Sign Gestures to Sinhala Language

Lasitha Lakmal Waduge

168273G

Degree of Master of Computer Science

Department of Computer Science

University of Moratuwa
Sri Lanka

May 2020

Declaration

I declare that no portion of the work referred to in the dissertation has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning. Further, all the work in this dissertation is entirely my own, unless referenced in the text as a specified source and included in the bibliography.

Signature:  Date:24/05/2020

Name: LasithaWaduge

Index No : 168273G

Signature of the supervisor: …………………………. Date: ………………..

Name: Dr. Indika Perera

Abstract

This research is based on the Sign language. However, there is a communication problem between deaf and normal persons unless there's no translator accessible.

This project tries to find out what are the situations in society that the disable person faces any difficulties. As an example, if he goes to a reception of a place, there are set of frequently used signs to communicate. So this project focuses on a specific situation and tries to identify what are the common symbols which need to be used by the disabled person. Since Sri Lankan sign language contains more than 2000 sign based words, in the initial step, this project tries to identify frequently used set of words needed by disabled people more often.

As the initial step of the proposed system, it needs to track the gestures of the disables person. Proposed system use to hardware device for Microsoft Kinect embedded camera. This will capture the video frames and track the hand gesture to generate the sign character. Tracked information will be sent to the gesture dictionary which is trained for range of hand gestures formerly. So gesture information input should match with dictionary information and identify the most matched word related with the gesture. Finally normal person will see the Sinhalese word related with the sign, sent by disabled person.

Acknowledgements

I am using this opportunity to express my gratitude to everyone who supported me throughout the course of this MSC research. I am thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during the research work.

I express my warm thanks to Dr. Indika Perera, for his support and guidance as the supervisor of the present research.

Further I would like to pay my gratitude to the authors of previous researches, books and articles in the field for their generosity in sharing their valuable knowledge.

I should thank the respondents of the research for giving their fullest support to understand their world of perceptions.

L.L. Waduge.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

DTW –Dynamic Time warping

NN – Nearest Neighbor

HMM – Hidden Markov model

IR – Infrared

SDK – Software Development KIT

API – Application Program Interface

SLT – Sign Language Translation

BSL - British Sign Language

Chapter 01

Introduction

1.1 Background

Many things such as the anatomy, habitual behaviors and communication methods differentiate human beings from animals. The use of a significant language in order to communicate with a fellow person can be set as an example to prove the previous statement. Throughout the history humans from different cultures and countries were able to develop various types of languages to speak to one another. However people with disability such as deafness cannot use a spoken language in the same manner as a person without any disability. Some people are born with this condition and some become disabled later in their lives due to unfortunate circumstances. World Health Organization, lot of people has disabling hearing loss, which is 5% of the total population [1]. Since this should not hold them back from communicating with the society, Sign language was developed as a solution to this problem. Sign language is mainly based on using hand gestures, facial expressions or body language to mediate a message as opposed to an oral language which uses sounds to convey a message across.

Similar to an oral language, sign languages also contain a several standards. It can be based on actual languages used by signers within regions and dialogue groups and also the notations with which sign languages were represented in written form. American Sign Language (ASL), Italian Sign Language (LIS) and British Sign Language (BSL) are examples for Sign Language standards around the world.

People with hearing or speaking disabilities in Sri Lanka use Sri Lankan Sign Language. This is also a variation of standard Sign Language. Sign language assists them when performing daily activities or interacting with the community while making them less dependent on someone else regarding such activities.

1.2 Motivation

Usually people with disabilities face a lot of issues when trying to perform their daily activities. Deaf or mute people often require assistance from another person to communicate their needs and wants. These difficulties can be fairly evident in public places. If they go to a bank, there's a good possibility, the person at the counter not understanding the exact need of the disabled person. Unless there's a translator this could not only happen in banks but in other public places such as hospitals, bus or train stations, police stations. This puts both parties in a difficult situation which could lead the disable person feeling humiliated or miserable about their lives and avoiding the interaction with the general crowd. And ultimately disable people are more likely to be isolated from the society.

With the advancement of the technology, geographical barriers are not a problem for the general crowd when it comes to communication. Plenty of social networking application have been developed through the years making the entire process easier. Technologies which use visual communication such as Skype will let two deaf people in different places have a conversation using sign language but this cannot be counted as a real solution for disabled people. Mainly these tools do not provide any methods to execute translation between an oral language and sign language. And also because of hardware and network related problems it has become a far more complex process for a deaf person to have a conversation with an ordinary person who is not aware of sign language.

By looking at the reasons mentioned above it can be clearly seen the need for a sign language translator. This should translate Sri Lankan Sign Language to Sinhalese language. Deaf people can use the translator to have a successful communication with ordinary people in public places.

1.3 Objectives of proposed system

The suggested system will be a real time translation method which is able of identifying a range of gestures of the Sri Lanka Sign Language and converting those signs into the matching Sinhalese word.

- Build a platform to allow smooth communication between deaf and ordinary persons by identifying basic set of signs in Sri Lankan sign language which is frequently used and translating them to Sinhalese Language.
- Provide ability to majority of Sinhalese people to easily interact with deaf community by providing meaning of each sign using Sinhalese characters.
- Explore the possibility of making the system usable in public places such as banks, hospitals, etc. So that deaf community will be able to achieve their day to day tasks more easily.
- Expand number of signs in the translator by introducing and easily trainable sign word mechanism within the system.

1.4 Scope of the system

The Sri Lanka sign language consists of more than 2000 signs [3]. Therefore it is quite difficult to build a system to identify all the signs within the given time frame.

WE අපි	**Come** එන්න
I මම	**Don't** එපා
Mother අම්මා	**Go** යන්න
Children ළමයි	**Where** කොහේද
Help උදව්කරන්න	**Train** දුම්රිය
Bus බසය	**Thank You** ස්තූතියි

Table 1.4 Hand gestures to identify words

The system will detect hand gestures to identify the above mentioned words. The words will be selected in a way to track hand movement. Nonetheless this project does not consider complex sign words which use not only hand gestures but also facial expressions and lip/finger movements.

The system mainly based on capturing hand sign which represent a word and the word will be then converted in its relevant Sinhalese word. Grammatical correction of a sentence will not be handled by the system. The system pays more attention to translating a single sign based word to its Sinhalese word in the most accurate way possible. The selected sign based words will be identified within short time intervals by proposed system.

Chapter 02

Background

This chapter will provide background studies done regarding the proposed system. It will present an overview of sign languages and also similar research studies done on the same domain. Further a summary on Microsoft Kinect sensor will be given in this section.

2.1 Sign Languages

There are just about lot of (100) sign languages can be found all in the word. These are several sign language standards that are being used in some countries in the world and those sign language standards are American Sign Language (ASL), British Sign Language (BSL) Mexican Sign Language (LSM), French Sign Language (LSF), Italian Sign Language (LIS), Irish Sign Language (IRSL), Australian Sign Language (Auslan), German Sign Language (DGS), Israeli Sign Language (ISL), and Spanish Sign Language (LSE) [2]. These sign languages standards have been used as a platform by other countries to create their very own local sign languages. Most of the spoken languages' corresponding sign languages have based on the above mentioned sign language standards. This is one of the main reasons which make it difficult to build a general translator for all the sign languages. Sri Lankan Sign Language has got some influence from the British sign language. Hand gestures are a representation technique of the British sign language.

This section will provide an in depth details on some common behaviors of sing languages. Due to different sign language having its own set of rules this section will not provide details on every single sign language that's being used at present. This will rather explain a several collective characteristics between dissimilar sign languages. The goal here is to illustrate the complexities such as considering all the variation when building a fully functioning Sign Language translator.

2.1.1 History of Sign Language

"Deaf people could not be educated. Without hearing, people could not learn" Greek philosopher, Aristotle had pronounced the above statement. [3] This particular statement was challenged by Geronimo Cardano, a physician of Padua, and he stated that deaf people could learn via sign communication. [3] Juan Pablo de Bonet, wrote the first book on teaching sign language, the manual alphabet in 1620, by supporting Geronimo's statement [3] sign language was already instinctively developed by deaf signers. The writing of the book was turning point in the history of sign languages development as it introduced independent set of hand gesture based signs for every individual character in English alphabet. Figure 2.1 shows set of the signs and corresponding English characters introduced by Pablo de Bonet.

Figure 2.1 Character based signs by Pablo Bonet[4]

The first free school for deaf people was founded by Abbé Charles Michel de L'Epee of Paris, in 1755 at France [3]. He developed sign system for French sign language. In 1960, William Stokoe of United States published a first linguistic book of American Sign Language. This book contained the first dictionary, and he organized the signs depending onits shapes (position of the hand, shape, motion, etc.) and notable depending on its translation of English. This publication marked the beginning of research about Sign Language linguistics, which then spread into more than 90 countries of the world [4].

2.1.2 Sri Lankan Sign Language

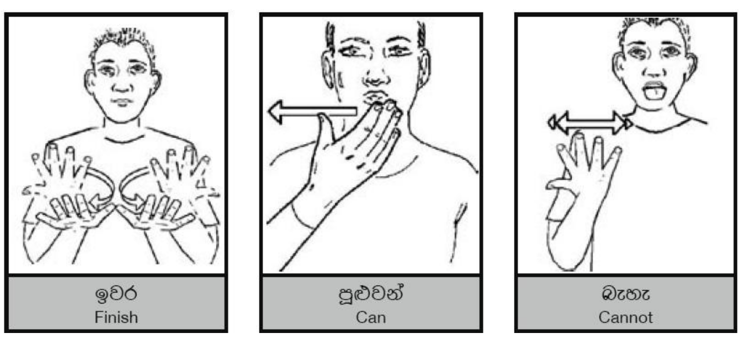
A British lady named Mary Chapmen introduced Sign Language to Sri Lanka by a in 1912. She was specialized in teaching for deaf and mute persons. Mary Chapmen lead the way to open the first deaf and mute school in Sri Lanka –"Rathmalana School for deaf and mute". Our Country has heavily built on the foundation of BSL because Mary chapmen was a teacher from Great Britain [5].

2.1.3Sri Lankan Sign Language used Gesture types

A gesture may be a human body motion or action that's meant to speak a message. In Sri Lankan signing, completely different mechanisms area unit accustomed represent signs. 3 main part will be known once it involves presenting sign primarily based word.

1. Hand Gestures
2. Lip Movements
3. Facial Expressions

A single above mentioned technique or a combination of those techniques will be used to convey a sign. Figure 2.2 shows some of the examples in Sinhala Sign language dictionary which uses above mentioned methods for presenting the sign. Picture (a) shows that the sign is conveyed only using hand gestures where picture (b) shows a combination of hand gestures and lip movements to convey the idea. Picture (c) is also an example for a sign, which uses a combination of lip movements and hand gestures, and kind of a facial expression is involved.



(a) (b) (c)

Figure 2.2 Sinhalese Sign word representation techniques [6]

2.1.4 Sign Language Hand gesture variations

Hand gestures play a major part when representing technique for sign language. The entire hand each finger can be useful when conveying some signs. The reason to use fingers is that finger variations express variety of meanings in sign language. According to Diane Brentari[7], the hand shape in sign languages can be analyzed with respect to four main sections.

1. Feature Geometry - Within structure of hand shape is best represented utilizing a particular chain of importance of highlights that catches the connections among the bases of hand shape.
2. H1/H2 - qualification of hand shape that catches the jobs of the two delivers a two-gave sign: the predominant hand, hand which used to speak to the genuine sign (H1) and the non-prevailing hand, hand which is inert during communication via gestures portrayal or offers additional help in depicting the motion (H2)
3. Selected/Unselected fingers - distinction of hand shape that captures unselected fingers which fingers are fore-grounded or back grounded . Chosen fingers can move during the explanation of a sign, and furthermore can expect a bigger range and progressively expand setups of joints. Unselected fingers can expect just two joint arrangements: completely open or completely shut. Figure 2.3 clarify the utilization of chose and un chose fingers in sign based word portrayal.
4. Selected fingers/joints - 2.3 shows the relationship among the joints and selected fingers while sign language words representation.

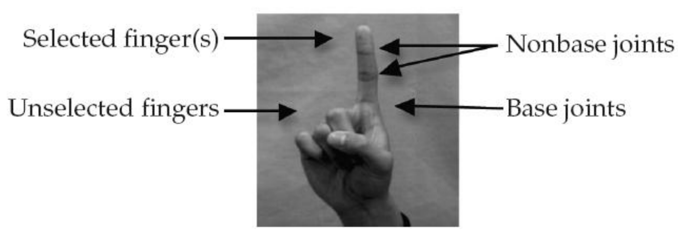


Figure 2.3Relationship between selected fingers [7]

2.2. Related Work

This section will focus on analyzing various types of research work done covering this area in recent past. Used methods and technologies will be compared to find out advantages and disadvantages of the past work.

