Sign Language Interpreter using Kinect Motion Sensor using Machine Learning.

Amol Potgantwar, Pragati Bachchhav

Abstract—Sign language is the route through which deaf and dumb people usually communicate with one another. It is seen that, impaired people find a difficulty while interacting with the society. Normal individuals are not able to understand their sign language. To bridge this gap, the proposed system acts as the mediator between impaired and normal people. This system uses Kinect motion sensor to capture the signs. The Kinect motion sensor captures 3 dimensional dynamic gestures. Thus this study is proposed for extracting features of dynamic gestures of Indian Sign Language (ISL). As American Sign Language (ASL) is popular and is used in research and development field, on the other hand ISL has been recently standardized and hence ISL recognition is less explored. The propose method extracts feature from the sign and converts it to the intended textual form. The method then integrates local as well as global information of the signs. This integrated feature leads to improving the system performance, the system serves as an aid to disabled people. Its application includes hospitals, government sectors and some multinational companies.

I. INTRODUCTION

Humans have one of the most important abilities which is communication. The process of communication starts from the birth of child where it starts to require things and express feelings. At least two humans must be involved for communication to happen. Furthermore they must use the same rules for naming and describing things. These rules of communication form the language. Oral communication is the most common form of communication.

It is based on speech and hearing, but when it comes to impaired people use a form of sign language. Sign language is the way through which deaf and dumb people can communicate with each other. A normal individual while observing a conversation in sign language might think that the conversation is very weird, comic or sometimes even dramatic. But the fact is that a normal person will never understand what the signs mean.

There are various transitions or phases of the sign language. A particular word is used in north India has a different name in south India but the meaning of those is same just the dictionary is different. This is common all over the world. In short words may vary but the meaning is the same. Sign language is mother tongue or the first language for the impaired individuals. The population of India is about 1.35 billion and is second most populated country in the world. The estimates tell us that there are over 1.5 million individuals who are deaf and dumb which is a huge number as compared to other countries.

II. SIGN LANGUAGE MYTHS AND FACTS

A. SIGN LANGUAGE MYTHS:

- Sign language is universal.
- Sign language only uses your hands.
- Sign languages are very easy to learning.
- It was invented by hearing people.
- To express the hidden meaning is not possible in sign language.

B. SIGN LANGUAGE FACTS:

- Sign languages have their own definite grammar.
- Sign language doesn’t only use signs to communicate. It uses hand movement, facial expression, body language and gestures to communicate.
- Sign language varies from country to country.
- Each sign is composed of five components. Any change in them will change the entire meaning of the sign.
- Deaf children born to normal parents need to learn sign language.

III. RELATED WORK

In [3] the Kinect device is used to develop a system which interprets the sign language into human understandable language. The study conducted experiments where the actions were done by a normal person. These actions performed by the person were a set of signs performed in a sequence which consisted of...
various hand movements and its location. As there are many languages in the world this system is specially designed for the Kosova Sign Language. They tested four parts: alphabets, numbers, four words and a sentence. The developed system was tested by one normal person and two impaired person. Thus, the system gave high accuracy and performance.

There are many systems that have been developed based on various technologies and algorithms. Some of them use Hidden Markov Models or others use Artificial Neural Networks classifiers. Here, XBOX 360 is used to solve all the problems that were faced in the previous algorithms. Basically it is a hardware which tracks all the actions and then sends it for training the data. It is very commonly used by people for playing games and performing other activities.

So, it has been put to further use in today's world of technology where everyone wants to perform tasks from one pacle with one click. It has a RGB-D camera which has three colour range used that is red, green, blue because other colours are formed using these basic colours. The system also verifies of giving a better performance than the traditional methods used [4].

In [5], the authors have developed a way in which we can align ASL video subtitles. This is done with the help of closed caption transcript. They have also used a discriminative model instead of generative model. Instead of using many words the system is using few words which give a high efficiency of the system. This helps to lessen the evaluation time of the words and gives output in a quick manner. First the system recognises the words or the sign language. Then alignment of the words is done in a proper understandable manner. After this the best sentence suited will be displayed which will have the exact or close meaning that of actions were performed as an input.

In 1991 Tomoichi Takahashi developed a system for real-time SLR that required from signer wearing of gloves. Gloves were connected with wires to the computer for transmission of hand configuration and joint angles [8]. HMMs can be used for online training mode. This was demonstrated in 1996 in a system that employed wired gloves for feature extraction and HMM for gesture recognition [9]. CyberGlove with 18 sensors and connected to computer through serial cable was used for transmission of 20 hand joints [9]. The system recognized 14 letters from sign language alphabet while training of the system with one or two examples was enough for sign recognition [9].

