Gesture Recognition and Conversion to Speech for Specially Abled

Mukul Chaudhari, Chinmay Mukhedker, Jaspreet Singh Pannu, Varun Prasannan, and Sonali Patil

Abstract There are around 1 million people who suffer from hearing and speech impairment. They might find it difficult to express their thoughts because of their limited capabilities. However, using the latest technologies we can overcome this problem and help the specially abled with their inabilities. The latest trends in human-computer interaction, artificial intelligence and machine learning made it possible to build a system which can act as a mediator for specially abled people. In this paper, a comparative analysis is provided with respect to few existing gesture recognition methods. The intent of this paper is to compare and analyse the existing methods of gesture recognition and conversion of text to speech to help find out the most efficient algorithm among the algorithms compared to other fellow researchers.

Keywords Gesture recognition · Human–computer interaction · Machine learning · Deep learning · Artificial intelligence · Contour detection · Convolution neural network (CNN) · Double channel CNN

1 Introduction

The diversified technology and its continuous growth have made it possible to interact with the computer without any physical touch by using gestures, facial expressions and even intellectual thoughts. Hand gestures are primarily divided into two parts: static gestures and dynamic gestures. Static gestures are the constant and still positions of the hand to indicate some sign or some sort of activity. Dynamic hand gestures are the gestures with continuous motion of the hand in a particular time frame. The mentioned gestures can be recognized in multiple ways. The main two ways are by using software and by using hardware.

M. Chaudhari · C. Mukhedker · J. S. Pannu · V. Prasannan (✉) · S. Patil
Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, Maharashtra 411044, India
e-mail: vprasannan4@gmail.com

S. Patil
e-mail: sonali.patil@pccoepune.org

V. K. Gunjan and J. M. Zurada (eds.), *Proceedings of 3rd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications*, Lecture Notes in Networks and Systems 540, https://doi.org/10.1007/978-981-19-6088-8_44

In the hardware, hand gloves wired with sensors are attached to the tip of the finger, and the motion is detected and parsed with the help of those sensors. In the software approach, there are multiple machine learning tools and algorithms that make it easy to detect the gestures through cameras.

2 Literature Survey

The survey provides an analysis of existing methods and technologies. For gesture detection, data gloves are the most used technology worldwide. But the latest trends in machine learning have made it possible to achieve better accuracy in gesture recognition using cameras. The approach by Aashni et al. [1] uses a camera to detect the gestures. The image frames are obtained in the form of video. In the later process, the frames are pre-processed to remove extra noises, useless background and to convert colours if required. Followed by this, contour detection is carried out to focus only on the required part of the image. Variable accuracy is seen in the approach with respect to the number of fingers used to show the gestures and type of background in the obtained frame.

In a convolutional neural network, the network is formed with multiple nodes, also called neurons. These nodes are connected to each other via links. Each link has its own weight, and as a whole, they produce the output. Generally for the visual classification of images, CNN is a better option than others. The pre-trained model MobileNet as proposed by Nishi et al. [2] states that the bottleneck (the second last layer, i.e. layer prior to last layer) does the main classification.

CNN is likewise a technique used for gesture reputation extra accurately. CNN usually acknowledges the entire neighbourhood functions after which it merges those neighbourhood functions at a better level. This technique is used to gather the all-inclusive traits of the photo and its structure, and then the properties of the pictures may be obtained. So in short, CNN has the upper hand in pixel price of the processing units.

In edge detection algorithm, by Wu et al. [3], an exceptionally speedy guided filter is followed to reconstruct the unique gesture image. Therefore, as a way to optimize the training method of gesture recognition, the double channel convolutional neural network is proposed. This shape is made from fairly impartial convolutional neural networks or CNNs. These two channels have their own two different inputs which usually have independent weights. These are linked to the relational layer, and a relational map is performed.

In orientation histogram approach by Deepali et al. [4], we find feature vector to classify the image. Firstly, the program reads the image database. Then, we resize all the images. The edges are being found by using two filters: one in x-direction and the other in y-direction. The divider method will give the gradient orientation. Then the image blocks are rearranged into columns using simulation software. The column values of the matrix are then converted to degree from radian.

Using optical character recognition and Festive Software, it consists of two models: image processing model and voice processing model. Image processing model will convert the image to text, and the voice processing model will convert the text to understandable speech.

