Novel Approach to Sign Language Recognition Using Kinect – Speak My Actions

Zamin Ali Khan*, Wajahat Muhammad, Muhammad Usman Shahid, Owais Ahmed Parkar, Muhammad Usman Altaf, Syed Ameer Hamza
Department of Computer Science & Engineering, IQRA University, Pakistan.

* Corresponding author. Email: dr.m.zamin@iqra.edu.pk
Manuscript submitted May 10, 2018; accepted July 8, 2018.
doi: 10.17706/ijcee.2018.10.3.248-253

Abstract: The paper discusses Sign Language Translating system using depth and coordinates sensing. It defines the major units of system; the first is based on learning from the impaired person and defining custom text accordingly; the second one is based on speech recognition and converting it into sign language. The third one is predefined American Sign Language translation based on actions performed by the impaired person. The system works completely wireless and detects signs and actions through the sensors placed in the system. The system performs submission on the entire axis of each coordinate and matches it with costs of each sign stored in the system. The sign with the nearest cost gets displayed to the screen and speaks out by the system on the behalf of impaired person.

Key words: Sign language recognition, impaired person, speech recognition, sensors, kinect, gesture recognition, 3D trajectories.

1. Introduction

Traditional ways of developing communication between impaired and normal people were based over several add-ons language to be learned by both the persons whether he is impaired or not. This dependency was easy for the impaired ones to get adopted their self in to it but unfortunately in the normal academic institute we don't have any special courses for this problem which results in un-employment for the people who suffers this natural challenge. Most people use gestures and body language in addition to words when they speak.

The number experiments are already performed by the many researchers to technically eliminate this dependency using machine learning algorithms. The system (Speak My Action) is like one of these systems.

The system is able to learn new signs and sentences from the users easily and speaks out the loud on the behalf of user; simultaneously it can work over the American Sign Language actions to provide ease to the users who are more comfortable with ASL.

The system connects Kinect sensor to get the maximum coordinates of the body and calculates the submission of all three axis (x-axis, y-axis, and z-axis) [1]. It calculates these submissions for every coordinate being detected by the Kinect sensor which results in numerous unique values. These values are then stored in the system with corresponding joint and axis name, and the submission of each joint/coordinate is stored as unique value for each joint in the system.

However the best existing Sign Language systems has limitations, Speak my Actions is a proper system with user and admins roles to manage signs and their meanings accordingly. Admin will have all the access

to create or modify users and then users/admin can add their own signs and meaning of their signs. Admin has all the authority to set voice for user that if user wants to have female or male voice when words are spoken as per requirement. Following are the main functionality and modules of Speak my Actions.

2. Proposed Design

Microsoft Kinect (Fig. 1) is used to preprocessing, body tracking, body recognition, and human activity analysis and hand gesture recognition. Visual information and depth of Kinect sensor opens up new opportunities to solve these problems [2]. The Kinect is connected to the power supply and connected to the system through USB cable using Kinect SDK and perform following processes such as object tracking and recognition, human activity analysis, hand gesture analysis, and indoor 3-D skeleton mapping [3].

As Sign Language has two main approaches, Data Glove and Image Processing [4], Speak my Actions is based on Image Processing to capture and gather human body coordinates.

Fig. 1. Microsoft Kinect version 2.

2.1. Custom Sign Language

In this module users have their own profile where they can add their own sign and meaning of sign respectively. It is not necessary that user had to add any standard sign; they can make their own signs. Users are required to train "Speak My Actions" with signs. Signs in this module are static, which means that actions must be still not moving.

Kinect allows us to locate various body locations like face, shoulders and other structural body parts in real time using depth mapping and 3D trajectory [5]. Coordinates of all body parts are captured with x-axis, y-axis and z-axis when every new sign is added. While standing in front of the sensor it will let you know whether your left hand or right is open or close by showing picture of hand in the screen. You can turn off the speech if you do not want to listen words. By performing different signs screen will show word mapped against the signs and will spoke out that word.

The below is the equation, used to calculate cost of every coordinate in x-axis, y-axis and z-axis.

$$\sum_{i=0}^{n} (Joints_n * 1000)$$

Submission of all axis and producing the result with 1000 will always give a large and unique value so there's a very less chance of two different signs having same value. Value of sign along with the meaning is then stored in the system. When user performs an action, the system will select the closest value with in a range. The more the signs the accurate will be the result.

2.2. ASL Sign System

American Sign Language is a natural language used as a sign language for impaired communities [6]. These signs are more complex than the custom signs because these signs contain continuous actions, gestures and postures. ASL is mostly spoken in North America, West Africa and Central Africa. ASL has approximately 6000 gestures along with finger spelling used to communicate [7].

