2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

Conversation of Sign Language to Speech with Human Gestures

Rajaganapathy. S[1], Aravind. B[2], Keerthana. B[3], Sivagami. M[4]

[1, 2, 3]*MS Software Engineering,* [4]*Assistant Professor*
VIT University – Chennai
INDIA.
rajaganapathy.s2010@vit.ac.in, aravind.b2010@vit.ac.in,keerthana.b2010@vit.ac.in,msivagami@vit.ac.in

Abstract

Inability to speak is considered to be true disability. People with this disability use different modes to communicate with others, there are n number of methods available for their communication one such common method of communication is sign language. Sign language allows people to communicate with human body language; each word has a set of human actions representing a particular expression. The motive of the paper is to convert the human sign language to Voice with human gesture understanding and motion capture. This is achieved with the help of Microsoft Kinect a motion capture device from Microsoft. There are a few systems available for sign language to speech conversion but none of them provide natural user interface. For consideration if a person who has a disability to speak can stand perform the system and the system converts the human gestures as speech and plays it loud so that the person could actually communicate to a mass crowd gathering. Also the system is planned in bringing high efficiency for the users for improved communication.

1. INTRODUCTION

Sign language is a system of communication using visual gestures and signs, as used by deaf and dumb people. There are various categories in the sign language like ISL (Indian Sign Language), ASL (American Sign Language), BSL (British Sign Language) and etc... But none of the sign languages are universal or international. A person should know the sign language to understand the language; this becomes complicated when a person who has inability to speak or hear wants to convey something to a person or group of persons, since most of them are not familiar with the sign language. Application development in

Human machine interaction and Natural user Interface has reached the crowning from the release of Microsoft's motion sensing gaming device Microsoft Kinect and its Software Development Kit [1] [2][3]. Humans however migrate towards technology advancements always expect flexibility in the way they use their system and machinery. At present lots of techniques and modulations are being introduced and are under research to minimize or simplify the complexity in sign language to speech. The paper is been proposed in the aim of minimizing all those complexions and to attain maximum accuracy in conversion of sign language to speech with gestures. Human gestures are an important sign of human communication and an attribute of human actions informally known as the body language. A lot of methods are being in use to track human gestures [2]. To get maximum accuracy and to bring out the system unique a lot of methods are attempted and best case is user defined actions (gestures) to control the system. For example consider a person who has the disability to speak wants to say "Hello" to a group of people who doesn't know sign language. The user stands in front of the system and waves the hands and system throws out the speech "HELLO". The use of Microsoft Kinect in the system is to track the human joints and gestures; the stream of input data to the Kinect will be the live action of human's gestures. Once the human skeleton is identified the system keeps track on the gestures and matches with the user defined gestures. Once if both the gestures suits the word is played.

2. LITERATURE RESEARCH

[3] Research in the sign language system has two well-known approaches are Image processing and Data glove. The image processing technique [4] [5] using the camera to capture the image/video. Analysis the data with static images and recognize the image using algorithms and produce sentences in the display, vision based sign language recognition system mainly follows the algorithms are Hidden Markov Mode (HMM) [6], Artificial Neural Networks (ANN) and Sum of Absolute Difference (SAD) Algorithm use to extract the image and eliminate the unwanted background noise. The main drawback of vision based sign language recognition system image acquisition process has many environmental apprehensions such as the place of the camera, background condition and lightning sensitivity. Camera place to focus the spot that capture maximum achievable hand movements, higher resolution camera take up more computation time and occupy more memory space. User always need camera forever and cannot implement in public place. Another research approach is a sign language recognition system using a data glove [7] [8].user need to wear glove consist of flex sensor and motion tracker. Data are directly obtained from each sensor depends upon finger flexures and computer analysis sensor data with static data to produce sentences. It's using neural network to improve the performance of the system. The main advantage of this approach less computational time and fast response in real time applications. Its portable device and cost of the device also low. Another approach using a portable Accelerometer (ACC) and Surface Electro Myogram (sEMG) [9] sensors used to measure the hand gesture. ACC used to capture movement information of hand and Arms. EMG sensor placed, it generates different sign gesture. Sensor output signals are fed to the computer process to recognize the hand gesture and produce speech/text. But none of the above methods provide users with natural interaction. This proposed system will be capable of performing the conversation without any wearable device instead using the human motion and gesture recognition.

3. MICROSOFT KINECT SENSOR

Kinect is a motion sensing device by Microsoft for the Xbox 360 video game console and Windows PCs. Based around a webcam-style add-on peripheral for the Xbox 360 console, it enables users to control and interact with the Xbox 360 without the need to touch a game controller,

Figure 1: Microsoft Kinect

through a natural user interface using gestures and spoken commands [10].

RGB camera, 3D depth sensing system, Multi-array microphone, motorized tilt

The basic parts of the Kinect are (Figure 1):

It interacts to the system by understanding human gestures.