A wireless glove that was designed by Ryan Petters in 2002, sensed hand movements involved in sign language and then transmitted them wirelessly to portable device, that displayed translated signs as lines of text [10].

A vision based approach was followed in [11] where feature extraction was done using color camera for recognition of ASL. Two experiments were run: the first required from the signer to wear colored gloves, while the second used hands skin color for feature extraction [11]. The extracted features were used as input for HMM algorithm that was used during sign recognition phase. The first experiment attained 99% word accuracy while the second 92% [11].

Different from previous works that were concentrated on extracting only manual features of sign language, the work by Kelly et al. [12] involved also non-manual features for continuous SLR. Features were extracted by single camera that tracked users hands and head. The user was required to wear colored gloves for better feature extraction. Extracted features were used as input to HMM for sign recognition. The system was tested with 160 video clips of fluent signer performing full un-segmented sentences. The achieved detection rate was 95% [12]. Tracking of non-manual features for SLR is done also in SignSpeak project [13] where face expression and head movements were extracted using vision based approach.

IV. PROBLEM STATEMENT

To design and develop a system with the use of Kinect Motion Sensor for impaired people which will help them to communicate with ordinary people and to also interact with the society by recognizing and interpreting sign language in to speech or text and vice versa. About 80 million deaf people use sign language as their mother tongue or first language. Many hearing people and some deaf blind people have sign language as the first language and mother tongue. Society treats the impaired people very differently and they hesitate to live a normal life. So, this project is a step towards solving their problems so that they can live a normal and happy life.

V. PROPOSED SYSTEM

A. Methodology

Sign language is a game of expressions which has a particular gesture attached to it. Every single word has a gesture related to it. Hence to form a meaningful sentence these gestures need to be combined which gives a meaning to the conversation. In the proposed system every gesture is interpreted word by word.

Here, there is a facility where the user can train the system to learn the gesture for new word. Like this many gestures and words will be stored in the training set of the database which will be used on a daily basis by the user.

While using the system user has to stand in a optimal position from the kinect sensor for the proper view and working of the system. Then, the initial position of human will be detected based on the depth image.

After the image is captured system will extract the joint points and normalize the points. For each sequence Si, in database it will calculate Euclidian distance Si with captured sequence. The output will in audio or text format that a normal user will understand.

Now, when a normal individual will communicate to the impaired individual the process is vice versa. The user will give audio as input that is he will speak and the data will be captured.

Then this speech will be matched with the trained data in the database and similar symbols and images will be searched for. Then, the whole speech will be interpreted to the impaired individual in sign language that he will understand.

This system is a step towards solving the problem of the impaired people and gives them a chance to live a better and normal life without any discrimination.

B. System Architecture

There are two modes in this system:

Mode 1 : Sign to Text/Audio Conversion

In this mode we have:
i. Skeleton Recognition
ii. Joint Recognition
iii. Distance Calculation
iv. Maintaining Dynamic Dictionary

Mode 2 : Audio to Sign conversion
In this mode we have:
i. Animated Sign Actions
ii. Audio to text conversion
iii. Text to sign conversion

The idea is develop a interpretation system using kinect sensor that will help the communication between the normal person and an impaired person to be better and smoother. It is like the events that take place in day to day life.

For instance two people from different countries are going to meet one another it is very obvious that the languages will be different so they will need a human interpreter who knows both the languages. He will interprete what they want to say to one another and there will be no difficulty. This concept is shown via figure 4.

Three components are involved during the process: the hearing impaired person, normal hearing person and module that consists of Kinect and a computer. The hearing impaired person will perform some actions in front of the Kinect and then Kinect will track the actions performed by hearing impaired person and transmit the sign information to the computer. Computer will analyze the actions that an impaired person performs and interprets its meaning as speech to normal hearing person.

C. Algorithm
According to the system there are various steps so that we can get the most accurate data. First will be the skeleton recognition and in that skeleton there will be joint recognition. When the user performs some actions they are normalized and data is trained and stored in the database. There are different algorithms used in this system as it will not work only on one algorithm.

1. Sign Dictionary
Here, we have a sign dictionary (database) where all the signs performed by the impaired person are stored. When the data is obtained and normalized, then we need to have a descriptor or token for each and every sign. The token should be able to depict in such a way that each and every sign is different from others and the data does not collide.

