Hand Gesture Recognition using Depth Data for Indian Sign Language

Shikha Singhal

Shalakha Singhal

Department of Electronics and Communication Engineering

National Institute of Technology, Rourkela-769008

May 2013

Hand Gesture Recognition using Depth Data for Indian Sign Language

A Thesis submitted in partial fulfillment of the requirements for the degree of

Bachelor of Technology

In
Electronics and Communication Engineering

By

Shikha Singhal

Roll No.: 109EC0244

Shalakha Singhal

Roll No.: 109EC0243

Under the Guidance of

Prof. Kamala Kanta Mahapatra

Department of Electronics and Communication Engineering

National Institute of Technology

Rourkela-769008 (ODISHA)

May 2013

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGG,

NATIONAL INSTITUTE OF TECHNOLOGY, ROURKELA- 769 008

ODISHA, INDIA

CERTIFICATE

This is to certify that the thesis entitled "**Hand Gesture Recognition using Depth Data for Indian Sign Language**", submitted to the National Institute of Technology, Rourkela by **Shikha Singhal, Roll No**. **109EC0244** and **Shalakha Singhal, Roll No**. **109EC0243** for the award of the degree of **Bachelor of Technology** in Department of Electronics and Communication Engineering, is a bonafide record of research work carried out by them under my supervision and guidance.

The candidates have fulfilled all the prescribed requirements. The thesis is based on candidate's own work, is not submitted elsewhere for the award of degree/diploma.

In my opinion, the thesis is in standard fulfilling all the requirements for the award of the degree of **Bachelor of Technology** in Electronics and Communication Engineering.

Prof. Kamala Kanta Mahapatra

Supervisor

Department of Electronics and Communication Engineering

National Institute of Technology-Rourkela,

Odisha– 769008 (INDIA)

Dedicated to

our Family

ACKNOWLEDGEMENT

We would like to convey our deepest gratitude towards our supervisor, Professor Kamala Kanta Mahapatra for his support and supervision, and for the valuable knowledge that he shared with us.

We would like to thank Mr. Kanhu Charan Bhuyan, my friends and juniors who have helped me to complete the thesis work successfully.

We would like to convey appreciation to our family members, for their encouragement and support.

We thank God for being on our side.

Shikha Singhal

Shalakha Singhal

ABSTRACT

It is hard for most people who are not familiar with a sign language to communicate without an interpreter. Thus, a system that transcribes symbols in sign languages into plain text can help with real-time communication, and it may also provide interactive training for people to learn a sign language. A sign language uses manual communication and body language to convey meaning. The depth data for five different gestures corresponding to alphabets Y, V, L, S, I was obtained from online database. Each segmented gesture is represented by its time-series curve and feature vector is extracted from it. To recognise the class of input noisy hand shape, distance metric for hand dissimilarity measure, called Finger-Earth Mover's Distance (FEMD) is used. As it only matches fingers while not the complete hand shape, it can distinguish hand gestures of slight differences better.

Keywords: Depth data, Hand gesture, Sign language, Segmentation, Time-Series Curve, Finger Earth Mover's Distance.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATION

- Full form of HCI is Human Computer Interaction
- Fig. stands for Figure
- FEMD stands for Finger Earth-Mover's Distance
- e.g. stands for exempli gratia which means "for example"
- MATLAB stands for Matrix Laboratory
- RGB stands for Red, Green, Blue
- HMM stands for Hidden Markov Model
- SVM stands for Support Vector Machine
- RBF stands for Radial-Basis Function

LIST OF SYMBOLS

- r_i denotes i^{th} cluster of signature R.
- t_j denotes j^{th} cluster of signature T.
- d_{ij} denotes the ground distance from cluster r_i to t_j
- f_{ij} denotes the flow from cluster r_i to cluster t_j

Chapter 1: Introduction

1.1 Introduction

Hand gesture recognition is of great importance for human-computer interaction (HCI), because of its extensive applications in virtual reality and sign language recognition. Despite lots of previous work, traditional vision-based hand gesture recognition methods are still far from satisfactory for many real-life applications. The quality of the captured images is sensitive to lighting conditions and cluttered backgrounds, because of the limitations of the optical sensors. Thus it is generally not able to detect as well as track the hands robustly. This largely affects the performance of hand gesture recognition. An effective way to make hand gesture recognition more robust is to use different sensors to capture the hand gesture and motion, e.g. through the data glove. Unlike optical sensors, such sensors are generally more reliable and are also not affected by lighting conditions or cluttered backgrounds. However, as the user has to wear a data glove which sometimes requires calibration, it not only is inconvenient for the user but also may hinder the naturalness of the hand gesture. Also, these data gloves are quite expensive. As a result, it is not a very popular way for hand gesture recognition. As a result of development of inexpensive depth cameras, like Xtion PRO LIVE sensor, new opportunities for hand gesture recognition are emerging. In spite of recent successes in applying Xtion PRO LIVE sensor for face recognition and human body tracking, using Xtion PRO LIVE for hand gesture recognition is still an open problem. Xtion Pro LIVE works well to track a large object, e.g. the human body. Due to the low resolution of the Xtion PRO LIVE depth map, of only 640×480, it is difficult to detect and segment small objects from an image, e.g., a human hand which occupies a very small portion of the image with

complex articulations. In such a case, segmentation of the hand is inaccurate and this may significantly affect the recognition step. It is also seen that contours have significant distortions that are local, along with pose variations. Due to the low resolution and inaccuracy of the Xtion PRO LIVE sensor, if two fingers are close to each other in the hand they may be indistinguishable. It is observed that classic shape recognition methods, such as skeleton matching and shape contexts, cannot perfectly recognize the shape contour with rigorous distortions. Visibly, recognizing noisy shapes is very challenging, especially if the number of gestures to be recognized are many. In order to tackle this problem, we use a novel shape distance metric called Finger-Earth Mover's Distance (FEMD). Finger-Earth Mover's Distance is explicitly designed for hand shapes. It considers each finger of a signature as a cluster. By testing on a 5-gesture dataset, this method is found to be accurate and efficient. It performs robustly to local distortions, scale and orientation changes.

1.2 Literature Review

Zhou Ren *et al.* The depth sensors, like the Xtion PRO LIVE sensor, have given rise to new opportunities for human-computer interaction (HCI). Although great progress has been made by using the Xtion PRO LIVE sensor in human body tracking and body gesture recognition, robust hand gesture recognition still remains a problem. Compared to the human body, the hand is a smaller object and has more complex articulations. Thus, a hand is easily affected by segmentation errors as compared to entire human body.