2.2.1 Sign Language recognition Methods

Rashmi D. Kyatanavar [8] has demonstrated set of vision based methods in recognizing and translating sign language with respect to gesture capturing, language recognizing and classification techniques. This research states that there are three main vision based techniques for sign language recognition as follows.

.

1. Gesture based communication acknowledgment dependent on Skin shading
2. Sign language recognition Using Custom Made Color Gloves
3. Sign language recognition Using Finger Detection

The primary technique will change over communication via gestures into voice signal by following head and hand motions. This framework proposes a basic signal extraction calculation for separating highlights from the pictures of a video stream. Each picture outline portioned into three districts which are head, left hand and right hand. A while later these portioned pictures changed over into paired pictures. For each edge, the region of items in sectioned paired picture is determined in highlight extraction stage. In this manner, each edge has three sectioned zones which are head region, left hand region and right hand region.

For each motion type there will be an alternate portioned region. Each fragmented zone is treated as a discrete occasion &discrete cosine change (DCT) is applied to it.

Initial 15 DCT coefficients considered as highlights. They will relate to each portioned zone. Mix of DCT coefficients from 3 portioned picture territories are utilized as feature vector for Neural Network (NN). A basic NN model has created for sign acknowledgment in this examination. The highlights registered from video stream are given as a contribution to this NN.

Second strategy has proposed a programmed vision based SLT framework. Uniquely designed shading gloves are utilized in this framework. These are standard gloves with explicit shading on palm and every fingertip. At first, picture obtaining has done utilizing a camera &custom-made shading gloves.Using color segmentation & image processing motion of hand tracked. Lastly, Artificial Neural Network (ANN) is used for gesture classification and translation. Camera used to obtain images and video of signer in image acquisition stage. Color-coded gloves were used by signers. The shading helps in extraction of information from sign pictures through shading division. In picture procurement, edges of pictures were gathered at reasonable interim, in this manner that casings were adequate to distinguish the development of hand. Shading division performed to expel impact of moving items in foundation moving following stage. Centroid of shading portioned area is found after movement following stage.Motion tracking phase included two steps:

- Hand posture extraction
- Tracking of hand movement

Location of each finger is determined in the Hand Posture Extraction phase. In the end Nearest Neighbor (NN) classifier has been used to determine the actual sign.

Third technique has presented a productive and quick calculation for recognizing the quantity of fingers opened in signal. Here the finger location performed utilizing the idea of limit following and fingertip recognition. This framework didn't have restrictions like hand ought to be impeccably adjusted to camera, utilization of extraordinary markers, utilization of info gloves on the hand, and so forth. As per this strategy, it was conceivable to distinguish the motion of a letter set if quantities of fingers opened. The point of this technique was to structure a framework that ought to naturally catch, perceive and make an interpretation of gesture based communication to the discourse for dazzle individuals. What's more, it ought to dissect and change over the discourse to either sign or printed show on screen for hard of hearing individuals.

According to results if this research, highest recognition rate received for using finger detection for sign language recognition where the accuracy rate was nearly 95%. Other than the methods provided above, there are number of methods available for Sign language recognition purpose. Hybrid method, which is a combination of above methods, can also use for SLT purpose.

2.2.2SimilarResearch Areas

Daniel Martinez Capilla[4] declared interpretation framework utilizing Microsoft Kinect XBOX gadget as the motion catching of the framework.The signs which were used to detect and recognize through the system did not belong to any standard sign language. The esearcher mainly paid his attention on coming up with systematic flow of steps and exploring the possibility of using the kinect device for sign language translation purpose. Kinect hardware device embedded camera used as the main device for data collection.

For the classification purpose two different classifiers were developed. The first one was Nearest-Group using with the DTW algorithm. This algorithm was used as a cost function. The second classifier was NN-DTW classifier. This has based on Dynamic Time Warping algorithm. The accuracy was 95.2% for understanding trained gestures. However the only drawback of the system there was no testing evidence with the system for the signs of a standard and acceptable sign language. Because there was not any acceptable sign language he introduces his own set of signs to be translated using the system. Finally the system needed to be checked for achieving the previous accuracy rate and also standard sign languages.

Yang Quan introduced [9] a Fundamental Sign Language Recognition framework that had the option to interpret an arrangement of signs into the normally utilized discourse language and the other way around. The framework thought to be introduced out in the open places, for example, air terminals, emergency clinics, and so on. Word reference of words contained explicit signs that permits hard of hearing client to transmit what the person in question needs. This sign language is bi directional translation which means that it is capable of translating symbols to speech and speech to symbols.

They had the option to make the framework increasingly hearty by joining hand motions acknowledgment with the lip developments. The exactness rate was 95.55% for perceiving 30 letters from Chinese manual letter set and they had the option accomplish this rate by utilizing a multi-fates .

Blunt Huang and Sandy Huang thought of a framework to decipher American Sign words utilizing kinect sensor specifically its skeletal tracker [11].The signs that were been chosen, involved arm movements, rather than individual finger movements. At normal skeletal tracking distance, the Kinect sensor does not resolve the hands or fingers clearly. This was the reason behind it. The tracked frames were stored using Cartesian coordinates. These stored frames provided to the classifier for identification reasons. The project used pre trained LIBSVM as the classifier. For the training of the LIMSVM individual date has been used.

Thad Starner and Alex Petland [10] thought of a Visual acknowledgment framework for American Sign Language interpretation. They were utilizing a jargon of 40 words which has been chosen among pronouns, action words, things and modifiers. This framework was presented for sentence level interpretation of ASL. A following camera has been utilized to watch client signs. A Hidden Markov model (HMM) based calculation had been utilized to dissect the client information sources and Sign language interpretation. The detailed exactness pace of the framework was 99.2%.

Group of Chinese scientists made an effort recently to translate the sign language to natural language using Microsoft kinect device. They presented new technique for dissecting the video outline caught through Kinect camera [12]. This framework additionally presented another calculation which can create and coordinate 3D direction. At first, the 3D direction depiction comparing to the info Sign language word produced by hand following innovation gave by Kinect Windows SDK.Considering the difference of hand motion speed, a linear re-sampling has done to get the normalized trajectory by averaging the accumulated length of the whole vector. This operation aims to normalize the trajectory of each word into the same sampling point. To give the recognition result matching scores computed according to the Euclidean distance measurement to. At the beginning the system was implemented recognize 239 Chinese sign language words.

The last usage comprises of two modes which were interpretation mode and correspondence mode. Under interpretation mode which used to distinguish every SL based word exclusively and move recognized words for sentence acknowledgment, 3D direction coordinating calculation was presented. So as to reenact the sign based word the specialized technique utilized for trading data utilizing SL and 3D symbol incorporated to the framework. This was a viewed as finished framework for interpreting gesture based communications.

Table 2.1 represents a summary of some of the latest research work carried out as mentioned above, with the intention of comparing the technologies, algorithms and the respective sign languages used in each work.

Name	Sign Language(word count)	Technologies/Algorithms	Accuracy %
Basic Sign Language Recognition system[9]	Chinese Sign Language	• SVM • Video Camera • Video Display Terminal	95.55%
Visual recognition system for ASL translation[10]	American Sign Language – 40 words	• Video Camera • Hidden Markov Model	99.2%
Custom Sign Translator[4]	Custom set of signs (Not belongs to any Standard SL) – 12 signs	• Microsoft kinect Sensor • Open NI/NITE framework • DTW algorithm	95.2%
ASL interpreter[11]	American Sign Language – 10 Signs	• Microsoft kinect Sensor • LIBSVM	97%
Chinese Sign Language Translator [12]	Chinese Sign Language – 239 signs	• Microsoft kinect Sensor • 3D Trajectory Algorithm	96.32%

Table 2.1Comparison of research projects carried out for sign language translation

Several projects on Sri Lankan Sign Language have been carried out in Sri Lanka. One of the projects was done by Information and Communication Technology Agency (ICTA) the goal here was to employ a web application language the public [5]. There are number of lessons where user can follow to gain knowledge about different signs and the meaning of each sign in different natural languages such as Sinhalese English and Tamil. This web application consists of useful collection of lessons for those who are interested in learning Sri Lankan sign language. This project excludes any mechanism for translating sign language into natural language in real time.

Hand region of the image will be extracted and further filtered to get the binary image of the hand once the image is sent to the system as input. Variety of feature extraction techniques applied on the image and later the corresponding meaning of the gesture is identified. This research focused on translating the language within minimum cost rather than using dedicated capturing devices for gesture detection. Therefore, it is impossible to use the proposed methodology for real time sign language translation.

Beside the project work mentioned above, there has not been a completed real time sign language translation system developed for the public to use for Sri Lankan sign language. Due to this reason, the proposed system will bridge the gap between Sri Lankan sign language and Sinhalese language.

2.3 Recently used technologies and platforms comparisons

Four most commonly used sign language recognition and translation technologies were identified after analyzing the past research work done on this area.

- Glove based technique[8]
- Hidden Markov Model (HMM)[8][10]
- Artificial Neural Networks(ANN)[8]
- Microsoft Kinect Platform[5][10][11]

In order to find out the most suitable technique to identify the sign languages, a research was done by a set of Indian Scientists [9]. The technologies mentioned above were evaluated in terms of three aspects: Accuracy of recognition, Cost of Implementation and number of occurrences each technique used for sign language recognition purpose. Results of the comparison are shown in Figure 2.4.According to the results, in terms of Accuracy in recognition Kinect platform shows a comprehensive difference compared to other techniques. But, according to the results, Kinect platform has not been used in many projects since it is one of the most recent hardware developments. Therefore, more research need to be carried out to find out the relevance of Kinect sensor for proposed study. In terms of implementation cost, glove based technique can be considered as the most cost effective solution.

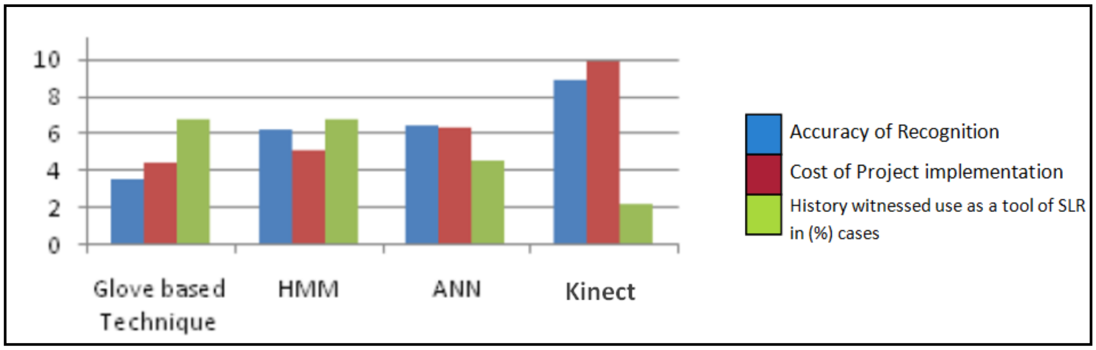


Figure 2.4 Comparison of technologies used for Sign Language recognition [9]

2.4 Use of hardware in gesture acquisition

Gesture acquisition is one of the main considerations in any sign language translation system since the accuracy of the detection depends on the way how gesture has been captured by the sensor or the camera. Related studies have shown various Hardware devices have been used for capturing purpose. Such devices are gesture gloves [8], Video camera or web cam [10], Microsoft Kinect Sensor [4] [11] [12].

2.4.1Kinect XBOX 360 Sensor

Microsoft Kinect XBOX sensor is one of the most recently used devices for gesture capturing. Even though this device has been introduced by Microsoft Corporation specifically for gaming, nowadays it has been used in areas such health, Virtual Reality, education, etc. [15]. This section will provide as to why Kinect is mainly suitable to use as the sensor for the proposed study.

Software platforms such as Open Kinect, OpenNI/NITE Middle-ware and Microsoft kinect SDK are the most famous platforms that have been developed to access the information provided by the sensor.

OpenKinect is an open community which builds free and open sources libraries which can be used with kinect sensor [16]. These libraries can be used in almost every platform such as Linux, Macintosh as in windows Platform etc.

