IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Sign Language Detection Using CNN

Mr. N. Anil Chakravarthy[1], S. Krishn Mohan[2], P. Vasudeva Rao[3], R. Dinesh Kumar[4], S. Swetha[5]
Dept. of Computer Science and Engineering, Raghu Institute of Technology
Visakhapatnam, Andhra Pradesh, India.

ABSTRACT

Sign Language Recognition (SLR) targets interpreting the sign language into text or speech, to facilitate the communication between deaf-mute people and ordinary people. This task has a broad social impact but is still very challenging due to the complexity and large variations in hand actions. Existing methods for SLR use hand-crafted features to describe sign language motion and build classification models based on those features. However, it is difficult to design dependable features that can adapt to the wide range of hand gestures. To address this issue, we propose a novel convolutional neural network (CNN) that automatically extracts discriminative spatial-temporal features from raw video streams without any prior knowledge, thereby avoiding feature design. Multi-channel video streams with color information, depth clues, and body joint positions are used as input to the CNN to integrate color, depth, and trajectory information to improve performance. On a real dataset gathered with Microsoft Kinect, we approve the proposed model and show how it beats conventional methodologies in light of hand-created highlights.

The created framework will be utilized as a learning device for gesture-based communication fledglings including hand recognition. The pictures were taken care of into a model called the Convolutional Neural Network for order (CNN). Keras was utilized to prepare the pictures. A uniform foundation and legitimate lighting are given. The's undertaking will likely foster an AI model that can order the different hand motions utilized in gesture-based communication fingerspelling.

1. INTRODUCTION

Humans require communication because it allows us to express ourselves. Speech, gestures, body language, reading, writing, and other methods are used to communicate. Speech is one of the most commonly used. A speech impairment is a disability that interferes with a person's ability to communicate verbally and through hearing. Affected people communicate using alternative methods such as sign language. Sign Language is the most natural and expressive mode of communication for people who are deaf or hard of hearing. Non-deaf people never try to learn sign language to communicate with deaf people. Deaf people are thus isolated. However, if a computer can be programmed to translate sign language to text format, the communication gap between the hearing and deaf communities can be bridged. Because most ASL signs are single-handed and thus have a lower level of complexity, most researchers in this field concentrate on ASL recognition. Another appealing aspect is the availability of a standard ASL database. Because Indian Sign Language relies on both hands more than ASL, an ISL recognition system is more complex. This proposed system recognizes the various alphabets of American Sign Language, which reduces noise and provides an accurate result.
The use of sign language for communication with the deaf is a hot topic in computer recognition research. This system implements efficient and quick techniques to identify hand gestures that illustrate an alphabet in Sign Language. There is currently a surge in focus on research into sign language recognition systems.

Although sign language has grown in popularity in recent years, non-sign language speakers still find it difficult to communicate with sign language speakers or signers.
This undertaking will likely foster an AI model that converts gesture-based communication to message, permitting underwriters and non-endorsers to impart all the more real. The images depicted beneath are all utilized in our undertaking. The proposed research expects to make a framework that can perceive static sign motions and make an interpretation of them into words. A fantasy-based approach is introduced that assembles data from the guarantor utilizing a web camera and can be utilized from a distance. The system was made to be a learning instrument for the individuals who need to dig further into the essentials of correspondence through signings like letter sets, numbers, and standard static signs. The defenders used Convolutional Neural Network (CNN) as the framework's recognizer and provided a white foundation as well as a specific area for hand image handling. The review covers both basic static signs and ASL letter sets (A-Z). The framework's ability to generate words by fingerspelling without the use of sensors or other external advances is one of the most intriguing aspects of this investigation.

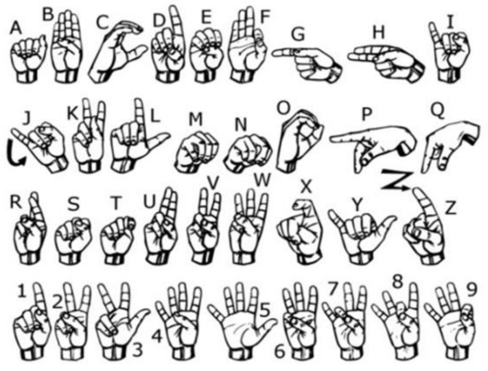

Fig 1: Finger Spelling American Sign Language[11]

2. LITERATURE SURVEY

These days, hand gesture recognition is a popular area of research. This is based on hand gesture recognition in sign language.

