input the new word. After writing the word, user then has to click the ok button and the textbox will be empty.

3.6.2.2 Select body position

When software starts to work, a time will appear. Than initial position detection will be done. This is very important in this project. First a picture of a standing man will appear. The user has to stand like that picture for about 5 seconds. One thing is necessary that user must always remain between 2-3 meters from the Kinect Motion Sensor Device. And another thing is that user has to stand so that centre shoulder must remain in the rectangular box.

3.6.2.3 Deliver gesture

Then the user has to deliver his gesture. The initial position and final position of each gesture must be same as the picture.

3.6.2.4 Find the Coordinate value

The user has to deliver the gesture of the same word for 3 times so that the software will get a real time gesture for each word. Than the software will calculate the three captured gesture and find the mean or most accurate value.

3.6.2.5 Store in database

When the software gets the three values, it calculates mean value from these and saves the more accurate value in the database along with the respective word. So storing the value in database is done. If user wants to input another value, he/she have to follow the same procedure from inputting the word to storing value.

3.6.3 Translation

If the user choose translation than user just have to give his gesture and the word correspond to that gesture will be printed.

3.6.3.1 Select body position

As stated earlier, the selection of initial body position of the user for translation mode is as same as the testing mode. So the calculation of the user will be done in this period.

3.6.3.2 Deliver gesture

In this section user has to deliver the appropriate gesture of sign language of any word which he/she wants to know. When delivering the gesture, every sequence will be stored in the linked list. After that the software will check every sequence which is matched in dictionary. If it is found, the respective word will be shown in the textbox. If it is not found, nothing will be printed.

3.6.3.3 Available in database

If the gesture is available in the database and all the value of gesture joint point is matched, then the respective word will be shown. If the available word is not found in the database, nothing will be print.

3.6.3.4 Print the word

Available output will be fetched from the database and show in the corresponding textbox.

3.6.4 Exit

If user doesn’t want any other function to work or wants to exit the software, he/she simply has to cancel the window. It will automatically stop the Kinect Sensor device too.

3.7 Conclusion

In this chapter, we have been described the whole procedure step by step from the Kinect sensor device to some real value to select every movement. In training mode, this value will be saved along with the word. And in translation mode, after matching with value, word will be showed. In the next chapter, the implementation of the system will be discussed.

Chapter 4
Implementation

4.1 Introduction

Implementation of Sign language is a challenging task as the gesture delivered by different person for the same word is slightly different. And also in this chapter the frontend task of the project with necessary diagram are given which will clearly describe the outcome of this project. The experimental setup of this project is also provided here. As stated section 2.7 we used the language C#, so we have written our code in xaml.cs file and the design section was written in xaml file.

4.2 Implementation of the project:

The implementation of the whole project has been described step by step in here.

4.2.1 Start

When user first runs the software, a new window appears in front of the user. Then two buttons will appear: Testing & Translate. User has to press any of the buttons then further procedure will work. Rather than user can simply stand in front of the sensor, 8 joint points will be shown and it will measure the height of the user and the distance from the Kinect. Before clicking any button , there shows some message such as please keep the distance between 2-3 meters Without sensor, program will run and show the staring window.

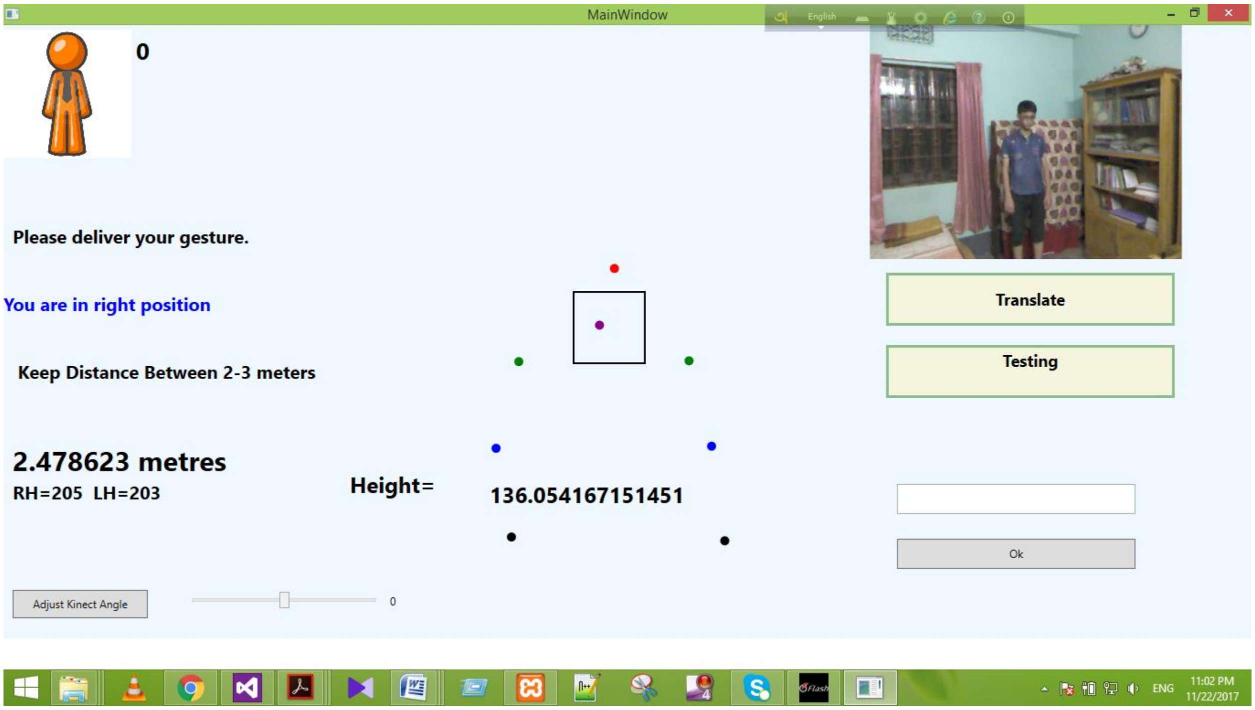


Figure 4.1: Starting window of the Project

4.2.2 Testing

As there are tons of word in any language, there are tons of sign language in respect to those word. In this project we have provided an advantage to train the machine like, user can input any sign language by his own which is currently unavailable in the database. For this user have to click the testing button.