OpenNI (Open Natural Interaction) [17] ,This framework can be considered as most stable open source framework up-to-date, for Kinect and similar sensor devices. This framework is not specifically designed for kinect sensor and some of the features are unable to access through this framework. However Open NI is able to provide all major sensor data of kinect sensor such as depth data, Color data and audio information for analytical purpose.Microsoft Kinect Software Development KIT is the official platform released by Microsoft forKinect sensor based application development. Main advantage of this SDK is that it contains comprehensive set of built in libraries and well written documentation which helps to build applications on the top of kinect sensor. This SDK is considered as the most stable platform to develop Kinect based applications.

Table 2.2 shows a comparison between these three technologies with respect to supported programming languages of each platform, usable operating systems and supported features.

As per the comparison, lib freenect platform supports variety of programming languages and can be used with many operating systems. However compared to other two platforms, features supported by this framework is less. Microsoft Kinect software development KIT supports number of features though it can be only used in Windows platform.

Platform	Programming languages	Operating systems	Features
Open Kinect / libfreenect [16]	C, Python,action-script, C#, C++,Java Javascript, Common Lisp	Linux, Windows, Mac OS X	• Color and Depth images. • Accelerometer data. • Motor and LED control. • Record color, depth and accelerometer data in a file.
OpenNI/NITE Middle-ware [17]	C, C++	Windows, Linux, Ubuntu	• User identification • Feature detection. • Gesture recognition. • Joint tracking. • Color and Depth images. • Record color and depth data in file.
Microsoft kinect SDK [15][18]	C++, C#, Visual Basic	Windows	• Capturing and processing the color image data stream • Processing the depth image data stream • Capturing the infrared stream • Tracking human skeleton and joint movements • Capturing the audio stream • Enabling speech recognition • Adjusting the Kinect sensor angle • Getting data from the accelerometer • Controlling the infrared emitter

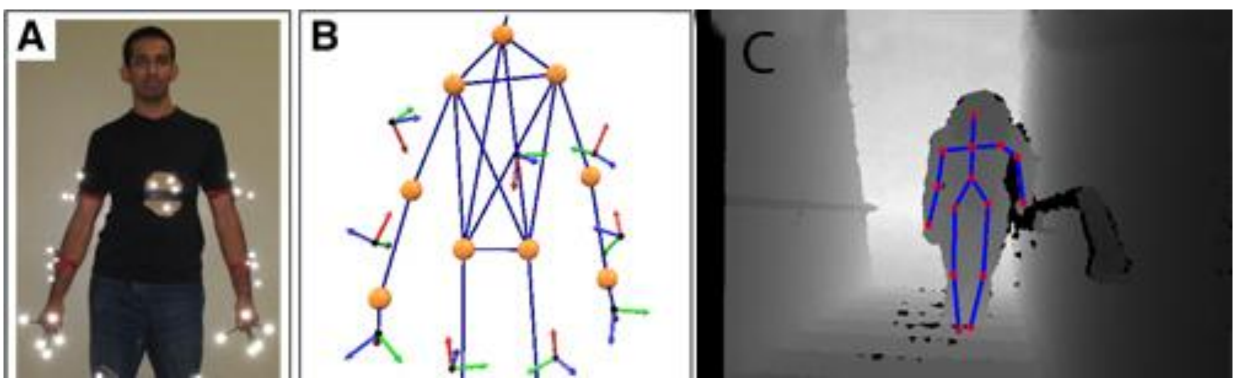
Table 2.2 Comparison between Kinect Software development platforms

Visual data captured by kinect sensor, can be divided into three main types as follows,

- **Color data** - The RGB camera of the kinect sensor produces a basic color video feed like any of the normal video camera or web cam. This camera is capable of capturing video streams using the different resolutions and frame rates.

- **Depth data** - Depth data generally defines the distance between a particular object in the captured environment and the kinect sensor. Sensor will analyze each pixel of the object and calculate the distance to each pixel from the sensor in millimeters.

- **Skeleton data -** Skeleton following is the handling of profundity picture information to build up the places of different skeleton joints on a human structure. For instance, skeleton following figures out where a client's head, hands, and focus of mass are. Skeleton following gives X, Y, and Z organizes values for every one of these skeleton focuses in 3 dimensional space.
- This information can be accessed by any of the software development frameworks that have been discussed earlier paragraph.Figure 2.5 shows the real-time previews of the above mentioned three types of visual information displayed in a kinect sensor based application. Image (a) represents the color data provided by kinect senor where image (b) is an example of the skeleton information. Since image consists of a human, skeleton data will be able to detect collection of joint positions in human figure. Image (c) is a real time preview of the depth data stream.



(a) (b) (c)

Figure 2.5 Real time preview of kinect visual information

2.5 Sinhalese Unicode

The proposed system is supposed to display the meaning of sing language using Sinhalese characters and this is one of the main goals of the system therefore it is necessary to understand the impact of Unicode characters on this project. This section gives an overview on Unicode character usage in a computer system

2.5.1 Importance of Unicode in a Computer

All the data stored in a computer are getting stored in binary format which can be represented by 1 and 0. To represent characters a separate way was needed as characters were also stored as numbers in a computer. There were not any issues when it came to storing characters in the English language. Issues arose when trying to store characters from different languages in the world. A mechanism was introduced to map as a solution to this problem. In the mechanism, character shapes of another language replaced the shapes of English characters. This method also had downsides when there were documents with multiple languages. Finally Unicode was introduced as solution to all the issues [19].

2.5.2 Sinhalese Unicode characters Usage

Sinhala was encoded in Unicode in 1997, despite the fact that the usage of Sinhala Unicode was not started until 2003 [19].Unicode territory dispensed for Sinhalese characters are in the middle of the scope of 3456-3583.In request to work with Sinhala Unicode work usefulness, it requires the accompanying parameters.

1. Sinhala Unicode font
2. Shaping Engine such as MS Windows, Linux etc.
3. Sinhala Unicode Keyboard driver for input and this is not necessarily required for Unicode text rendering. [19]

2.6 Summary

This chapter aimed at providing background information on different aspects related to the proposed system. The first section gave advanced descriptions on Sri Lankan sign language and various types of sign languages that are being used around the world. In the second section similar research carried out under the same domain as well as different types of technologies used to build sign languages translation systems were discussed. The comparison of various technologies was done on section three. Section four mainly focused on Microsoft Kinect XBOX 360 sensor camera which is a piece of hardware specifically built for gesture acquisition. The final section discussed about the Unicode characters in computer systems and also Sinhalese Unicode in computer systems.

Chapter 03

Design and Methodology

This chapter contains detailed description about proposed design of the software application. Furthermore, Systematic flow of the design and specific design considerations of each internal system module will be described in detail.

3.1 Overview of Proposed System

The proposed system suggests a solution for a successful communication method between deaf or mute persons with an ordinary person if they do not possess any knowledge in sign language. This software based solution will convert Sri Lankan sign symbol into Sinhala. The main features of the system is to capture gestures through a camera and identify the respective sign based word. Afterwards the gesture will be processed to recognize the meaning of the sign. Finally the identified word will be displayed to the user in Sinhalese letters in an embedded display of the system. Therefore by displaying the word represented by the sign language, ordinary people can tend to the needs of the disabled person by taking appropriate actions.

.

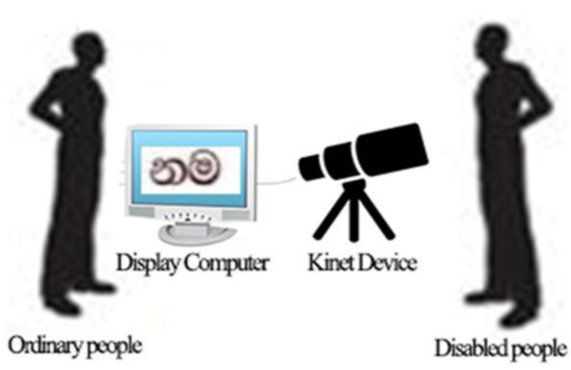


Figure 3.1 Advanced flow of the System

3.2 System Architecture

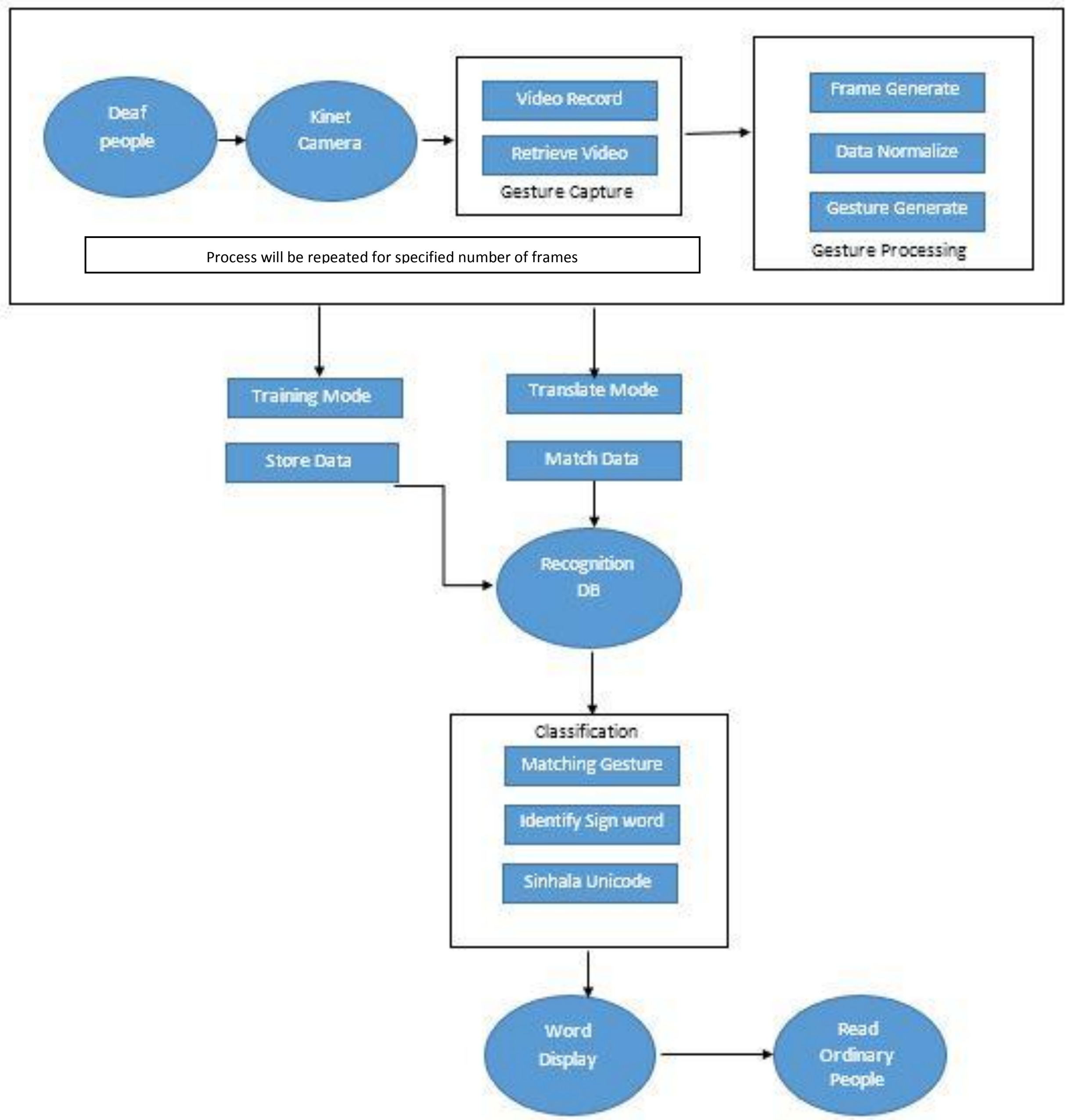


Figure 3.2 System architecture

End to end flow of the system will be studied in this section. Figure 3.2 shows the detailed design architecture of the system. As exposed in the Figure, a deaf person starts making the gesture in front of the gesture acquisition hardware device. This marks as the starting point of the system. As the gesture observation device for the project Microsoft Kinect XBOX 360 sensor camera is going to be used.

In order to track the hand movements, each casing of the video will be recovered after the camera starts recoding the hand gestures of the person. These isolated frames will be analyzed by the system. The information retrieved after analyzing the frames will be directed to the preprocessing module. This module will extract the information which wants to be used for sign recognition. A "feature frame" will be created from data extracted from each frame. It is details of the feature frames before translating the sign into Sinhalese due to the variations of the users who performed the gesture such as length of arm, differences in height and other differences that could occur while performing the gesture. Normalization of date will assure that factors mentioned above will not have any effect on the final output.

