Interactive Sign Language Interpreter using Skeleton Tracking

A. Muhammad[1], M.F. Addenan[2], M.M. Latiff[3], B. Haris[4], S.S. Surip[5], A.S.A. Mohamed[1]
[1, 2, 3]School of Computer Sciences, Universiti Sains Malaysia,
11800, Penang, Malaysia.
[4]Faculty of Business and Management, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.
[5]School of The Arts, Universiti Sains Malaysia,
11800, Penang, Malaysia.
sufril@usm.my

***Abstract*—The aim of this paper is to introduce an interactive communication system that will benefit both people with hearing and verbal difficulties to convey in the form of the sign language naturally. The idea is to provide two ways of communication between two users by converting sign language to voice and text and provides means of returning communication feedback whereby the other party can speak or key-in text and translates it into sign language movement performed by a three-dimensional (3D) model. A Microsoft Kinect device is used to captures the sign movements by optimizing the skeleton tracking algorithm to understand specific hands movements and dictates using the pre-recorded gesture library to digitized voice and using the same apparatus, speech is translated back into sign language. Research leads in helping the disables have been carried out extensively and majority focuses on only using single type of motion sensing technology such as TOBII (eye tracking) and LEAP (leap motion) which are either costly or limited to a small workable space. Microsoft Kinect technology would be a genuinely equipment used to create a cost-effective and capable technology prototype that enables sign-language communication between signer and non-signer, thus, offers translation into Bahasa Malaysia text.**

***Index Terms*—Kinect; Language Translator; Motion; Sign Language; Skeleton Tracking.**

I. INTRODUCTION

In this world, there exist people who loss the hearing and speaking abilities, and a sign language is the primary form of communication. Sign language exists in various forms for hundreds of years and is literally as other language, has accents and linguistic components [1,2]. There are many approaches for translating and interpreting sign language to aid in communication between the signer and non-signer. The common one for example are the sign language interpreter, voice-to-text transcription and the conventional pen and paper. Each of these method has its own pros and cons in terms of cost and convenient [4-6].

Sign language is based on gestures instead of using spoken words. However, unlike other existing communication methods, the number of disable people who use sign language continues to grow and in general, the number of normal people who are literate with the sign language are getting lesser, which proves to be a barrier for both parties to communicate [5].With the intensive research on the robustness of computer and animation technology application in three dimensional (3D) games tracking algorithms, joint tracking algorithm was introduced to track body parts and joints [7] in real-time. Since the tracking of joints is advantageous in a character signing for animated 3D movement, most 3D games record the movement of joints in the database. This recorded data will be applied for changing of character's position and movements [8] and display quickly.

This paper is based on the techniques to develop using the Kinect sensor to automatically track the hands movement and facial gesture to aid in communication between the signer and non-signer. This paper also represents a discussion on Kinect sensors and skeleton tracking algorithm. Our experiments show that the differential structure proposes the process for developing the prototype to detect the movement of the body parts and joints with a Kinect sensor and these functions are capable enough to be used as a standard deterministic optimization method, which the computational requirements are much smaller than those of probabilistic ones and can nevertheless yield very good results even in difficult situations.

This system could even open up more job opportunities at tourist confirmation counter for people who are deaf. Imagine an information kiosk, say, at an airport, and rather than the person seeking information being deaf, imagine that the person staffing the information kiosk was deaf as shown in Figure 1. Now, a non-signer could come to that kiosk and ask questions and could use the system to help them communicate. This system can also be implemented in the universities itself, for instance, in the lecture halls in Bahasa Malaysia to break the boundaries between the lecturers that were deaf or mute and the students. It can also be applied to teleconferencing via Sign Language Interpreter System to enhance the system for users to fully utilize the facilities to the maximum level.

II. RELATED WORK

Detection of the joints of human movement and motion of body parts is still a major problem because the depth of the human body cannot be determined through the use of conventional camera. There are some innovative ways [9] the

Microsoft Kinect are being used for instance to produce high-quality 3D scans, providing a low-cost home rehabilitations solution or stroke victims or categorized as a disabled person and also for surgeons to use gestures to manipulate images via the Kinect. Furthermore, there has been several studies related to this problem, a group of dedicated researchers from United States had created the Kinect Sign Language that understands the gestures of sign language and converts them to alphabets [10].

