JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

HAND GESTURE RECOGNITION AND VOICE CONVERSION FOR DEAF AND DUMB

P.Raguraman [1], Ratala Radhika [2],Pulakandam Sruthi[3], Jampala Pavan[4], Naru Chenchi Reddy[5], Syed Sohail[6]

[1]Assistant Professor, Department of Computer Science and engineering

[2,3,4,5,6]B.Tech., Scholars, Department of Computer Science and engineering ,[1,2,3,4,56]Qis College Of Engineering And Technology,Ongole.

Abstract-The use of gesture-based sign language detection systems in human-computer interaction is critical for the advancement of deaf-to-hearing community communication.Here, those who are deaf or hard of hearing may express themselves and communicate with others who are also deaf.Hand gestures have proven to be a fruitful research avenue, and they've been put to good use in the field of sign language interpretation (SLR). This means that SLR is heavily dependent on the identification of hand gestures, since the sign word is a sort of expressive gesture.Hand gestures vary widely in their complexity and variety, which may have a significant impact on their dependability and capacity to be recognised. This research proposes a deep learning approach that includes a feature fusion convolutional neural network to address the issue of sign word identification. A camera and a preprocessed hand gesture picture are used to collect the live video input in the proposed system. Conversion to binary, erosion and hole filling are all part of the pre-processing. The features are extracted from preprocessed photos using two CNN channels. The feature fusion is done at the fully connected layer, and the softmax classifier uses this feature for gesture categorization. Real-time recognition of the signs of fifteen frequent words has been set up in our laboratory setting. According to the findings of the experiments, gesture-based sign word recognition has a high level of recognition accuracy compared to current methods.

Keywords— hand gesture, segmentation, sign word, convolutional neural network (CNN)

I. INTRODUCTION

The deaf population and others who are hard of hearing rely on Sign Language Recognition (SLR) as a vital communication tool. Each sign language motion has a distinct connotation. Hand shapes, body motions, and even face expressions are all used to convey different aspects of the language. Increasing communication with the deaf population is now reliant on human-based translation services, which is troublesome and expensive due to the reliance on human talent. Since hand gesture-based sign word recognition offers a strong and reliable interface that enables individuals to obtain instant response from the symptoms of the signer as text without translation services, we are concentrating on this technology. To convey their ideas and emotions, people utilise hand gestures. This helps to reinforce the information conveyed in vital communication. Hand gestures However, human-computer interaction (HCI) relies heavily on hand tracking and gesture detection [1-3]. In order to develop human computer interaction, we may accomplish human expression by detecting gestures, which can lessen the impassivity among the deaf and the general population. Research on the recognition of sign language has been carried out by a large number of experts in the past. In order to understand the hand motion, geometric elements like shape and edge are stressed [4-5].

II. RELATEDWORKS

Image processing, on the other hand, is very difficult and time consuming, and the identification accuracy of these systems is worse than that of ours. Researchers from Bangladesh have developed an artificial neural network that can recognise Bengali sign language [6]. However, the system has only been tested on a known dataset, and it has not been tested on light. Hand gesture recognition based on bone data was suggested by the author in [7]. Large amounts of computing are required to recognise the gesture and palm position.Recent advances in image identification and computer vision using deep learning algorithms [8–9] have shown impressive results. It utilises a nonlinear network model with numerous hidden layers to reach high levels of accuracy. Use of CNN's two input channels to identify hand gestures was pioneered by Chinese researcher Xiao Yan Wu.[11] provides an overview of cutting-edge hand gesture and sign language recognition technologies.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

Pre-processing and segmentation, together with feature extraction and classification, are used to identify motions and recognise sign language. In this study, we suggest a hand gesture-based human-computer interface for better communication with the deaf population and the general public. Photos are analysed and various characteristics in the images are extracted using a deep learning approach such as CNN. Softmax classification is used to categorise recognition results once feature fusions are finished at the fully linked layer.This paper was arranged in the following manner. In the second section, we describe the process and provide a model for it. Section 3 explains the experimental data and analyses of this system. Section 4 summarises the study's findings and discusses its future directions.