The identifier data generated by the system will be used to complete two tasks. Those are for training purposes and translation to Sinhalese. There should be pre-trained gesture identifier data included in the system in order to translate a sign to Sinhalese. To do this Training mode can be used to store an adequate number of sign identifiers details in the sign dictionary database. When the gesture dictionary is filled with sufficient amount data, the system will be ready to execute comparison to identify the gesture.

The classifier will check the entire dictionary to search for a matching gesture identifier in order to do the translation. When a matching gesture identifier is found the system will retrieve the name of that entry. The retrieved name will be then converted into its Sinhalese characters by using Unicode. The converted Sinhalese name will be displayed to the user in an embedded display device in the system.

3.3 Specific design aspects

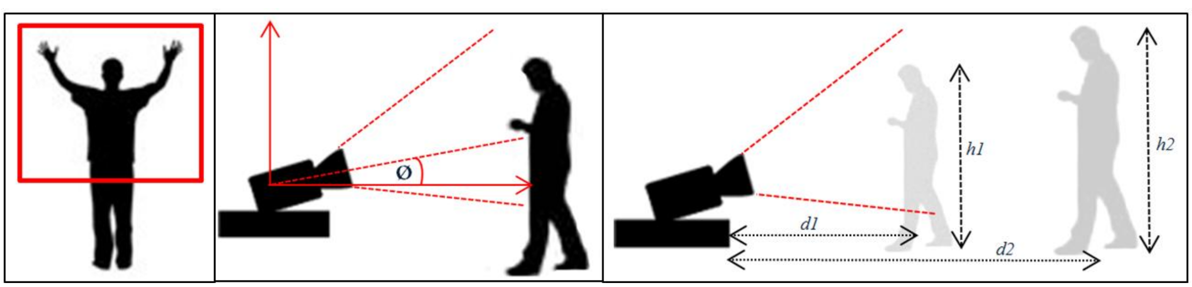
Purpose of this section is to discuss some of the important design considerations in detail. Each sub section will focus on a specific area of the system design and discuss the important concepts behind designing each module.

3.3.1 Gesture acquisition

In order to observe the gestures Kinect sensor camera will be used. The accuracy of the translated word mainly depends on capturing the gesture movements correctly; therefore this is one of important functionalities of the system. Gesture observation is greatly dependent on the user who performs the action and also the position of the sensor device. Also these other factors mentioned below have an impact when identifying a gesture accurately.

- Configuration: Hand shape when doing the sign.
- Position: parts involved with the sign (shoulder, arms, hand).
- Motion: Movement of the hand when doing the sign (straightly, swaying, circularly).
- Capturing window: Area of the human which has been captured by the camera.
- Context: Where the sign is done. Distance and angle from Camera, Materials and objectsin environment…etc.

Figure 3.3 shows the use of some factors affected in observing the gesture through the camera.



(a) Capture window (b) Camera angle (c) Distance between user and camera

Figure 3.3 camera and user positions in gesture acquisition

The system will be received two types of visual information by Kinect sensor and these will be used to identify the gesture. These two types of visual information are Skeleton data and Color data stream .Color data stream will provide real-time preview of the context while skeleton video provides the twenty joints identified from the human skeleton. These two visual information need to be put together in order to generate a final visual output. This final visual frame will be redirected to the Gesture preprocessing stage for extracting the information

3.3.2 Gesture Preprocessing

The frame received from gesture acquisition stage will be the input for gesture preprocessing process. The main objective is to analyze the received frame and store the data which is used for gesture identification. This stage has 3 main steps.

1. Generate feature frame
2. Data Normalization
3. Generate gesture identifier

Internal design flow of above mentioned steps can be described as follows.

Generate feature frame:

Further this step, expected visual info from the detector are filtered by remove unessential details. As shown in Figure 3.4, Kinect is capable of recognizing twenty joint positions of the human skeleton.

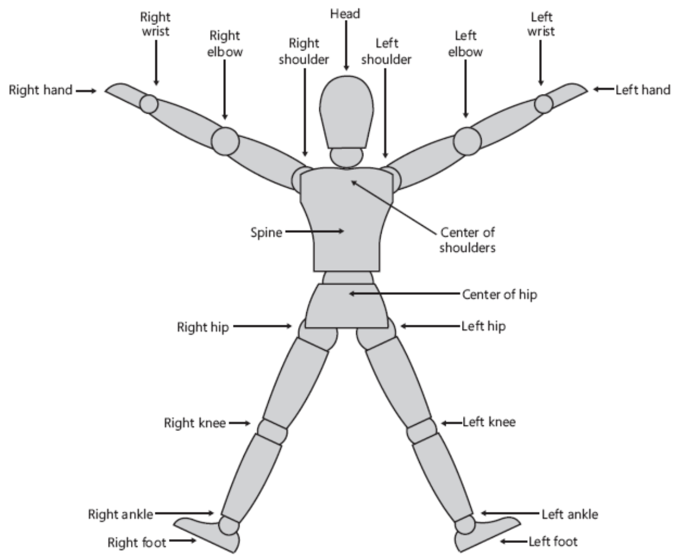


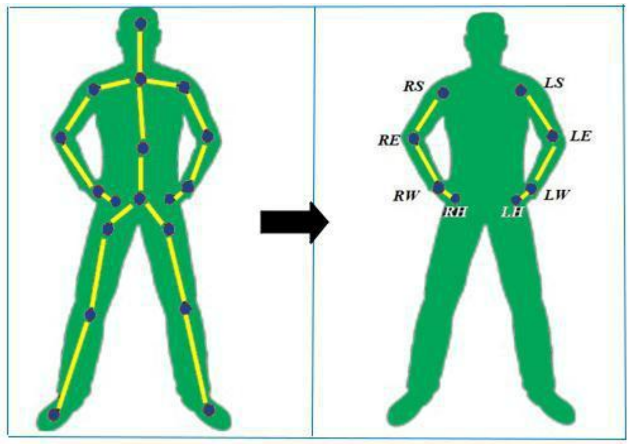
Figure 3.4 Skeleton joint positions provided by Kinect sensor [15]

When performing a particular gesture all the points of the human skeleton will not be used therefore it is absolutely necessary to identify the joints which want to be analyzed for gesture recognition.

Left body joints	Right body joints
Left Elbow (LE)	Right Elbow (RE)
Left Hand (LH)	Right Hand (RH)
Left Wrist (LW)	Right Wrist (RW)
Left Shoulder (LS)	Right Shoulder (RS)

Table 3.1 selected set of skeleton point names for feature frame

Picture (a) of Figure 3.5 display all the skeleton joint data sent by Kinect camera and picture (b) shows the selected skeleton joints for the feature frame.



(a) (b)

Figure 3.5 Joint position extractions from skeleton data

Feature extraction from each joint begins after all the relevant joins are filtered. The user will be performing the gestures in a three-Dimensional environment. Therefore the gesture performance context can be represented by 3D Cartesian coordinate system. Each point of this environment can be represented by coordinates of X, Y and Z axis. (Refer to Figure 3.6). Therefore, a gesture performed by a user can be represented as a series of 3 dimensional coordinates.

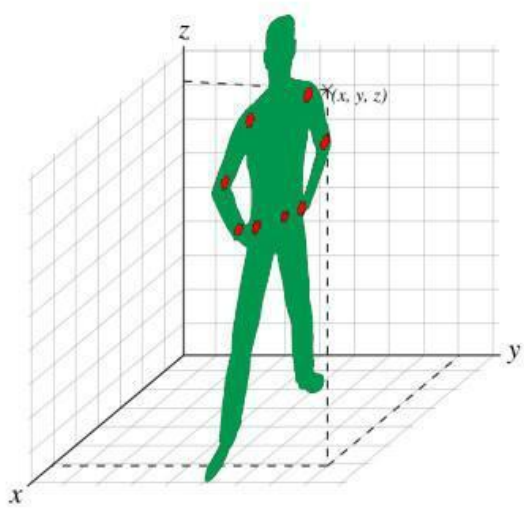


Figure 3.6 skeleton joint representation using 3d coordinates

Data Normalization and gesture identifier generation

There is going to different types of users therefore user's height, length of hands and other joint positions will vary from one user to another user. Their position in the environment when performing the gesture could also be different from one another. Due to this reason the same set of coordinate points cannot be guaranteed. To avoid this, the system needs to have a proper data normalization mechanism before saving the final coordinate points in the database Figure 3.3-picture (c) shows, variations of the users' height and different positions that users may select, while performing the gesture. System consists of coordinates which is independent from the behavior of the user after the data has been normalized. These normalized data will be used to generate the gesture identifier, which can be either used for train the gesture dictionary or match the data in the database in order to detect the corresponding sign.

Store data and generate gesture dictionary

Storing data is one of the important modules of the system as the stored data of the gesture identifier is going to result in successful gesture detection. Two factors will be considered by the when storing the data. Putting away individual motion information and putting away all signal information. An individual motion information source hold the information related with the single sign (Refer to 3.7 Picture-a). It is conceivable to have numerous examples has a place with a solitary motion name. Mix of such individual motion records will produce the last motion word reference (Refer to 3.7 Picture-b)

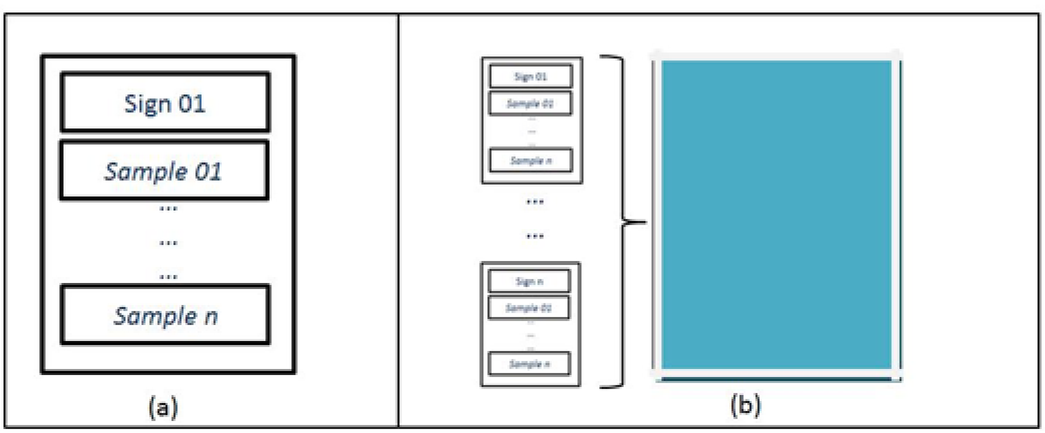


Figure 3.7 Process of storing data in system

3.3.3 Classification and gesture Identification

Framework design has a module for distinguishing the put away motion. Essential for this module is a very much prepared signal word reference. Caught motion identifier subtleties can be utilized to group the subtleties. Framework will explore all through the motion word reference so as to locate the most appropriate coordinating example of motion information. Order calculation will be utilized for this (Classification Algorithm will be examined in Chapter 4).