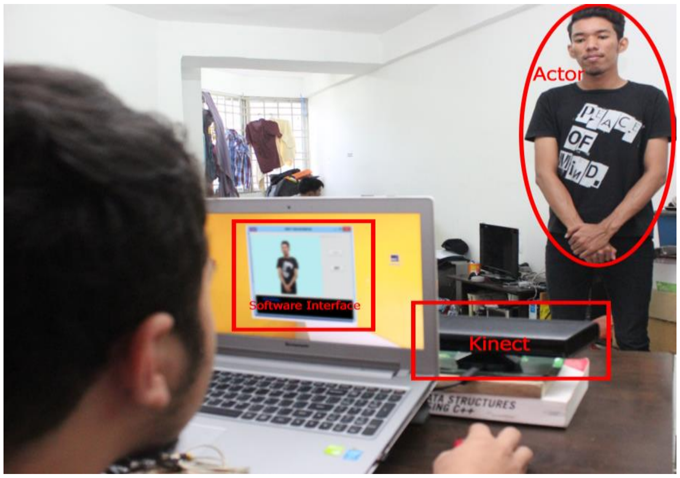

Figure 1: Examples of possible usage of Kinect for sign language interpreter

III. METHODOLOGY

The proposed Sign Language Interpreter would be a two ways communication for both the signer and the non-signer. Using the Natural User Interfaces (NUI) li-brary which comes with the standard SDK detects the presence of humans in front of the sensor. Kinect itself can view up to 4 peoples and accurately track the movement of them. When Kinect track a person, it then provides a skeleton made up of key points on the user as shown in figure 2. A deterministic optimization scheme using the real-time data to reduce the computational cost and better accuracy.

This skeleton feature may detect the basic scenario of using the *Skeleton Stream* where in default mode or seated mode, and displaying an updated visuali-zation 30 frames per second. Once detected, the further process is continued. The first process is pre-composing. Using the exact Skeletal Tracking Algorithm re-leased by Microsoft, the body movements are detected and categorized into vari-ous basic human actions and role such as walking, running, jumping, kicking and more. The second process is information management. In earlier process, the sys-tem already composed by some information of basic human action and role based on body movement.

This process involves the interaction between the administrator and the data-base itself. The administrator can also update the database by being a subject himself. As for the normal subject, they do not have the privilege to access the database. So, any action that they do in front of the Kinect will be captured by implementing the algorithm mentioned above and it will be categorized into any basic human activity and role that match with the pattern. Those patterns are stored in the database, while the administrator can converse natively using speech and translated into gesture using avatar showing the relevant sign language (see Figure 3).

The third process is post-composing. As stated before, the system can detect any basic human activity and role. For instance, the system not only can detect one action, but it can also detect more than one at the same time (e.g. Waving hand while running and punching while kicking). All of the action will also be stored in the database platform. The tracking algorithm developed over Natural User Interface (NUI) and by manipulating the algorithm, the system can be configured to begin with a signer stand in front of effective area of Kinect coverage and do the gestures of Sign language while for the non-signer the system can capture speech and translated into avatar performing relevant sign language movements, which works like a two way communications (see Figure 4).

Next the application interprets the movement through an appropriate algorithm as shown in figure 5, showing the output of the interpreted gestures in the form of text and voice to be displayed to the non-signer.

Figure 2: Joints tracking of a person constructed into a number of skeleton lines