C.W.Ng and S.Ranganath interpret a user's gestures in real-time using hand segmentation to extract binary hand blobs. The shape of blobs is represented using fourier descriptors. This fourier descriptor representation are input to radial-basis function(RBF) networks for posture classification.

N.Tanibata *et al.* obtain hand features from a sequence of images. This is done by segmenting and tracking the face and hands using skin colour. The tracking of elbows is done by matching the template of an elbow shape. The hand features like area of hand, direction of hand motion, etc. are extracted and are then input to Hidden Markov Model(HMM).

D.Kelly *et al.* recognise hand postures used in various sign languages using a novel hand posture feature, eigen-space Size Function and Support Vector Machine (SVM) based gesture recognition framework. They used a combination of Hu moments and eigen-space Size Function to classify different hand postures.

H. K. Nishihara *et al.* (US patent, 2009), generate silhouette images and three-dimensional features of bare hand. Further, classify the input gesture by comparing it with predefined gestures.

Daniel Martinez Capilla used 8-dimensional descriptor for every frame captured by Microsoft Kinect XBOX 360 and compared the signs by dynamic time-warping(DTW).

Jagdish L. Raheja *et al.* This paper describes a novel method of fingertips and centre of palms detection in dynamic hand gestures generated by either one or both hands without using any kind of sensor or marker. We call it Natural Computing as no sensor, marker or color is used on hands to segment skin in the images and hence user would be able to do operations with natural hand.

Yi Li This system consists of three components: Hand Detection, Finger Identification, and Gesture Recognition. The system is built on Candescent NUI project, which is freely available online. Open NI framework was used to extract depth data from the 3D sensor.

1.3 Motivation

Human beings have been gifted, by nature, with voice that allows them to interact and communicate with each other. Hence, spoken language becomes one of the main attributes of humans. Unfortunately, not everybody possesses this capability due to the lack of one sense, i.e. hearing*(Daniel Capilla).* In India, there are around 5 to 15 million deaf people (ref. www.def.org). Sign language is the basic alternative communication method between deaf people and several dictionaries of words or single letters have been defined to make this communication possible. It is hard for most people who are not familiar with a sign language to communicate without an interpreter. Thus, system that transcribes symbols in sign languages into plain text or audio can help with real-time communication. It may also provide interactive training for people to learn a sign language*(Yi Li)*.

1.4 Objective of our thesis

- To develop a system that recognises hand gesture which distinguishes between five different hand gestures signifying five alphabets in Indian Sign Language.
- Analysis and comparison of different hand gestures using MATLAB.

1.5 Work done and Thesis Organization

1.5.1 Work done

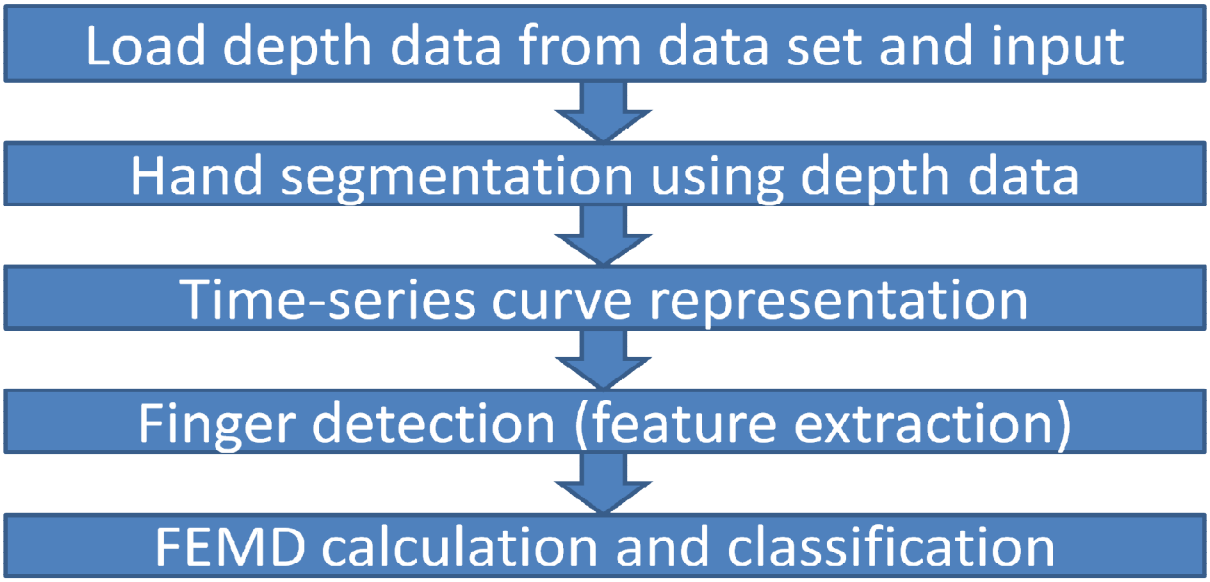

Fig. 1 : Flow diagram of work done

- A database containing depth data in .txt format for many different gestures was used.
- Thresholding was done for a certain depth interval to obtain segmented binary image.

- Hand contour was generated and time-series representation for the hand gesture was obtained.

- In time-series representation:
    1. Horizontal axis denotes the angle between each contour vertex and the initial point relative to the center point, normalized by 360◦.
    2. Vertical axis denotes the Euclidean distance between the contour vertices and the center point, normalized by the radius of the maximal inscribed circle.
- A finger is a segment in the time-series curve, whose height is greater than a threshold hf .
- For 2 signatures, T and R, their FEMD distance is calculated.

1.5.2 Thesis Organization

Chapter 1 includes Introduction, Literature Review, Motivation, Work done and Thesis organisation.

Chapter 2 includes System description and Description of different steps carried out in our project.

Chapter 3 includes Simulation Results and Discussion.

Chapter 4 includes Future Scope and Conclusion.

Chapter 2: System Description

2.1 Introduction

Previous works have been focused on sign language recognition systems. They can be arranged into the gesture/action recognition field, which is a complex task that involves many aspects such as motion modelling, motion analysis, pattern recognition, machine learning, and even sometimes psycho-linguistic studies. They mostly utilized 2D information and only a minority of them worked with depth data (3D). The Centre for Accessible Technology in Sign (CATS) is a joint project between the Atlanta Area School for the Deaf and the Georgia Institute of Technology. The system required an ASL phrase verification to enable interaction. The important citation here is the project that they are developing today. They developed a system called Copy Cat as a practice tool for deaf children to help them to improve their working memory and sign language skills. They are working on a Kinect-based ASL recognition system. In fact, it was after being in contact with this company when the brake on my project's goal was put. Although they did not provide the details of their implementation, they are using the GT2K gesture recognition toolkit and they also use Hidden Markov Models. They are trying to build a system capable to recognize the whole ASL dictionary.

