An Effective System to Acoustically Convey Sign Language

T.C. Sandanayake, G.J.Y. Lenagala, W.L.A.D.V.A. Liyanaarachchi, A.D.K.R. Atukorala, R.S. Weerarathna,
Faculty of Information Technology, University of Moratuwa, Katubedda, Moratuwa, Sri Lanka

***Abstract*—Having 5% of the world's population suffering from speaking disability, building up an effective communication between them and the ordinary community is essential. The people of the community of deaf use sign language as their communication method. In other words sign language is employed by the hearing-impaired to communicate with each other. There are different sign languages used by different cultures and the communities across the globe. It is very important to have a mechanism for this special group to communicate and blend with the ordinary people specially in working day to day. These special groups of people are contributing much in the development of the economy and other social activities. Therefore, there is a need for a mechanism to bridge the communication gap between this special group and the ordinary community. This research study aims to develop an effective mechanism to acoustically convey sign language to the ordinary community using current technological advancements. The research study has identified the communication issues prevailing between the two parties and has developed an effective methodology to convey sign language used by Sri Lankan deaf community. The system has been designed in such a way that the sign or body gesture of the people with speaking disabilities are captured and converted into voice in English which can be easily understood by ordinary people. This research work also can be extended to many sign languages used by different deaf communities around the world which will contribute to the enhancement of communication in the business, education, technological and sociological platforms.**

***Index Terms*—sign language, communication, deaf user, kinect device**

I. INTRODUCTION

Communication is one of the most important basic human needs since it is nearly impossible to spend a single day without the use of communication. Ability to communicate with other parties is needed for almost all the tasks in every human's day-to-day life. Sign language is a visual way of communicating which uses hand gestures, facial expressions and body language by the people with disabilities in hearing and speaking. According to Wikipedia groups of deaf people have used sign languages throughout history across the globe [1]. One of the earliest written records of a sign language is from the fifth century BC, in Plato's Cratylus, where Socrates says: "If we hadn't a voice or a tongue, and wanted to express things to one another, wouldn't we try to make signs by moving our hands, head, and the rest of our body, just as dumb people do at present?" [2]. Visual communication methods have been around for thousands of years and nowadays there are hundreds of different types of sign languages in use across the world [3]. But according to the researches 5% of the world's population is suffering with speaking disabilities [2]. From that, a certain amount can communicate using sign language but the sign language can be understood only by a very limited amount of ordinary people.

The ability to communicate effectively is necessary to carry out the visions and thoughts of people. Sign languages are usually developed and used in deaf communities, which can include interpreters and friends and families of deaf people as well as people who are deaf or hard of hearing themselves [4]. A sign language is a language which, instead of acoustically conveyed sound patterns, uses visually transmitted sign patterns to convey meaning by simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions to express fluidly a speaker's thoughts [4]. Across the globe the sign language has been developed and spread having many different varieties. This special communication method is not only used by the deaf people, it is also used by people who can hear, but cannot speak.

Hundreds of sign languages are in use around the globe and are mixed and merged with the different cultures of deaf. There is a misconception among many people that sign language is somewhat similar to the spoken language in different cultures [4]. Sign language has been made of a combination of shapes and movements of different parts of the body, which are to communicate words and sentences to audience [3]. These parts include face and hands. This research study is aiming of developing a model to acoustically convey sign language to present an efficient and accurate mechanism to transcribe text or speech, thus the "dialog communication" between the deaf and hearing person will be smooth.

This research work helps both the speechless people and people who have speaking abilities to build up an effective communication. The software developed converts the motions of the speechless person (motions which are done according to the sign language) in to words and enable the other party to hear the message through speakers of the computer. The research study has a methodology where different signs of disable person captured and build a data base consist of those signs where it converts in to speech when the need arises and vice versa.

II. LITRETURE REVIEW

During the past few decades several approaches have been undertaken to build and effective communication between the differently able people and the people who are fortune enough to hear and speak. The core of all these approaches is to identify the signs performed by the differently able people.