This is shown in the figure 5 that when the man is performing actions they are stored in the dictionary. The first action depicts good now every time a person does the action for the word good then the proper word should only be displayed. This, is done to avoid any loss of action or word and a proper communication can be done without any barrier and confusion.
3. Dynamic Time Warping algorithm (DTW)

DTW is introduced in very early in time period. Sometimes, same action is performed with different velocity. So, all the actions with its time frame are compared and a mean and proper data is displayed by traing the data properly. This does not cause any omission of words or actions and the system is dynamic.

**VII. RESULT ANALYSIS**

The analysis is based on the signs that are performed and the words that are said and stored in the system. The accuracy is calculated by the algorithm gives the nearest answer. The visuals, graphs and diagrams gives the idea of the system. Now in figure 6 the person is standing in front of the kinect sensor and performing some actions in this case it is the HELLO word. As, the action is performed it is recorded and stored in the system and then we can name the action as hello. There are different ways in which this action can be performed, hence the system takes in the nearest possible to the stored action and gives the output accordingly. This figure demonstrates the first module that is the sign to audio translation.

**VI. MATHEMATICAL MODEL**

A. Sign to audio conversion/ impaired person to normal person:

Let S be the solution
So, let S = { I, F, O }
where, S = Set of inputs, functions, outputs
Now, I is the set if inputs
i.e., I = { I₁ }
Let,
I₁ = gestures/signs/actions
Now, F is the set of functions
i.e., F = { F₁, F₂, F₃ }
Let,
F₁ = record action
F₂ = normalization
F₃ = gesture classification

Now, O is the set of output
i.e., O ={ O₁ }
Let ,
O₁ = audio/speech

B. Audio to sign conversion/ normal person to impaired person:

Let S be the solution
So, let S = { I, F, O }
where, S = Set of inputs, functions, outputs

Now, I is the set if inputs
i.e., I = { I₁ }
Let,
I₁ = speech/audio
Now, F is the set of functions
i.e., F = { F₁, F₂ }
Let,
F₁ = speech recognition
F₂ = speech to text conversion

Now, O is the set of output
i.e., O ={ O₁ }
Let ,
O₁ = display sign

**Fig. 6 Demonstration of HELLO word**

Now, the figure 7 demonstrates the audio to sign interpretation.

**Fig. 7 Demonstration of word : EXCITED**

Here, figure 6 and 7 look the same but there is a difference. In this window we speak a word for instance EXCITED so this word will be processed and then the output will be the sign that is stored for the word 'excited'.
As, the body is detected in front of the kinect, it works in co-ordinates like x-axis, y-axis and z-axis. This is the simplest way as it detects the joints like the elbow, shoulder, wrist, etc to give the most accurate output. In the graph there are four words as input and their accuracy is also shown. Now for word hello we use our elbow and wrist therefore it has 2 axis x and y. Same goes for the word bye. But, for the word 'what' all the 3 axis are used. When testing was done it is seen that what word has the higheest accuracy. There are many components that can be studied in sign language as a character, number, word and sentence. In this system the word and sentence accuracy is the component that are focused on.

VIll CONCLUSION

We, got to know that there are many myths regarding sign language which are not true and have facts that reveal the truth about sign language. It is commonly seen that the impaired are treated differently in our society as people think that if they are disabled the they cannot live like a normal human being. But, the world is fair to all and many ways and technologies are coming up to help these people so that they can lead a normal life. In the proposed study, solution to one of the problems of the deaf and dumb people is presented. These people communicate in sign language and a normal human cannot understand it. Here, is an interactive system which is developed with the use of kinect motion sensor to perform sign language interpretation. The signs is a combination of hand movements, shape, location, etc. which is captured by the kinect and then the actions are trained and stored. When proper recognition method is followed then continuous communication is done without any confusion. Even if the system is new but a new user will not find it difficult to work with it. The system is developed for dynamic and uninterrupted communication between the normal user and impaired one.

REFERENCES

1. Shihab Shahriar Hazari, Asaduzzaman, Lamia Alam and Nasim Al Goni, “Designing a Sign Language Translation System using Kinect Motion Sensor Device,” in International Conference on Electrical, Computer and Communication Engineering (ECCE), February 16-18, 2017, Cox’s Bazar, Bangladesh.

2. Pooja Gupta, Dr. Ambuj Kumar Agrawal, Dr. Shahnaz Fatima, “Sign Language Problem And Solutions For Deaf And Dumb People,” at 3rd International Conference on System Modeling & Advancement in Research Trends (SMART).

3. Eden Mustafa, “Sign Language Interpretation using Kinect,” at Conference Paper, September 2014.

4. D. M. Capilla, “On Sign Language Translator Using Microsoft Kinect XBOX 360,” at M. Sc. Thesis, University of Tennessee, Knoxville, USA.