The flow is as follows: Read the object as we take the picture with the camera and convert the image into grayscale. Then we determine the region of interest, OCR processing is initiated, and the text is achieved as output. The output of OCR is stored in .txt format, and Festive software is being used to convert text to speech in different languages [5].

The algorithm by Guo et al. [6] uses data gloves as its base. Initially, it uses the least square technique, and also the calculation of Hausdorff distance is done. The curve fitting method has finished before and so the shapes of the curves of the same gesture were similar, and the algorithm could transform a motion recognition method into a curve recognition method. The benefits of implementing this method are robustness, high accuracy, and reliability. The foremost downside of this method would be the speed and potency of the system. Once the recognition result is applied, the HCI is increased within the virtual scene.

In the study by Nascimento et al. [7], a method has been developed to control the playback function in Netflix using gesture recognition with the help of a smartwatch-like device. Here, the user performs an already defined gesture, and after it is processed by the smartwatch-like device, the command is sent to the communication platform and is executed on the Netflix application. A continuous gesture recognition algorithm is used and is instantaneously executed in Netflix.

The method proposed by Pinto et al. [8] detects the hand gestures using convolutional neural network (CNN). Later analysis of the results is done using cross-validation technique. In the proposed method, images are taken as input using a camera. The obtained images could have variable formats, scenarios, or backgrounds. So to bring the uniformity in the images, they are sent for pre-processing. During pre-processing, the images are passed through certain operations as follows:

- Colour segmentation: In this, the algorithm identifies uniformity in colours in the image and identifies the cluster of pixels in the image.
- Morphological operations: In this, digital images are processed on the basis of their shapes. The two specific operations in this case are erosion and closing. Erosion removes pixels on the object boundaries. Closing first adds the pixel on object boundaries and then removes it simultaneously.
- Contour generation: This is used to focus on the required area in the image.
- Polygon approximation: In this, the exact required area is detected and used. A logical AND operation is performed on the resultant pre-processed image so that it maintains the information accommodated on the fingers and surface of the hand. The images are then fed to CNN to train the model, and the results are stored. The stored results are then analysed using cross-validation techniques.

The approach by Bhagat et al. [9] has used image processing and deep learning to identify the gestures. The proposed method focuses mainly on Indian sign language translation. Different static images are trained using CNN. The area of focus is on

Indian sign language alphabets and numbers. A five-layer CNN model along with computer vision techniques is used in the method. The results showed the accuracy of 99.81%. With this, a method is proposed to identify the dynamic gestures too. For dynamic gestures, convolutional LSTM is used to train the video input of gestures. Accuracy in case of dynamic gestures is found to be 99.08%.

Varun et al. [10] The system proposed in this paper consists of acquiring the gesture specimen, processing the gesture specimen, gesture recognition at runtime and a control system. Then each and every image in the specimen data provided by the dataset will be worked on by the proposed system. The colour images are converted into black and white images which are called masks. Then input is converted into computer understandable images by our proposed system.

Shelke et al. [11] The points of defect are identified in gesture. With the help of these points, the number of fingers present in that particular gesture is evaluated. The obtained result of the gesture is supplied to a 3D CNN one by one to identify and recognize the current gesture. The system assists skeletal structure detection, skin colour detection, adjusting lighting and camera effects. Then we will implement hand localization using histogram clustering method. The resultant image is compared with the trained dataset using R CNN.

3 Comparative Analysis

Table 1 gives comparison of a few hand gesture recognition techniques and their conversion to text and speech based on parameters such as accuracy while testing accuracy in real time. With this, remarks have been added to each of the techniques.

Table 1 summarizes that the technique used in [9] is the most accurate one, but it lags behind in terms of speed. This CNN technique is only suitable and useful for static images. Technique used in [6] is the most accurate one for dynamic gestures. But data glove is not economical as compared to the other ideas.

4 Conclusion

In this paper, we have analysed some papers and the techniques used in them and shortlisted those in such a way that one would get a perfect idea of which algorithm will be most efficient for them.

Here we can conclude that a lot of advancements are needed in the field of hand gesture recognition systems, and more accurate and faster methods are needed to be developed for helping in communication using them. The recent updates in technologies have made it easier to achieve the same with just using softwares. The use of hardware devices like sensors and data gloves gives improvised accuracy, but it is costly to implement. The paper clearly identifies and analyses different existing approaches with their own pros and cons.

