Sign Language Recognition-A Survey of Techniques

Udit Barde, Archana Ghotkar

Abstract*: Sign language is the only method of communication for the hearing and speech impaired people around the world. Most of the speech and hearing-impaired people understand single sign language. Thus, there is an increasing demand for sign language interpreters. For regular people learning sign language is difficult, and for speech and hearing-impaired person, learning spoken language is impossible. There is a lot of research being done in the domain of automatic sign language recognition. Different methods such as, computer vision, data glove, depth sensors can be used to train a computer to interpret sign language. The interpretation is being done from sign to text, text to sign, speech to sign and sign to speech. Different countries use different sign languages, the signers of different sign languages are unable to communicate with each other. Analyzing the characteristic features of gestures provides insights about the sign language, some common features in sign languages gestures will help in designing a sign language recognition system. This type of system will help in reducing the communication gap between sign language users and spoken language users.

Keywords: American Sign Language, gesture recognition, Indian Sign language, sign language recognition, Sign language translation

I. INTRODUCTION

Sign language is a form of visual communication for speech and hearing-impaired people which involves the simultaneous movement of hand gestures and different body parts. The sign languages are complete languages which have their own grammar and vocabulary, and are Sign language consist of manual and non-manual components. Manual components include the movement of body parts like hand, head, etc. The hand movements include combinations of hand shape, hand orientation with respect to other body parts. Non-manual components include the movements of eyes, lips, etc. Facial features are also crucial in interpreting emotions or questions in sign languages. Some signs in sign languages use both hands for depicting sign, whereas some signs can be simply depicted by using single hand. The right hand is considered as the dominant hand and left hand is non-dominant hand. It is difficult for deaf people to access the information available in any form other than gestures because of their language problem and hence this hinders their normal social life. Like there are different spoken languages in the world, there are different sign languages in different countries. They are developed independent of the spoken language in a particular region. For example, the people in Britain and USA speak the same language i.e. English, but still British Sign Language (BSL) and American Sign Language (ASL) are not similar. Since there is no universal sign language developed which can be used by deaf people around the world, different sign languages are used in different countries or regions. The hearing-impaired person in America who know ASL may not understand BSL. Some countries adopt features of American Sign Language into their sign languages, but the actual signs are different. The American Sign Language is a completely distinct language containing all the important features like rules for pronunciation, formation of words, and word order. The difference between Sign languages and spoken languages is how gestures are done, like asking a question rather than making a statement. For example, in spoken English, speakers raise the pitch of their voices and adjust word order, while asking a question; ASL users ask a question by raising their eyebrows, widening their eyes, and tilting their bodies forward. [13]. Figure 1. Shows the number of sign language users for each sign languages, and the available registered human interpreters of the language. Sign multilingualism is very rare, there are very less people in the world who can interpret more than one sign language. The number of American Sign Language interpreters is comparatively more than Indian sign language and British sign language. As the number of human interpreters is very low than the number of sign language users, a computer interpreter could help in bridging the gap in communication of the hearing and speech impaired people.

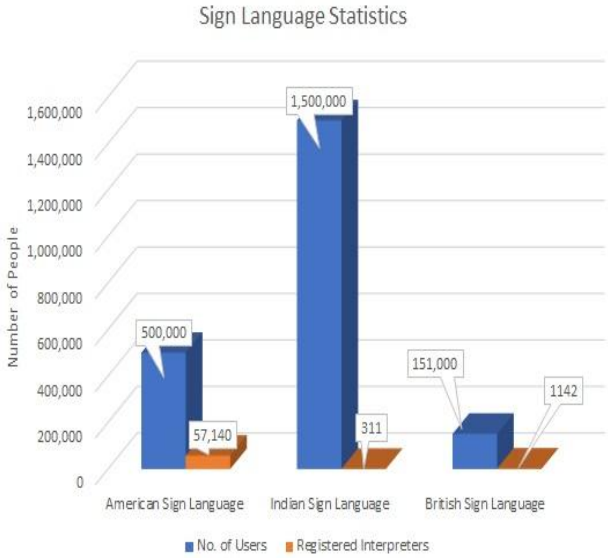

Figure 1. Statistics of sign languages [14][15][16][17]

Revised Manuscript Received on June 22, 2020.