A tracking algorithm is used to determine the Cartesian coordinates of user's hands and nose using Hidden Markov Model (HMM) algorithm as HMM's have been proven to be very effective in the speech recognition [8]. ASL sign system requires depth sensing to capture the continuous sign of user and a medium of communication with database to recognize the sign.

Eigen values and Eigen vectors are a part of linear transformations. Eigen Value Weighted Euclidean Distance Based Classification Technique was able to recognize two hand gestures with an improved accuracy rate of 97%. Moreover, experiment was carried out with bare hands and computational time was also less thus removing the difficulties faced by use of the hand gloves with sensors [9]. As ASL Module uses standard signs so a number of basic signs are already trained into the system. Training part consists of the following steps (Fig. 2)

- User is required to record a video clip performing an ASL Sign.
- With Visual Gesture Builder mark the action as true where user is performing the standard ASL sign and mark all the other actions as false.
- Save the clip and make build of the program.

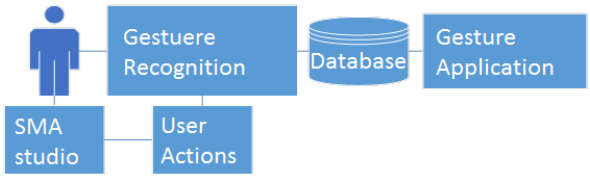

Fig. 2. SMA architecture.

2.3. Speech to Sign Module

The module will be consisting of list of signs stored against the words delivered by the normal person. Which will then, be displayed to the impaired person in the form of sign in order to understand the words being spoken by the person.

In this module, when the normal person will say any word, "Speak My Actions" will convert that word into signs accordingly. These actions will be displayed in "Speak My Actions" where impaired person can see those signs.

2.4. Sentence Builder

After capturing coordinates of impaired person, recognizing the action performed and getting the word mapped against that action from data source, the next step is to make a sensible sentence combining recent words to give a better experience to the end person about what message is supposed to be delivered. This is a core module of Speak my Actions which combines a group of words containing different word classes, let's say verbs, pronouns, nouns, adverbs, adjectives etc., that expresses a thought in the form of statement, feelings, asking's, most probably the idioms and phrases, instructions and the most important thing is to start first word with capital letter of the sentence. As there are different types of sentences (i.e. Affirmative, Negative, and Interrogative etc.), the system may vary tone and probably the nature of next coming word at once, just to express a variety of feelings, emotions and thoughts.

The most important tasks is to get to know the nature of the word whether it is acting like a verb or a noun, adverb or an adjective and the nature of whole sentence is based on selecting and determining the correct nature of word. Moreover, after detecting the nature of word we have to use correct helping verbs to make the statement sounds more smooth and consistent. Since statements in English language are quite simple and structured, we use hard programming tactics to get this work done.

Sentence Builder is a separate module, which we have integrated with our system. Whenever any word is fetched from data source against any action this event is fired to a make statement from words. Information and characteristics of words are stored in database along with their classes.

2.5. Speech to Text

The module is completely based on converting the speech from the normal person to the generated grammar text; the module focuses over understanding of human language and converting it to written text or statement, which supports other modules in their proper functionalities.

The module is input from the normal person interacting with the system, which helps him to communicate information to the impaired person. Person interacting with the system provides input in the form of voice through Kinect Sensor, which then is evaluated by the algorithm, and respective text is generated based on the evaluation.

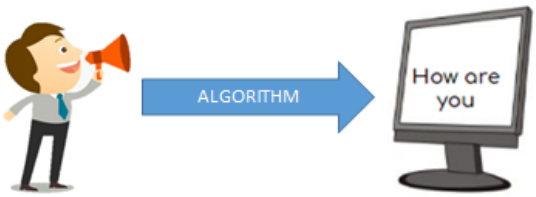

Fig. 3. Speak my actions.

The algorithm starts with initializing the language dictionary for comparing the voice input. The algorithm uses the Kinect Sensor voice input and detects the person speech if it has arrived. After detecting the speech algorithm identifies the best match word against the speech and then it displays the best matched word to the screen and sends it to the next module of the system.

Speak my Actions (Fig. 3) is a novel approach towards Sign Language Recognition and American Sign Language. There are many sign language projects that work on HMM and Neural Networks. Moreover Sign Recognition System has also been developed for Spanish Language using HMM with 91 signs [10]. But Speak my Actions is a complete system with American Sign Language and Custom Sign Language alongside a proper User and Admin Panel to manage things accordingly. Speech to Sign Module is a unique approach in such projects making it a two way communication application and making special communities close to common people.