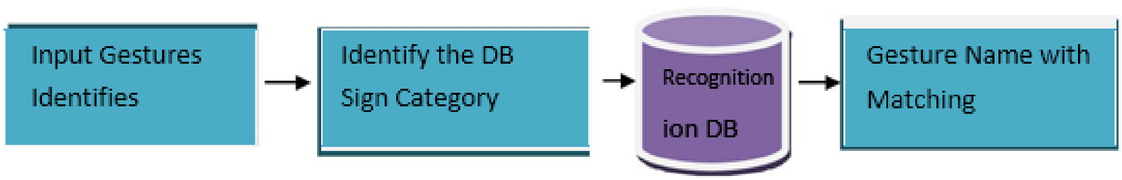


Figure 3.8 classifying input gesture

One of the primary functionalities of the framework is to show the recognized motion name in Sinhalese characters. This particular module in the framework will change over the English arrangement of characters into relating Sinhalese characters which has the comparative articulation. This module isn't intended to complete any interpretations among English and Sinhalese words. It will take an assortment of English characters as an information and afterward those English characters will be changed over to its Sinhalese characters without shifting the method of articulating of the underlying English word. Essential procedure of this module has appeared in Figure 3.10

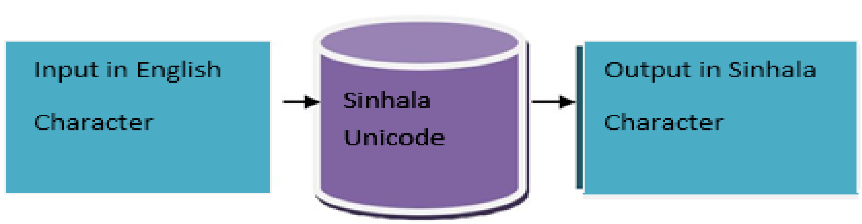


Figure 3.9 Sinhala Unicode module Flow

3.4 System Interfaces

This section provides an overview of the interfaces available in software application. As shown in Figure 3.10, System can be divided into three main interfaces where each interface has specific set of functionalities

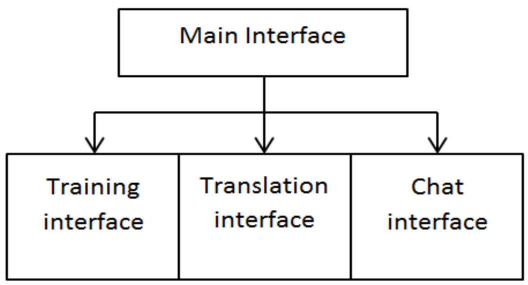


Figure 3.10 Main interfaces of the system

3.4.1 Train mode interface

Objective of this interface is to allow user to train the system prior to the translation of any gestures. In order to recognize gestures more accurately, it is important to train this system with variety of users and multiple gesture samples for a specific sign.

Functions of Train mode interface:

Display real-time visual information- Interface is designed with an embedded window which is capable of displaying real time video and skeleton data streams received from the Kinectsensor.

Allow user to input a gesture name and record the multiple gesture samples. – Through this interface, user can have to type the name of the gesture he wishes to train. Since the system provides the sign word in Sinhalese language,this interface is connected with the Sinhalese Unicode module (refer to section 3.4.1). Once the name of the gesture has been converted to Sinhalese, gestures can be captured through the interface and data samples will be stored in separated text files. Users are not restricted to assign single sample for a specific gesture. Instead, System allows training single gesture using multiple samples.

Generate and update final gesture dictionary – Training mode interface allows user to create the final gesture data file by combining all the individual gesture data files together. Generated gesture file will be used in Translation mode for gesture detection.

View and test gesture data file – Users can recheck whether the trained gesture is trained properly by reloading the data back to the interface. If the gestures are not trained up to the desired level, data can be removed and new chunk of data can be trained via the interface.

3.4.2 Translation Mode Interface

Proposed System consists of the translation mode interface which will be used by deaf user and ordinary user for the sign language to Sinhalese translation. This interface reflects the core functionality of the entire system since all the necessary components needed for sign to Sinhalese conversion embedded to this interface.

Functions of Translation mode interface:

Display real-time visual information–As in the train mode interface, this interface design also contains an embedded window which is capable of displaying Video and skeleton data streams received from the Kinect sensor.

Load final gesture dictionary – In order to detect the gestures system needs to refer to the final gesture dictionary created during the training mode. Therefore this interface design will be interacted with gesture dictionary and load all pre-trained gesture samples related with each sign.

Sign language to Sinhalese translation and display the identified word in Sinhalese –This interface will be connected with classification, database and gesture preprocessing modules in order to identify the matching word belongs to the gesture.

Chapter 04

Implementation

4.1 Components of developer environment

The carefully decided hardware and software developments platforms after observing the Literature survey will be listed down in this section. These technologies were selected for the development of the proposed system after examining the past research work done on this domain.

4.1.1 Microsoft Kinect XBOX 360 sensor

Kinect XBOX 360 sensor device will be used as the gesture observation device of the system. The data belonging to the gesture performs by the user will be captured by the integrated RGB camera and the IR emitter of the sensor. The advantages of selecting Kinect over other video cameras for gesture acquisition are listed below

Figure 4.1 Kinect Device

- Ability to provide clear information during various environmental conditions – Kinect sensor camera is capable of producing accurate visual data in different conditions such as low light, crowded, etc.
- Reduce preprocessing time taken – Sensor information can be directly accessed using different SDK platforms. Therefore it avoids the additional actions need to be taken for extracting required data from input. This ultimately affects to the overall response time of the system.

- Ability to access human skeleton data – It is very important to find out the human skeleton for gesture recognition. Kinect sensor is capable of detecting human skeleton among various captured data. This is a very difficult task in most of the other cameras.

4.1.2 Selection of software platforms

- So as to get to caught information of Kinect, framework will utilize Kinect SDK version 1.8. This SDK will be upheld for all the highlights of Kinect sensor and it gives a steady stage to creating Kinect based applications. (Refer to section 2.4.1).
- Microsoft VS 2010 will be utilized to fabricate the product utilizing visual C# programming language. Moreover, Windows Presentation Foundation (WPF) will be utilized since it will be exceptionally bolstered for speaking to visual data. Framework will be bolstered for working with Windows based stage, explicitly Windows 7 and Windows 8 working frameworks.
- So as to create application, Network Comms [20] libraries, has been utilized. These libraries incorporate a devoted arrangement of functionalities which is fit for making a fundamental application utilizing C#. Fraud programming form 6 [21] will be utilized to create reenacted 3d models which expected to coordinate with application. Since Poser programming is explicitly intended for creating human character activitys, it supports to make recreations of communication through signing based signals utilizing human figure.

4.2 Gesture observation

The caught information by Kinect sensor should be gotten to at the underlying usage. For this proposed framework skeleton information stream and shading information stream will be utilized. Sensor gives these information utilizing two separate libraries, Skeleton Frame and Color Frame in Kinect SDK. The framework will utilize two distinctive showcase layers to adjust information streams together and show as a solitary yield.

To capture the gestures belongs to a single sign the system will use a buffer of 32 frames under the assumption that all gestures used in the system can be performed within thirty two frames. Proposed project will use this restriction for the ease of comparing the real-time gesture and the samples in the trained database. Figure 4.2 shows set of sample gesture frames belongs to a sign which has been performed within 32 frames.

Figure 4.2Distribution of a gesture against frames

4.3. Skeleton point extraction

Not all the skeleton points will be used for gesture detection. Eight points will be extracted from the frame for further analysis. (Refer to the section 3.3.2). System extracts 3 dimensional coordinates from each selected skeleton received from Kinect sensor and maps each point in a 3D point of windows for further processing. Figure 4.3 shows an example mapping of the left hand skeleton coordinates into the point 3d representation of windows environment.

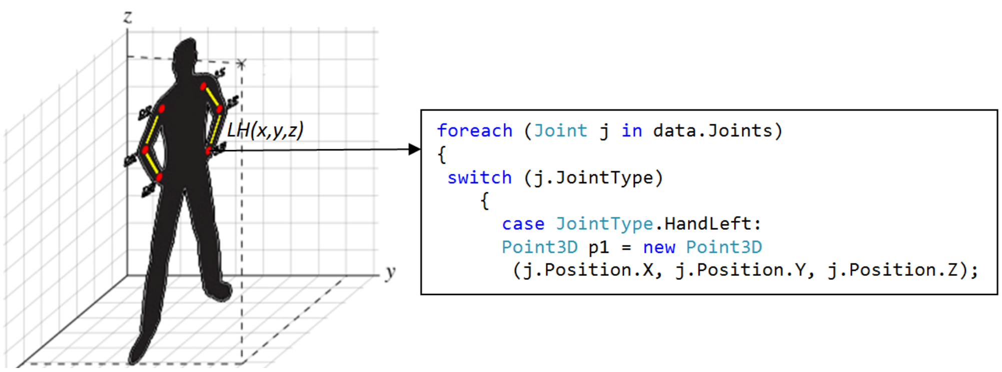


Figure 4.3 mapping real world coordinates into program

4.4 Data Normalization

As discussed in Section 3.3, it is impossible to guarantee the physical behavior of users who performs the gesture. Correct normalization want to be done for extract coordinates from Kinect, before storing them in data file or perform gesture matching.

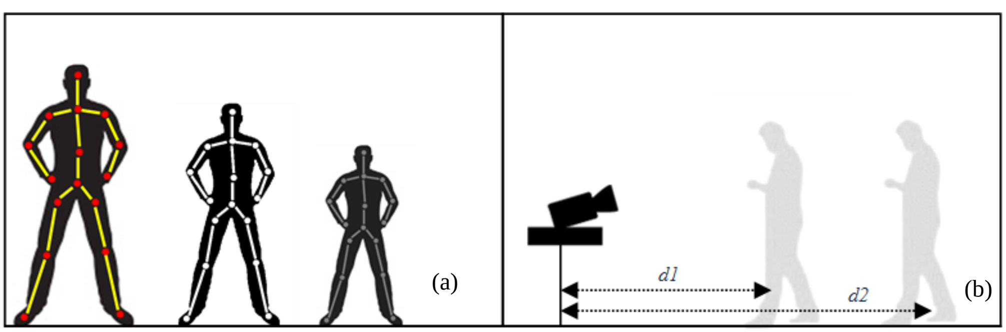


Figure 4.4 variations of the heights and distance from camera in gesture performing

As appeared in the Figure framework couldn't straightforwardly store the Cartesian directions of every client skeleton since it exceptionally relies upon the signal execution setting of the client. Client might be in various tallness (Figure 4.4 picture - a) or the good ways from the camera to the client might be extraordinary (Figure 4.4 picture - b). In such circumstance, framework may need to consider huge measure of tests has a place with a similar signal thinking about every conceivable variety. This prompts excess information stockpiling of the framework and may set aside long effort for characterization. In this manner framework executes a two-advance standardization process before putting away information.

4.4.1 Adjustment of center point based on user

Given the set of joints $K = \{LH, LE, LW, RH, RE, RW\}$,consider a point $M_{x',y',z'}$ as the new center point where x', y' and z' are new center point coordinates. Assuming x_1, y_1 *and* $z_1$are the observed coordinates from Kinect sensor for Left shoulder, *LS* and Right shoulder *RS*,

$$M_{x'} = (LS_{x_1} + RS_{x_1}) / 2$$
$$M_{y'} = (LS_{y_1} + RS_{y_1}) / 2$$
$$M_{z'} = (LS_{z_1} + RS_{z_1}) / 2$$

Given Joint*i* in the joint Collection *J*, considering the new coordinates of *K (i)* are x_2, $y_2$and z_2

$$K(i)_{x_2} = K(i)_{x1} - M_{x'}$$
$$K(i)_{y_2} = K(i)_{y1} - M_{y'}$$
$$K(i)_{z_2} = K(i)_{z1} - M_{z'}$$

This arrangement will be accomplished for all the joints in the joint assortment expecting focal point of the organize framework is the center purpose of client's shoulders. Figure 4.5 shows impact of focusing process. Picture (a) speaks to the directions before focusing procedure and picture (b) speaks to the skeleton organize portrayal in the wake of focusing process.

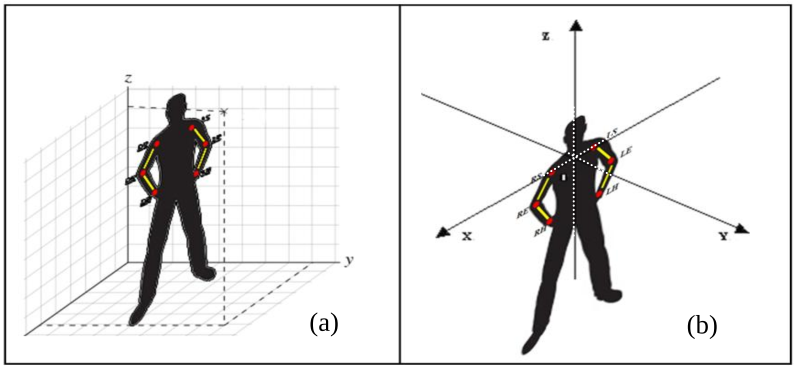


Figure 4.5 Change of coordinates through centering process

4.4.2 Normalization of coordinates

When the focusing procedure is finished, a client explicit arrange framework will be made. Anyway still it is beyond the realm of imagination to expect to store the information in information document since the information should be client autonomous preceding stockpiling. In this way framework performs standardization for every facilitate dependent on the length of two shoulder focuses under the supposition of the skeleton focuses are even. This progression will guarantee that the created yield doesn't rely upon the physical size of the user.