***** Correspondence Author

Udit Barde, Computer Engineering, Pune Institute of Computer Technology, Pune, India. E-mail: uditbarde16@gmail.com

Archana Ghotkar, Computer Engineering, Pune Institute of Computer Technology, Pune, India. E-mail: aaghotkar@pict.edu

II. METHODOLOGY

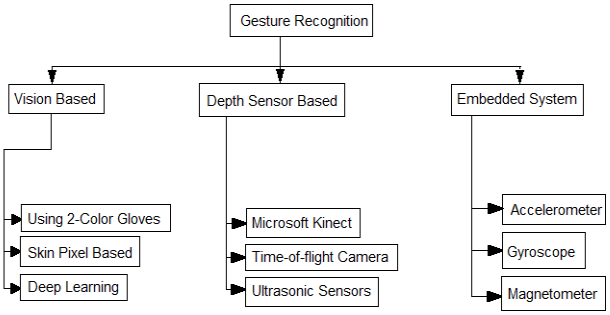

Figure 2. Sign language recognition techniques

Gesture recognition

In this step the gesture made by the signer is captured using different techniques. The insights about the gesture are obtained using various techniques and gesture is predicted. The insights about gestures include the body pose data, hand pose data, hand shape, hand orientation, hand location with respect to the body, and hand movement. After predicting the gesture correctly, it can be translated into intermediate language.

Gesture recognition techniques

Vision Based: As the name suggests, the vision-based sign language recognition techniques require processing of data in the form of images, videos. Input is provided using different cameras, which include gestures, signs made by multiple users. Data is collected stored for further processing. The gestures made by users need to be identified by the computer for translating into desired form. To identify the gestures first hand segmentation is done. The features used in gesture recognition can be:

- Shape representation: contours, chain code, polygon, compactness, B-spline, Fourier descriptor, etc. [1]
- Region based: Area, Euler number, convex hull, geometric moments, etc. [1]

After feature extraction from preprocessed gesture images, recognition can be done using different supervised & unsupervised machine learning techniques.[1]

- Supervised: K-nearest neighbors, naïve Bayes, neural networks, support vector machines.
- Unsupervised: K-means clustering, Fuzzy K-means, Mutual neighborhood, etc.

Vision based gesture recognition can be done using techniques like glove based and pixel based.

Glove based: In this method a glove of specific color is worn by the signer while performing sign in front of camera. Both hands can use single colored glove or can use different colors for different hands for different applications. Advantage of this method is, it is easy to implement, and hand detection can be performed by segmenting the color of glove from frames. Finding proper color of glove for available camera is challenging, complex gestures cannot be recognized, hand orientation is difficult to identify.

Pixel based: In this method the hand segmentation is done based on the skin color of humans performing the sign in front of the camera. No extra glove is required for detecting the hand from images. Advantages of this method are, this method is more robust, can be applied to direct videos, but identifying accurate skin color is a challenge. Humans have a lot of variety in skin color, if the dataset contains data from a large group of people belonging to different ethnicity, finding skin color is difficult. Hand orientation can be identified by identifying skin color of palm and back of palm. Raheja, J.L., Mishra, A. & Chaudhary, A proposed a dynamic hand gesture recognition technique in real-time which would benefit the teaching and communication with hearing-impaired people. The system used a real-time video captured in 30fps, by applying skin filtering hand region was extracted. The segmentation was done based on skin pixels from extracted image in HSV color space. Hu-moments and motion trajectory were extracted from image frame. Authors also used depth information in parallel for more accurate results. SVM was used for classification of gestures. The proposed system gave 97.5% recognition rate for selected 4 signs of Sign Language. Out of 80 signs 78 are correctly classified. [2] A. S. C. S. Sastry, P. V. V. Kishore, D. Anil Kumar, E. Kiran Kumar proposed a sign-to-sign language translator, for conversion of sign languages of different countries. Similarity between different sign languages was found using histogram of oriented gradients (HOG) and support vector machine (SVM) from a database of finger spelled sign language from 30 different countries' sign languages. The multi class classification rate using SVM and HOG features was 95%. [8]