These approaches differ from one to another based on the method they have used to identify the sign language. Sign language differs from country to country or region to region. Therefore, a global approach to identify the signs is not possible. People have tried on different approaches to identify the sign language based on a specific region. British Sign Language is used by deaf people in the United Kingdom and spread in some other areas of the world. This language makes use of hand gestures, finger spelling, lip patterns and facial expressions [3]. American Sign Language is a language that employs signs made with the hands and other movements, including facial expressions and postures of the body [4]. In Sri Lanka the sign language has been made of using other sign languages used around the world. With the rapid improvement of technology, researchers have paid more attention to this field because of its necessity. Modern development of computer technology has facilitated this in numerous ways. Following are some of the approaches undertaken by different researchers over past few years to identify sign language.

A. *Microsoft uses Kinect to interpret sign language from deaf people*

Developers at Microsoft Research have been using the Xbox 360 gaming peripheral to read sign language from deaf users, and translate it into spoken text [5]. The technology can not only turn sign language into words spoken by a computer, but also do the reverse. A non-deaf user can speak or type words into the Kinect translator. The system will then motion the words in sign language using a virtual avatar shown on a display.

B. *Recognition of American Sign Language using gesture segmentation*

This research study has carried out to automatically translate static gestures of alphabets in American Sign Language [6]. Feature extraction and neural networks have been used to identify the gestures. The system deals with images of bare hands, which allow the user to interact with the system in a natural way. An image is processed and converted in to a feature vector that will be compared with the feature vectors of a training set of signs. The signs involving movements of the body cannot be recognized through this system since only the images of static gestures are processed to identify the signs.

Use of sensor gloves

Sensor gloves are also used to convert sign language to voice in real time. The problem so far with sensor gloves for translating American Sign Language has been that they only interpret letters, which are signed by quick finger gestures, whereas sign language also has more complex gestures for words and meaning [7]. It uses flex sensors for the fingers, and an accelerometer in an attempt to interpret arm gestures.

C. *Indian Sign language recognition using Eigen value weighted Euclidean distance based classification technique*

This research study has been designed for 24 letters of the Alphabet with 240 static images; 10 images for each letter. The alphabets 'H' and 'J' were not considered as they were dynamic gestures. This research work does the skin filtering first to segment the hand from the background, by separating the skin colored pixels from the non skin colored pixels [8]. After that the hand cropping is done and converted it to a binary image..

D. *Behaviour analysis based on coordinates of body tags*

This research was aiming at lengthen the independence of elderly people by detecting falls and other types of behavior indicating a health problem. The tags are worn on the body to get the coordinates to analyze the behavior and the coordinates will be detected using radio sensors. To get accuracy above 90%, four tags were enough to recognize the activities and two tags to identify the walking anomalies [9]. Machine learning technologies have been used in this research work.

E. *Three dimensional dynamic analysis of whole body movements*

This research has been undertaken to do a biomechanical analysis of human body activities. The human activities are evaluated by motion camera system with external reflexive markers [10]. This model can provide a mathematical description of the kinematics of body segments and conduct kinetic analysis. Motion analysis systems use skin attached markers to record unconstrained 3D kinematics of the body segments [11].

III. THE RESEARCH PROBLEM

According to the statistics, 5% of the world's population suffers from speaking disability. Since communication is one of the most important needs of people, this is an issue that should be addressed [5]. Speaking disability not only makes people uncomfortable, it also reduces the opportunity of differently abled people to express themselves, their thoughts and ideas to the others. This problem has to be solved by reducing the communication gap between the differently abled people and the people who are able to talk and hear.

Finding a proper method to recognize sign language and converting it into a format (text and voice) which can be understood by the all is the research problem identified here. There are numerous approaches undertaken by different researchers, but there is a need for a cost effective, simple to use and an accurate solution to recognize Sign language used in the Sri Lankan context..

