# A Study of Sign Language Recognition Techniques

Deven Nabar

Dwarakdas J. Sanghvi College of Engineering, Mumbai, India

Vaibhav Gandhi

Dwarakdas J. Sanghvi College of Engineering, Mumbai, India

Vedant Gandhi

Dwarakdas J. Sanghvi College of Engineering, Mumbai, India

Prachi Tawde

Dwarakdas J. Sanghvi College of Engineering, Mumbai, India

Corresponding author: Deven Nabar, Email: nabar.deven@gmail.com

Humans communicate with one another in order to share their views, emotions, and stories with other people. For people suffering from communication disorders, it is not the case. Deaf-mute people can interact owing to sign language. The idea of this research is to build a system for identifying sign language that allows people with speech impairments and regular people to communicate, bridging the communication barrier. Furthermore, being able to recognize common sentences that are used frequently becomes crucial for conversations to flow effectively. Such a system is essential in services like banking where privacy would be a top priority.

**Keywords**: Sign Language, OpenPose, Hand Gesture Recognition, Keypoint Detection, Sign Language Recognition.

2021. In Rahul Srivastava & Aditya Kumar Singh Pundir (eds.), *New Frontiers in Communication and Intelligent Systems*, 129–135. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.52458/978-81-95502-00-4-16

## 1 Introduction

According to a survey conducted by the World Health Organization (WHO), more than 5.5% of the global populations have some form of hearing loss [1]. The number of individuals suffering with this impairment was 46.6 crore in March 2018, and it is projected to reach 90 crore by the next 30 years. In addition, according to the 2011 Indian census, 70 lakh Indians suffer from communication disorder [2]. Most of them are unable to communicate since they are unable to talk or hear, limiting their communion. People with communication disorders interact using sign language to exchange thoughts and feelings. Texting, writing, using visuals, are just a few of ways that enabled and disordered persons interact. However, the disabled prefer communicating using gestures as they believe they can express their thoughts and ideas better throughgestures.

Few people suffering from this disorder can interpret the text of their language. Even so, these individuals are at a major disadvantage in a variety of situations, such as job, school, and social events [3]. Even though several sign languages exist, most people are unaware of them. As a result, communication with the deaf and dumb becomes increasingly difficult.

This project's major objective is to identify bank-related sign gestures in the Indian Sign Language lexicon and create an alternate communication medium for hearing-impaired people. The development of such a useful application was motivated by the fact that it would aid in increasing social awareness and reducing the isolation of these people from banking operations. Due to the difficulty in complicated gesture patterns, there has been less study in this sector of ISL [4-5].

In this paper, we have review edacious techniques related to Indian-Sign Language recognition and studied and analyzed various methodologies related to human pose estimation algorithms. The rest of the paper is organized as follows: Section 2 talks about the Literature Review related to existing systems and methodologies. In Section 3, we have presented the analysis of our findings. Section 4 goes over our proposed system. In Section 5, we have provided our conclusion for the paper.

## 2 Existing Systems

### A. Literature Related to Existing Systems

### i. Sign language translation using Microsoft's Kinectsensor

This model converted sign language that consisted of simple gestures to speech or voice by using Microsoft's Kinect sensor that helps in sensing different motions [6]. The program takes input when the sensor is switched on; whenever human gesture is detected, the person's skeletal information is captured using the 20 joints. Different skeleton frames are created from the stream of inputs, with every frame consisting of gestures. The gestures received are compared to a gesture set that has been previously defined. The word corresponding to the input gesture is provided as an output to the windows narrator if the present skeleton frame follows the predetermined gesture pattern. The speech is delivered by the narrator. The Kinect requires its own power supply. It is connected to the computer through a USB port.

### ii. Sign language recognition system based on FCM

This system included a camera to record the gestures of users [7]. For the users' convenience, the system depended on a portable unit. The system received unprocessed videos shot against a dynamic back drop as an input. To ensure that all the videos were identical in size, the image frames are adjusted. Facial expressions and hand gestures were the key points to detect in a sign language and for extracting key features and classifying videos they used Open CV. For detecting hand gestures, they used a fuzzy c-means clustering algorithm, in which they assigned points to data items that are similar and grouped them together. After iteratively updating them, centers of each data point are returned by the algorithm which helps in classifying newgestures.