III. PROPOSED SYSTEM ARCHITECTURE

An isolated sign word recognition system based on hand gestures is shown in Fig. 1. A webcam is used to take a picture of the area of interest (ROI) in a live video frame. The steps involved in recognising hand gestures are outlined in the following paragraphs.

A.Segmentation of Hand Gestures

Using data pre-processing methods, hand motions may be identified and separated from input pictures. The YCbCrcolour space is transformed from the RGB colour space to create an input picture from a grayscale image. Luminance (Y) and chrominance (Cb and Cr) are represented by the YCbCr. The grayscale image's pixel values range from 0 to 255, where 0 is normally black and 255 is white. We can process the picture in grayscale using binary images since 128 establishes a threshold value and the pixel values are specified as 0-127 to 0 and 128-255 to 255. Using erosion, we may erase pixels in the foreground by erasing their boundary areas. As a result, the size of the foreground pixels decreased, and the holes in the region became wider. In the end, we'll fill in those gaps and accept the hole picture we used to extract the feature as final proof. Input picture preprocessing is shown in Fig. 2.

C. Extracting and Classifying Feature Sets

The Convolutional Neural Network (CNN), a prominent family of machine learning algorithms, has grown greatly in technical developments in human-computer interaction [12]. A convolutional, pooling, and fully connected layer, activation function and a classifier are included in the CNN shown in Figure 3. Two channels receive input from gesture pictures and segmented images. The convolutional level is responsible for detecting the input's local characteristics and advancing the convolution through a predetermined number of kernel steps. Weight parameters are determined by processing each level one at a time in a convolution kernel. It is at the output level convolutional pooling layer that this function change takes place.Because of this, it is possible to attain a greater degree of invariant characteristics.In contrast, the data pooling layer may minimise the data layer while retaining the feature data. To classify numerous characteristics conveyed by multiple features, the fully connected layer links to the most recent level of pooling and classifier. It is suggested that the feature descriptor utilise the proposed mechanism to offer comprehensive information on the last completely linked layer. Using Flatten, we were able to accept one-dimensional input into the fully linked layer. Finally, the softmax classifier receives all the characteristics from the fully linked layer. As an activation function, the Relu function is used.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

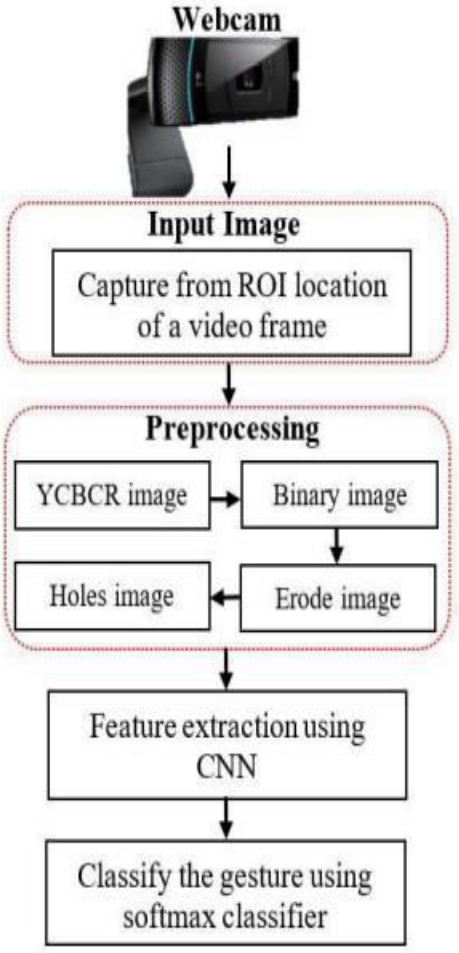

Fig. 1 Overall architecture of sign word recognition system.