2.2 Sign Language

Nowadays, one can find a wide number of sign languages all over the world (more than 50) and almost every spoken language has its respective sign language. American Sign Language (**ASL**), Mexican Sign Language (**LSM**), French Sign Language (**LSF**), Italian Sign Language (**LIS**), Irish Sign Language (**IRSL**), British Sign Language (**BSL**), Australian Sign Language (**Auslan**), German Sign Language (**DGS**), Indian Sign Language (**ISL**), and Spanish Sign Language (**LSE**) are just a few of them. Among all this large list, American Sign Language is cur-rently the best studied of any sign language and its grammar has been successfully applied to several other sign languages such as British Sign Language, which is not closely related to ASL.

The goal is to provide the reader with a basic knowledge about the sign languages. This section is not going to get into details of a single sign language because each one has its own rules. The following section will attempt to give a general description of the shared characteristics among the different sign languages: origin, phonology, and syntax (for the last two, contains a easy-to-understand description). By doing so, people who are not familiar with them will realize how complex it would be to design a whole Sign Language Translator and why the decision to simplify the system without taking into account these characteristics was made in the version of the system introduced here.

2.2.1 Origin of sign language

Sign language is mainly taught to deaf people, but its origin dates from the beginning history. In fact, gestures are the native way that kids have to express their feelings until they learn

spoken language. Moreover, several hearing communities have used various sign languages to communicate with other ethnic groups that use entirely different phonologies (e.g. American Indians from the Great Plains).

The starting real study of sign languages is relatively younger compared to spoken languages. It dates from 1960, but today there is not an exact definition of their grammar. There is not yet a tradition in the use of a common transcription system that let us guess how young these disciplines are. There is an obvious quantitative and qualitative advance since the beginning of sign language linguistics, but there are still some methodological problems like the definition of a tool to transcript any sort of sign language.

2.2.2 Phonology

In spoken language, the phonology denotes the study of physical sounds present in human speech (called phonemes). Similarly, the phonology of sign language can be defined. Instead of sounds, the phonemes are considered as the different signs present in a row of hand signs. They are analyzed taking into account the following characteristics:

- Configuration: Hand shape while doing the sign.
- Orientation of the hand: Where the palm is pointing.
- Position: Where the sign is being done (mouth, forehead, chest, shoulder).
- Motion: Movement of the hand while doing the sign (straightly, swaying, circularly).
- Contact point: Dominant part of the hand that used to touch the body (palm, fingertip, back of the fingers).
- Plane: Where the sign is being done, depending on the distance with reference to the body (first plane is the one with contact to the body and fourth plane is the most

remote one).

- Non-manual components: Refers to the information provided by the body (facial expression, movements of the shoulders or lip movements). For exam-ple, when the body leans front, it use to express future tense. When it is leaned back, expresses past tense. Also, non-manual signatures show grammatical in-formation such as question markers, negation or localization, conditional clauses, and relative clauses.

2.2.3 Morphology

Spoken languages have both in inflectional and derivational morphology. The former refers to the modification of words to express different grammatical categories such as tense, grammatical mood, aspect, person, number, grammatical voice, gender, and case. The latter is the process of forming a new word based on an existing word (e.g. happi-ness and un-happy from happy). Sign languages have only derivational morphology because there are no inflections for number, tense or person.

Hereafter the main parameters those deal with the morphology are summarized:

- Degree: Known also as mouthing, its the action to make what appear to be speech sounds to give emphasis on a word. For example, the sign "man tall" express "the man is tall". If this signature comes with a syllable "cha", the phrase becomes "that man was enormous".
- Reduplication: Repeating the exact sign several times. By doing so, the sign "chair" is prepared by repeating the verb "sit".

- Compounds: When a word is expressed as the combination of two different words. For example, the verb "agree" is composed by the verbs "think" and "alike" (one sign is executed just after the other).
- Verbal aspect: Verbs can be expressed in different methods so that we can say "to be sick", "to be often sick", "to get sick", "to be continuously sick","to be sickly", "to be very sick", etc. Many of these include reduplication.
- Verbal number: To express singular or plural verbs. Reduplication is also used to express it.

2.2.4 Syntax

As is with spoken language, syntax takes importance in sign languages also. It is primarily expressed through a combination of word order and non-manual features.

It is defined by:

- Word order: For instance, a full structure in ISL is [topic] [subject] [verb] [object] . Topics are indicated with non-manual features, and give a great deal of flexibility to ISL word order. Within a noun phrase, the word order is noun-adjective and noun-number.
- Topic and main clauses: Sets of background information those will be discussed in the following main clause. The eyebrows are raised during the development of a topic. People are prone to want to set up the object of their concern first and then discuss what happened to it. In English, we generally do this with passive clauses: "my cat was chased by a dog". In ISL, topics are used

with similar effect "[my cat] ^{TOPIC} dog chase".

- Negation: Negated clauses may be signaled by shaking the head at the time of the entire clause. Another way to negate a clause is to put the manual sign "none" or "not" at end of sentence. For example, in English: "I don't have any dogs", in ISL: "DOG I HAVE NONE".
- Questions: The questions are signed by lowering the eyebrows. The questioner bends forward slightly and extends the time of the last sign.
- Conjunctions: There is no different sign in ISL for the conjunction "and". Instead, multiple phrases or sentences are combined with a short pause between them.

2.2.5 Conclusion

Once the basic knowledge about the sign languages has been introduced, its interesting to discuss what the designed system takes into consideration and reason why the other characteristics aren't feasible for the project.

The main linguistic characteristics used by the system are a part of the Phonol-ogy section. The joint positions are tracked, that means that the following parameters are considered:

- Position: Where the sign was being done (mouth, forehead, chest, shoulder). For every frame, we can know the position of those joints.
- Motion: Movement of the hands while doing the sign (straightly, swaying, circularly). The positions of the joints was tracked along the frames.
- Plane: Where the sign was done, depending on the distance with reference to the body.

2.3 Xtion PRO LIVE

2.3.1 Specifications

- Power consumption of the device is within 2500mW.
- It works best within a distance of 80cm to 350cm.
- The field of view is 45^{o} V, 58^{o} H and 70^{o} D (Vertical, Horizontal, Diagonal).
- The depth in image size is QVGA (320x240) - 60 fps and VGA (640x480) - 30 fps.
- The resolution is SXGA (1280x1024).
- The dimensions are 0.18 x 0.035 x 0.05 m.