IV. AIM & OBJECTIVES OF THE RESEARCH

The aim of this research study is to develop an effective system to acoustically convey sign language for deaf people. This is an effective solution to enhance communication between the differently abled people and the people who are able to talk and hear. Major objective of this research study is to identify the sign language used by the Sri Lankan deaf community. Also the study has to convert the identified Sign language in to human understandable format that is text and voice.

V. FRAMEWORK OF THE RESEARCH

The research study was developed aiming at the people who have speaking or hearing disabilities. The main functionality of the study is to provide the message that the disabled person wants to deliver as it is to the ordinary community who does not understand the sign language. The message is expressed through body language by the deaf person and the study uses the Kinect device to detect the signals. The device is used to capture the pre identified coordinates of the human body or finger movements perform by the disabled person. Then the detected signal is sent to the computer which holds the main application created.

The system has used C# as its programme language. The main application is created in such a way that will identify the detected signal and convert it into meaningful text which appears on the computer screen. The programs written are capable of matching the particular sign into a meaningful message in English. The study has used the stored words, phrases and letters in English in a database.

The system also converts the text message into voice which can be heard through the speakers of the computer. The basic framework of the proposed solution is showed in the Figure 1 and it indicates how the communication takes places between two persons who is able and disable in hearing and speaking. The figure 1 shows how the kinect device is being capturing the sign language of the deaf person as the input signal and the output signal being visible to the ordinary user.

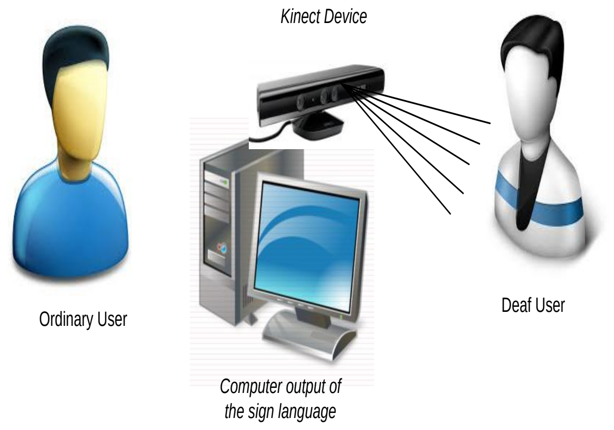

Fig 1. Basic framework of the proposed solution

As per stated in the aim of the study in above the research study is targeting to develop a cost effective communication system to build the gap between deaf person and an ordinary person using modern technological advancements.

VI. REASERCH APPROACH

The first approach of the research study was to identify the sign language used by the Sri Lankan deaf community. In order to gather required information, students of the Ceylon Deaf and Blind School at Ratmalana were closely monitored and studied for their use of sign language. The research team had to learn the sign language virtually along with the meanings that they use to communicate. The observations was taken with the help of instructors and teachers who can hear and speak English language. The next step was to define the human body coordinate metric to capture different signs and signals sent by the user, using Kinect device. Figure 2 shows the high-level architecture of the system.

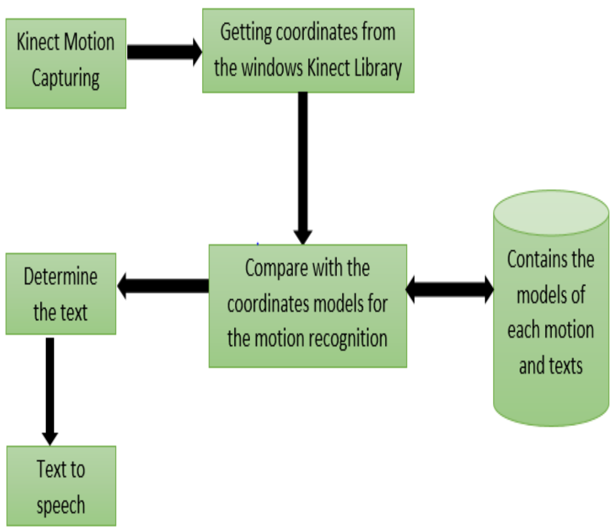

Fig 2. High level architecture of the proposed system

The device is capable of capturing the movements of the body and the selected points also can be seen as the points of a human skeleton using the application [13]. The following figure 3 shows the process diagram of the research study. This process diagram much more elaborate on the signal processing mechanism of the proposed software solution. The observations were collected using the group of users who are reading their Bachelors degree in the Faculty of Information Technology.