- Software can use video, sound and gesture recognition driven interactions

- The Kinect sensor contains a high quality video camera which can provide up to 1280x1024 resolution at 30 frames a second

- The Kinect depth sensor uses an IR projector and an IR camera to measure the depth of objects in the scene in front of the sensor

- The Kinect sensor contains four microphones

    - These can perform background noise cancelling

Kinect is able to capture the surrounding world in 3D by combining the information from depth sensors and a standard RGB camera. The result of this combination is an RGBD image with 640x480 resolution, where each pixel is assigned color information and depth information (however some depth map pixels do not contain data, so the depth map is never complete). In ideal conditions the resolution of the depth information can be as high as 3 mm, using 11-bit resolution. Kinect works with the frequency 30 Hz for both RGB and depth cameras. On the left side of the Kinect is a laser infrared light source that generates electromagnetic waves with the frequency of 830 nm. Information is encoded in light patterns that are being deformed as the light reflects from objects in front of the Kinect. Based on these deformations captured by the sensor on the right side of RGB camera a depth map is created as shown in figure 2. According to Prime Sense this is not the time-of-flight method used in other 3D cameras.
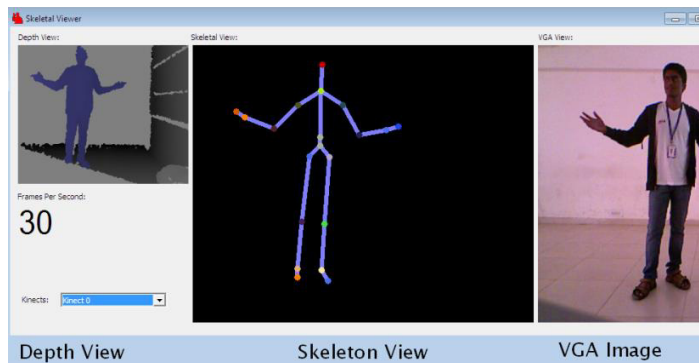

Figure 2: Depth, skeleton and VGA view

4   PROPOSED WORK

The proposed work is producing speech/ voice to sign language with simple human gestures and motion sensing technology with the help of Microsoft's Kinect sensor. This paper started its initiations in the vision to successfully minimize the human machine interaction and to take up the Natural User Interface at the forefront. The initial phase of the project began in controlling a simple PowerPoint presentation with

gestures like moving hands from left to right or moving from right to left to move between PowerPoint slides. The next phase of the project is controlling electrical appliances with human gestures. Now the paper is motivated on understanding the sign language and producing speech.

The system is already trained with the understanding of human gesture movements. Based on the body gestures the inputs are taken and the speech is produced. The system is stably designed to identify 20 of human joints like (head, hand_right, hand_left and so on...) as shown in figure 3. For sign language understanding the gestures are the key aspect, using the identified 20 joints the system keeps in track of all the gestures that the user performs. As on the paper the system uses a general sing language for interpretation since sign languages like ASL and BSL use fingers for certain words. The position of each joint is given as an offset from the Kinect sensor as shown in figure 4. X for left and right, Y for up and down and Z is to calculate the distance between the user and the sensor. The values are given in millimetre.
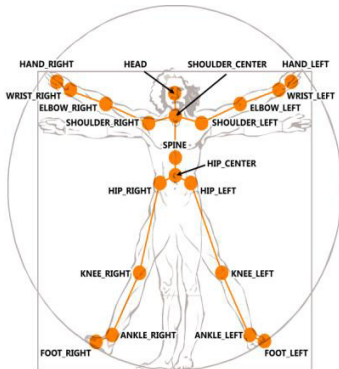
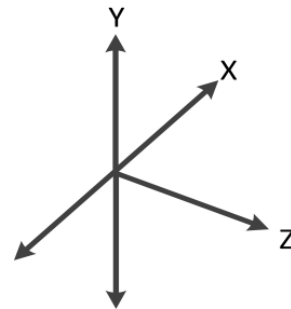Figure 3: Skeleton Understanding by Kinect with its joints   Figure 4: Coordinate Axis

Once the user gesture matches the predefined gesture the system throws out the corresponding word as a text to the windows narrator. The system uses windows narrator for converting the generated text to speech. The detailed working of narrator is shown in figure 5.
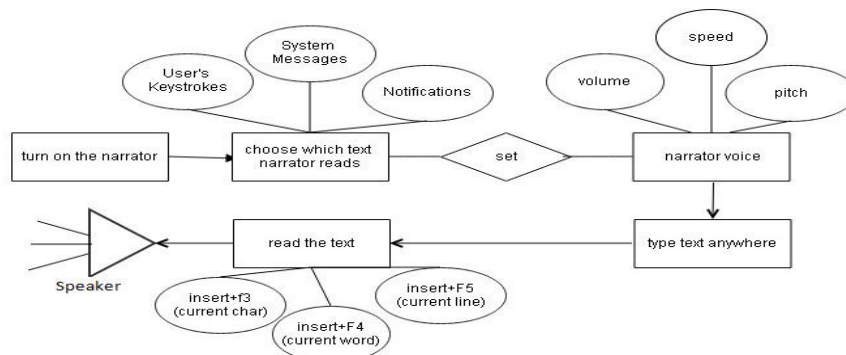
Figure 5: Windows narrator