The sensor comprises of a 3D/depth sensor, a RGB camera and a microphone array. The depth/3D sensor consists of an IR laser projector combined with a monochrome CMOS sensor and makes the sensor to process 3D scenes in ambient light condition of any kind. The projector shines a grid of IR on the field of view and a depth map is created on the basis of the rays that the sensor receives from reflections of objects in the scene. The depth map specifies the distance of the object surfaces from the viewpoint of the cameras. This technology was developed by Israeli PrimeSense and interprets the 3D scene information from a continuously-projected IR structured light. The resolution is 320x240 with 16 bits of depth and frame rate of 30fps. This kind of sensing system is considered as a structured light system because it calculates the depth map of the scenes by projecting a structured array of IR dots. The optimal depth sensor ranges from 1.2 to 3.5 meters. The RGB camera, which has a 32-bit high-color resolution of 640x480 with 30fps frame rate . It gets a 2-D colour videos of the scenes. An array of two microphones located at the bottom of the horizontal bar. It enables speech recognition along with acoustic source localization, echo cancellation and ambient noise suppression. The data stream in those cases is 16-bit at 16kilo-Hz.

2.3.2 Application areas

The Xtion PRO LIVE uses IRsensors, adaptive depth detection technology, colour image sensing and audio stream to capture a users' real-time image, voice, movement making user tracking more precisely. The Xtion PRO LIVE development solution comes with a group of developer tools to make its easier for the developers to create their own gesture-based applications without the need of writing complex programming algorithms.

1. Gesture detection: The Xtion PRO LIVE development solution tracks people's hand movements without having any delay, which turns your hand into a controller. It has more than 8 predefined poses which allows you to push, circle, click, wave and much more– perfect to use controlling a user interface. Based on these functions, it can be widely developed into various types of applications, saving developer’s effort and time for developing software.

2. Whole body detection: The Xtion PRO LIVE development solution allows developers to track a users' whole body motion, which makes it ideal for whole body gaming, while supporting multiple player recognition.

3. RGB: Xtion PRO LIVE enables colour (RGB) image sensing. With RGB, Xtion PRO LIVE can capture the users’ images, which is useful for human detection, digital signage, security systems and more applications to be created.

4. Audio: Audio stream allows to support for voice control and any other voice recognition applications. Video conferencing at the office and home are also possible together with the RGB functions.

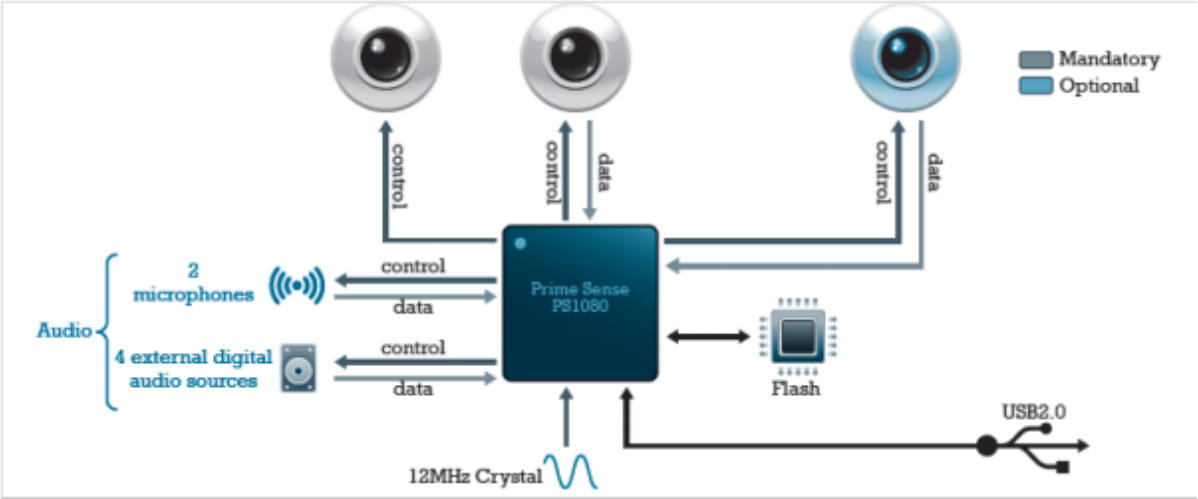

Fig. 2: Basic architecture of the structured light 3D sensor

Chapter 3: Techniques used, Simulation results and Discussion

3.1 Techniques used

Five different gestures are considered corresponding to five different English alphabets in Indian Sign Language. The alphabets used are Y, V, L, S, I. Firstly; each gesture/signature is segmented to extract the hand. The hand is the closest object to the camera. This property used to segment the hand in the image by considering a certain depth interval. The result is a binary segmented image with black segmented hand in white background. The contour of the segmented hand is found and a maximum inscribed circle is found for the corresponding contour. The next step is time-series curve representation of the signature. The time-series curve records the relative distance between each contour vertex and a center point. For the segmented hand portion, a center point is found using distance transform. In distance transform, for every pixel, the distance of the pixel from the closest non-zero pixel is calculated. The center point is the pixel for which this distance is the minimum. The bottom-leftmost pixel of the segmented hand is chosen as the initial point. In time-series representation:

1. Horizontal axis denotes the normalized angle between each contour vertex and the initial point with respect to the center point. The normalization is done by 360◦.
2. Vertical axis denotes the normalised Euclidean distance between the contour vertices and the center point. The normalization is done by the radius of the maximal inscribed circle.
3. A finger is a segment in the time-series curve representation, that has height greater than a threshold height h*f*.

Let R={(r1, wr1), ..., (rm, wrm)} be the first hand signature and T ={(t1, wt1), ..., (tn, wtn)} be the second hand signature with m and n clusters respectively. Here ri and tj are the cluster

representatives ri=[ria, rib], where $0 \leq$ ria $<$ rib ≤ 1 and wri and wtj are the weights of the clusters found by the normalized area within the finger segment.