This user group learnt sign language before getting the readings of the research study. According to the figure 2 , the high-level architecture of the research study. the process of converting the input signal of sign language convert in to an English language work or a phrase as a voice output. Figure 3 shows the process diagram of the system which given below.

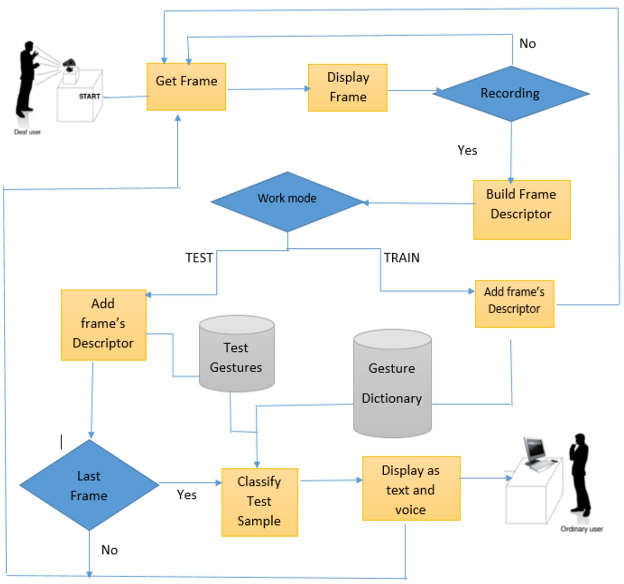

Fig 3. Process diagram of the system

According to figure 3, the deaf user is in front of the camera and shows a sign to capture. When the user wants to record a sequence of signals, system capture all of them and if not the system asks the camera to get the next frame. There are two major processors in the system. The first one is to capture images and obtaining the data of the Joints of Interest (JoI) required for the frame descriptor. The second process consists of normalizing the captured data and then builds the frame descriptor. The system is performing on the 'training' mode when the user adds up a new sign to the system. In the training mode, the frame descriptor is added to the correspondent file of the dictionary of signs. When the user requires to translate a sign, two processors are executed and the frame descriptor is added to the current test sample. Then the test sign is compared using a classifier with the signs stored in the sign dictionary and the corresponding output is displayed so that the ordinary user will get the corresponding word in English.

Open Natural Interaction (OpenNI) is a multi-language, cross-platform frame-work that defines APIs for writing applications utilizing Natural Interaction [14]. The main purpose of OpenNI is to form a standard API that enables communication with both vision and audio sensors and vision and audio perception middleware [14]. This OpenNI/NITE is able to track up to 15 joint positions of a human body. According to the signs of the proposed default dictionary for the system, only 4 joints out of the 15 resulted to be significant for the description of a sign. They are both hands and both elbows of a person. The other joints such as shoulders, the knees, and the feet will not be tracking since those joints remain static during the signal generation. Some other joints such as head and torso joints are required for the normalization and the sign modelling steps. The figure 4 below shows the positions of the joints which measures in the process. They are Left Hand (LH), Left Elbow (LE), Right Hand (RH), Right Elbow (RE) and Head (H). In order to start a recording, an initial position is established. The recording will be finished when the user remains with a static position for a given number of consecutive frames.