5. A. Farhadi, D. Forsyth, “Aligning ASL for statistical translation using a discriminative word model,” University of Illinois.

6. S. Young, “ HTK: Hidden Markov Model Toolkit VLS,” Cambridge Univ. Eng. Dept., Speech Group and Entropie Research Lab. Inc.,Washington DC.

7. P. Mekala, Y. Gao, J. Fan and A. Davari, “Real-time sign language recognition based on neural network architecture,” in Proc. 2011 IEEE43rd Southeastern Symposium on System Theory, Auburn, AL, 2011, pp. 195-199.

8. T. Takahashi and F. Kishino, “Hand gesture coding based on experiments using a hand gesture interface device,” ACM SIGCHI Bull., vol. 23, no. 2, pp. 6774, 1991.

9. C. Lee and Y. Xu, “Online, interactive learning of gestures for human/robot interfaces,” in Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on, 1996, vol. 4, pp. 29822987.

10. T. K., “Glove lends the deaf a hand,” USA Today, 2002.

11. T. Starner and A. Pentland, “Real-time american sign language recognition from video using hidden markov models,” in. in Motion-Based Recognition, Springer, 1997, pp. 227-243.

12. D. Kelly, J. Reilly Delanney, J. Mc Donald, and C. Markham, “A framework for continuous multimodal sign language recognition,” in Proceedings of the 2009 international conference on Multimodal interfaces, 2009, pp. 351358.

13. P. Dreuw, J. Forster, Y. Gweth, D. Stein, H. Ney, G. Martinez, J. V. Llaha, O. Crasborn, E. Ormel, W. Du, and others, “Signspeak—understanding, recognition, and translation of sign language in video,” in Proceedings of 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, 2010, pp. 2223.

14. J. Atwood, M. Eicholtz, and J. Farrell, “American Sign Language Recognition System,” Artificial Intelligence and Machine Learning for Engineering Design, Dept. of Mechanical Engineering, Carnegie Mellon University, 2012.

15. L. Pigou, S. Dieleman, P.-J. Kindermans and B. Scharuwuw, “Sign Language Recognition Using Convolutional Neural Networks,” in proc. European Conference on Computer Vision, 6-12 September, 2014

16. D. Aryanie, H. Heryadi, “American Sign Language-Based Finger-recognising Spelling Recognition using k-Nearest Neighbors Classifier,” in Proc. 3rd International Conference on Information and Communication Technology, 2015, pp. 533-536.

17. R. Sharma, Y. Nemani, S. Kumar, L. Kaneand P. Khanna, “Recognitonof Single Handed Sign Language Gestures using Contour Tracingdescriptor,” in Proc. World Congress on Engineering (WCE 2013), vol.2, July 3 - 5, 2013, London, U.K.

18. D.Martinez, “MSc Thesis - Sign Language Translator using MicrosoftKinect XBOX 360 TM,” VIBOT 5, Department of Electrical Engineering and Computer Science, Computer Vision Lab, University of Tennessee.

19. Pavlovic, V.J. and Sharma, R. and Huang, T.S., “Vision-Based Gesture Recognition: A Review”, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997.

20. Wu, Ying and Huang, Thomas S., “Visual interpretation of hand gestures for human-computer interaction: a review”, at Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction, 1999.

21. D. M. Gavrila, “The Visual Analysis of Human Movement: A Survey,” at Computer Vision and Image Understanding, 1999.

22. Yang Quan and Peng Jinye, “Application of improved sign language recognition and synthesis technology in IB”, Industrial Electronics and Applications, 2008. ICEA 2008. 3rd IEEE Conference on.

23. Dai Jun, “Gesture Recognition Reach Based on High-Order NMI,” Master Dissertation, ShanHai Maritime University, 2004.

24. Yang Quan, “Sign Language Letter Recognition Algorithm Based on 7 Hu Invariant Moments”, Master Dissertation, Xi’ans University of Architecture and Technology, 2007.

25. Thad Starner and Joshua Weaver and Alex Pentland, “Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998.

26. J. L. Raheja, A. Chaudhary, and K. Singal, “Tracking of Fingertips and Centers of Palm Using KINECT”; 2011 Third Int. Conf. Comput. Intell. Model. Simul., pp. 248–252, Sep. 2011.
Sign Language Interpreter using Kinect Motion Sensor using Machine Learning.