Consider d_{ij} is the ground distance from cluster r_i to t_j . d_{ij} is defined as the minimum moving distance for interval [r_{ia}, r_{ib}] to totally overlap

$$d_{ij} = \begin{cases} 0, & \mathbf{r}_i \text{ totally overlap with } \mathbf{t}_j, \\ \min(|\mathbf{r}_{ia} - \mathbf{t}_{ja}|,\ |\mathbf{r}_{ib} - \mathbf{t}_{jb}|), & \text{otherwise.} \end{cases}$$

For two signatures, R and T, their FEMD distance is defined as

$$\mathrm{FEMD}(R,\ T) \quad = \quad \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}}$$

where f_{ij} is the flow from cluster r_i to cluster t_j and is found by minimising the following objective function:

$$\arg\min \mathrm{WORK}(R,\ T,\ \mathbf{F}) = \arg\min \sum_{i=1}^{\overline{m}} \sum_{j=1}^{\overline{n}} d_{ij} f_{ij},$$

$$s.t. \begin{cases} f_{ij} \geq 0 & 1 \leq i \leq \overline{m},\ 1 \leq j \leq \overline{n}, \\ \sum_{j=1}^{\overline{n}} f_{ij} \leq w_{\mathbf{r}_i} & 1 \leq i \leq \overline{m}, \\ \sum_{i=1}^{\overline{m}} f_{ij} \leq w_{\mathbf{t}_j} & 1 \leq j \leq \overline{n}, \\ \sum_{i=1}^{\overline{m}} \sum_{j=1}^{\overline{n}} f_{ij} = \min(\sum_{i=1}^{\overline{m}} w_{\mathbf{r}_i},\ \sum_{j=1}^{\overline{n}} w_{\mathbf{t}_j}). \end{cases}$$

We have five classes and one image per class as a template of the class. We use template matching for recognition, i.e., the input hand is recognized as the class with which it has the minimum dissimilarity distance:

$$c = \text{argmin } c \;\; \text{FEMD}(H, Tc),$$

where H is the input hand; Tc is the template of class c; FEMD(H, Tc) denotes the proposed Finger-Earth Mover's Distance between the input hand and each template.

The time-series curve for the input hand signature as well as template signatures is formed and finger detection is also performed in each of the representations (this is nothing but feature extraction). Then, the Finger-Earth Mover's distance is foundbetween the input hand gesture and each of the template signatures. The class having minimum FEMD is chosen as the recognised class for the given input hand gesture.

3.2 Simulation Results

The template image for the class of alphabet V is as shown. First of all the depth data for the template is loaded, followed by hand segmentation. Then, hand contour is found and maximum inscribed circle is plot for the segmented hand. Lastly, the time-series curve is plot for the template image.

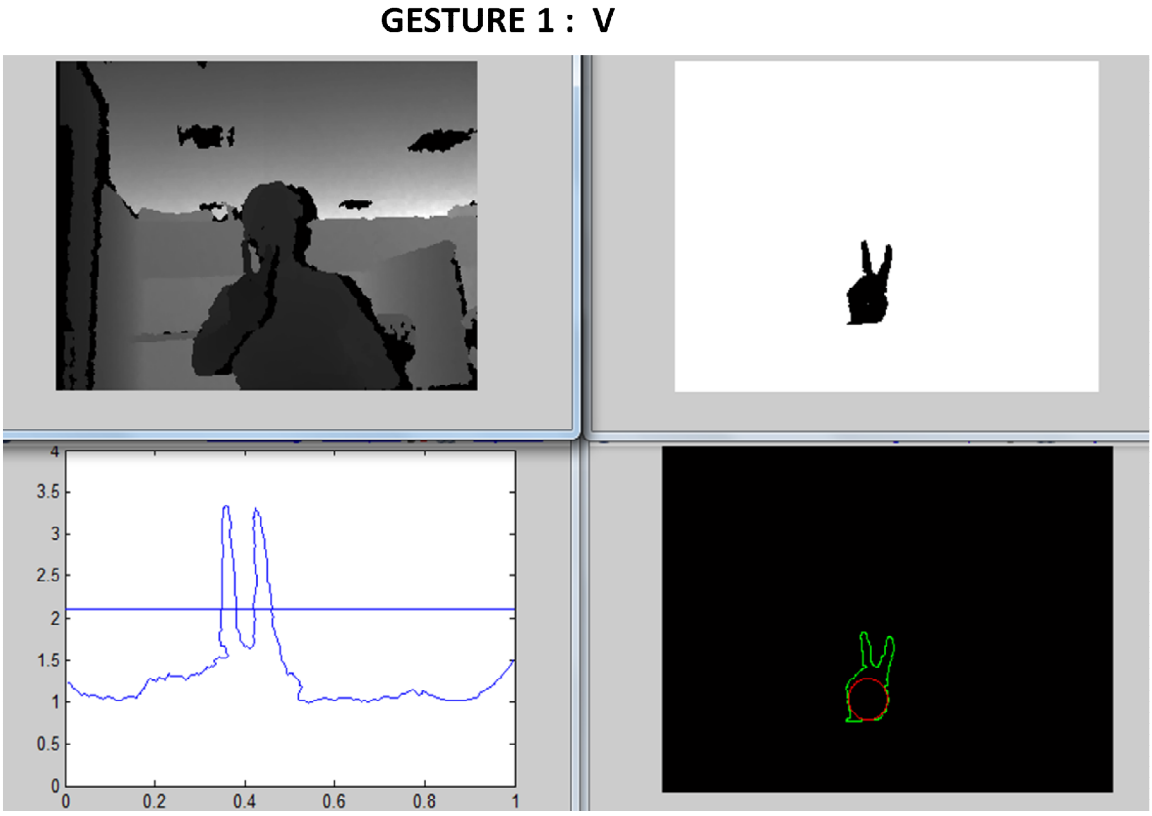

Fig. 3: Clockwise from top-left : Depth image, Segmented binary image, Hand contour and maximum inscribed circle, Time-series curve

The template image for the class of alphabet Y is as shown. First of all the depth data for the template is loaded, followed by hand segmentation. Then, hand contour is found and maximum inscribed circle is plot for the segmented hand. Lastly, the time-series curve is plot for the template image.