Framework computes the separation between the left shoulder point and right shoulder point dependent on Euclidean separation of 3d facilitates. A short time later, the good ways from focus point to each chosen organize will be estimated. A standardized facilitate will be determined by separating the joint explicit good ways from the separation between shoulders. This procedure can be appeared from following advances.

Given two points $P_{(LS)}$, $P_{(RS)}$belongs to Left shoulder and right shoulder respectively, considering the distance between $P_{(LS)}$ and P $_{(RS)}$is $d_{(s)}$

$$d_{(s)} = \sqrt{(X_{(LS)} - X_{(RS)})^2 + (Y_{(LS)} - Y_{(RS)})^2 + (Z_{(LS)} - Z_{(RS)})^2}$$

Where $X(i)$, $Y(i)$ and $Z(i)$ are the coordinates of given point i

Given the set of distances $D = \{d_{LW}, d_{RW}, d_{LE}, d_{RE}, d_{LH}, d_{RH}\}$, the normalized set of distances $D_{norm}$is obtained as follows.

$$\sum_{i=1}^{n} D_{norm} = \frac{d(i)}{d(s)}$$

Where n is the quantity of good ways from D and d(s) is the separation between two shoulders determined in past segment .Figure 4.6 shows the good ways from each chose joint to the balanced focus point.

When the standardization procedure is finished, it will produce the applicant signal identifier outline which can be utilized for capacity or examination reason.

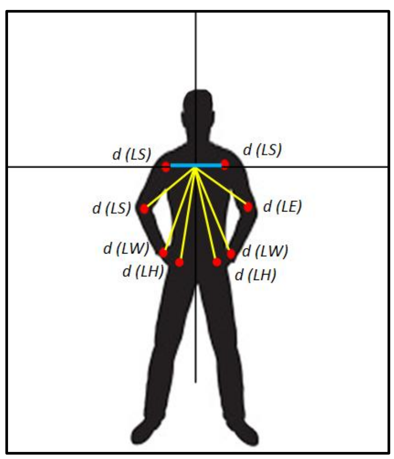


Figure 4.6 distance from center point to selected joint positions

4.5 Continuous frame Capturing and storage of data

This project assumes that a gesture belongs to a given sign word can be performs within 32 frames. Therefore before storage of the data or perform classification, system go through the process described in Section 4.4, for 32 frames and extract the features of each frame independently.

When the information has been separated it is absurd to expect to begin the signal recognition straightforwardly since there ought to be a particular measure of information should be put away for correlation reason. Inside the executed framework, content document based capacity will be utilized to store the extricated data. Figure 4.7 shows the example information record utilized for putting away the data has a place with a particular motion. This information comprise of the name of the signal and the standardized arrangement of X, Y, Z facilitates for each chosen purpose of the skeleton. Data of thirty two such examples for a solitary preparing signal will be accessible on an information document as appeared in Figure 4.7

System generates individual data file for each gesture and combines all data files into a single gesture dictionary. It possible to update the existing dictionary or create a completely new dictionary from the scratch, based on the accuracy of recorded sample gestures.

```
@අම්මා
-0.751576942301981
-1.84811726532769
-0.169771688029978
-0.725259150223472
-1.55518793148086
-0.128890407668435
-0.620527930963863
-0.859112185985556
0.0963657700289838
0.578047787697646
-0.836614125248765
-0.0899271114778043
0.606062300420178
-1.56962097335625
-0.230113884497999
0.613240843289824
-1.79049343622798
-0.264643470580127
~
```

Figure 4.7 format of a sample data storage text file

4.6 Classification

This will be the most significant module of the framework since the accomplishment of the framework relies upon the way how the motion will be perceived. Before grouping, framework has been prepared utilizing assortment of motion tests. During the underlying phase of advancement, five preparing tests have been utilized per motion and word reference of 15 sign words have been utilized in the framework. Along these lines last motion word reference comprise of seventy five prepared examples through and through.

Task done by the classification module can be summarized as follows.

- Observe the normalized frame information and trained information into the module.
- Compare the coordinate information in both samples and retrieve the gesture name with highest matching probability.
- Send the selected gesture to the user Interface for display purpose.

As per this task, it is required to look at the constant organize information and pre-prepared data in motion word reference. Information which has the most noteworthy coordinating grouping will be chosen as the signal. Since this procedure follows the correlation between two groupings (Real time arrange information and pre-prepared example information), Dynamic Time Warping calculation will be utilized with improvements. Moreover, closest neighbor classifier will be utilized to pick the best coordinating signal name dependent on the DTW order results. Figure 4.8 demonstrates the normal usefulness to be performed utilizing DTW calculation.

As shown in the figure 4.8, received real time gesture will be compared against each gesture sample of the dictionary, using Dynamic Time Warping algorithm. For each comparison, algorithm generates a value which called as DTW distance. Input gesture will be classified into gesture sample with the minimum distance and recognized as the matching gesture.

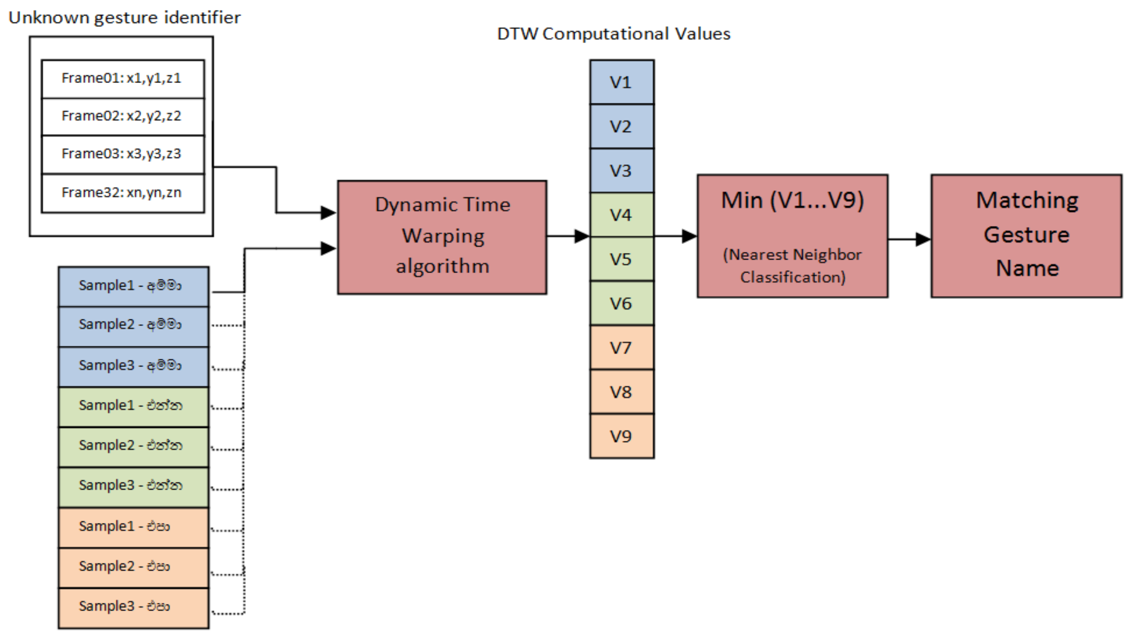


Figure 4.8 Flow of classification process

4.6.1 Classification And Dynamic Time Warping Algorithm

Success of the entire system depends on this critical part of the project which is the method to identifying the gesture performed by the user. Throughout the preliminary stage development of the system, it has been trained with a selection of gestures before the classification phase.

The classification module will be used to compare the sign performed by the user and training sample. Dictionary will be generated before the classification process. A summation the objectives of the classification module are mentioned below.

- Detecting the normalized frame information and trained information into the module
- Retrieve the name of the gesture with the top most identical probability after comparing coordinate data of the signed sample and gesture identifiers in the dictionary.
- Send the retrieved name of the gesture to the user interface to be displayed.

The venture is assumed look at the continuous organize information and the pre-prepared data in motion word reference. The information which has the most noteworthy coordinating succession in the motion word reference will be recognized as the motion caught by the gadget. An improved Dynamic Time Warping calculation will be utilized in this venture on the grounds that the framework manages a correlation between two arrangements which are ongoing direction information and information tests prepared ahead of time. Besides so as to choose the most fit signal name dependent on the aftereffects of the DTW characterization, closest neighbor classifier will be utilized. Figure 4.8 demonstrates the normal usefulness that should be finished utilizing DTW calculation.

Dynamic time warping calculation is a period arrangement calculation grew initially for discourse acknowledgment. It targets adjusting two arrangements of highlight vectors by distorting the time pivot iteratively until an ideal match (as indicated by an appropriate measurements) between the two groupings is found [20].It measures the likeness between two successions dependent on a cost capacity of the amount it needs to "twist" the focuses forward/in reverse so as to have them line up. This arrangement strategy is frequently utilized in time arrangement characterization.

Dynamic time warping calculation is a period arrangement calculation grew initially for discourse acknowledgment. It targets adjusting two successions of highlight vectors by twisting the time hub iteratively until an ideal match (as indicated by a reasonable measurements) between the two groupings is found [22].It measures the likeness between two arrangements dependent on a cost capacity of the amount it needs to "twist" the focuses forward/in reverse so as to have them line up. This grouping arrangement technique is frequently utilized in time arrangement order. Calculation 4.1 shows the fundamental arrangement of steps in Original Dynamic Time twisting calculation.

Algorithm 4.1 Dynamic Time Warping algorithm

```
functionDTWDistance (char s[1..n],char t[1..m])
    declareintDTW[0..n, 0..m]
    declareinti, j, cost
    fori := 1 to m do
    DTW[0, i]  ← ∞
    end for
    fori := 1 to ndo
    DTW[i, 0]  ←∞
    end for
    DTW[0,0] ←  0
    fori := 1 to ndo
    fori := 1 to mdo
            cost← d(s[i], t[j])
        DTW[i, j]←cost+ minimum(DTW[i– 1,j],
                                    DTW[i; j - 1],
                                    DTW[i– 1, j - 1])
    end for
    end for
    returnDTW[n,m]
    end function
```

This algorithm takes two input parameters for comparison, input gesture sample and a sample from trained gesture dictionary. Then the algorithm will create the DTW matrix which compares the values of two input parameters and fill up the matrix based on the cost of the selected indexes of the input parameters. The cost can be calculated according to the behavior of the selected parameters. Table 4.1 and 4.2 shows two sample DTW matrixes generated from the algorithm. Difference of the two selected values has been used as the cost function.

These values have been used only for demonstration of the algorithm. These are not the values generated through proposed project) According to the example in table 4.1, it assumes two input arrays to the algorithm are *P*={1,2,3,4,5} and *Q*={3,5,6,7,8}. Once the matrix is generated, it returns the calculated dtw value as 10. Example shown in Table 4.2 takes two input parameters into account which is *P*={1,2,3,4,5} and *S*={1,2,2,3,4,6} and the observed value through the algorithm is 2. Therefore it is possible to conclude that out of sample *Q* and *S*, most matching sample with input *P* is sample *S*, since the minimum generated value is 2 among the comparisons.

Q / *P*		**3**	**5**	**6**	**7**	**8**
	0	∞	∞	∞	∞	∞
1	∞	2	6	11	17	23
2	∞	3	5	9	14	20
3	∞	3	5	8	12	17
4	∞	4	4	6	9	13
5	∞	6	4	5	7	**10**

Table 4.1 Sample matrix of DTW algorithm (01)

S / *P*		**1**	**2**	**2**	**3**	**4**	**6**
	0	∞	∞	∞	∞	∞	∞
1	∞	0	1	2	4	3	8
2	∞	1	1	1	2	4	7
3	∞	3	2	2	1	2	5
4	∞	6	4	4	2	1	3
5	∞	10	7	7	4	3	**2**

Table 4.2 Sample matrix of DTW algorithm (02)

Despite the fact that Dynamic Time Warping calculation suits for this venture it is difficult to utilize the underlying adaptation of the calculation since it has been executed for contrasting a solitary component of qualities in a given edge. This task utilizes 18 measurements for each frame(X, Y, Z directions of chose six joint positions).Therefore Euclidian separation will be utilized to figure the expense of a given two purposes of the sources of info.

Given *P* and *Q* input arrays and calculation logic below.

$$\text{Cost}=\sqrt{\sum_{i=1}^{18} (P_i - Q_i)^2}$$

Where given joint position *i* is a coordinate point.

4.6.2 Nearest Neighbor Classifier

. Algorithm 4.2 shows steps of developed nearest neighbor classifier.