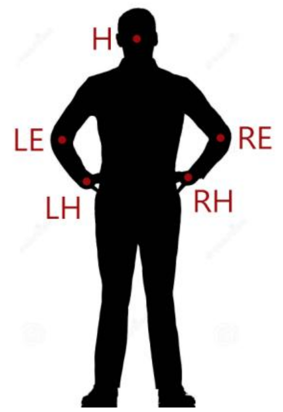

Fig 4 Position of the joints

The table I gives the comparison between the current approaches and the proposed system. The table 1 given below shows the similarities and differences of existing systems and proposed solution in various measures.

TABLE I : COMPARISON BETWEEN THE CURRENT SYSTEMS AND THE PROPOSED SYSTEM

Current Systems	Proposed Solutions
Translation is mostly limited to text.	Translate into both text and voice.
Body movements cannot be recognized. Only the images of static gestures are use to process the signs	Signs can be recognized dynamically.
Only interpret letters of alphabet, signed by quick finger gestures.	Considered letters as well as words in interpretation.
Mostly targeted only the adult users	Anyone can use specially the children.
Movements of whole body considered in the analysis in translation	Hands, Elbows, and Head movements are considered.

In the comparison of the system, the study has used following as the measures. They are translation output format, capturing of body movements, translation elements, target group of users.

VII. RESERCH IMPLEMENTATION AND RESULTS

The proposed system must be robust to size and position of the deaf user. Hence, the normalization becomes one of the most indispensable steps of the implementation. The data normalizer uses a feature of the Open NI named Scale Invariant Feature Transform (SIFT) was introduced to extract distinctive invariant features from images that can be used to carry out reliable matching between various views of an object or a scene [15]. The system allows the deaf user to provide signals in different physical locations in a closed area. The data must be stored accordingly to that position of the user. A slight variation in depth can cause a considerable variation of the X and Y values [15]. The distances between one joint and another one can drastically vary depending on the position of the user. Instead of directly storing the Cartesian coordinates X, Y, and Z coordinates are being captured since it normalize all the joint spherical coordinates with respect to the position of the torso. But the system considered that the position are remains constant always through the normalization process [16]. In system it has considered the values: the radial distance of that point from a fixed origin(r), polar angle measured from a fixed zenith direction (θ) and angle of its projection on a reference plane (ϕ) in identification of the signal [16]. According to the user position these values are input for the system to continue the further functionality.

The system has considered variant of the user's positions and variant of the user's size. It has to be capture and translate the correct sign irrespective of the height of the user. Also user can move during the signal generation process and the system should capable of identifying such signals. Figure 5 shows the different user positions during the signal generation process.

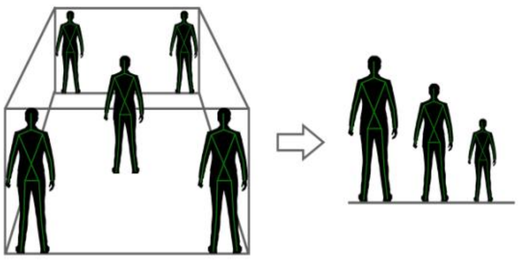

Fig 5. Different user positions and sizes

The distance from one joint to another change significantly depending on the user's size but there is no way to add the samples for all the possible user's sizes to the dictionary [16]. Therefore the dictionary is built to have the same sign described for different user's sizes according to the relative distance to the torso joint and the two angles θ and ϕ that describe the orientation of this distance.

Sign descriptor based on the spherical coordinates values for every joint. Once the coordinates obtained, the next step is building a descriptor for each sign. The descriptor must be able to describe and should be unique and different from the other descriptors of the dictionary. Hence it will give the corresponding word in English when the deaf user inputs a sign as the output.