The Hand gesture recognition procedure has eight steps to evaluate the effective outcome: image acquisition, skin color-based segmentation, background removal, canny edge detection, PCA feature extraction, image classification with a support vector machine classifier, data training, and data testing SLR research has been ongoing for several decades. Many studies employed sensor-based devices such as Sign Speak. Gloves linked via Bluetooth. Sign Speak has been found to be 92% accurate. Electromyography (EMG) sensors, RGB cameras, Kinect sensors, and leap motion controllers, or their combinations, are other motion sensors that can capture signs. Although these sensors provide accurate data measurement parameters, they do have limitations. The first is cost; because they require large datasets with diverse sign motion, high-end computers with powerful specifications are required. The second consideration is aesthetics; because the sensors are attached to a user's fingers and palms, the user may have difficulty configuring the device; ambient lighting conditions or backgrounds in real-world settings may also be an issue. As a result, many scientists switched from sensor-based to visual-based SLR.

For hand gesture segmentation, the hand skin threshold method is used. The system would not produce good results due to the lighting conditions, skin color interference, and complex backgrounds that increased noise. Skin color detection methods are classified into three types: explicit range methods, nonparametric methods, and parametric methods. Based on the assigned color range, the explicit range method divides the class of pixels into skin-based and non-skin-based types. This technique is popular because of its straightforward approach and reasonable computation rate. This technique, however, is only suitable for a wide range of skin tones. Balbin et al. used colored gloves to easily identify the hands by setting an exact range of the hand skin color threshold (color of the gloves). The input images were processed using a variety of image processing methods or steps to recognize the hand gesture. The first step is pre-processing, which involves converting images to grayscale and denoising them with a median filter. Before moving on to feature extraction, the color of the hand gloves was detected and isolated from the background. Pattern recognition was then applied to the image. The system used Kohen self-organizing maps, a neural network that can unsupervised learning to identify patterns and group datasets. These studies propose a complex yet manageable skin color thresholding process; it can be seen that when only the signer's bare hands are used, the system struggles to recognize the gesture due to various impediments such as noise. Other studies used colored gloves to solve the problem. Still, the current study proposed a system that can recognize static sign language without gloves or hand markings while still producing acceptable results.

3. PROBLEM STATEMENT

We are unable to communicate with deaf and dumb people without assistance in any situation. Dumb people communicate using hand signs, so normal people have difficulty recognizing their language through signs. As a result, systems that recognize various signs and convey information to ordinary people are required. Building a system that recognizes sign language will allow the deaf and hard-of-hearing to communicate more effectively with modern technologies.

4. OBJECTIVE

Gesture-based communication is the essential language of the hard of hearing and deaf, too as the individuals who can hear yet can't genuinely talk. A complex however complete language incorporates hand development, looks, and body stances. Gesture-based communication is anything but a widespread language. Every nation has its communication through signing. Each communication through signing has its syntax, word request, and elocution rules.

The issue emerges when almost senseless individuals endeavour to convey in this language with individuals who are new to its punctuation. To comprehend them, a programmed and intuitive mediator should be created.

5. METHODOLOGY

The framework will be run on a work area with a Full-HD web camera. The camera will take pictures of the hands, which will then be taken care of into the framework. It ought to be noticed that the underwriter will acclimate to the size of the edge so the framework can catch the endorser's hand direction. Figure 4 portrays the framework's calculated system. At the point when the camera has caught the client's signal, the framework arranges the test and analyzes it to the motions put away in a word reference, and the comparing yield is shown on the screen for the client.

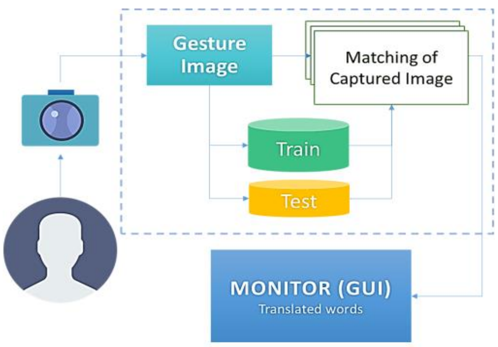

Fig. 4. Conceptual framework.