Algorithm 4.2:Nearest Neighbor Classifier

```
Function NNClassifier (testSample, dictionarySamples [1..n])
    declare double mindist ← ∞
    declare dist,threshold
    declare string classification
    for i := 1 to n do
        declare Tsample←dictionarySamples[i]
        if d(testSample[length],Tsample[length]) <threshold
            dist ← DTWdistance (testSample,Tsample)
            if dist ≤ mindist then
                mindist ← dist
                Classification ←gestureName[i]
            end if
        end if
    end for
return classification
```

This classifier takes the generated gesture dictionary and the real-time gesture samples as inputs. Since the dictionary contains collection of samples classifier calls the DTWdistance function (Refer to algorithm 4.1) for calculating the DTW value for each sample in the gesture dictionary along with the received data. Prior to the calling of DTWdistance function, it computes the distance between the last coordinates of the gesture sample from the dictionary and the real-time sample. If the calculated distance is greater than the defined threshold value, classifier does not compute the DTW distance for the selected pair. This step will be performed to avoid comparing the gestures with larger variance and to minimize the response time.

4.7 English to Sinhalese character Conversion

One of the fundamental targets of the framework is to give the perceived signal name utilizing Sinhalese characters. Framework had the option to utilize assortment of normal articulations for change of the English characters into Sinhalese. Prior to the arrangement of the signs, User can type the name of the movement in English characters which will be changed over into the individual Sinhalese characters which gives the practically identical explanation.

For this reason, framework utilizes assortment of ordinary articulations for every Sinhalese character in the letters in order, thinks about the client input and plays out the transformation. Every individual character of the client information will be coordinated with the applicable customary articulation and if a match discovered, specific English character will be changed over into relating Sinhalese character looked at by the ordinary articulation. Figure 4.9 shows test set of Regular articulations utilized in Conversion module. So as to build up this module, framework alluded to the example C sharp class record referenced in [23].

```
vowelsUni[0] = "ඌ"; vowels[0] = "oo";
vowelsUni[1] = "ඕ"; vowels[1] = "o\\)"
vowelsUni[2] = "ඕ"; vowels[2] = "oe";
vowelsUni[3] = "ආ"; vowels[3] = "aa";
vowelsUni[4] = "ආ"; vowels[4] = "a\\)"
vowelsUni[5] = "ඈ"; vowels[5] = "Aa";
vowelsUni[6] = "ඈ"; vowels[6] = "A\\)"
vowelsUni[7] = "ඈ"; vowels[7] = "ae";
```

Figure 4.9 Sample regular expressions used for English to Sinhalese conversion

Chapter 05

Evaluation

During this chapter, the system will be evaluated under different aspects and test results are mentioned to reflect the system accuracy. Furthermore, obtained results will be analyzed to discuss the levels of success in the system.

5.1 Evaluation criteria

During the initial stage the system will contain selected sign based words, Altogether, there will be **seventy five** samples available in the training dictionary. Each trained gesture sample has been recorded along 32 frames where each frame consists of the normalized three dimensional skeleton coordinates. Furthermore users with height variances used for system testing and also system has been tested from different distances between camera and the user in order to find out the amount of success in data normalization. Above mentioned factors will take into account when calculating the overall accuracy rate of the system.

System evaluation criteria can be summarized as shown in table 5.1

Cause	No of Tests
Individual sign based assessment	225
Effect of the distance from the camera	50
Users with physical height variations	50
Speed of the gesture based evaluation	50

Table 5.1 System Evaluation criteria

5.2 System testing based on selected signs

As mentioned in Chapter 01,system will be used 15 initial signs from the Sri Lankan Sign language and during the testing stage each sign has been tested against 15 test samples to evaluate the success level of each sign in terms of word recognition. Accuracy rate for each test sample was calculated according to following equation.

$$\text{Accuracy rate} = \frac{\text{Number of samples passed}}{\text{Number of total samples}} \times 100$$

Table 5.2 shows the results received from the system evaluation during this phase.

Sign	Total samples	Passed	failed	Accuracy
We (අපි)	15	12	2	80%
I (මම)	15	14	1	93.3%
Mother(අම්මා)	15	15	0	100%
Children(ළමයි)	15	15	0	100%
Bus Halt(බස් නැවතුම)	15	12	3	80%
Train(දුම්රිය)	15	13	2	86.6%
Bus(බසය)	15	12	3	80%
Tired(මහන්සි)	15	14	1	93.3%
Come(එන්න)	15	15	0	100%
Don't(එපා)	15	13	2	86.6%
Go (යනවා)	15	14	1	93.3%
Where (කොහෙද)	15	14	1	93.3%
What time(වෙලාව කීයද)	15	15	0	100%
Help(උදව් කරනවා)	15	15	0	100%
Thank You(ස්තුතියි)	15	14	1	93.3%
Overall Status	**225**	**208**	**17**	**92.4%**

Table 5.2 Accuracy rates based on individual signs

Concurring the outcomes appeared in Table 5.2 framework can accomplish a general exactness pace of 92.4% out of 225 test tests. Framework had the option to accomplish a sensible exactness rate for the greater part of the signs and the rate was bit low for certain motions. Above outcomes were gotten via preparing equivalent number of tests for each motion. In this way results shows that a portion of the motions need all the more preparing tests because of the intricacy and variety of motion design.

5.3 Effect of the user and the distance range

Purpose of this phase of the evaluation is to test the success rate of used normalization procedure in the system. Main Purpose of the Normalization process was to ensure that the final result of the system the physical behavior does not depend on of the user or the distance from Kinect camera to the gesture performing context.

5.3.1 Evaluation by different users

Sample No	User height (in cm)	Total samples	Passed	failed	Accuracy
01	134.6	10	5	5	50%
02	152.4	10	7	3	70%
03	167.1	10	9	1	90%
04	172.3	10	8	2	80%
05	180.3	10	8	2	80%

Table 5.3 Accuracy rates based on user heights

During this phase of one of the principle issues happened was to test the framework against the clients who didn't know about the communication via gestures motions. In this manner the achievement paces of the clients were exceptionally reliant in the manner they play out the motion. Anyway framework was able in perceiving greater part of the motions as referenced in the Table 5.2 with the detailed most noteworthy precision pace of 90%. Despite the fact that the outcomes show that the precision rate is legitimately corresponding with the stature, last yield will dependent on the exactness of the individual signal perform by the client to the measure of each examples for the preparation database.

Therefore, it can be concluded that the system is functioning independent of the physical height of the user due to the normalization procedure described in Section 4.4.2.

5.3.2 Impact of the good ways from Camera

One of the significant contemplations in the standardization stage is that the yield ought not rely upon the good ways from the camera to the motion performing setting. Along these lines the framework is assessed against various good ways from the Kinect camera so as to play out the signal. During the assessment stage five irregular signal examples of the word reference, were in use and performed from ten unique scopes of the separations among camera and the client. Separation scope of 0-160(inches) was considered for assessment.

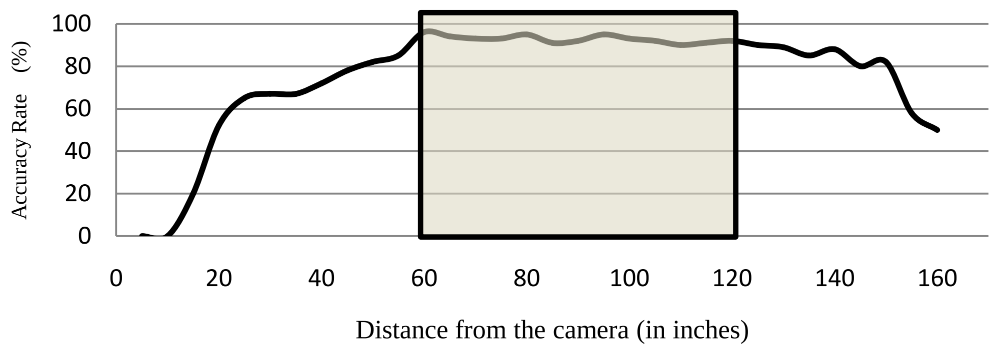


Figure 5.1 Accuracy rates based on user heights

According to Figure 5.2, system does not show a reasonable accuracy rate when the distance between the user and the camera is considerably low (i.e. 5-15 inches). Main reason behind this observation is the capturing window of the camera is not that accurate to capture the necessary frames of the user. However, system was able to recognize the simple gestures even though the user is very near to the camera but was unable to recognize the complex set of gestures.

According to the outcomes, framework shows the most elevated exactness rate with out there extend between 60-120 creeps with the revealed most noteworthy pace of 94% (Highlighted territory in Figure 5.2). In the middle of this separation go, the catching window of the camera had the option to catch the whole human skeleton. Despite the fact that the undertaking doesn't consider every distinguished skeleton, during the referenced scope of separation camera was increasingly exact in recognizing and constant drawing of the skeletons.

Acknowledgment level began diminishing, when the separation go surpassed measure of 140 inches, since camera couldn't recognize a portion of the skeleton focuses. In any case, framework given a type of exactness level in perceiving certain arrangement of signals.

This result reflects that the system is capable of recognizing the gestures in high level of accuracy rate independent from the position of the user and distance from the camera, meaning the normalization procedure described in section 4.4.1 is functioning properly in the system.

5.4 Outcome of the speed of performing gesture

One of the fundamental difficulties that the framework confronted is failure to foresee the speed of the signal performed by the client. As depicted in usage section Dynamic Time twisting calculation was skilled in coordinating the motion autonomous from the speed of the signal. In this way the framework has been assessed utilizing five motion tests acted in different paces. Motions have been chosen to conceal all mind boggling and straightforward motion ways. Table 5.4 shows the chose signals for the testing.

Sample No	Gesture Name
01	I (මම)
02	Where (කොහෙද)
03	Bus Halt(බස් නැවතුම)
04	Come(එන්න)
05	Bus(බසය)

Table 5.4 selected gestures for gesture speed test

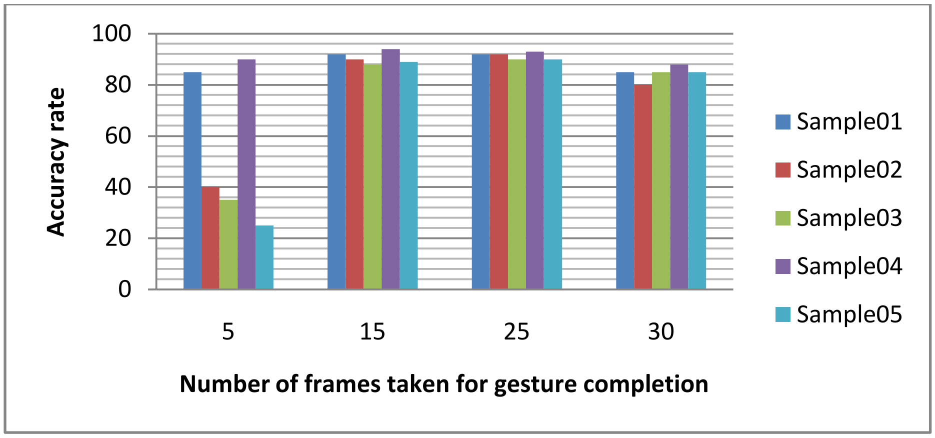


Figure 5.2 Speed of the gestures accuracy rate

As indicated by the chart appeared in Figure 5.2, a large portion of the motions are not appropriately recognized when the motion finish outlines are extremely low.

Beginning from outline 15 onwards practically all signals shows high exactness rate mirroring that DTW calculation is equipped for perceiving the motion even with fluctuating paces. One significant perception during this assessment stage is that acknowledgment of certain examples doesn't rely upon recorded edges in the motion word reference. Along these lines it isn't important to record a similar motion with various speeds in preparing stage. Be that as it may, during the preparation stage, it is in every case better to prepare the signal in moderate casing rate on the grounds that DTW calculation which utilized for grouping reason in the framework has the ability of contrasting the motions and fluctuating rates.

Chapter 06

Conclusion

6.1 Conclusion

Introductory objective of the task was to structure a framework which is equipped for deciphering Sri Lankan Sign Language into Sinhalese language. Also, this examination has concentrated on the chance of sending the made an interpretation of words to an incapacitated or ordinary individual a good ways off with a coordinated application in the framework. When the related work under this space was checked, it turned out to be certain that, there were diverse research has been completed to interpret nation explicit gesture based communications into individual communicated in language. Various ideas and advancements were utilized for this reason, for example, Hidden markov model and Neural Networks.