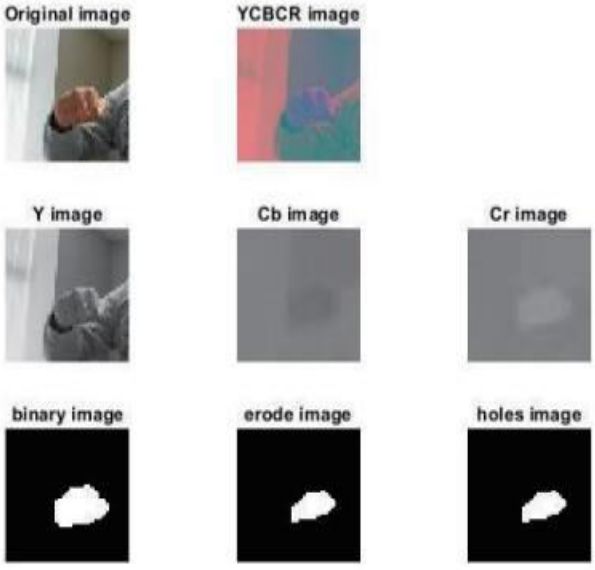

Fig. 2 Preprocessing steps of an input image.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

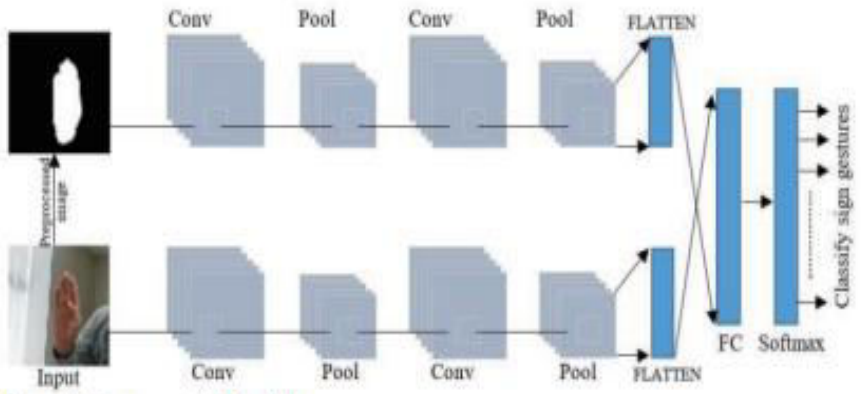

Fig. 3 Proposed Model.

IV. RESULTS AND DISCUSSION

The suggested architecture is able to identify a wide range of sign movements in a complete test. Setup and implementation of the many suggested architectural designs may be found in this section.

A. Dataset for Experimentation

The suggested method's efficacy will be assessed via the usage of hand gesture-based sign word graphics.

Using a camera, the dataset collects photos of fifteen separate motions. A total of 900 pictures are collected for each move.

Sign word recognition thus utilises a total of 13,500 pictures. In order to build a gesture database, three participants are needed. Each subject provided us with a total of 300 photographs for each move. The pictures of the gestures have a resolution of 200x200 pixels. Preprocessed photos are given for feature extraction from the input images. Figure 4 represents the symbol of a sign word's gesture pictures, while Figure 5 provides an example of a gesture image divided. On a machine with Intel Core i5-2400, 3.10 GHz processor and GTX 1080 Ti graphics card, the experiment was carried out.

B. Experimental Evaluation

In this part, we examine gesture-based distinct sign word recognition. The complete dataset was trained using the suggested CNN architecture. We trained on 80% of the datasets and tested on the remaining 20%. The retrieved features are then used in the classification process once they have been trained. However, sign word recognition has an average accuracy of 96.96 percent. "Fine" and "Call" gestures are accepted 100% of the time by the system. The identification accuracy confusion matrix is shown in Fig. 6. For training and testing, we used the approach of lowering the feature dimension [13] and the calibration of skin models [14]. Figure 7 compares the accuracy of recognition between the two groups. Table I compares the accuracy of gesture recognition with the most current techniques. Because our suggested approaches outperform the state-of-the-art methods in terms of accuracy, we believe this is due to the implementation of an efficient segmentation strategy. CNN feature fusion extracts efficient features that boost classification accuracy, thus enhancing the accuracy. Vectors in pre-trained networks provide the pre-processed pictures and auxiliary features for real-time performance, and the gesture is identified based on the most active output outcomes. * A total of 23.03 ms is needed to analyse a picture using this suggested architecture, which may be used to identify in real-time. Webcam captures images at a rate of 33.33 milliseconds per frame. This is why the identification accuracy of various sign movements was tested by comparing the results of different users. Users were requested to do the gesture of sign word fifteen times, each time requiring 10 repetitions.

Fig. 8 shows a user's simulation results. As seen in Fig. 9, the average recognition accuracy for all fifteen sign motions is shown here.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

Fig. 4 Samples of sign word dataset images.