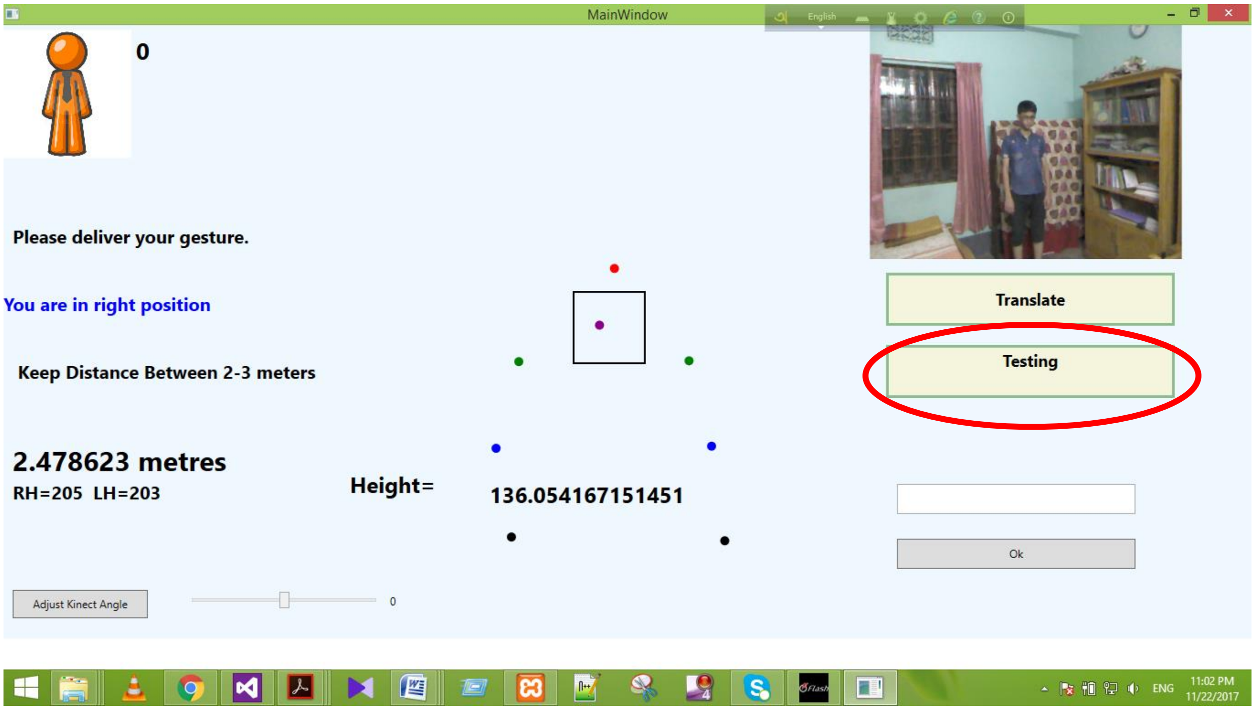


Figure 4.2: Choosing the Testing Button

4.2.2.1 Input the Word

After clicking testing a text will appear saying "Write the Word below". User has to put the word below the text in the textbox. Then clicking the "ok" button, word will be saved in database.

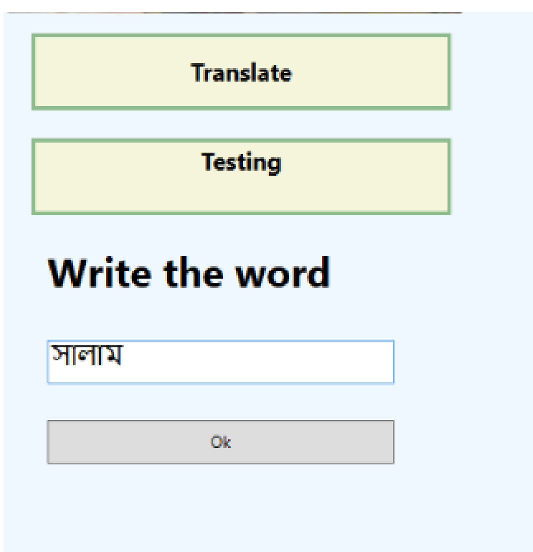


Figure 4.3 : Input the Word

4.2.2.2 Select Initial Body Position:

In the first section, we will discuss about the testing. When sensor shows up and it gets the user in the right position, a timer turns on. When timer starts to count time from 4 to 0, in meantime user have to stand in a position almost 5 second. Actually it is called our initial body position selection. User has to stand in front of the sensor in between 2 to 3 meters as we declared this position as right position. When time will equal one, height and distance will be measured. Timer will stop when it reaches to zero and all the calculation will be showed it in the window.

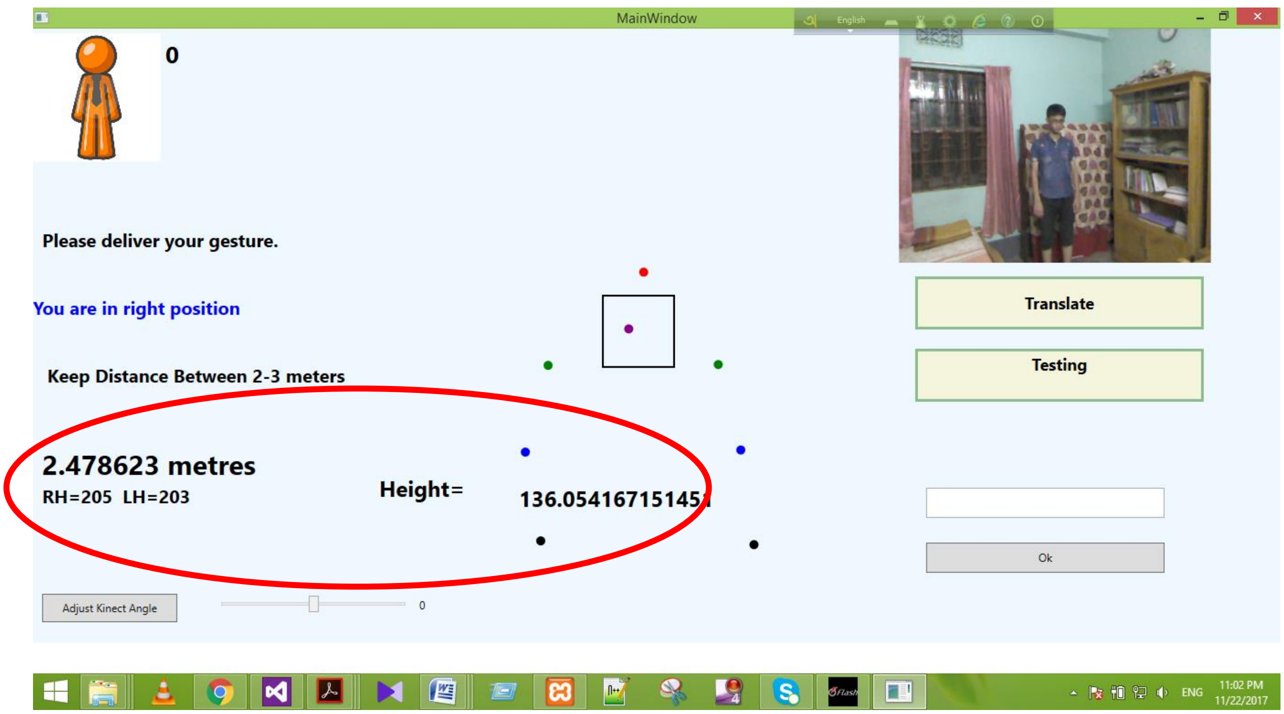


Figure 4.4: Initial Body Position

4.2.2.3 Deliver the Gesture:

User now has to deliver his/her gesture. A thing has to keep in mind that the initial position and the final position of the gesture will be the same. In the meantime the movement will be captured and saved the database.

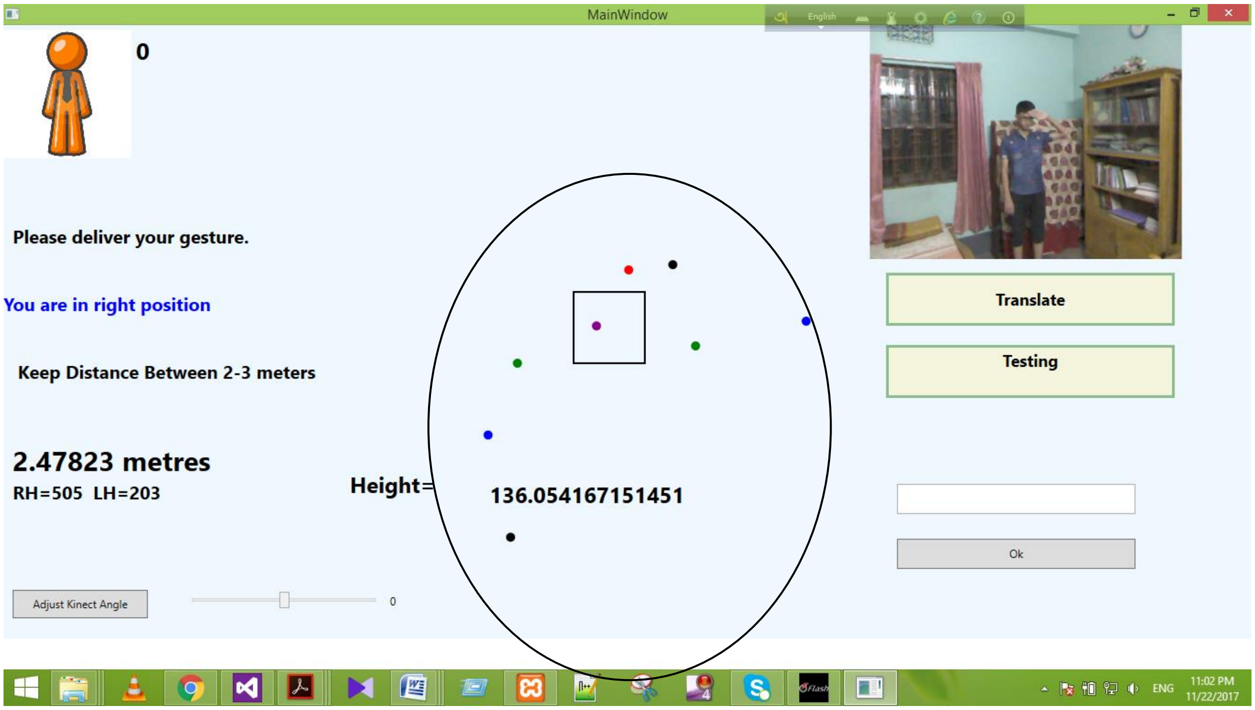


Figure 4.5: Delivering the Gesture.