### iii. Indian sign language gesture recognition and formation of sentences

The system focused on the recognition of continuous dynamic Indian Sign Language [8]. The data set used included a collection of signs for performing the hand gestures. These hand gestures were performed using single or both hands. Here, the dataset consisted of a total of 10 sentences. The sentences were made of a few types of gestures (two-four) which were each either static or dynamic. The gradient-based key frame extraction method was used to extract the start frame or final frame of each gesture. The change in gestures was displayed by the momentous difference in gradient at the end of one of the gestures and the start of another one. Key- frame is responsible for breaking down each sentence into sequences of individual gestures and is essential to finding the meaning of the sentence. Processes like DWT, Orientation histogram, and Principal Component Analysis (PCA) were used for extracting characteristics from those frames that comprised of relevant and importantgestures.

## B. Literature Related to Methodology /Approaches

### i. Pose Estimation using Open Pose:

To figure out and identify the parts of the human body with subjects in the frame, the described system involved part affinity fields (PAFs). Regardless of the amount of individuals in the input, such a bottom-up approach delivers great accuracy and performance. Estimations of the body part location and PAFs were simultaneously enhanced along all training stages in prior research. They showed that are vision with only PAF, instead of a combination of body part position refinement [9] and PAF, resulted in a significant improvement in both durability and its efficiency during runtime. They also demonstrated the first body to feet key point detector. They demonstrated that, when compared to executing the components sequentially, the merged detector not only lowered inference time but also sustained the accuracy of detection of every component independently. OpenPose, the first open-source real-time system for multi- person 2D pose recognition, comprising torso, feet, arms, facial and hand key points, was released because of this work.

### ii. Hand gesture recognition using deeplearning:

This research proposed a unique method for detecting dynamic hand gestures by using numerous learning algorithms for breaking down of hand key points, local and global feature descriptions, as well as sequence feature globalization and detection [10]. The suggested system is sorely tested on a difficult dataset consisting of 40 dynamic hand gestures made by 40 people in an unconstrained environment. The findings reveal that the suggested approach beats current best practices, proving its efficacy.

The OpenPose framework was utilized to recognize and estimate hand regions in this investigation. For gesture space estimation and normalization, a comprehensive face identification method and the principle of body parts ratio were used. The characteristics of the form of the hand and the features of the global body configuration were learned separately using two 3DCNN instances. To incorporate and globalize the extracted local features, MLP and auto encoders were used. For classification, the SoftMax function was utilized. Furthermore, they examined domain adaptation and ran large-scale comprehensive experiments to maximize the level of information transfer to lower the training cost of the 3DCNN module. The suggested system was put to the teston challenging sign-language gestures in a dataset from the real- world.