As mentioned in the theses proposal, this system was built to recognize fifteen sign based words in Sri Lankan sign language and provide the meaning of the respective sign using Sinhalese characters. Gestures which were presented using complex methods such as lip movements, facial expressions or finger movements were not detectable by the system.

Another initial goal was to make the system easily trainable. Within this system, total of fifteen signs are presented, but the list can be increased by using the Training mode of the interface. Only constraint is that the user must take into account which kind of features, the implementation is using when adding a new sign so that the system will be able to recognize it. One of the reasons why this project is restricted into fifteen signs is because of the data collection and the processing time. Once the capacity of final gesture dictionary is very high system takes more processing time to compare samples of the dictionary.

The results from the analysis have shown an accuracy rate of 92.4% for detecting overall collection of words in the created gesture dictionary. Findings fromthis study suggest that an increased test samples may improve the accuracy rate of recognizing gestures. While detecting some of the gestures, threshold value, which was an input parameter in DTW algorithm, played an important role since the changing the threshold value affected to some of the selected sign based gestures.

6.2 Improvements

Even though the project was able to satisfy most of the expected objectives, there are few areas which need some improvements. One such area is response time of the system. Implemented system interfaces highly deals with the real time video and skeleton information to be displayed continuously. At the same time system performs gesture matching based on the captured coordinates. Therefore system tends to take considerable time to provide the identified word. This draw back can be reduced by using a machine with high processing power.

It is possible to use a professional database management system for storing the recorded gestures since it can be an effective way for comparing and filtering the data rather than using a text file database. Since querying and filtering can be done with ease by using such database management system, system response time and accuracy rate may increase.

6.3 Future Work

So far the system has been designed and developed to detect and translate few sign based words in Sri Lankan. In order to use the system as a professional sign language translator, more gestures need to be added to the gesture dictionary. The system can be further improved to recognize grammatically correct sentences rather than recognizing individual words.

During this implementation, signs which based only on hand gestures were used for translation. In addition there are gestures in sign language which uses facial expressions, lip movements and finger movements. It is possible to integrate these features into the system as well since Kinect sensor provides ability to detect facial expressions as well.

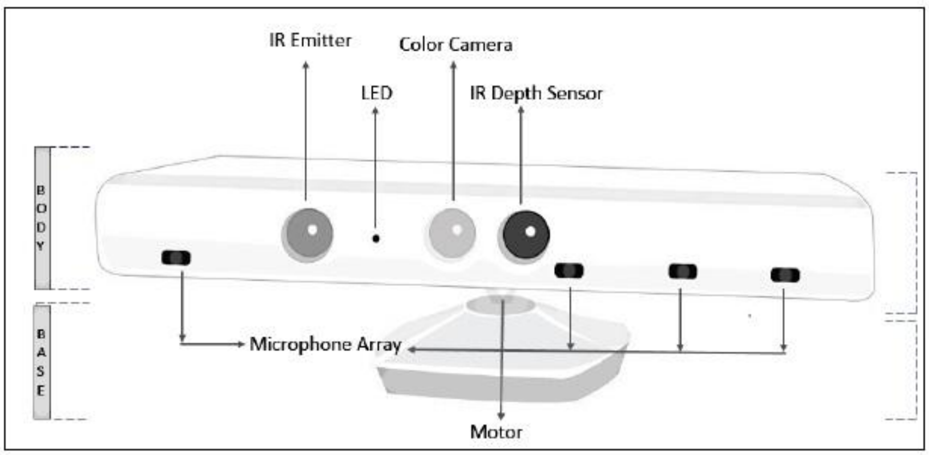
In order to obtain more use from the system, it need to be used more public places like banks, hospitals, etc. A proper survey must be carried out to identify the possibility of installing this system in such places.

During this study, the system has been implemented to use in windows based platform. If the system can be used in more platforms such as Linux or Macintosh, more audience can be addressed through the system. It is possible to design a mobile version of the system as well, but in such situations the system architecture need to be modified since it may not feasible to use Microsoft Kinect sensor as the gesture acquisition device in mobile platform.

Appendix A

Microsoft Kinect Hardware Sensor

This is a level gadget with profundity sensors, shading camera, and a lot of amplifiers with all made sure about contained by a little, level box. The level box is appended to a little engine filling in as the base that empowers the gadget to be tilted an even way. Figure A.1 shows the key components of kinect device.

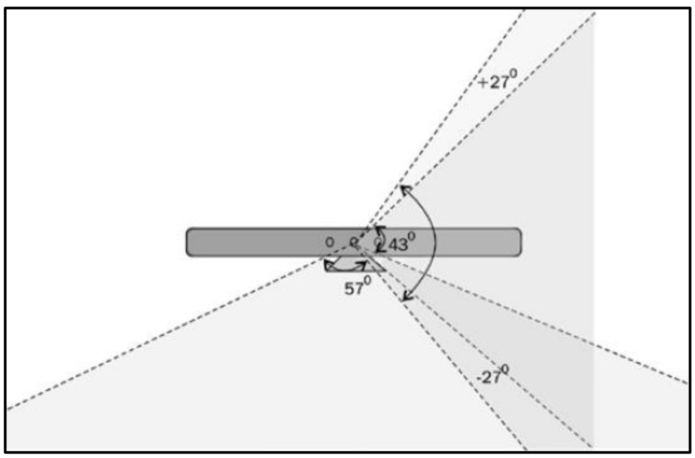


FigureA.1 Kinect sensor components [15]

- Color camera– This segment is liable for communicable and gushing the shading video information. The Kinect shading stream underpins a speed of 30 edges for each second (FPS) at a goals of 640 x 480 pixels, and a most extreme goals of 1280 x 960 pixels at up to 12 FPS. The estimation of casings every second can differ contingent upon the goals utilized for the picture outline.

- Infrared (IR) emitter and IR profundity sensor – Kinect sensors include of an IR producer and an IR profundity sensor. These two work together to get things going. The IR producer considered, yet it's an IR projector that continually radiates infrared light in a "pseudo-arbitrary dot"[15] design over everything before it.

- Tilt motor – As shown in Figure A.1, Kinect device consist of two major parts called base and body Figure A.2 shows an illustration of the angle being changed when the motor is tilted



FigureA.2 Angle variations of Tilt Monitor [15]

Appendix B

Selected Code Listings

```
privatevoid Initialize()
        {

if (kinectSensor == null)
return;
kinectSensor.ColorStream.Enable(ColorImageFormat.RgbResolution640x480Fps30);
kinectSensor.ColorFrameReady += kinectSensor_ColorFrameReady;
kinectSensor.SkeletonFrameReady += kinectSensor_SkeletonFrameReady;
kinectSensor.SkeletonFrameReady += SkeletonExtractSkeletonFrameReady;
Skeleton2DDataExtract.Skeleton2DdataCoordReady += NuiSkeleton2DdataCoordReady;
kinectSensor.DepthStream.Enable(DepthImageFormat.Resolution640x480Fps30);
kinectSensor.DepthFrameReady += kinectSensor_DepthFrameReady;
kinectSensor.SkeletonStream.Enable(newTransformSmoothParameters
            {
                Smoothing = 0.5f,
                Correction = 0.5f,
                Prediction = 0.5f,
JitterRadius = 0.05f,
MaxDeviationRadius = 0.04f
            });


kinectSensor.Start();

kinectDisplay.DataContext = colorManager;
//kinectdepthImage.DataContext = depthManager;
skeletonDisplayManager = newSkeletonDisplayManager(kinectSensor, kinectCanvas);
        }

privatevoidWindow_Loaded(object sender, RoutedEventArgs e)
        {
try
            {
//listen to any status change for Kinects
KinectSensor.KinectSensors.StatusChanged += Kinects_StatusChanged;       //loop through
all the Kinects attached to this PC, and start the first that is connected without an
error.

foreach (KinectSensorkinectinKinectSensor.KinectSensors)
                {
if (kinect.Status == KinectStatus.Connected)
                    {
kinectSensor = kinect;
// MessageBox.Show("Kinect Device Connected", "S2S Converter", MessageBoxButton.OK,
MessageBoxImage.Information);
break;
                    }
```

```
                }
if (KinectSensor.KinectSensors.Count == 0)
MessageBox.Show("No Kinect found", "S2S Converter");
else
                {
Initialize(); // Initialization of the current sensor

                }
            }
catch (Exception ex)
            {
MessageBox.Show(ex.Message);
            }
            _lastTime = DateTime.Now;
            _dtw = newDtwGestureRecognizer(12, 0.6, 2, 2, 10);
            _video = newArrayList();
dtwTextOutput.Text = _dtw.RetrieveText();
        }
```

References

[1] Population of sign language users. [Online].available: http://www.who.int/mediacentre/factsheets/fs300/en/

[2] Different Types of Sign Languages. Sign Language Communication Inc.[Online]. Available:http://signlanguagecommunication.com/deaf-awareness-training/types-of-sign-languages/

[3] Rod R. Butterworth and Mickey Flodin, *The Perigee Visual Dictionary of Signing*. Berkley Publishing Group, 1995

[4] Daniel Martinez Capilla, "Sign Language Translator using Microsoft Kinect XBOX 360" , M.S. thesis, University of Tennessee at Knoxville -USA, IL, 2012.

[5] Sri Lankan Sign Language. Information and Communication Technology Agency Sri Lanka.,[Online]. Available: http://www.lankasign.lk/index.html

[6] Adam Stone and Monika Rego, *An Introduction to Sri Lankan Sign Language.* Rohana Special School, 2007

[7] Marc Van Oostendorp, Colin J. Ewen, Elizabeth Hume and KerenRice,*Companion to Phonology. Vol I*, New York/oxford:WileyBlackwells, 2011

[8] Ms. Rashmi D. Kyatanavar, Prof. P. R. Futane ,Comparative Study of Sign Language Recognition Systems.*International Journal of Scientific and Research Publications, Volume 2*.June 2012

[9] Yang Quan and PengJinye. Application of improved sign language recognition and synthesis technology in ib. In *Industrial Electronics and Applications,2008.ICIEA 2008.3rd IEEE Conference* on, pages 1629 -1634, June 2008.

[10]Thad Starner and Alex Petland(1995), Visual Recognition of Sign Languages Using Hidden Markov Model, Massachusetts Institute of Technology, USA. [Online].Available: http://www.cc.gatech.edu/~thad/p/031_10_SL/visual-recognition-of-asl-using-hmm-95.pdf

[11] Frank Huang and Sandy Huang,” Interpreting American Sign Language with Kinect” [Online]. Available: http://cs229.stanford.edu/proj2011/HuangHuang-InterpretingAmericanSignLanguageWithKinect.pdf

[12] Xiujuan Chai, Guang Li, Yushun Lin, ZhihaoXu, Yili Tang, Xilin Chen(January 2013). Sign language Recognition and Translation with Kinect. Institute of Computing Technology China., [Online].Available: http://vipl.ict.ac.cn/sites/default/files/papers/files/2013_FG_xjchai_Sign%20Language%20Recognition%20and%20Translation%20with%20Kinect.pdf

[13] H. C. M. Herath, W. A. L. V. Kumari, W. A. P. B. Senevirathne, M. B. Dissanayake Image based Sign language Recognition System for Sinhala sign language. in*Proc. SAITM RSEA -2013,* April 27, 2013

[14] ParanjapeKetki Vijay, NaphadeNilakshiSuhas, ChafekarSuparnaChandrashekhar&DeshpandeKetakiDhananjay, Recent Developments in Sign Language Recognition : A Review, *International Journal on Advanced Computer Engineering and Communication Technology* pp2278-5140, Volume-1, Issue – 2, 2012

[15] Abhijith Jana, *Kinect for Windows SDK Programming Guide,* PACKT publishing, 2012

[16] Open Kinect Project.[Online]. Available: http://openkinect.org/wiki/Main_Page.

[17] Open NI Project.[Online]. Available: http://www.openni.org/.

[18] David Katuhe, *Programming with the Kinect for Windows Software Development KIT*, Microsoft Press, 2012

[19] Samaranayake, V. K., Nandasara, S. T., Dissanayake, J. B., Weerasinghe, A.R., Wijayawardhana, H. *An Introduction to UNICODE for Sinhala Characters*, University Of Colombo School of Computing. 2003

[20] JoCrielandElenaTsiporkova,“Gene Time Expression Warper: a tool for alignment, template matching and visualization of gene expression time series”, *Oxford Journal for Bio informatics,*Vol. 22 no. 2 2006, pages 251–252,