E. Song, H. Lee, et al, developed a system for hand tracking which used Thermal Images. Thermal sensors are less sensitive to the illumination changes, and other visual constraints. For capturing the images, authors used FLIR S65 and Microsoft LifeCam. The thermal cameras can detect the thermal radiation emitted by human body and visualize it into images. The images were captured in 23-37 degree Celsius with resolution of 640 x 480. The authors achieved 86.8% precision rate and 82.4% accuracy. [9] In vision-based gesture recognition, the challenge of making algorithm scale invariant, skin color invariant, is very important. Using different error estimation techniques, the system must be optimized and the system needs to achieve real-time recognition.[1]

Depth based: In this method, a specialized camera which can capture depth information as well as RGB information from the environment is used. Microsoft Kinect is one of the depth sensors available. The signers hand movements are recorded using this depth sensor, which enables to identify the movements made in the all three axes i.e. X, Y, Z. Sensors like MS Kinect have features to create a digital human skeleton from the depth information. This greatly helps in gesture recognition. These sensors are expensive and have some interfacing constraints such as distance from sensor, angle from sensor. Kumar, P., Saini, R., Roy, P.P. et al. presented a position invariant sign language recognition (SLR) framework which could recognize occluded sign gestures. Most of the SLR systems require the signer in front of capturing device and fail when position is varied. The skeleton information which is captured through Kinect sensor is processed through affine transformation. Hand segmentation is done after transformation which gave 3 features- angular features, velocity features and curvature features. Hidden Markov model was used for gesture recognition. The system recorded an accuracy of 81.29% for single handed gestures, 84.81% for double handed gestures and 83.77% for combined gestures. [3]

Hisham, B. & Hamouda introduced a dynamic Arabic Sign Language recognition system in which a Microsoft Kinect sensor was used. It was found that the system would help in reducing the barriers faced by deaf sign language users in terms of health education. 42 different gestures related to medical field were recorded using the sensor from two volunteers at different position. The system used Kinect SDK which detects the body joint positions. It was observed that only 10 joints out of 25 were used for performing signs. Joint position normalization was performed and angle between joints was calculated. Classification was done using decision tree and Bayesian network with accuracy 91.18% and 92.5% respectively. [4]

Molina, J., Pajuelo, J.A. & Martínez have proposed a real-time dynamic hand gesture recognition technique for providing natural and good interaction with computer applications. 495 videos were recorded using TOF camera, 11 users were asked to execute 5 repetitions of 9 gestures. Arm model based on human anatomy was proposed wherein the trajectory of hand motion is captured. Dynamic Time Warping (DTW) is used for comparing hand trajectories at different speeds. Results showed that the computational cost of this technique is much less that improved real-time HCI. Accuracy of 95.1% and 78% was recorded in 2 different testing scenarios, that is, 2D and 2.5D. [5]

Hasler, B.S., Salomon, O., Tuchman, P. et al. A real-time culture specific gesture translator using body tracking and gesture recognition was proposed, for bridging the gap between different cultures and enhance mutual understanding. The authors captured the video using MS Kinect sensor, the 11 joints' orientation data was used as input features for Hidden Markov Model (HMM). The weighted average of the output generated by HMM was compared with threshold for translation of that gesture. Authors used only 2 greeting gestures and some random movements for testing, the system was not evaluated for accuracy. [6]

Deep Learning: In this method, artificial neural network is used for gesture recognition. The neural network is trained with large dataset of human body, hand images to detect body pose from images. After training, the image containing human body, hand is passed to the neural network. The neural network generated output in the form of probability map of key points of human body, which includes head, shoulders, elbows, wrist, etc. The coordinates of key points are then printed on the image.