Fig. 5 Example of the preprocessed images of different sign gestures.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

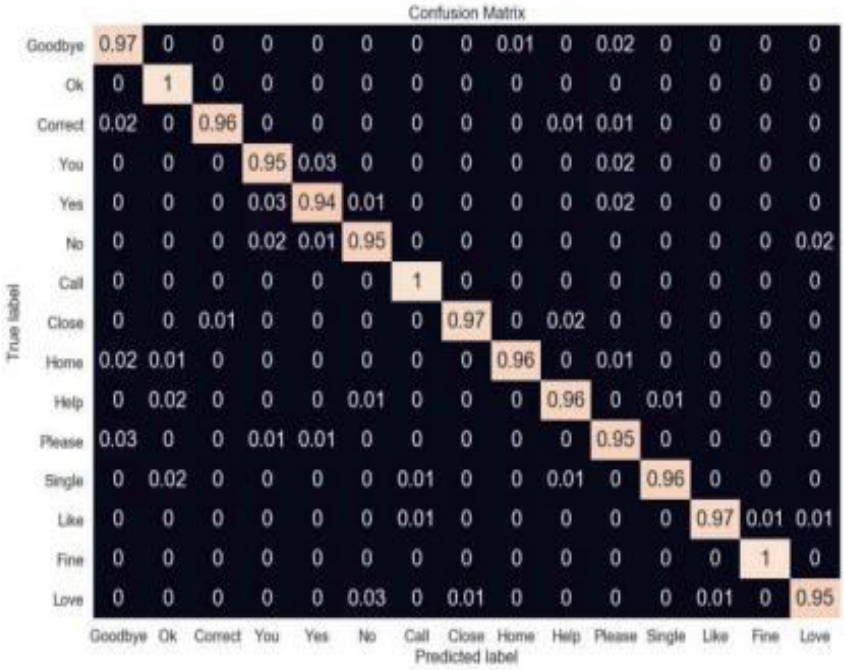

Fig. 6 Average recognition accuracy of different sign word of all users.

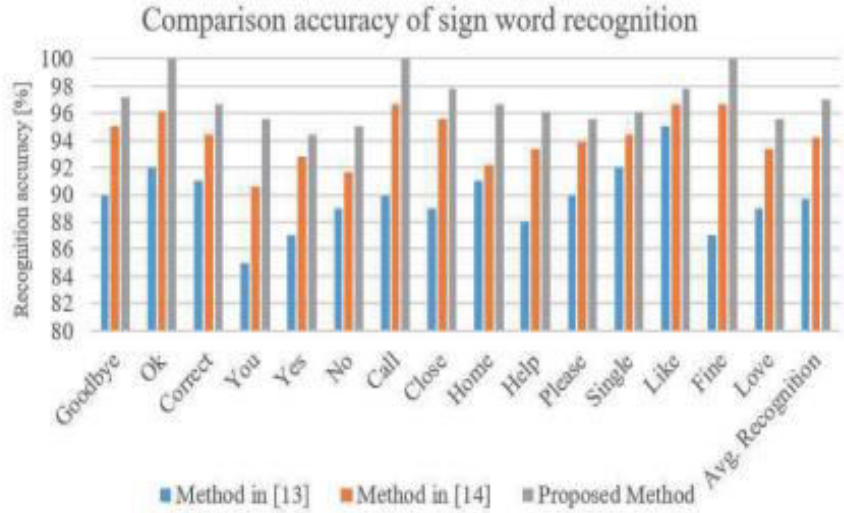

Fig. 7 Comparison of recognition accuracy of isolated sign word.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

TABLE I
COMPARISON OF GESTURE RECOGNITION ACCURACY

Model	Method	Function	Gestures	Accuracy
Ref. [13]	DWT & SVM	Hand Gestures	15 Gestures	89.67% (Evaluated using our dataset)
Ref. [14]	CNN	Hand Gestures	7 Gestures	95.96% (From [14])
Ref. [14]	CNN	Hand Gestures	15 Gestures	94.22% (Evaluated using our dataset)
Proposed	Segmentation & CNN	Sign Word	15 Gestures	96.96%

Fig. 8 Simulation of sign word recognition of a user.