In classification of the sign the system uses Dynamic Time warping (DTW) algorithm [17]. By using DTW computer is able to find an optimal match between two given sequences with certain restrictions. In simply it is an algorithm for measuring the similarity between two sequences, which may vary in time or speed. The Microsoft Kinect XBOX 360TM is proposed to solve the problem of sign language translation [16]. By using the tracking capability of this RGB-D camera, a meaningful 8-dimensional descriptor for every frame is introduced here. In addition, an efficient Nearest Neighbor DTW and Nearest Group DTW are developed for fast comparison between sign languages.[16]. The default samples which are already stored in the database, belongs to the same user and are executed at the same position. In order to test the system, ten test signs are collected randomly and those tested signs were given in the table II as given below.

TABLE II LIST OF FEW SIGNS TESTED

Sign No	***Meaning of the sign***
Sign1	Mother
Sign2	Love
Sign3	Doctor
Sign4	Want
Sign5	Sick
Sign6	Hungry
Sign7	Eat
Sign8	You
Sign9	I
Sign10	Hello

The accuracy of the system has to be evaluated and tested for the different implemented approaches and configuration of parameters is analyzed. In the testing process the study has used a group of ten users in different sizes [16].

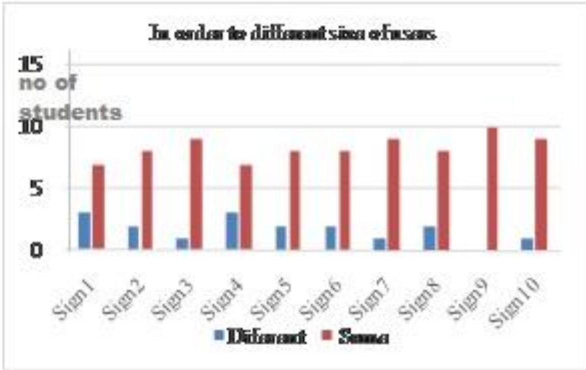

Fig 6. Sign test result for different user sizes

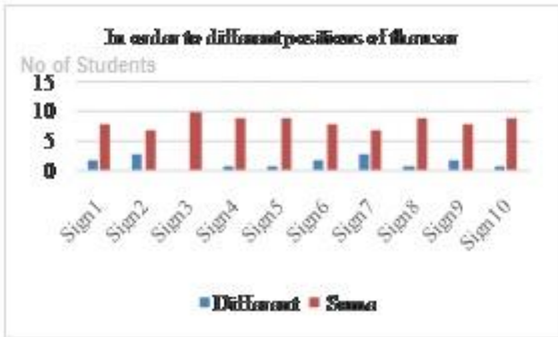

Fig 7. Sign test result for different user positions

For each sign which is tested using the sample group was done and calculated as to how many users did the signs differently than the default signs stored in the dictionary. These tested signs were tested according to the similarity and the differences according to the figure 6 given below. The figure 6 shows that there is a significant relationship with the tested signs for the different sizes of users and only very limited numbers has shown the difference.

The same test has been conducted while the users are being physically moving in different locations. The study has observed and calculated the number of users who did the signs differently than the default sign in the dictionary. The figure 7 shows that there is a significant relationship with the tested signs for the different positions of users and only very limited numbers has shown the difference. According to the above figures 6 and 7, the accuracy of the system is 84%. By combining the 10 signs from the default dictionary, basic sentences have been identified as well as the list can be easily increased. These sentences consist of basic sentences such as "I love you", "I am sick", "I am hungry" have tested. The communication between a deaf user and an ordinary user will become possible in business meetings, hospitals, supermarkets, etc by using these sentences.

VIII. Conclusion

This research study was aiming at developing an effective system to acoustically convey sign language of deaf people, where they will be able to communicate with ordinary people in English. Even though several works have been proposed previously they have mostly used probabilistic models such as Hidden Markov Models or Artificial Neural Networks classifiers, therefore, the Microsoft Kinect XBOX 360TM was proposed to solve the problem of sign language translation in this study. By using Kinect, a wide number of signs can be recognized and the system has the potential to provide a computationally efficient design without sacrificing the recognition accuracy compared to other similar studies. To make this system work with all Sign Language signs, some other features such as the finger position or shape of the hand has to be considered. In addition to that new ways to detect the initial frame of a gesture will make the system more automatic. The last future improvement refers to the computational cost. Although the current system works in real time, its computational cost could be improved by reducing the number of dimensions from the descriptors to those that are most meaningful.