4.2.3 Translation

In translation mode, all of the procedure is similar to that of testing mode. The only difference occurs when the gesture ends. Unlike testing mode where the gestures are stored in a gesture database dictionary, in translation mode, the gesture currently stored in a linked list is compared to the gestures in the database dictionary. If it matches to any sequence of dictionary, it shows the corresponding word.

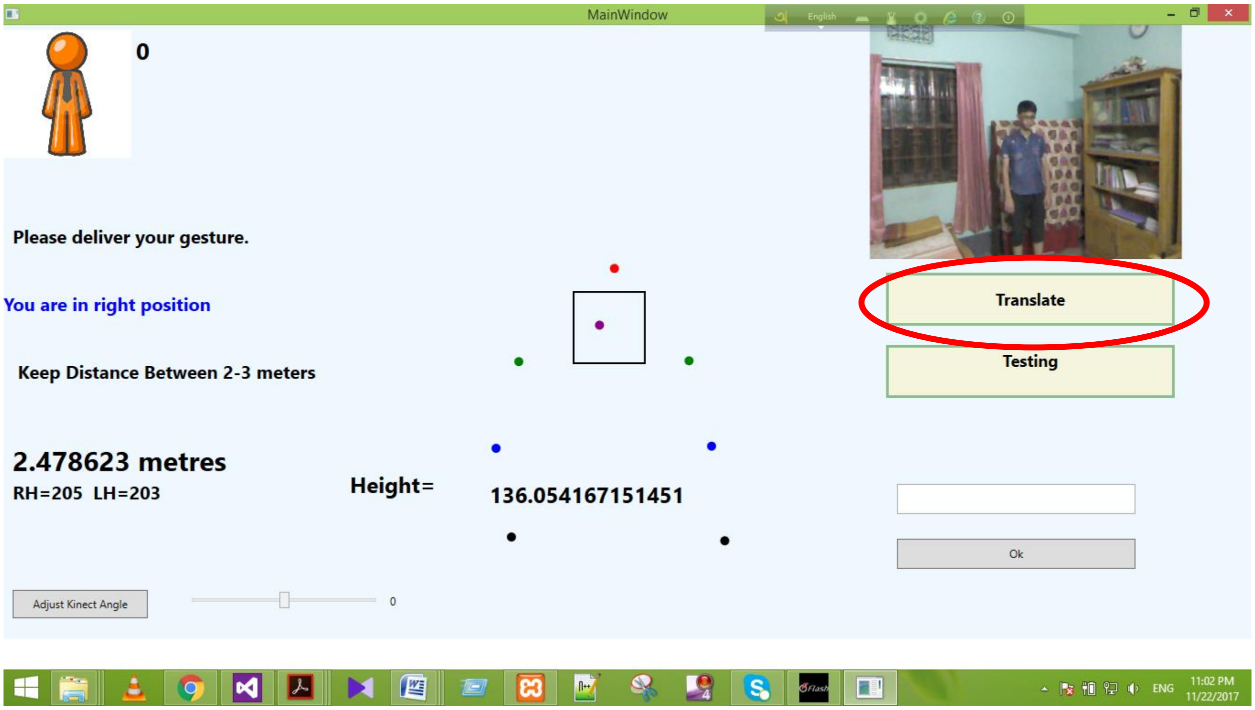


Figure 4.6 : Choosing the Translate Button

4.2.3.1 Select Initial Body Position:

As stated earlier, the selection of initial body position of the user for translation mode is as same as the testing mode. So the calculation of the user will be done in this period.

4.2.3.2 Deliver the Gesture

In this section user has to deliver the appropriate gesture of sign language of any word which he/she wants to know. When delivering the gesture, every sequence will be stored in the linked list. After that the software will check every sequence which is matched in dictionary. If it is found, the respective word will be shown in the textbox. If it is not found, nothing will be printed.

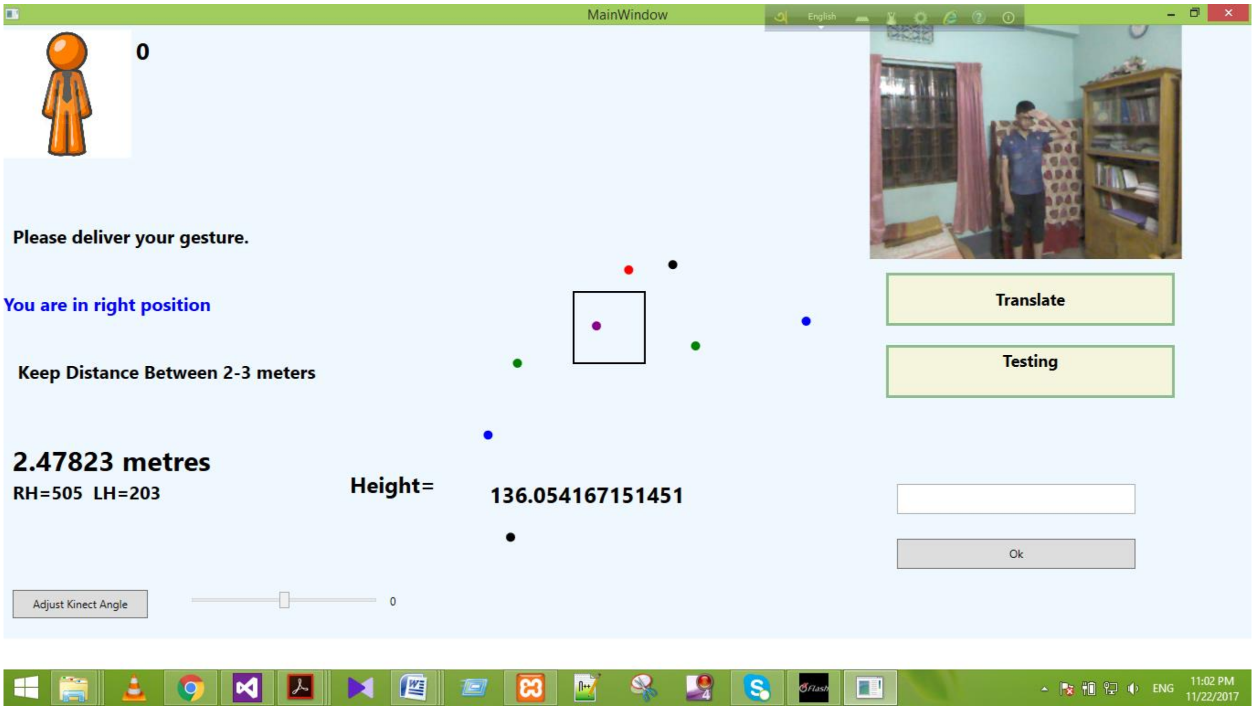


Figure 4.7 : Delivering the Gesture.

4.3.3.3 Show the Word

The desired output will be shown in the textbox if the gesture is matched in any word of the dictionary. If the user wants other gesture to deliver, first have to take initial position and then deliver it.