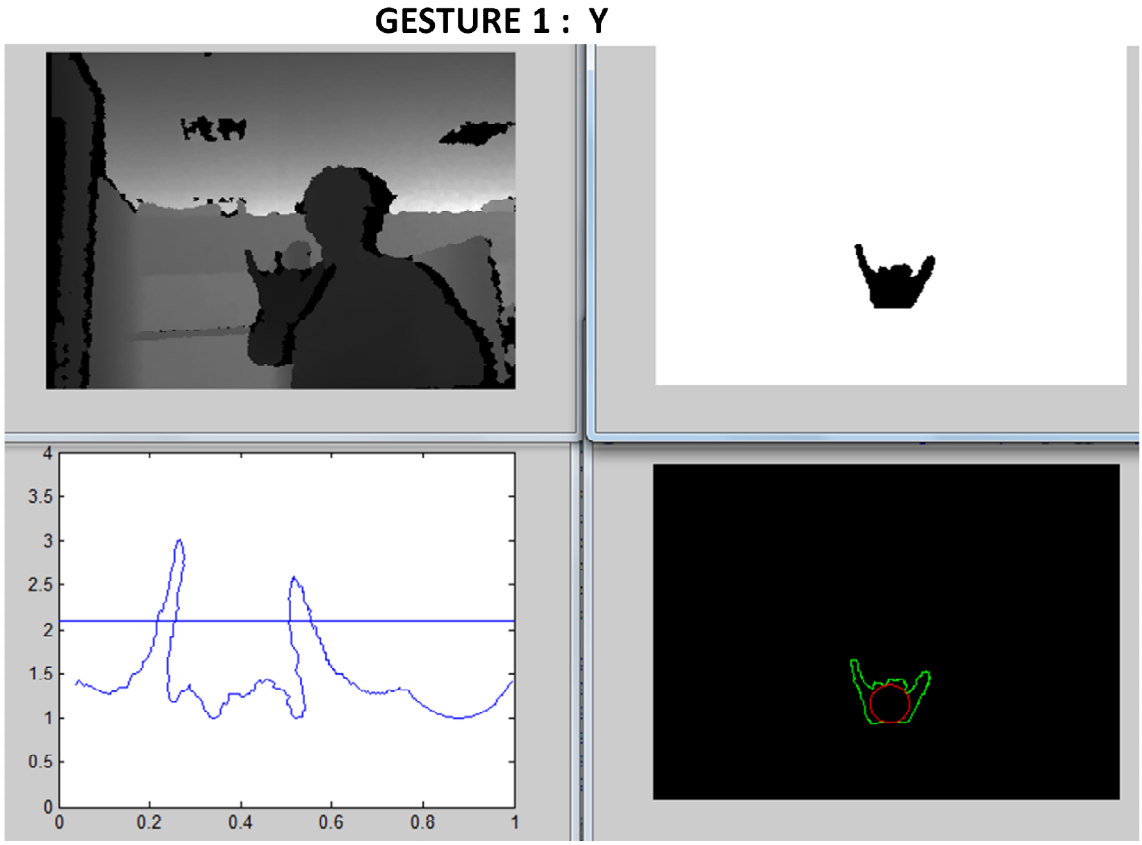

Fig. 4: Clockwise from top-left : Depth image, Segmented binary image, Hand contour and maximum inscribed circle, Time-series curve

The template image for the class of alphabet L is as shown. First of all the depth data for the template is loaded, followed by hand segmentation. Then, hand contour is found and maximum inscribed circle is plot for the segmented hand. Lastly, the time-series curve is plot for the template image.

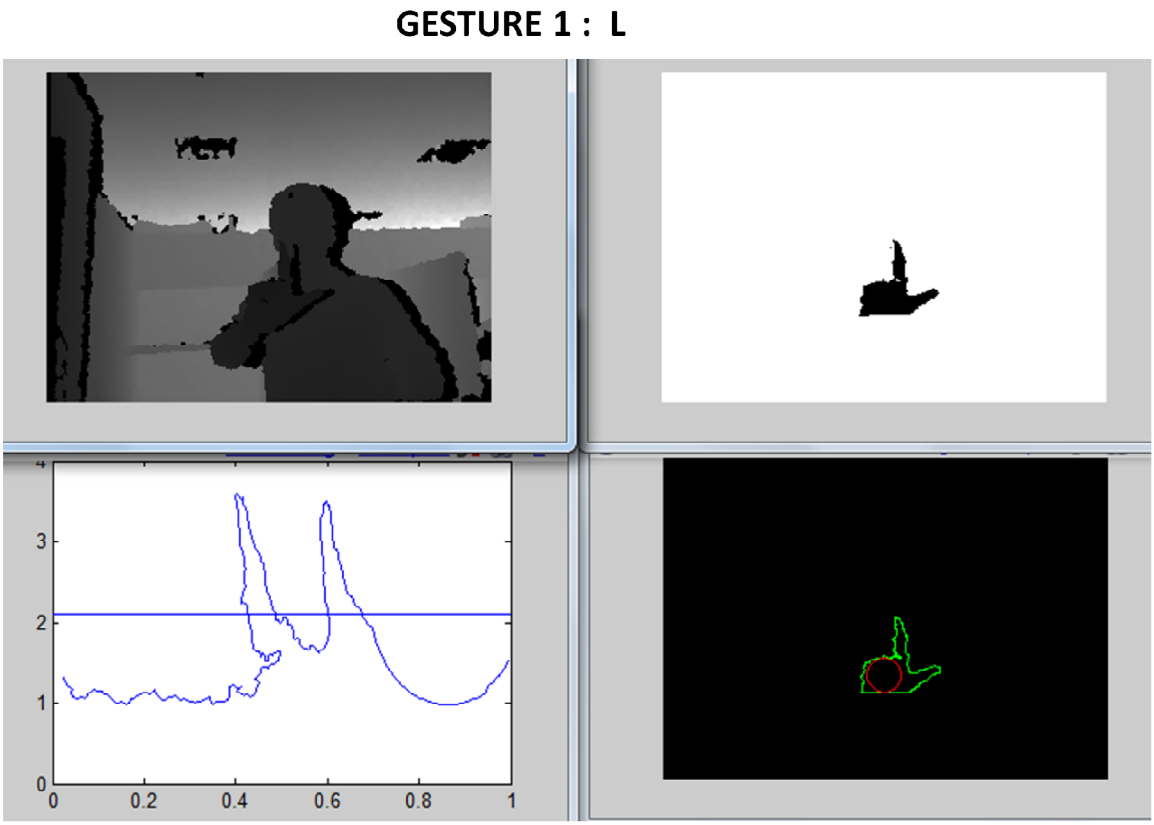

Fig. 5: Clockwise from top-left : Depth image, Segmented binary image, Hand contour and maximum inscribed circle, Time-series curve

The template image for the class of alphabet S is as shown. First of all the depth data for the template is loaded, followed by hand segmentation. Then, hand contour is found and maximum inscribed circle is plot for the segmented hand. Lastly, the time-series curve is plot for the template image.