AUTHORS PROFILE

Dr. Amol Potgantwad has completed Ph.D and is H.O.D of Computer Department, Sandip Foundation, Nashik. The focus of his research in the last decade has been to explore problems at near field communication and its various applications, in particular interested in applications of mobile computing, wireless technology, near field communication, image processing and parallel computing. He has registered 5 patents on indoor localization system for mobile device using RFID and wireless technology, RFID based vehicle identification system and access control into parking, Standalone RFID card based healthcare system. Since last 5 years he is working with the organization Sandip Institute of Technology and Research Center. His focus has always been on nurturing and guiding the students in different domains and projects.

Pragati Bachchhav has completed B.E in Computer Engineering from Sandip Foundation, Nashik. Has been a part of WSN based Infrastructural Health Monitoring and Audit System project in B.E. Had patent application on the same project which is also recognized by central government of India. She has been working with multinational company named Accenture as Quality assurance engineer. She has good skill set in manual and automation testing domain. Along with, she has served many projects within deadline. She has got many recognitions by her team for working hard and efficient on the different projects.

27. N. Pageault and R. Bowdlen, “Spelling it out: Real-time asl fingerspelling recognition,” Comput. Vis. Work. ( ICCV ... , pp. 1144–1149, 2011.
28. C. Keskin, F. Kiraq, Y. Kara, and L. Akarun, “Real time hand pose estimation using depth sensors,” ... Depth Cameras Comput. ... , pp. 1228–1234, 2013.
29. Y. F. Admasu and K. Raimond, “Ethiopian sign language recognition using Artificial Neural Network,” in Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on, 2010, pp. 995–1000.
30. T. Takahashi and F. Kishino, “Hand gesture coding based on experiments using a hand gesture interface device,” ACM SIGCHI Bull., vol. 23, no. 2, pp. 67–74, 1991.
31. C. Lee and Y. Xu, “Online, interactive learning of gestures for human/robot interfaces,” in Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on, 1996, vol. 4, pp. 2982–2987.
32. T. K., “Glove lends the deaf a hand,” USA Today, 2002. [Online]. Available: http://www.usatoday.com/news/science/2002-01-16-deafglove.htm. [Accessed: 01-Dec-2013].
33. T. Starner and A. Pentland, “Real-time american sign language recognition from video using hidden markov models,” in in Motion-Based Recognition, Springer, 1997, pp. 227–243.
34. T. Starner, J. Weaver, and A. Pentland, “Real-time american sign language recognition using desk and wearable computer based video,” Pattern Anal. Mach. Intell. IEEE Trans., vol. 20, no. 12, pp. 1371–1375, 1998.
35. C. Vogler and D. Metaxas, “ASL recognition based on a coupling between HMMs and 3D motion analysis,” in Computer Vision, 1998. Sixth International Conference on, 1998, pp. 363–369.
36. J. Weissmann and R. Salomon, “Gesture recognition for virtual reality applications using data gloves and neural networks,” in Neural Networks, 1999. IJCNN’99. International Joint Conference on, 1999, vol. 3, pp. 2043–2046.
37. P. Vijay and N. Suhas, “Recent Developments in Sign Language Recognition: A Review,” irindia.in, no. 2, pp. 21–26, 2012.
38. T. Kelly, J. Reilly Delannoy, J. Mc Donald, and C. Markham, “A framework for continuous multimodal sign language recognition,” in Proceedings of the 2009 international conference on Multimodal interfaces, 2009, pp. 351–358.
39. P. Dreuw, J. Forster, Y. Gweth, D. Stein, H. Ney, G. Martinez, J. V. Ll manifestations, 2013.
40. F. Huang and S. Huang, “Interpreting American Sign Language with the Robot Assisted Sign Language,” in Advanced Computer Science and Information Systems(ICACISIS), 2012 International Conference on, 2012, pp. 213–218.
41. F. Trujillo-Romo and S.-O. Caballero-Morales, “3D data sensing for hand pose recognition,” in Electronics, Communications and Computing (CONIELECOMP), 2013 International Conference on, 2013, pp. 109–113.
42. M. Ozszt and M. Wysocki, “Polish sign language words recognition with Kinect,” in Human System Interaction (HSI), 2013 The 6th International Conference on, 2013, pp. 219–226.
43. H. Takimoto, J. Lee, and A. Kanagawa, “A Robust Gesture Recognition Using Depth Data.”
44. SigmaRD, “SigmaRD sign language recognition using Kinect,” Vimeo. [Online]. Available: http://vimeo.com/45073111. [Accessed: 01-Dec-3].
45. B. W. Boehm, “A spiral model of software development and enhancement,” Computer (Long. Beach. Calif.), vol. 21, no. 5, pp. 61–72, 1988.
46. M. Fowler and J. Highsmith, “The agile manifesto,” Softw. Dev., vol. 9, no. 8, pp. 28–35, 2001.