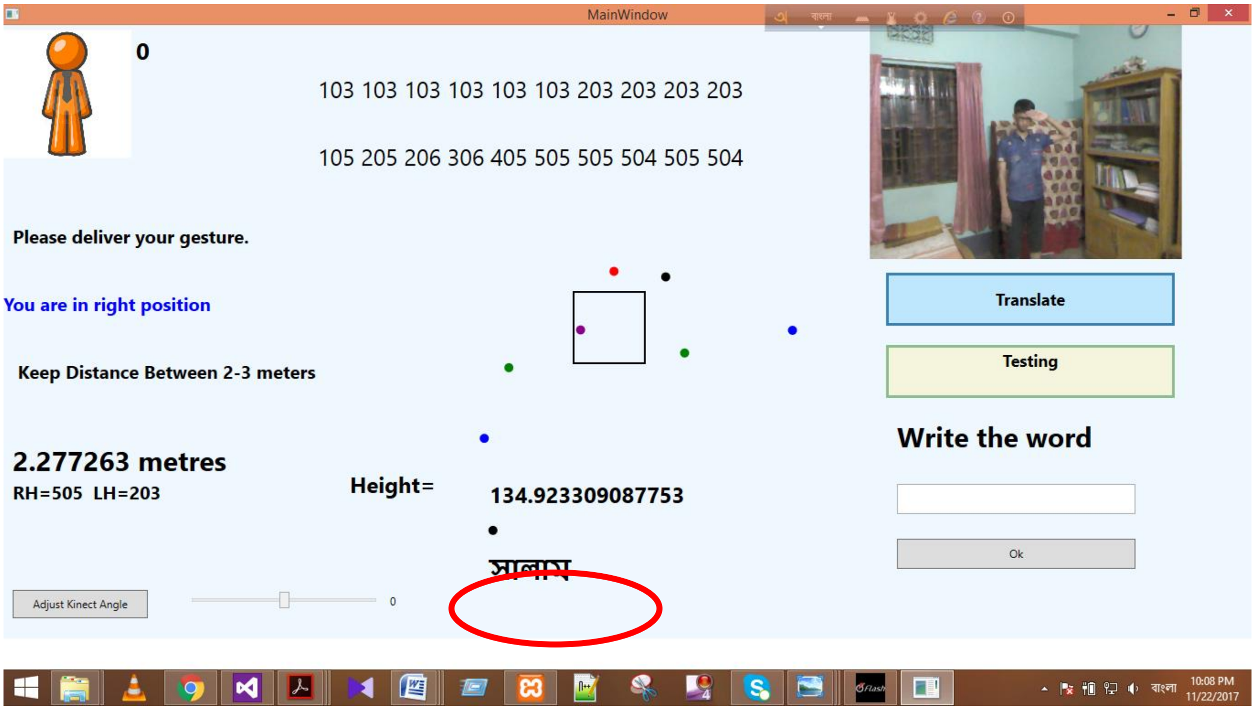


Figure 4.8: Showing the Respective Word

4.2.4 Exit

If user exits the window, Kinect sensor will also be stopped. User simply can close the software by exiting main window.

4.4 Conclusion

In this chapter, we have provided the total working procedure of the software along with the proper example and pictures. In the next chapter, we will discuss about the experimental result.

Chapter 5
Experimental Result & Discussion

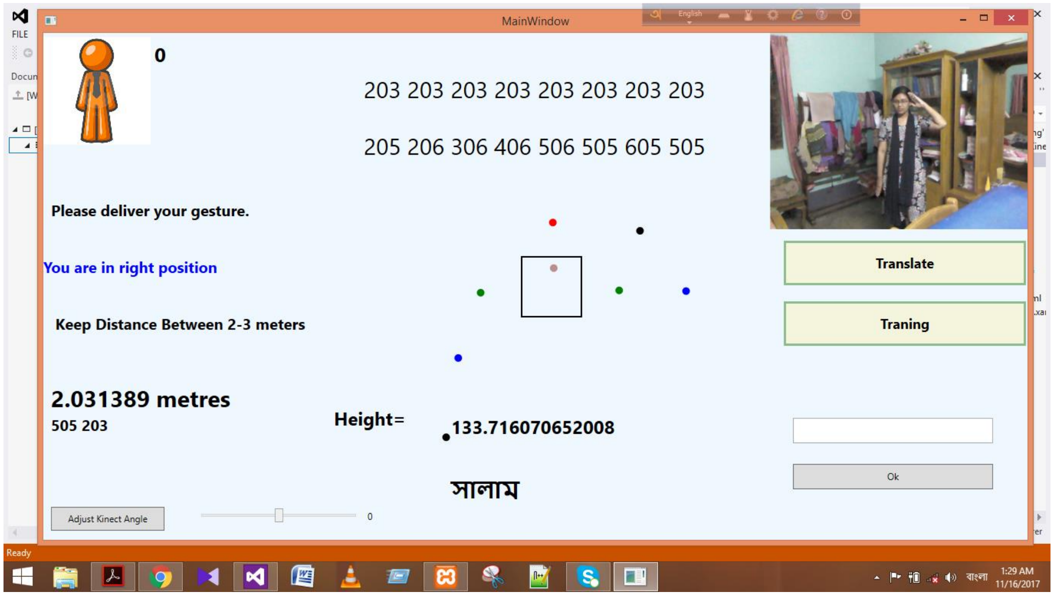
5.1 Introduction

In this chapter, it is analyzed about the extent that this project is successful in achieving the objectives defined at the beginning of this paper. At first some sample gesture is trained by the machine in this project. Those gestures are saved in the database. Then these gestures are delivered by different user. With considering different user by comparing different value we tried to get an accuracy of the project how it is successful or unsuccessful. We also here define that by determining the successful rate and unsuccessful rates. To illustrate our experimental results we needed to calculate the accuracy of the recognition of each sign.