J. Joy, K. Balakrishnan, Sreeraj M proposed SignQuiz, a cost-effective web-based application for learning finger spelled signs of Indian sign language. The application captured images of signs made by users and sent them to a server where recognition of sign was done. Alphabet learning algorithm is proposed for learning new alphabets in sign language. Two pre-trained neural network models Nasnet and InceptionV3 were used for sign language recognition which produced 97% and 99% accuracy respectively. [7]

Simon T, Joo H, et al, presented a technique for detecting key points such as joints of hand from single image. The used multi-view bootstrapping for increasing the accuracy of detection, which includes the use of multi-camera system. This multi camera system is used to train the detectors for key points which get occluded. The

Table 1. Literature review

Authors	Year	Technique	Worked on	Translation Into
Hasler, B.S., et al	2014	Kinect Sensor for Skeleton points, HMM for classification	1 East Asian greeting gesture	Western greeting gesture
Raheja J.L., et al	2015	Skin pixel based hand segmentation, SVM for classification	4 Signs of Indian Sign Language	English
Molina, J. et al	2015	Depth information using TOF Camera, Selecting point of interest on hand,	9 Gestures corresponding to directions	English
K. Kaur and P. Kumar	2016	JA SigML Player for animation of HamNoSys into Indian Sign Language	250 English Words	Indian Sign Language
Kumar, P., et al	2017	Kinect Sensor for skeleton, HMM for classification	30 Signs of Indian Sign Language	English
A. S. C. S. Sastry et al	2018	HOG & SVM features for finding similarity between images.	Alphabets from sign languages of 30 countries	Sign Language
M. G. Grif and A. L. Prikhodko	2018	OpenPose Library for body pose detection, hand keypoint detection	Russian Sign Language	Russian
Hisham, B. & Hamouda	2019	Kinect Sensor for skeleton, Decision tree, Bayesian network for classification	42 Signs of Arabic Sign Language	Arabic
J. Joy et al	2019	Artificial Neural Network for Sign Language recognition	20 Finger Spelled Alphabets of Indian Sign Language	English

proposed method runs Realtime on RGB images and videos, thus enabling to detect hand key points in-the -wild. [11]

X. Deng et al., proposed a joint hand detection technique, using a convolutional neural network. The authors used in-plane rotation to improve hand detection. The rotation framework aligns the regions detected in upward direction, then detect the hands in rotated images. De-rotation layer was designed to make the images in original direction. The average precision the authors obtained on the EgoHands dataset is 75.7%, and on Oxford hand dataset is 48.3%. [12]

Data Glove: This technique includes usage of a glove made-up of sensors like accelerometer, gyroscope, and magnetometer. The specialized hardware is integrated into a glove which is then worn by the signer. As the signer makes gestures, these sensors calculate the movement, angle, velocity of fingers. This technique can detect almost perfect gestures made by the signer. Complex gestures involving both hands, finger movements can be recognized. But the required hardware is very expensive and needs to be worn at all times for recognition, and creating multiple gloves for multiple people is expensive.

T. L. Baldi, et al., presented glove for enhanced sensing and touching (GESTO), a wearable glove for sensing and actuation applications. The glove consists of inertial and magnetic sensors for hand tracking. For receiving feedback, they used various cutaneous devices. The presented system does not suffer from occlusions and lighting problems. The gloves consist of magnetic, angular rate and gravity (MARG) sensors, which contains, magnetometer, accelerometer, and gyroscope. Orientation estimation error is $3.06^{o} \pm 0.12^{o}$. [10]

III. CONCLUSION

This work provides a study on sign language translation techniques. We observe that a data glove will provide the best accuracy in gesture recognition, followed by a 3D sensor. 2D videos provide the most inexpensive solution in gesture recognition. Using the techniques, a sign language translator can be designed, for minimizing the gap between the hearing and speech impaired people and spoken language users. Different techniques of gesture recognition can be applied for achieving more accurate and near real time sign language recognition system.