3. Conclusion

The Speak my Actions reduces the communication barrier between the normal person and impaired community. Perspective of each and every module in Speak my Actions works according to the needs of impaired person and normal person. Whereas the system is completely wireless, it is easy to carry from one place to another. So far in this project, we have used Kinect which works on the concept of depth sensing to get body coordinates of the user. Although Kinect is actually buildup of such microprocessors that would map human skeleton but our major work was done to control Kinect's functionality, make use of it in recognizing human body whether static or in motion and to map different signs with different meanings. Using depth sensing, Kinect calculates the submission of all three axis (x-axis, y-axis, and z-axis) on every other sign. These signs are then stored in the system with respective meanings. Nonetheless, a number of researches have been done on sign language recognition in recent years but Speak my Actions is a unique system. It is a simple, portable system that has all the possible features to decrease the barrier between impaired community and normal person.

References

[1] Wang, L., Kang, S., & Wang W. (2015, November). Kinect-based situated measure design of motor coordination performance. *Academia Journal of Educational Research, 3(11)*, 44-49.

[2] Han, J., *et al.* (2013, June). Computer vision with Microsoft Kinect sensor: A review. *IEEE Transactions on Cybernetics*.

[3] Shotton, J., *et al*. (2013, January). Real-time human pose recognition in parts from a single depth image. *Megazine Communications of the ACM, 56(1)*, 116-124

[4] Gunasekaran K., & Manikandan, R. (2013, May). Sign language to speech translation system using PIC microcontroller. *International Journal of Engineering and Technology (IJET), 5*, 1024-1028

[5] Yang, H. D. (2014, December). Sign language recognition with the Kinect sensor based on conditional random fields. *Sensors, 15*, 135-147.

[6] Strong, M., & Prinz, P. (1997). A study of the relationship between American sign language and English literacy. *Journal of Deaf Studies & Deaf Education, 2(1)*, 37.

[7] Malenfant, H. (2013). American sign language: Culture, community, & identity. Senior theses, Trinity College, Hartford, USA.

[8] Shrott, R. (2017, October 15). Sign language recognition with HMM's feature engineering, model design, implementation and results. *Machine Learnings*.

[9] Parcheta, Z. (2017, June). Sign language gesture recognition using HMM. *Pattern Recognition and Image Analysis*.

[10] Singha, J., Das, K. (2013, February). Indian sign language recognition using Eigen value weighted Euclidean distance based classification technique. *International Journal of Advanced Computer Science and Applications, 4*, 188-195.

M. Zamin Ali Khan is currently a head of Computer Science and Computer Engineering Department at IQRA University. He has received the B.E. in electrical engineering from NED University, Karachi, Pakistan and the MS in electrical and computer engineering from Concordia University, Montreal, Canada and the PhD in computer science. He has more than 18 years of experience of teaching and industry. He had worked in Victhom Human Bionics, Canada as an engineer scientist. Currently. He is a senior member of Pakistan Engineering Council, Canadian Engineering Council and IEEE. His research interest includes VLSI, digital design, and digital signal processing and analog front end of wireless devices.

Usman Shahid is currently a head of Technology and Innovation Department at Hudson Pharma. He has received the bachelors in computer science (BSCS) from IQRA University, Karachi, Pakistan. He has more than 3 years of experience in software industry. During his career he has designed numerous complex algorithms including brain signals processing, big data handling and sign language processing using depth sensors. He has also worked as a business intelligence developer and designed various data models for multinational food and pharma companies.

Owais Ahmed Parkar is currently a senior system analyst at Human Resource Service Global BPO. He has received his BSCS (bachelors of computer science) from IQRA University, Karachi, Pakistan. He has strong skills to administer an SQL database and all the task that entails. He dreams to pursue his career in IT or data science.

Syed Ameer Hamza is currently working as software engineer at Vortechs Innovations. He is graduated from IQRA University and is pursuing master's degree from SZABIST University now. Ameer Hamza developed different web applications during his job career. His technology are stack node JS, angular JS, PHP, LARAVEL, MySQL, and mongo DB.

Muhammad Usman Altaf is currently a database engineer at Sybrid Private Limited. He has received his BSCS (bachelors of computer science) from IQRA University, Karachi, Pakistan. He has strong skills in database world. He is works in different technologies like SQL, MySQL, SSRS, SSIS, C#, and ASP.NET.

Wajahat Muhamad is currently running his own e-commerce business. He has received his BSCS (bachelors of computer science) from IQRA University, Karachi, Pakistan. He has skills in C# and Database. He has formerly worked as business development executive at Axact.