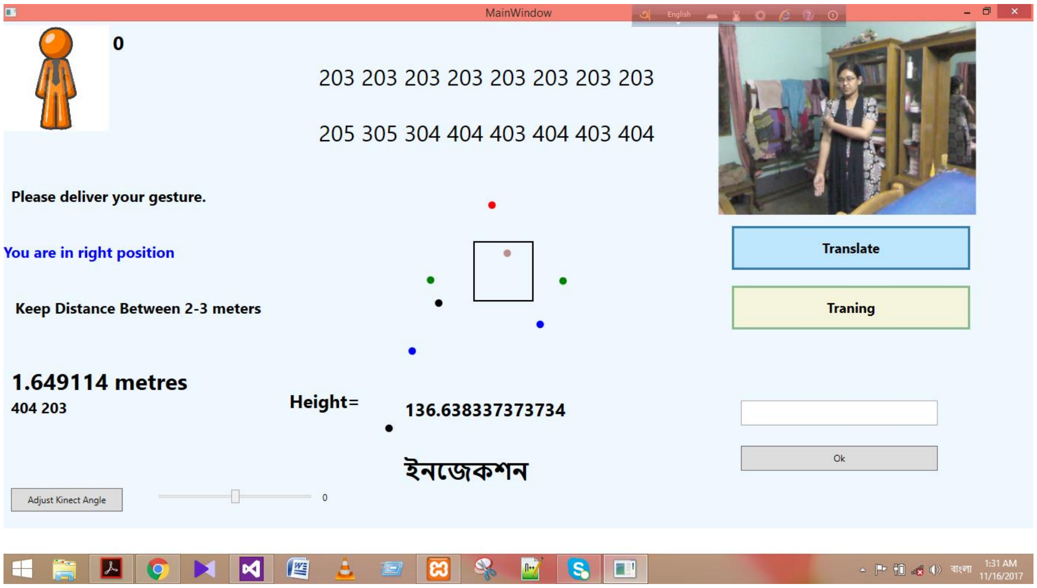
5.2 Experimental Results

5.2.1 Sample Output

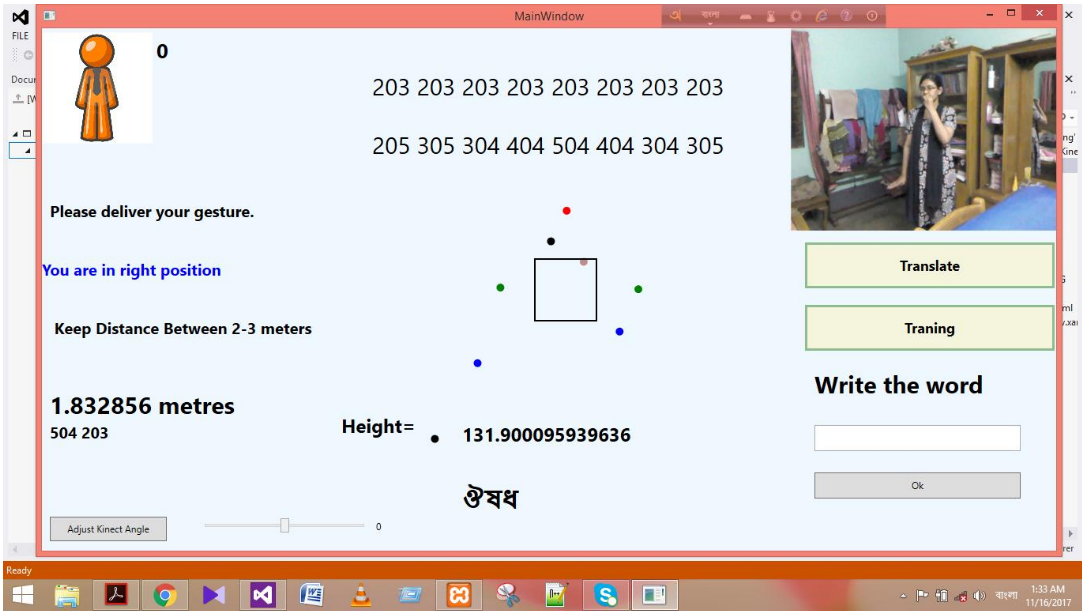
Here we intended to design a automated system which translates Bangla sign language into Bangla text. Our goal is trying to recognize the gestures of user when they perform sign language In front of the Kinect sensor and give them corresponding word as an output. Here some output of this system :


MainWindow
0
203 203 203 203 203 203 203 203
205 206 306 406 506 505 605 505
Please deliver your gesture.
You are in right position
Keep Distance Between 2-3 meters
2.031389 metres
505 203
Height=
133.716070652008
সালাম
Translate
Traning
Ok
Adjust Kinect Angle

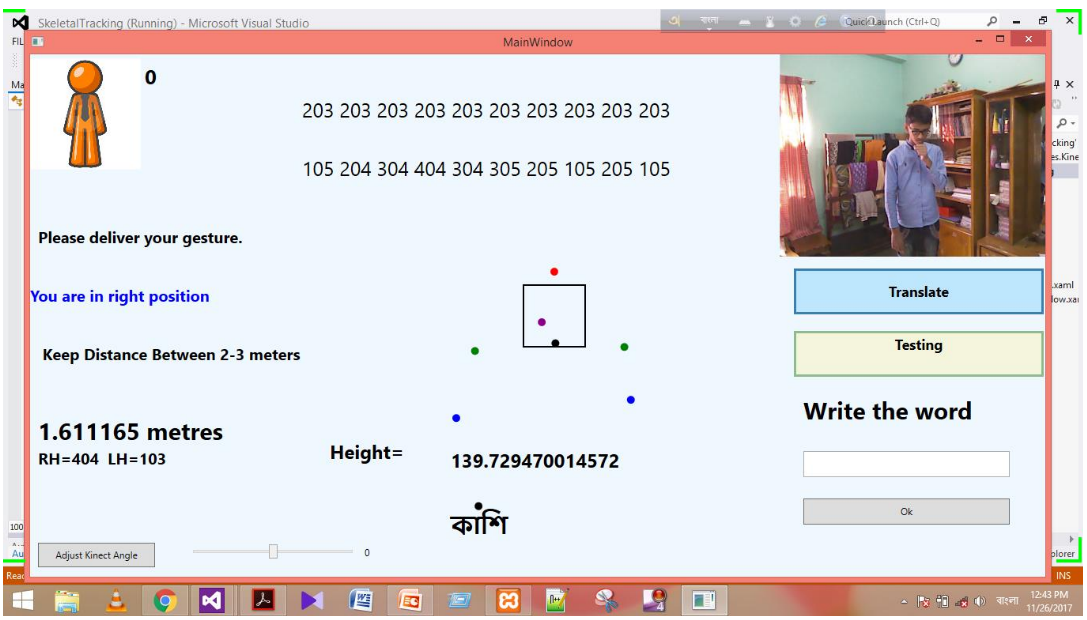
(a)


MainWindow
0
203 203 203 203 203 203 203 203
205 305 304 404 403 404 403 404
Please deliver your gesture.
You are in right position
Keep Distance Between 2-3 meters
1.649114 metres
404 203
Height=
136.638337373734
ইনজেকশন
Translate
Traning
Ok
Adjust Kinect Angle

(b)

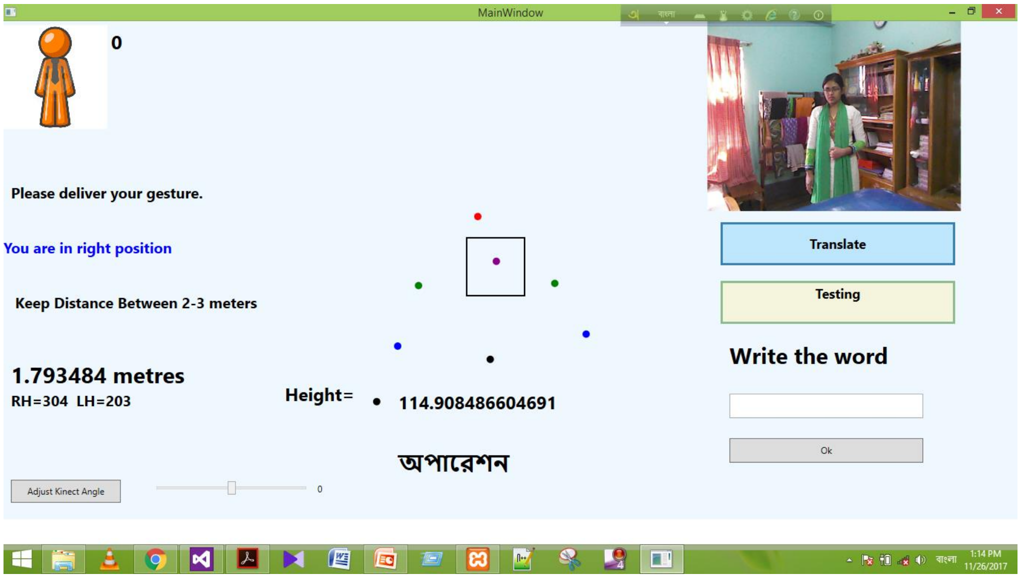

MainWindow
0
203 203 203 203 203 203 203 203
205 305 304 404 504 404 304 305
Please deliver your gesture.
You are in right position
Keep Distance Between 2-3 meters
1.832856 metres
504 203
Height=
131.900095939636
ঔষধ
Translate
Traning
Write the word
Ok
Adjust Kinect Angle
Ready

(c)