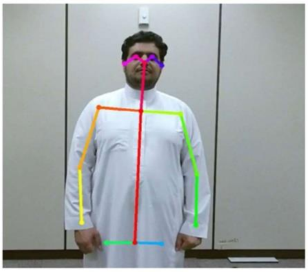

**Fig. 1**. Upper body keypoints detected by OpenPose.

### iii. An Automated Social Interaction Measurement for Children on the Autism Spectrum

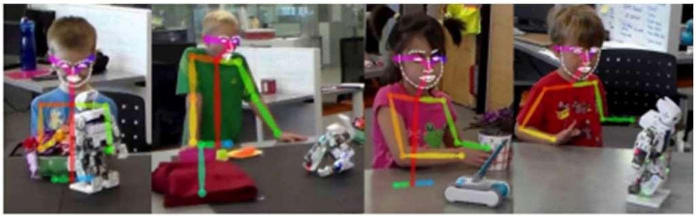

**Fig. 2**. Facial keypoints and upper body skeleton extracted by OpenPose

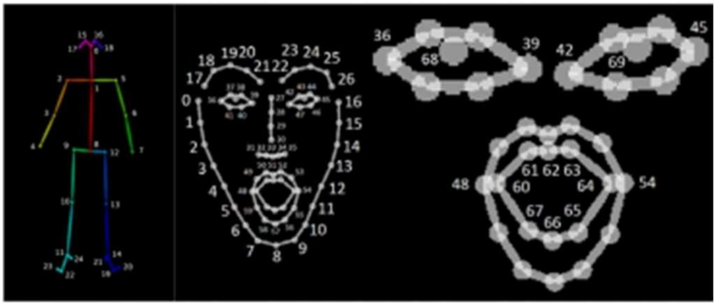

**Fig. 3**. Facial and skeletal keypoints generated by OpenPose

In this paper, the researchers gathered video recordings of the interaction between children and robots with a video recorder positioned at a corner of the space. As the children maneuvered about, the camera was repositioned. Using the OpenPose library, they were able to retrieve sound and 2-D motion data. Even though OpenPose provides full-body information, they were only able to access and use upper body data since it was meant for children who were near the table which was blocking their lower body view (figure 3). Along with all this, facial expressions were also captured and data points were gathered from the video. Six intended behaviour patterns were identified to capture the ground truth of the child's interaction with the robotic systems. These included vocalizations, self-initiated interactions, eye gaze focus, smiling, triadic interactions and imitation. This system predicts autism-like behaviours early in a child's life. The state of the child can be classified by utilizing a combination of facial gestures as well as upper body data to train a CNN [11].

## 3 Analysis

It was found that the accuracy for the paper "Conversation of Sign Language to Speech with Human Gestures" [6] was found to be extremely high, reaching a maximum accuracy of 90%. The reason behind this as stated in the paper was that the system recognized words and not complete sentences. In addition, the system was explicitly trained to identify only a hundred words. Another disadvantage of this technique is that it necessitates the use of a separate device (MicrosoftKinect) to record the user's movements. Due to the additional costof using and maintaining such components, this would not be practicable.

The accuracy for the paper [7] was around 75%. Also, another drawback of this paper was that the system could only recognize 40 words which are low for a system to be used in the real world. This paper used a fuzzy c-means clustering algorithm for detecting hand gestures. But using this algorithm adds to the time complexity of the entire system.

The accuracy of the paper [8] was the highest of 94%. This system used various distance classifiers like Manhattan distance, Minkowski distance, Euclidean distance, etc for classifying the hand gestures. The database consisted of only 10 sentences, which is why the accuracy of this system was so high. These sentences were primarily simple sentences with not more than 4 words in most of them. Even though the dataset was limited, the time-complexity of this system was high as it used gradient-based key frame extraction for splitting the gestures into a sequence of signs, and features of each sign were then extracted.

**Table 1.** Comparison of various papers

| Papers | Accuracy | Results |
|---|---|---|
| Paper 1:- "Conversation of Sign Language to Speech with Human Gestures [6]" | Up to 90% | Recognized only 100 words |
| Paper 2:- "Real-Time Recognition of Indian Sign Language [7]" | 75% | Recognized 40 words |
| Paper 3:- "Continuous Indian Sign Language Gesture Recognition and Sentence Formation [8]" | 94% | Recognized only10sentences |

## 4 ProposedMethodology/Approach

Existing systems require a lot of computational power for pre-processing stages like removal of noise, altering the contrast of the video frames, etc. Some systems also require additional equipment/devices like gloves, IR gadgets [6, 12- 14]. This directly affects the overall time complexity, thereby reducing the efficiency of thesystem.

Top-down approaches suffer from early commitments, where the person is first detected and then the components are detected. Since our system would not necessarily be provided with the full body view, OpenPose that uses a bottom-up approach is ideal in such a case [9].

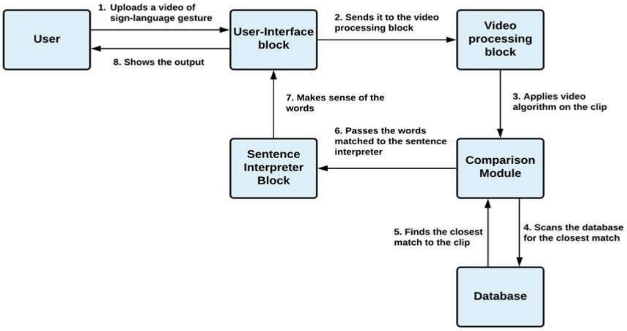
User
1. Uploads a video of sign-language gesture
User-Interface block
8. Shows the output
2. Sends it to the video processing block
Video processing block
3. Applies video algorithm on the clip
7. Makes sense of the words
6. Passes the words matched to the sentence interpreter
Sentence Interpreter Block
Comparison Module
5. Finds the closest match to the clip
4. Scans the database for the closest match
Database

**Fig. 4**. Proposed System Architecture

Hence, we propose a system using OpenPose, where the user will upload a video, consisting of sign language gestures, which will be fed to the user interface. The system will then send it to the video processing block where a skeleton, containing key points of the body, will be overlayed on the input video and this processed video will further be sent to a comparison system.