ACKNOWLADGEMENT

The research team appreciates the great effort proofreading this research paper by Mrs. M. Kalahepathirana with an special attention. A special thanks goes to Mr. Cristopher from the Ceylon Deaf School at Rathmalana for the enormous support rendered at the beginning of the research study in information and data gathering.

References

[1] Wikipedia, *Sign Language*,

[2] http://en.wikipedia.org/wiki/Sign_language, Retrieved on 05/09/2014

[3] B. Dirksen. *Open your eyes: Deaf studies talking*. University of Minnesota Press. ISBN 0-8166-4619-8, 2008

[4] NHS, *Choices / Communication*, Retrieved on 08/09/2014 http://www.nhs.uk/CarersDirect/guide/communication/Pages/Signlanguage.aspx

[5] Disabled World, *Sign Language and Deaf Communication Methods and Information*, http://www.disabled-world.com/disability/types/hearing/communication/, Retrieved on 13/10/2014

[6] PC World, *Microsoft uses Kinect to interpret sign language from deaf people,* http://www.pcworld.com/article/2059880/microsoft-uses-kinect-to-interpret-sign-language-from-deaf-people.html, Retrieved on 12/10/2014

[7] V.S. Kulkarri, S.D. Lokhendle, *Appearance Based Recognition of American Sign Language Using Gesture*, International Journal on Computer and Engineering, Vol 2, No 3, pp 560-565, 2010

[8] S.A. Mehdi, Y.A. Khan, *Sign language recognition using sensor gloves*, 9th International Conference on Neural Information Processing, Vol 5, 2002.

[9] J. Singha, K. Das, *Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique,* International Journal of Advanced Computer Science and Applications, Vol. 4, No. 2, 2013

[10] M. Lustrek, B.Kaluza, E. Dovgan, B. Pogorelc, M. Gums, *Behavior Analysis Based on Coordinates of Body Tags*, 2009

[11] Swami, A.B. Ramesh, D. Ranganathan, J. Sequeira, P.Rmadorai, *Kinect 2 Connect-Software Requirements Specification*, Retrieved on 13/10/2014 http://www.acsu.buffalo.edu/~abhinavs/Phase%203%20SRS%20Document%20Final.pdf

[12] Y. Zhang, S.M. Hsiang, *A New Methodology for Three-dimensional Dynamic Analysis of Whole Body Movements*, International Journal of Sports Science and Engineering, Vol. 2 (2): 87-93. 2008.

[13] X. Chai, G. Li, Y. Lin, Z. Xu, Y. Tang, X.Chen, M.Zhou, *Sign Language Recognition and Translation with Kinect*, Microsoft Research Asia, the FiDiPro Program of Tekes and Natural Science Foundation contracts 61001193 and 60973067, 2012

[14] D. Martinez, *MSc Thesis - Sign Language Translator using Microsoft Kinect XBOX 360TM, VIBOT 5.* , Department of Electrical Engineering and Computer Science, Computer Vision Lab, University of Tennessee, 2012

[15] F. Bellet , *RPM resource openni*, http://www.rpmfind.net/linux/rpm2html/search.php?query=openni, Retrieved on 15/09/2014

[16] D. Lowe, *Distinctive Image Features from Scale-Invariant Keypoints*, International Journal of Computer Vision, Volume 60, Issue 2, 11, pp 91 – 110, 2004

[17] D. Capilla, *Sign Language Translator using Microsoft Kinect XBOX 360TM*, Department of Electrical Engineering and Computer Science - Compute Vision Lab Uni. of Tennessee, http://www.dani89mc.com/documents/SignLanguageTranslator_DaniMC_paper.pdf, Retrieved on 22/09/2014