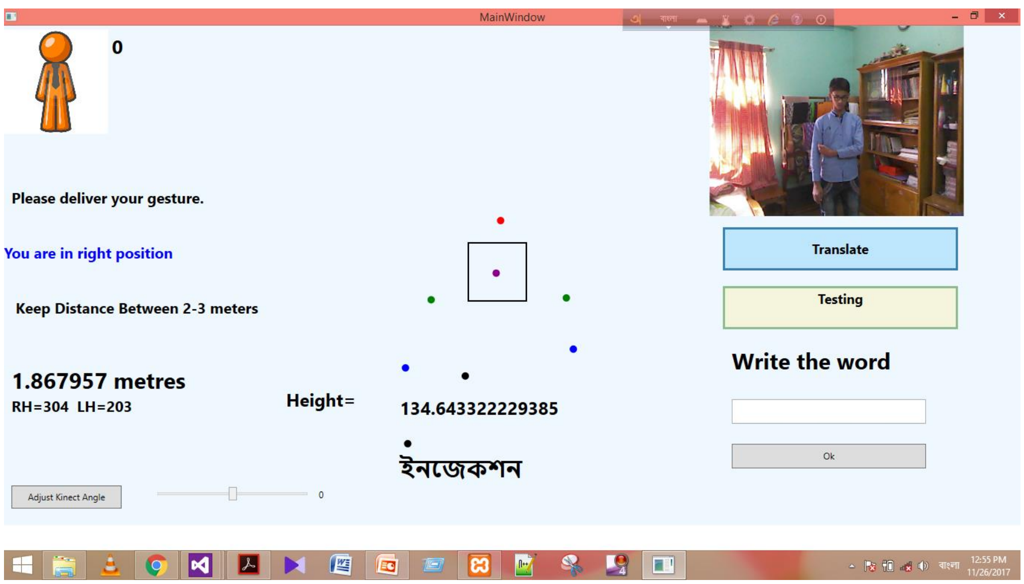

SkeletalTracking (Running) - Microsoft Visual Studio
MainWindow
0
203 203 203 203 203 203 203 203 203 203
105 204 304 404 304 305 205 105 205 105
Please deliver your gesture.
You are in right position
Keep Distance Between 2-3 meters
1.611165 metres
RH=404 LH=103
Height=
139.729470014572
কাশি
Translate
Testing
Write the word
Ok
Adjust Kinect Angle

©

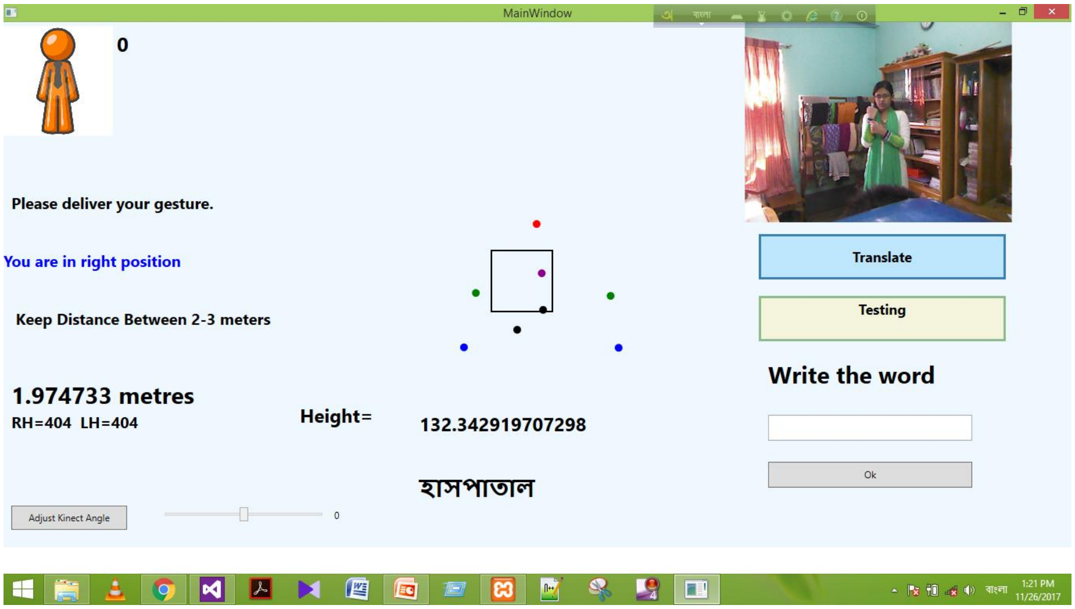
(d)


MainWindow
0
Please deliver your gesture.
You are in right position
Keep Distance Between 2-3 meters
1.793484 metres
RH=304 LH=203
Height=
114.908486604691
অপারেশন
Adjust Kinect Angle
0
Translate
Testing
Write the word
Ok

(e)


MainWindow
0
Please deliver your gesture.
You are in right position
Keep Distance Between 2-3 meters
1.867957 metres
RH=304 LH=203
Height=
134.643322229385
ইনজেকশন
Adjust Kinect Angle
0
Translate
Testing
Write the word
Ok

(f)



(g)

Figure 5.1: Sample output images

5.2.2 Success Rate:

We trained our system to recognize some different gestures. These gestures are delivered by 7 different users. Each of them delivers this gesture and the success rate is made depending on accurately detection of those gestures.

Table 5.1: Success rate for every person upon different gesture

Number	**Person**	**Gesture**	**Successful?**
1	1	সালাম (Salam)	Yes
2	2	সালাম (Salam)	Yes
3	3	সালাম (Salam)	Yes
4	4	সালাম (Salam)	Yes
5	5	সালাম (Salam)	Yes
6	6	সালাম (Salam)	Yes
7	7	সালাম (Salam)	Yes

8	1	হাসপাতাল (Hospital)	No
9	2	হাসপাতাল (Hospital)	Yes
10	3	হাসপাতাল (Hospital)	Yes
11	4	হাসপাতাল (Hospital)	No
12	5	হাসপাতাল (Hospital)	No
13	6	হাসপাতাল (Hospital)	Yes
14	7	হাসপাতাল (Hospital)	No
15	1	ইনজেকশন (Injection)	Yes
16	2	ইনজেকশন (Injection)	Yes
17	3	ইনজেকশন (Injection)	Yes
18	4	ইনজেকশন (Injection)	Yes
19	5	ইনজেকশন (Injection)	Yes
20	6	ইনজেকশন (Injection)	Yes
21	7	ইনজেকশন (Injection)	No
22	1	ঔষধ (Medicine)	No
23	2	ঔষধ (Medicine)	Yes
24	3	ঔষধ (Medicine)	Yes
25	4	ঔষধ (Medicine)	No
26	5	ঔষধ (Medicine)	No
27	6	ঔষধ (Medicine)	Yes
28	7	ঔষধ (Medicine)	Yes
29	1	গলাফোলা(Mumps)	No
30	2	গলাফোলা(Mumps)	No
31	3	গলাফোলা(Mumps)	Yes
32	4	গলাফোলা(Mumps)	Yes

33	5	গলাফোলা(Mumps)	No
34	6	গলাফোলা(Mumps)	Yes
35	7	গলাফোলা(Mumps)	Yes
36	1	অপারেশন (Operation)	Yes
37	2	অপারেশন (Operation)	No
38	3	অপারেশন (Operation)	No
39	4	অপারেশন (Operation)	Yes
40	5	অপারেশন (Operation)	No
41	6	অপারেশন (Operation)	Yes
42	7	অপারেশন (Operation)	Yes
43	1	কাশি (Cough)	No
44	2	কাশি (Cough)	Yes
45	3	কাশি (Cough)	No
46	4	কাশি (Cough)	Yes
47	5	কাশি (Cough)	No
48	6	কাশি (Cough)	Yes
49	7	কাশি (Cough)	Yes
50	1	আস (Come)	Yes
51	2	আস (Come)	Yes
52	3	আস (Come)	Yes
53	4	আস (Come)	Yes
54	5	আস (Come)	Yes
55	6	আস (Come)	Yes
56	7	আস (Come)	Yes
57	1	আনা (Bring)	Yes

58	2	আনা (Bring)	Yes
59	3	আনা (Bring)	Yes
60	4	আনা (Bring)	No
61	5	আনা (Bring)	No
62	6	আনা (Bring)	Yes
63	7	আনা (Bring)	Yes
64	1	মাথা (Head)	Yes
65	2	মাথা (Head)	Yes
66	3	মাথা (Head)	Yes
67	4	মাথা (Head)	Yes
68	5	মাথা (Head)	Yes
69	6	মাথা (Head)	Yes
70	7	মাথা (Head)	No
71	1	হাত (Hand)	Yes
72	2	হাত (Hand)	Yes
73	3	হাত (Hand)	No
74	4	হাত (Hand)	No
75	5	হাত (Hand)	Yes
76	6	হাত (Hand)	Yes
77	7	হাত (Hand)	Yes

5.2.2 Success Rate of Different People

In this section the sample gesture along with its successful rate for each different person are discussed. From table 5.1 the experimental data has found. For each different person the successful rate can be mentioned as following theory:

SR= SW/TW*100%

Here,

SR=Success rate for each different person

SW=Total number of gesture successfully recognized by the system

TW= Total number of gesture delivered by the user.