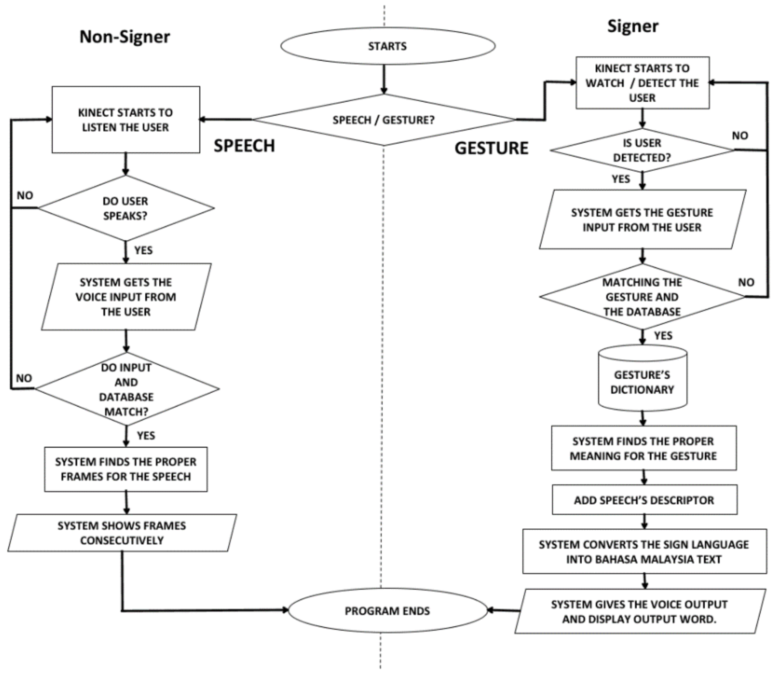

Figure 3: Kinect sign language interpreter flow chart

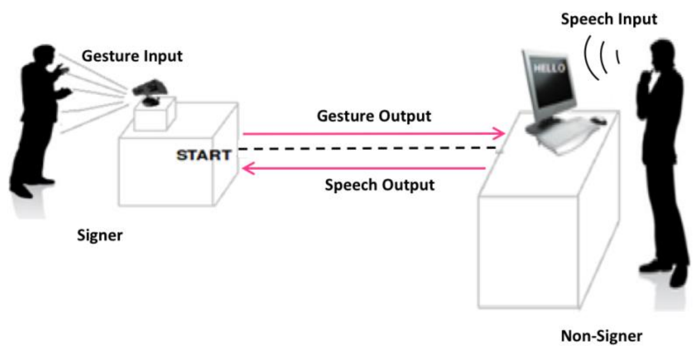

Figure 4: Two-ways communications between the signer (deaf and mute) and the non-signer

IV. Discussion And Conclusion

Kinect rapidly expanded to PC's and became a platform for programmers and companies to develop new and innovative products. Our project will focus more on joint skeleton tracking technique using all the sensors available within Kinect. The main objectives are to enable approximate motion of body parts to be cap-tured, reconstructed and displayed the joint skeleton in the virtual environment and surroundings to comply with the market demand, such as motion capture, fa-cial recognition, voice recognition and so forth. These are just a few of the many features Microsoft made available to everyone with the release of the Kinect and the Kinect SDKs.

The use of devices other than Kinect could have been considered, being that probably most of the research will be done based on its RBG camera. But the fact that Kinect is able to integrate all these different technologies in one small device, making it a very powerful tool. Therefore, and by taking into consideration all the features and possibilities Kinect offers, we believed that this device will be suitable in completing our proposed project. We also believed that if the finding of an integral of Kinect is successful, besides being a suitable device for our experiment, it will be a device that, due to its characteristics, will allow researchers to go much further in their investigations, enabling them to explore new and different areas.

References

[1] William C. Stokoe, Dorothy C Casterline, and Carl G Croneberg. "A Dictionary of American Sign Language on Linguistic Principles" Silver Spring, Md. New Edition [Linstok Press], 1976.

[2] William C. Stokoe. " Sign Language Structure [microform] " Scientific Washington, D.C. Journal, [ERIC Clearinghouse], 1978.

[3] Yin-Poole, W. (2010). Source: MS quadrupling Kinect accuracy. Eurogamer. Re-trievedfromhttp://www.eurogamer.net/articles/2010-12-17-source-ms-quadrupling-kinect-accuracy.

[4] Kathryn Woodcock, Steven L. Fisher. " Occupational Health and Safety for Sign Lan-guage Interpreters " Workplace Safety and Insurance Board Research Advisory Council Grant, [Ryerson University], 2008.

[5] Frank Huang, Sandy Huang. "Interpreting American Sign Language with Kinect " Journal of Deaf Studies and Deaf Education, [Oxford University Press], 2011.

[6] Cynthia Roy, Melanie Metzger. "Researching Signed Language Interpreting Research through a Sociolinguistic Lens " The International Journal for Translation & Interpreting Research, [Oxford University Press], 2014.

[7] Isard, M., Black, A: CONDENSATION – conditional density propagation for visual tracking. International Journl of Computer Vision 29 (1998) 5-28.

[8] Sidenbladh, H., Black, M.J., Fleet, D.J., : Stochastic tracking of 3D human gestures using 2D image motion. European Conference on Computer Vision (2000).

[9] Choo, K., Fleet, D., People Tacking using hybrid monte carlo filtering. International conference on computer vision, Vancouver, Canada (2001).

[10] Cao Dong, Ming C. Leu and Zhaozheng Yin, American Sign Language Alphabet Recognition Using Microsoft Kinect,Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on (2005).