REFERENCES

1. Ghotkar, Archana & Gajanan, K & Kharate, Gajanan. (2014). Study of vision-based hand gesture recognition using Indian sign language. International journal on smart sensing and Intelligent System.
2. Raheja, J L., Mishra, A. & Chaudhary, “A. Indian sign language recognition using SVM.” Pattern Recognit. Image Anal. 26, 434–441 (2016)
3. Kumar, P., Saini, R., Roy, P.P. et al., “A position and rotation invariant framework for sign language recognition (SLR) using Kinect”, Multimedia Tools Applications (2018) 77: 8823.
4. Hisham, B. & Hamouda, “Supervised learning classifiers for Arabic gestures recognition using Kinect V2”, Springer Nature Applied Sciences, 2019 1: 768.
5. Molina, J., Pajuelo, J.A. & Mart´ınez, “Real-time Motion-based Hand Gestures Recognition from Time-of-Flight Video”, Journal of Signal Process Syst (2017) 86: 17.
6. Hasler, B.S., Salomon, O., Tuchman, P. et al., “Real-time gesture translation in intercultural communication”, AI & Soc (2017) 32: 25.
7. J. Joy, K. Balakrishnan, Sreeraj M, SignQuiz: “A Quiz Based Tool for Learning Fingerspelled Signs in Indian Sign Language Using ASLR”, IEEE Access Vol 7, 2019, pp 28363-28371.
8. A. S. C. S. Sastry, P. V. V. Kishore, D. Anil Kumar, E. Kiran Kumar, “Sign Language Conversion Tool (SLCTooL) Between 30 World Sign Languages”, Smart Computing and Informatics, 2018, pp 701-711.
9. E. Song, H. Lee, J. Choi and S. Lee, "AHD: Thermal Image-Based Adaptive Hand Detection for Enhanced Tracking System," in IEEE Access, vol. 6, pp. 12156-12166, 2018.
10. T. L. Baldi, S. Scheggi, L. Meli, M. Mohammadi and D. Prattichizzo, "GESTO: A Glove for Enhanced Sensing and Touching Based on Inertial and Magnetic Sensors for Hand Tracking and Cutaneous Feedback," in IEEE Transactions on Human-Machine Systems, vol. 47, no. 6, (Dec. 2017) pp. 1066-1076,
11. Simon T, Joo H, Matthews I, Sheikh Y. Hand keypoint detection in single images using Multiview bootstrapping. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition 2017 (pp. 1145-1153).
12. X. Deng, Y. Zhang, S. Yang, P. Tan, L. Chang, Y. Yuan, and H. Wang, “Joint Hand Detection and Rotation Estimation Using CNN”, IEEE Transactions on Image Processing, vol. 27, no. 4, (April 2018) pp. 1888-1900.
13. American Sign Language [Internet]. NIDCD. 2019 [cited 25 December 2019]. Available from: https://www.nidcd.nih.gov/health/american-sign-language
14. Jane E. Johnson and Russell J. Johnson. 2008. Assessment of Regional Language Varieties in Indian Sign Language. SIL Electronic Survey Reports 2008-006. 1-121.
15. "Interpreters and translators: employees U.S. 2018 | Statista", Statista, 2020. [Online]. Available: https://www.statista.com/statistics/320340/number-of-employees-in-interpreting-and-translating-services-us/. [Accessed: 06- Mar- 2020].
16. ANNUAL REPORT. New Delhi: INDIAN SIGN LANGUAGE RESEARCH AND TRAINING CENTER, 2017, p. 14.
17. "Help & Resources - British Deaf Association", British Deaf Association, 2020. [Online]. Available: https://bda.org.uk/help-resources/#statistics. [Accessed: 06- Mar- 2020]

AUTHORS PROFILE

Mr. Udit Barde completed B.E. in Computer Engineering from M.E.S College of Engineering, Pune in 2017, is currently pursuing M.E. in Computer Engineering from Pune Institute of Computer Technology, Pune.

Dr. Archana Ghotkar has received B.E degree in Computer Science and Engineering in 1998 from Dr. BAMU university, M.E degree in Computer Engineering in 2000 and Ph.D. in Computer Engineering from Savitribai Phule Pune university. She is an associate professor in the department of computer engineering at Pune Institute of Computer Technology in Savitribai Phule Pune university, INDIA. Her main research interest includes Computer vision, Pattern recognition, Human computer interaction and Data mining. She is a member of ISTE and CSI