Now from the experimental data it is seen that from person 1 to 7 has given 7*11=77 gestures and sometimes the gestures are successfully recognized and sometimes the system failed to recognize it respectively. Thus following the above theory, the chart with successful rate is found given in following.

Table 5.2: Successful rate of different people

Person	Total Number of Given Gesture	Total Number of Given Successful Gesture	Successful Rate (%)
1	11	7	63.63
2	11	9	81.81
3	11	8	72.72
4	11	7	63.63
5	11	6	54.54
6	11	11	100.00
7	11	8	72.72

Now if the average successful rate is considered over this chart than mean value will be
(63.63+81.81+72.72+63.63+54.54+100.00+72.72)/7 = 72.71 %

5.2.3 Success Rate of Different Gesture

Previously result for different person is given. Now the gesture for different word is different. Some are simple & some are complex. Some required using of two hands and some required to use only one hand. In which one hand is used like 'সালাম (Salam)', it's given the most accurate value. Time is also variant for different gesture. Now the word chosen as experimental data for this project include most of these criteria. Though much complex gesture is neglected. Here the accuracy rate for each word are given.

Table 5.3 : Accuracy rate for different word

Word no	Word	Number	Total Number	Accuracy rate
1	সালাম (Salam)	7	7	100
2	হাসপাতাল (Hospital)	7	4	57.14
3	ইনজেকশন (Injection)	7	6	85.71
4	গলাফোলা(Mumps)	7	4	57.14
5	ঔষধ (Medicine)	7	5	71.43
6	অপারেশন (Operation)	7	4	57.14
7	কাশি (Cough)	7	4	57.14
8	আস (Come)	7	7	100.00
9	আনা (Bring)	7	5	71.43
10	মাথা (Head)	7	6	85.71
11	হাত (Hand)	7	5	71.43

Now if the average successful rate is considered over this chart than mean value will be (100+57.14+85.71+57.14+71.43+57.14+57.14+100+71.43+85.71+71.43)/11 = 74.02 %

5.3 Evaluation

As mentioned earlier in this chapter our experimental evaluation for different gesture and different person are discussed. Here two types of result are discussed: For different gesture and for different person.

5.3.1 Different Persons Evaluation

The gestures given by different person for the same words are slightly are different. This is due to size and shape of different persons is different. It is already being mentioned that this project will appear for different height of person in different way. The gesture is stored depending the grid appear behind the person. The height of person is calculated and size of grid depends on that height. But size of people is not variable on height only, but also it depends on weight and shape of body structure of different person. Here it is seen that for the 6th person all gesture deliver was successful as this person was identical to the trainer. For the 2nd, 3rd and 7th person the successful rate were 81.81% and 72.72%. They were almost same in height and weight to the first person. Thus we got a good successful rate from these two persons. In case 1st and 4th person the successful rate was 63.63%. Here this person was skinny and taller in height than the first person. In this case we also got an expected successful rate which is more than the average result. In case of 5th person we got 54.54% accuracy. This big person was fat and tall than the first person. That is in case of fat or more weighted person we get some error result.

5.3.2 Different Gesture Evaluation

During analyzing the result it is seen that the successful rate doesn't only dependent on person but also on gesture. The successful rate for some gesture are perfect where in case of some gesture a very few successful rate is found. This happens due to the following reason:

- Since the RGB camera requires good lighting conditions, a low light condition may result in a capturing failure. Figure 5.1 shows how different lighting conditions affect the measurement accuracy. Although the same Kinect device captured the same gestures performed by the same tester, the trajectory in the good lighting condition is much smoother than that in the bad lighting condition.

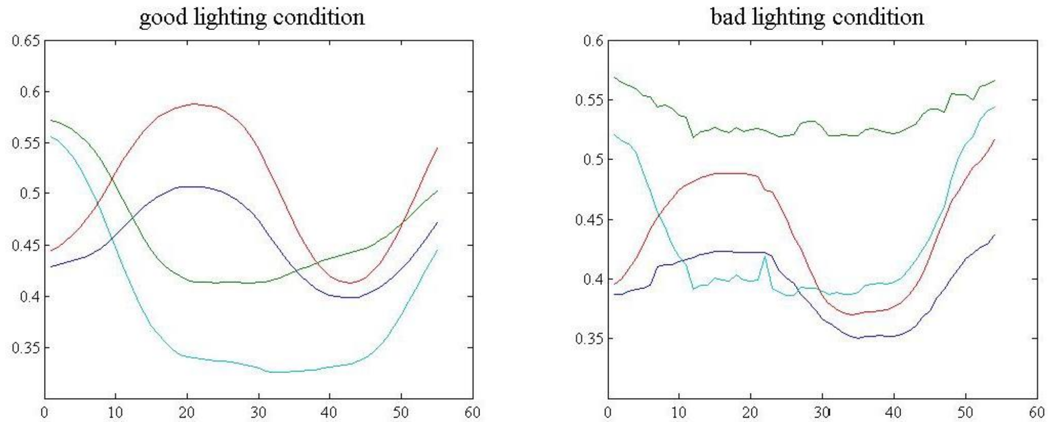


Figure 5.2: Same Gesture Measured in Good and Bad Lighting Conditions

- When a detecting point or body joint overlap on one another an error result is found. This is due to when Kinect Motion Sensor Motion captures a joint and calculates the depth data there occurs some problem. Sometimes when a detecting point overlaps on one another the points fluctuate for a small portion of time. Thus the depth data calculated is not correct. As a result the gesture stored in the database or the gesture is comparing with the database are incorrect. Ultimately an error result is found.

- Sometimes the gesture detected by the device is not same as the gesture delivered. This is because the gesture captured by the device is frame by frame in a fixed time interval. Thus the gesture captured and the gesture delivered is slightly different. For this in a very few case incorrect data is found. This is applicable for both training and translation mode.

5.4 Result

A total of 7 people participated to test our system, each of different body structure. We trained our system to recognize 11 Bangla sign word gesture. Depending on these two criteria success rate is found which already being mentioned. If the results are shown in graph chart than we get the following charts