V. FUTURE SCOPE AND CONCLUSION

In this study, we used a convolutional neural network to construct a sign word recognition system based on dynamic hand movements. Hand gestures and sign word recognition may be detected by a series of preprocessing, feature extraction, and classification stages. The photos of the hand gestures have been preprocessed using YCbCr conversion, grayscale image selection, binarization, erosion, and filling in of the missing pixels. The features are extracted using a deep learning approach, such as CNN with feature fusion, then sent to the classifier for classification. These motions are categorised using a softmax classifier. Additionally, a camera is used to construct a gesture-based sign-word recognition system in real time. As a consequence, gesture-based sign word recognition has a higher acceptance rate (96.96 percent) and so produces better outcomes than current methods.Hand gesture-based sign words and unique sentences interpretation algorithms for the sign language recognition system will be developed in the next phase of the project.

REFERENCES

[1] F.S. Chen, C.M. Fu and C.L. Huang. "Hand gesture recognition using a real-time tracking method and hidden Markov models." Image and vision computing, vol. 21, no. 8, pp. 745-758, Aug. 2003.

[2] G. Marin, F. Dominio, and P. Zanuttigh. "Hand gesture recognition with jointly calibrated leap motion and depth sensor." Multimedia Tools and Applications, vol. 75, no. 22, pp. 14991-15015, Nov. 2016.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

[3] P. Kumar, H. Gauba, P.P. Roy, and D.P. Dogra."Coupled HMM-based multi-sensor data fusion for sign language recognition." Pattern Recognition Letters, vol. 86, pp. 1-8, Jan. 2017.
[4] D. Lifeng, R. Jun, M. Qiushi, W. Lei, "The gesture identification based on invariant moments and SVM[J]." Microcomputer and Its Applications, vol. 31, no. 6, pp. 32-35, 2012.
[5] Y.H. Sui, Y.S. Guo, "Hand gesture recognition based on combing Hu moments and BoF-SURF support vector machine." Application Research of Computers, vol. 31, no. 3, pp. 953–956, 2014.
[6] M.A. Rahim, T. Wahid, M.K. Islam, "Visual recognition of Bengali sign language using artificial neural network." International Journal of Computer Applications, vol. 94, no. 17, Jan. 2014.
[7] M.A. Rahim, J. Shin, and M.R. Islam, "Human-Machine Interaction based on Hand Gesture Recognition using Skeleton Information of Kinect Sensor." In Proceedings of the 3rd International Conference on Applications in Information Technology, ACM, pp. 75-79, Nov. 2018.
[8] T. Yamashita, T. Watasue, "Hand posture recognition based on bottom-up structured deep convolutional neural network with curriculum learning." IEEE international conference on image processing (ICIP), pp 853–857, Oct. 2014.
[9] Y. Liao, P. Xiong, W. Min, W. Min, and J. Lu, "Dynamic Sign Language Recognition Based on Video Sequence with BLSTM-3D Residual Networks." IEEE Access, 2019.
[10] X.Y. Wu. "A hand gesture recognition algorithm based on DC-CNN." Multimedia Tools and Applications, pp. 1-13, 2019.
[11] M.J. Cheok, Z. Omar, and M.H. Jaward."A review of hand gesture and sign language recognition techniques." International Journal of Machine Learning and Cybernetics, vol. 10, no. 1, pp. 131-153, Jan. 2019.
[12] S.F. Chevtchenko, R.F. Vale, V. Macario, and F.R. Cordeiro."A convolutional neural network with feature fusion for real-time hand posture recognition."Applied Soft Computing, vol. 73 pp. 748-766, Dec. 2018.
[13] R.A. Bhuiyan, A.K. Tushar, A. Ashiquzzaman, J. Shin, and M.R. Islam, "December. Reduction of gesture feature dimension for improving the hand gesture recognition performance of numerical sign language." IEEE 20th International Conference of Computer and Information Technology (ICCIT), pp. 1-6, Dec. 2017.
[14] H.I. Lin, M.H. Hsu, W.K. Chen, "Human hand gesture recognition using a convolution neural network." In IEEE International Conference on Automation Science and Engineering (CASE), pp. 1038-1043, Aug. 2014.