The system initializes once the sensor is turned on; once a human gestures enters the frame the skeleton data of the user is tracked with the 20 joints and their coordinates. The stream of inputs are obtained as individual skeleton frames and each frame has a posture or a gesture. The obtained gestures are matched with the predefined gesture input set. If the current skeleton frame matches the predefined gesture pattern the corresponding word for the gesture is thrown as text to the windows narrator. The narrator produces the speech. The Kinect is connected to the power supply with a power cable and central processing unit of the computer through a USB port; program developed in .Net platform takes control of the Kinect and its input.

The input stimulator in the application virtualizes the key stokes of the generated word which in turn becomes the windows narrator to produce the sound.
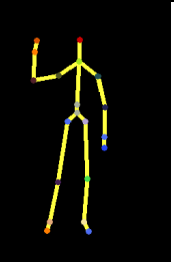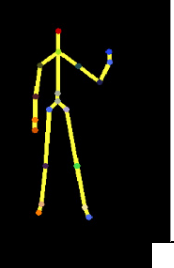
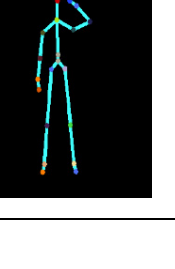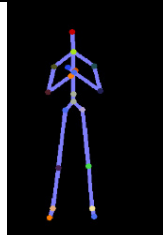| Human Gestues : Input Stream | Kinect : Receive Gestures | Application : Match Gestures | Application: Generate Text | Narrator: Receive Word | Spearker: Produce Speech |

Figure 6: Sequential process flows.

5   PERFORMANCE ANALYSIS

A series of experiments are being conducted to evaluate the system's returns and limitations.
In a test done for a sample of 100 spells for different signs. Accuracy up to 90 percent has been achieved.
The below given tabular column shows a few random tests.

Table 1: Random Sample Test Particulars

Sign	VGA Image	Skeleton Understanding	Users	Response time (Seconds)	Result
Hello			1	1.2	Pass
Hi			1	1.3	Pass
Good Morning			1	1.2	Pass

Welcome			1	1.1	Pass

Limitations- The sensor has an optimal range between 40cm and 4M from the sensor. The sensor cannot recognize the human objects beyond this range.
Another major limitation is the Natural Language Processing system that interprets the generated text to speech. Increase in efficiency of the NLP system would intern increase the efficiency of the whole system. The gesture tracking is limited only to 2 individuals.
I.e. Commands to the system cannot be imposed by more than two from n number of individuals.

6   CONCLUSION

The paper is not only aimed at converting the sign language into voice, it's well known that inability to speak and hear is one major challenge for human race. To overcome these disabilities there are a lot of research and development going in different fields. The paper is aimed to minimize the major complexions in the system for further extensions, the sensor comes with the feature of face recognition and voice recognition and therefore the next phase of the project would be adding face recognition for capturing the expressions which in turn increases the productivity of the application by adding a little more accuracy. Also for people with partial voice disabilities the speech recognition system will do the further enhancement in speech systems for the disable people.

REFERENCES

[1] Jungong Han, Enhanced Computer Vision with Microsoft Kinect Sensor: A Review, IEEE TRANSACTIONS ON CYBERNETICS.
[2] Microsoft Kinect SDK,http://www.microsoft.com/en-   us/kinectforwindows/.
[3] Gunasekaran. K, Manikandan. R, International Journal of Engineering and Technology (IJET): Sign Language to Speech Translation System Using PIC Microcontroller
[4] RiniAkmeliawatil, Melanie PO-LeenOoi et al,"Real-Time Malaysian Sign Language Translation using Color Segmentation and Neural Network. Instrumentation and Measurement Technology Conference Warsaw", Poland.IEEE.1-6,2007
[5]Yang quan,"Chinese Sign Language Recognition Based On Video Sequence Appearance Modeling",IEEE.1537-1542,2010
[6]Wen Gao and GaolinFanga,"A Chinese sign language recognition system based on SOFM/SRN/HMM. Journal of Pattern Recognition".2389-2402,2004
[7]Nicholas Born.,"Senior Project Sign Language Glove", ELECTRICAL ENGINEERING DEPARTMENT. California Polytechnic State University,1-49,2010
[8]Kirsten Ellis and Jan Carlo Barca."Exploring Sensor Gloves for Teaching Children Sign Language. Advances in Human-Computer Interaction".1-8.,2012
[9]Yun Li, Xiang Chen et al,"Sign-Component-Based Framework for Chinese Sign Language Recognition Using Accelerometer and sEMGData",IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 59, NO. 10,2695-   2704, 2012
[10] Kinect camera, http://www.xbox.com/en-US/kinect/default.htm.