Figure 5.2: Successful rate depending on person

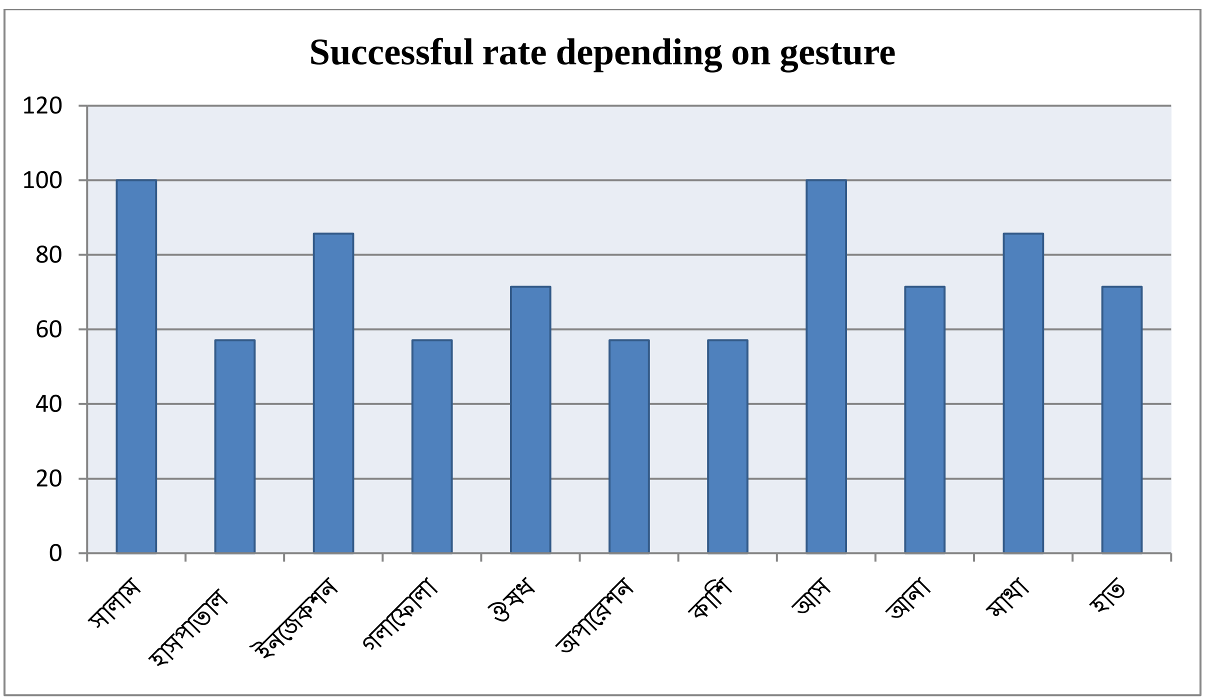


Figure 5.3: Successful Rate Depending on Gesture

5.5 Conclusion.

The primary focus of the work is to develop a system which can train sign language and translate also. For this purpose different person and different gesture are taken as experimental data to find the accuracy rate of this system. By experimenting on those people depending on those words the successful rate is calculated. So here it is seen that though the project can’t give a hundred percent accuracy but it gives a good successful rate.

Chapter 6
Conclusion & Future Recommendation

6.1 Conclusion

In this project mainly three things has being done. Those are: Capture the gesture, Store the database or compare with database and show the output. In this project some sample word as taken as input word for testing and several persons have being used to find an accepted result. The data found from those has being calculated and a theory has being deduced to find the accuracy rate. Though this project has got some challenges in different sections like detection of gesture, capturing of gesture, real time depth value and variation of output for different word & also for different person. These drawbacks have being removed as far as possible. In fact of those this project can play a vital role in future communication system. A revolution took place in science and technology in the last generation. One of the reasons behind this revolution is rapid development of communication. But still in some case there are some shortcomings in communication. The main purpose of this project is to play a vital role in our present communication system and also to implement more developed version of this project to make the future communication system as fast and efficient as possible in case of both deaf and general people.

6.2 Future Recommendation

As there is no end point in development, we can always make a system a better one. There are some certain scopes for improvement in our developed project. We have marked some areas to bring into consideration for improvements and future research.

For future work of this project many recommendation is got from different person. If these following things are fulfilled than the project will be a complete one. The future recommendations are:

- Detection of finger.
- Output in voice.

- Make the gesture detection more accurate
- Working with much larger training dataset.
- Developing more efficient algorithm for the calculation procedure to produce better result.

Bibliography

[1] M. T.Hoque, Md. R.Tauwaby, Md. F. Kabirz, F. Sarkerx, M.N. Huda and K. A.Mamunk, “Automated Bangla Sign Language Translation System : Prospects, Limitations and Applications” In Proc. 5th *International Conference on Informatics, Electronics and Vision (ICIEV,2016)*

[2] M. A.Rahim , “Visual Recognition of Bengali Sign Language using Local Binary Pattern Compared with ANN” ,In Proc *Global Journal of Computer Science and Technology: F Graphics & Vision,*ISSN: 0975-4172(Online) & 0975-4350(Print), Vol. 14,no.2 ,2014

[3] D. M. Capilla, MSC Thesis On Sign Language Translator Using Microsoft Kinect XBOX 360, University of Tennessee (Knoxville - USA).

[4] F. M. Rahim, T. E. Mursalin, N. Sultana, “Intelligent Sign Language Verification System – Using Image Processing, Clustering and Neural Network Concepts” *International Journal of Engineering Computer Science and Mathematics*, ISSN:0976-6146, Vol.1 No.1,2010.

[5] D. S. H. Pavel, T.Mustafiz, A. I. Sarkar, and M. Rokonuzzaman, “Geometrical Model Based Hand Gesture Recognition for Interpreting Bengali Sign Language Using Computer Vision”, in Proc *International Conference on Computers and Information Technology*, 2003.

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

[7] T. E. Starner, MSC Thesis on Visual Recognition of American Sign Language Using Hidden Markov Models, Massachusetts Institution of Technology.

[8] K. Deb, H. P. Mony & S. Chowdhury. "Two-Handed Sign Language Recognition for Bangla Character Using Normalized Cross Correlation.", Global Journals Inc. (USA) 12, no. 3,2012.

[9] D. S. H.Pavel, T. Mustafiz , A.I.Sarkar, M.Rokonuzzaman. "Modeling of Bengali sign language expression as dynamic 3D polygons for developing a vision based intelligent system for dumb people” In Proc. National Conference on Computer Processing of Bangla, 2004.

[10] S. Rahman, N. Fatema, and M. Rokonuzzaman, "Intelligent Assistants for Speech Impaired People," In: *International Conference on Computer and Information Technology*, East West University, Dhaka, Bangladesh. 2002.

[11] A.M. Jarman, S. Arshad, N. Alam and Md J. Islam, "An Automated Bengali Sign Language Recognition System Based on Fingertip Finder Algorithm" In: *International Journal of Electronics & Informatics*, ISSN: 2186-0114 http://cennser.org/IJEI Vol.4, No.1, 2015.

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

[13] "Sign Language"[Online].Available : https://www.britannica.com/topic/sign-language WRITTEN BY The Editors of Encyclopædia Britannica

[14] "Sign Language" [Online].Available : http://en.wikipedia.org/wiki/Sign_language.

[15] N. S. Juthy, B. G.Shaham and M. S. A.Shojol. "Ishara Bhashay Jogajog (Communicating through Sign Language)," Dhaka: Center for Disability inDevelopment, 2005.

[16] "Bengali language - wikipedia, the free encyclopedia," [Online].Available: https://en.wikipedia.org/wiki/Bengali language, (Visited on 09/30/2015).

[17] Z. Halim PhD & G. Abbas PhD, "A Kinect-Based Sign Language Hand Gesture Recognition System for Hearing- and Speech-Impaired: A Pilot Study of Pakistani Sign Language" ISSN: 1040-0435 (Print) 1949-3614 (Online) Journal homepage: http://www.tandfonline.com/loi/uaty20

[18] "Kinect" [Online].available : https://en.wikipedia.org/wiki/Kinect

[19] Bangladesh Sign Language committee. "Bangla Sign Language Dictionary".Dhaka: National centre for special education & Ministry of social welfare, 1994.

[20] " Kinect skeleton tracking" [Online].available: https://msdn.microsoft.com/en-us/library/

[21]"Static gesture recognition" [Online]. available: http://www.sciencedirect.com/science/article/pii/S0031320316000157

[22] “Dynamic gesture of sign word” [Online]. available:
https://www.speechbuddy.com/blog/language-development/5-ways-sign-language-benefits-the-hearing/

[23] S.S.Hazari,Asaduzzaman,L.Alam,N.A.Goni, “Designing a sign language translation system using kinect motion sensor device” ,In Proc. Electrical, Computer and Communication Engineering (ECCE), INSPEC Accession Number: 16851191, 2017

[24] “How does the Kinect work?”[Online].Available :
https://users.dickinson.edu/~jmac/selected-talks/kinect.pdf

Appendix

As mentioned before we have used eleven signs to train our system during the testing mode. Information on how to perform these signs is given below. All of the sign was taken from the Bangla sign Language Dictionary.