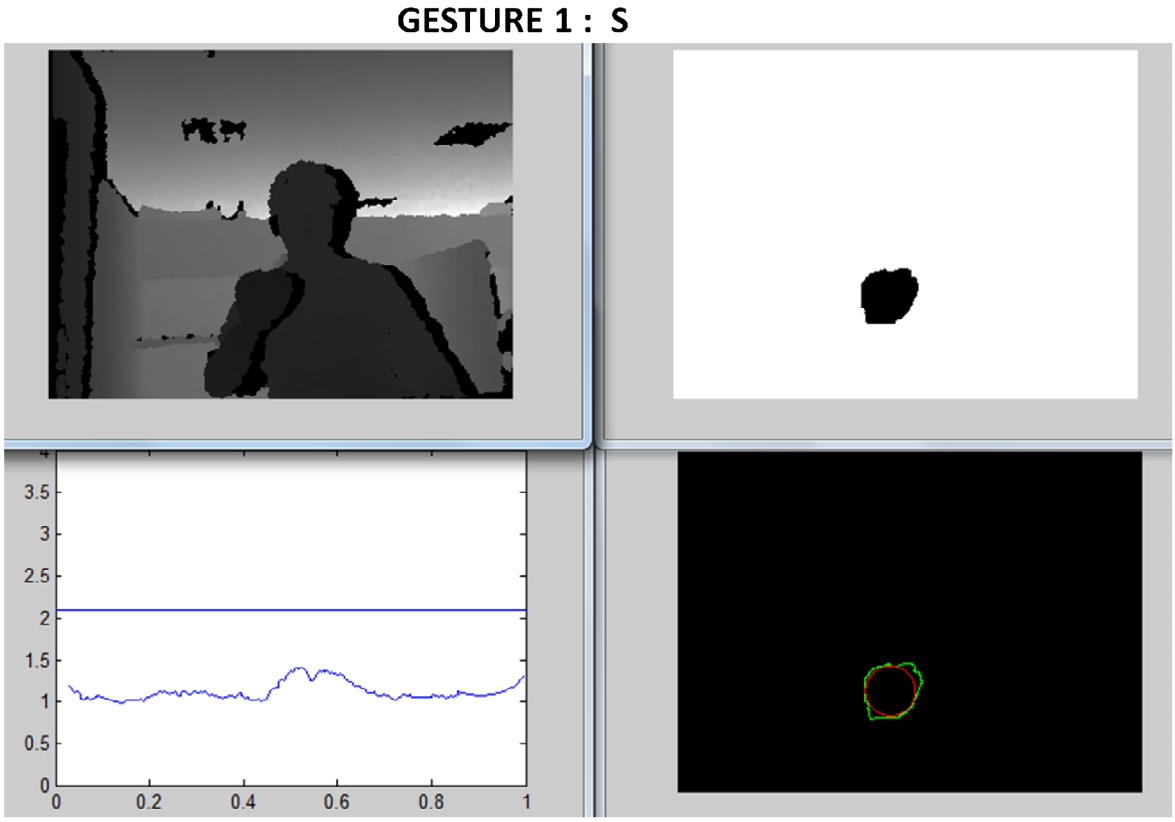

Fig. 6: Clockwise from top-left : Depth image, Segmented binary image, Hand contour and maximum inscribed circle, Time-series curve

The template image for the class of alphabet I is as shown. First of all the depth data for the template is loaded, followed by hand segmentation. Then, hand contour is found and maximum inscribed circle is plot for the segmented hand. Lastly, the time-series curve is plot for the template image.

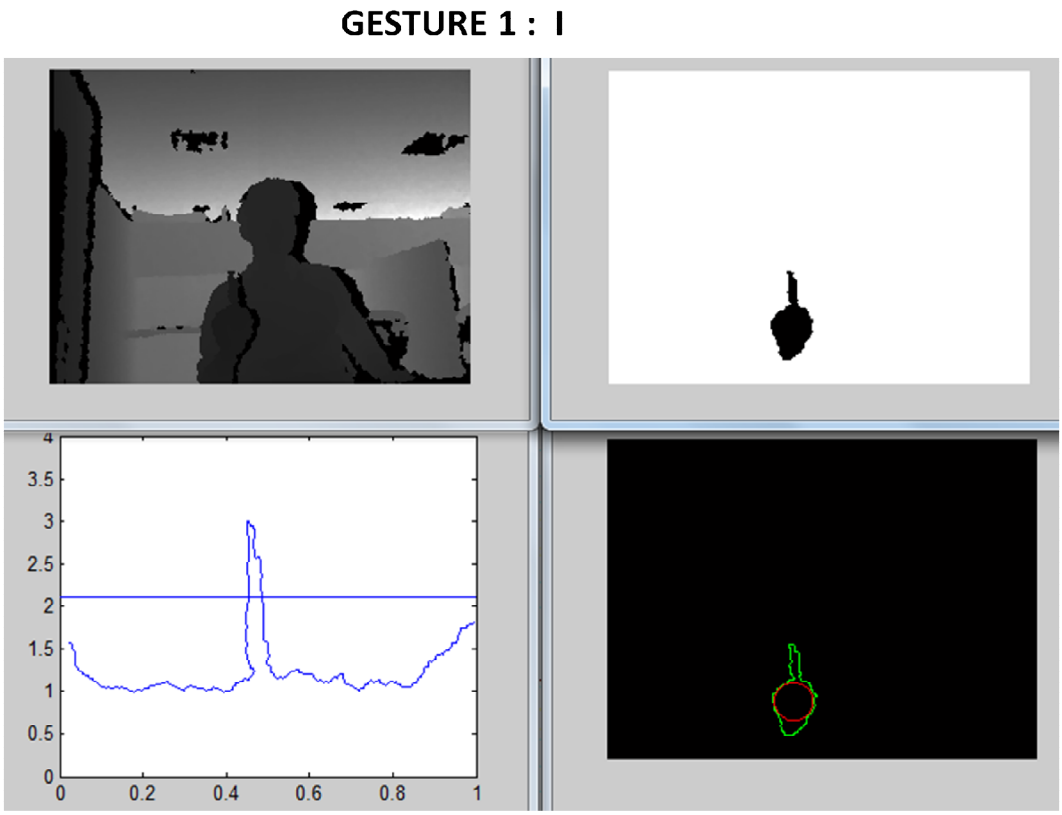

Fig. 7: Clockwise from top-left : Depth image, Segmented binary image, Hand contour and maximum inscribed circle, Time-series curve

The image for the input hand gesture is as shown. First of all the depth data for the input gesture is loaded, followed by hand segmentation. Then, hand contour is found and maximum inscribed circle is plot for the segmented hand. Lastly, the time-series curve is plot for the input hand gesture image.