Here the database will be scanned to find the closest match to the given clip. After finding the closest match the comparison module forwards the words matched (to the clip) to a sentence interpreter block. Here the words passed relating to the clip will then be processed to form a meaningful sentence which will be shown as the final output to the user.

**Fig. 5.** Video Captured By Camera

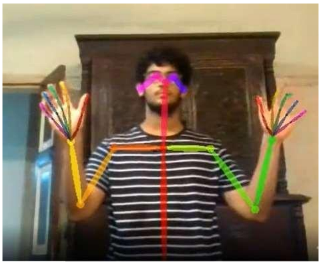

**Fig. 6.**Openpose Keypoint Detection and Skeleton generation on the Captured Video

## 5 Conclusion

We have reviewed several papers related to Indian Sign Language detection as well as papers related to systems using OpenPose algorithm for key point detection. Systems pertaining to Indian sign language detection had a few disadvantages regarding either their accuracy or efficiency. Furthermore, reliance on external devices for capturing inputs is not convenient and is not feasible in most cases. By taking input via the integrated camera unit, we eliminate any use of external equipment or gadget. Hence, we propose to use a bottom-up approach using the OpenPose algorithm to solve theseproblems.

# References

[1] Deafness and hearing loss (n.d.), Retrieved from https://www.who.int/health-topics/hearing-loss#tab=tab_1

[2] D. Verma and P. Dash, "Disabled Personsin India," Retrieved from http://mospi.nic.in/sites/default/files/publication_reports/Disabled_persons_in_India_2016

[3] V. K. Verma and S. Srivastava, "Toward Machine Translation Linguistic Issues of Indian Sign Language", *Adv. Intell. Syst. Comput. Speech Language Proc. Human-Machine Communs,* pp. 129-135, 2017.

[4] R. Savant and J. Nasriwala, "Indian Sign Language Recognition System: Approaches and Challenges", in *Conf. Emerg. Innov. Inform. Tech.: Prospects and Challenges*, 2019.

[5] S. Chakraberty, "Challenges in building an app to interpret Indian sign language", https://www.livemint.com/news/business-of-life/challenges-in-building-an- app-to-interpret-indian-sign-language-11608476748623.html, 2020.

[6] S. Rajaganapathy et al., "Conversation of Sign Language to Speech with Human Gestures," *Procedia. Comp. Sci.*, vol. 50, pp. 10–15, 2015.

[7] H. M. Mariappan and V. Gomathi, "Real-Time Recognition of Indian Sign Language," in *Int. Conf. Comput. Intell. Data Sci.*, 2019.

[8] K. Tripathi and N. B. G. Nandi, "Continuous Indian Sign Language Gesture Recognition and Sentence Formation," *Procedia Comp. Sci.*, vol. 54, pp. 523–531, 2015.

[9] Z. Cao et al., "Realtime Multi- person 2D Pose Estimation Using Part Affinity Fields," in *IEEE Conf. Comp. Vision Pattern Recog*., 2017.

[10] M. A. Hammadi et al., "Deep Learning-Based Approach for Sign Language Gesture Recognition with Efficient Hand Gesture Representation," *IEEE Access*, vol. 8, pp. 192527– 192542, 2020.

[11] H. Javed, W. Lee and C. H. Park, "Corrigendum: Toward an Automated Measure of Social Engagement for Children With Autism Spectrum Disorder—A Personalized Computational Modeling Approach," *Front. Robot. AI*, vol. 7, 2020.

[12] S. Y. Heera et al., "Talking hands-An Indian sign language to speech translating gloves," in *Int. Conf. Innov. Mech. Industry Appl.,* 2017.

[13] P. Das et al., "Analytical Study and Overview on Glove Based Indian Sign Language Interpretation Technique", in *Michael Faraday IET Int. Summit, 2015.*

[14] K. Mehrotra, A. Godbole and S. Belhe, "Indian Sign Language Recognition Using Kinect Sensor", *Lecture Notes in Comp. Sci. Image Anal. Recogn.,* pp. 528-535, 2015.