5.1 Image Processing Hand Skin Color Detection

For further developed skin variety acknowledgment, the underwriter was encouraged to have a clear foundation for the hands, as this will make it simpler for the framework to distinguish the skin tones. cv2.cvtColor was utilized for skin location. HSV pictures were made by changing over RGB pictures to HSV pictures. The HSV outline was given by the cv2.inRange work, which took as contentions the lower and upper reaches. The cv2.inRange work created the cover. The skin district of the casing was thought to be the white pixels in the cover delivered. Notwithstanding the way that dark pixels are overlooked, the cv2.erode and cv2.dilate capacities eliminate little districts that might be misleading positive skin locales. From that point forward, the part was utilized to perform two emphases of disintegrations and enlargements. At last, the covers were smoothed with a Gaussian haze.

5.2 Adding Network Layers

The objective of this review is to plan an organization that can successfully order a picture of a static communication through signing motion to its comparable text by a CNN. To accomplish explicit outcomes, we utilized Keras and CNN engineering containing a bunch of various layers for the handling of preparing information.

5.3 Training the System

The training for character and SSL recognition was done separately; each dataset was divided into two: training and testing. This was done to see the performance of the algorithm used. The network was implemented and trained through Keras and TensorFlow.

5.4 Testing Procedure

Before recognizing the signs, the user must calibrate the light to ensure that the skin masking of the hand is detected and has less noise; this can be accomplished by moving the lampshade sideways. It is preferable if the light does not shine directly on the hand. Because the system is light sensitive, determining the proper location of the lamp should be considered. If the edges of the hand in the masking are detected, the user can start using the translator. The hand should be in front of the camera for the signs to be recognized. The hand can only be detected if it is inside the box visible on the screen of a computer monitor. Because each person's hand is different in size, a user may move his or her hand back and forth to fit inside the virtual box. The user should then wait for the system to generate the desired textual equivalent of the signs. It is also advised that the user's hand remain motionless until the system generates the output.

5.5 Creating Own Dataset

Constant picture catch with Python was utilized to gather the datasets for static SLR. Pictures were consequently trimmed and changed over completely to a 50 x 50 pixel high contrast test. Each class had 2000 pictures that were then flipped on a level plane as shown in figure 5.

Python incorporates various libraries for picture and video control. One of them is OpenCV. OpenCV is a huge library that gives an assortment of capacities to picture and video errands. Utilizing OpenCV, we can catch video from the camera. It permits you to make a video catch object, which is helpful for catching recordings with a webcam and afterward carrying out the ideal strategy on that video. To get a video catch object for the camera, call cv2.VideoCapture().

Use the cv2.cvtColor() method to convert an image from one colour space to another. More than 150 color-space conversion methods are included in OpenCV. The HSV colour space is mostly used for object tracking. In Python, the inRange() function is used to mask the given image by specifying lower and upper bounds and then displaying the resulting image as an output on the screen.

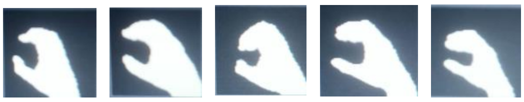

Fig.5. Sample dataset of flipped images for 'C'.

6. ADVANTAGES OF PROPOSED SYSTEM

(i) The model allows for two-way communication, which facilitates interaction between normal people and the disabled.
(ii) Using convolution, this architecture integrates color, depth, and trajectory information.
(iii) Easy to interface.
(iv) Flexible.
(v) More number of gestures can be predicted using this model.

7. RESULT DISCUSSION

The model captures the sign images from the user and predicts the respective letter corresponding to the captured sign image. Below is the list of sample outputs generated by the model while capturing images from the user. The accuracy of the model is found to be nearly 90 per cent.