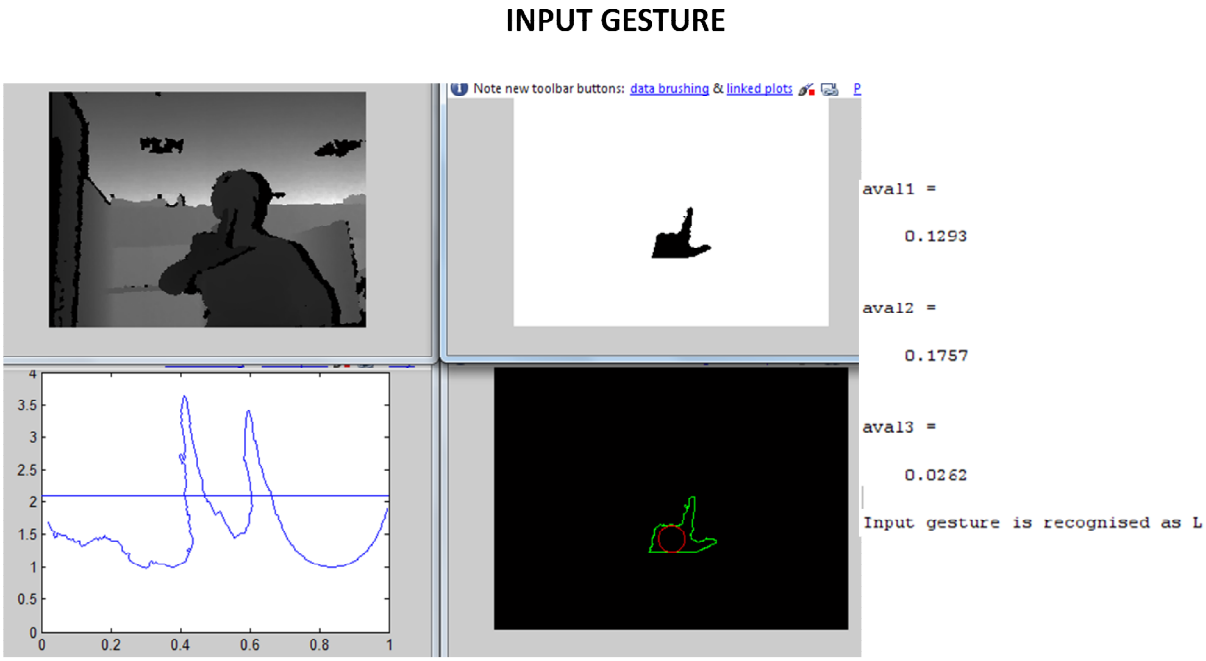

Fig. 8: Clockwise from top-left : Depth image, Segmented binary image, FEMD values and recognition result, Hand contour and maximum inscribed circle, Time-series curve

3.3 Discussion

Once the depth data for the templates and the input hand gesture have been loaded and the time-series curve is plot for each signature, the next step is to recognise the class of the input hand gesture. The time-series curve for the input hand signature as well as template signatures is formed and is used for finger detection is also performed in each of the representations (this is nothing but feature extraction). Then, the Finger-Earth Mover's distance is found between the input hand gesture and each of the template signatures. The class having minimum FEMD is chosen as the recognised class for the given input hand gesture.

For classification purpose, first the number of fingers in the input hand gesture is found. If the number of fingers is zero, then without any FEMD calculation, the input gesture is recognised as belonging to the class of alphabet S. Otherwise, if the number of fingers is one, then the input gesture is recognised as belonging to the class of alphabet I, without calculation of FEMD from the templates. Otherwise, if the number of fingers is two, the FEMD calculation is done for the input feature vectors, from each of the templates. Thereafter, the class with minimum FEMD is chosen as the recognised gesture.

The result for one such input gesture is shown. The number of fingers was found to be two and so, the FEMD from templates of classes V, Y, L are found as 0.1293, 0.1757 and 0.0262 respectively. The minimum FEMD is 0.0262 corresponding to the class L and hence, the input hand gesture is recognised as belonging to the class of alphabet L.

Chapter 4: Future scope and Conclusion

4.1 Future scope

The project involves distinguishing among five different alphabets of English language. Future work may include recognition of all the English alphabets. Further, we may move on to recognising words, from as large a dictionary as possible, from Indian Sign Language. Another method to improve the performance is by using a more accurate method for finger detection from the time-series curve like near-convex decomposition.

4.2 Conclusion

- Hand gesture recognition system was implemented for five gesture classes (Y, V, L, S, I).
- The system was tested for three different inputs per class.
- Out of 15 testing images, 14 were recognized correctly.
- Thus, an accuracy of 93.33% was acheived.

REFERENCES

[1] Zhou Ren, Junsong Yuan and Zhengyou Zhang, “Robust Hand Gesture Recognition Based on Finger-Earth Mover’s Distance with a Commodity Depth Camera,” MM’11, November 28–December 1, 2011, Scottsdale, Arizona, USA.

[2] N.Tanibata, N.Shimada and Y.Shirai, “Extraction of Hand Features for Recognition of Sign Language Words,” *International Conference on Vision Interface,* pp.391-398, 2002.

[3] D.Kelly, J.McDonald and C.Markham, “A person independant system for recognition of hand postures used in sign language,” *Pattern Recognition Letters,* Vol.31, pp.1359-1368, 2010.

[4] H. K. Nishihara *et al.*, Hand-Gesture Recognition Method, US 2009/0103780 A1, date of filing Dec 17, 2008, date of publication Apr 23, 2009.

[5] Daniel Martinez Capilla, Sign Language Translator using Microsoft Kinect XBOX 360, Master dissertation, EE dept., The University of Tennessee, 2012.

APPENDIX

Power Consumption	below 2.5W
Distance of Use	Between 0.8m and 3.5m
Field of View	58° H, 45° V, 70° D (Horizontal, Vertical, Diagonal)
Sensor	RGB& Depth& Microphone*2
Depth Image Size	VGA (640x480) : 30 fps QVGA (320x240): 60 fps
Resolution	SXGA (1280*1024)
Platform	Intel X86 & AMD
OS Support	Win 32/64 : XP , Vista, 7 Linux Ubuntu 10.10: X86,32/64 bit Android(by request)
Interface	USB2.0
Software	software development kits(OPEN NI SDK bundled)
Programming Language	C++/C# (Windows) C++(Linux) JAVA
Operation Environment	Indoor
Dimensions	18 x 3.5 x 5 cm

Specifications of Structured Light 3D Sensor