Table 1 Average accuracy of some hand gesture recognition techniques

Title	Average accuracy	Remark
1. Hand gesture recognition for human–computer interaction	92.28% with plain background and 64.85 with non-plain background	The analysis clearly shows that the accuracy is less in case of a non-plain background as compared to plain background
2. Indian sign language converter using convolutional neural networks	Accuracy of 96% for the testing phase with images and 87.69% for the images taken in real time	This analysis shows that for the images taken in real time, the accuracy might deteriorate as compared to just training image sets
3. Hand gesture recognition algorithm based on double channel CNN	98.02% with DC-CNN and 97.04% with SC-CNN	The analysis clearly shows that the accuracy is less in case of DC-CNN as compared to SC-CNN
4. Hand gesture recognition on Indian sign language using neural networks	Accuracy achieved is 93.32% with real-time images	The analysis shows that the accuracy of copy gestures is less than cut, open, close, and refresh gestures
5. A novel method for data glove-based dynamic gesture recognition	Rate of statistical recognition using several types of gesture has been achieved 98%	This study concludes that the accuracy of the system does not deviate with different shapes and sizes of hands
6. Static hand gesture recognition based on convolutional neural networks	For CNN with different layers, the average accuracy is found to be 96.21%	The method proves that performing segmentation and other techniques, CNN gives more accuracy with less computational cost
7. Indian sign language gesture recognition using image processing and deep learning	A five-layer CNN model used in this showed the accuracy of 99.81% for static images	The results show that using a multilayer CNN model has a positive impact on the accuracy

References

1. Haria A, Subramanian A, Asokkumar N, Poddar S, Nayak JS (2017) Hand gesture recognition for human computer interaction. Procedia Comput Sci 115(2):367–374. https://doi.org/10.1016/j.procs.2017.09.092
2. Intwala N, Banerjee A, Gala N (2019) Indian sign language converter using convolutional neural networks. In: IEEE 5th international conference for convergence in technology (I2CT), pp 1-5. 1109/I2CT45611.2019.9033667
3. Wu X (2020) A hand gesture recognition algorithm based on DC-CNN. Multimedia Tools Appl, vol 79. https://doi.org/10.1007/s11042-019-7193-4
4. Kaushik D, Bhardwaj A (2016) Hand gesture recognition on indian sign language using neural network
5. Venkateswarlu S, Duvvuri B, Kamesh K, Jammalamadaka, S, Rani R (2016) Text to speech conversion. Indian J Sci Technol 9. https://doi.org/10.17485/ijst/2016/v9i38/102967. PY—2016/10/01

6. Guo X et al. (2017) A novel method for data glove-based dynamic gesture recognition. In: 2017 International conference on virtual reality and visualization (ICVRV), pp 43–48. https://doi.org/10.1109/ICVRV.2017.00018
7. Nascimento TH, Soares FA, Nascimento HA, Vieira MA, Carvalho TP, de Miranda WF (2019) Netflix control method using smartwatches and continuous gesture recognition. In: 2019 IEEE Canadian conference of electrical and computer engineering (CCECE), pp 1–4. https://doi.org/10.1109/CCECE.2019.8861610
8. Pinto RF, Borges CD, Almeida A, Paula IC (2019) Static hand gesture recognition based on convolutional neural networks. J Electr Comput Eng. pp 1–12. https://doi.org/10.1155/2019/4167890
9. Bhagat NK, Vishnusai Y, Rathna GN (2019) Indian sign language gesture recognition using image processing and deep learning. In: 2019 Digital image computing: techniques and applications (DICTA). IEEE, pp 1–8
10. Varun KS, Puneeth I, Jacob TP (2019) Hand gesture recognition and implementation for disables using CNN'S. pp 0592–0595. https://doi.org/10.1109/ICCSP.2019.8697980
11. Shelke T, Nerkar V, Nandanwar T, Barapatre N, Umare D (2021) Development of real time hand gesture recognition for dumb and deaf using machine learning and deep learning. Int J Creative Res Thoughts (IJCRT) 9(3):2107–2114, ISSN: 2320-2882