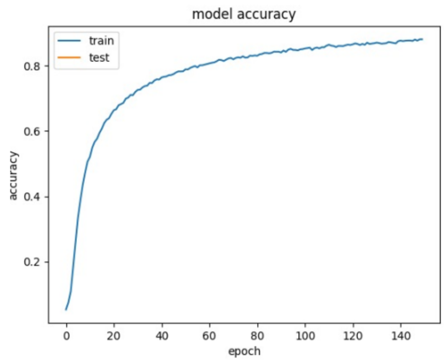

Fig. 7.1 Model Accuracy

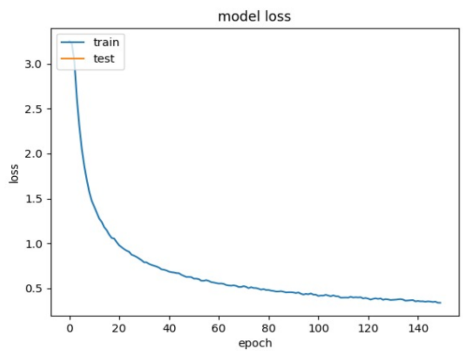

Fig. 7.2 Model Loss

7.1 MODEL RECOGNITION OUTPUTS

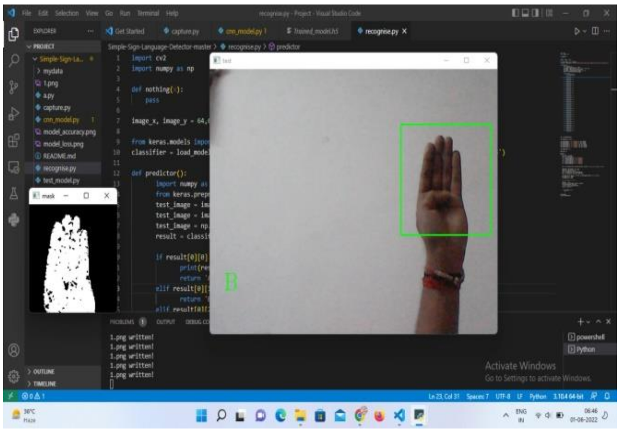

Fig. 7.1.1 Letter ‘B’

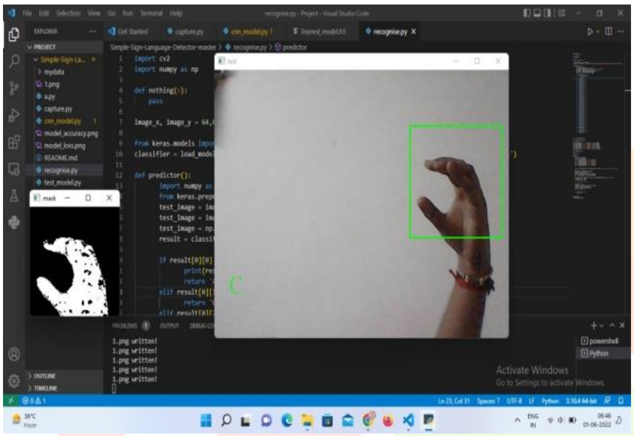

Fig. 7.1.2 Letter ‘C’

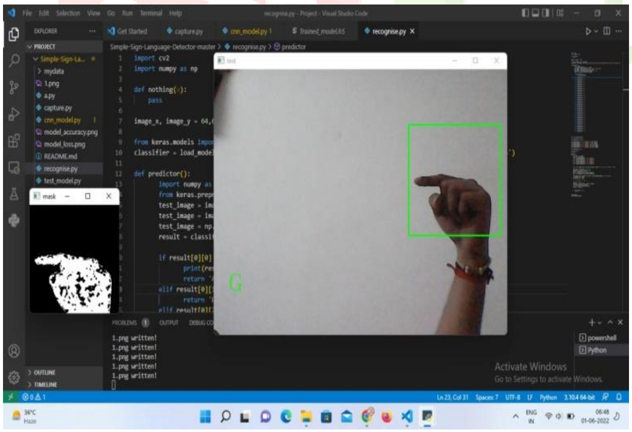

Fig. 7.1.3 Letter ‘G’

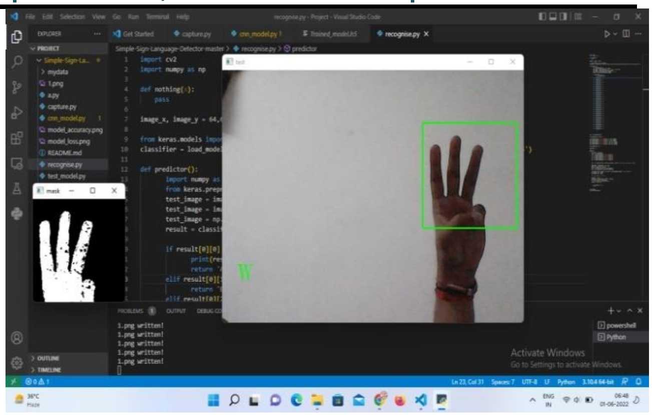

Fig. 7.1.4 Letter ‘W’

8. CONCLUSION

The task's principal objective was to make a framework that could make an interpretation of static communication via gestures into its comparing word same, which included letters and essential static signs, to acquaint clients with the basics of gesture-based communication.

The analysts conceived an assessment methodology and led a progression of tests to guarantee the meaning of the framework's non-underwriter functionalities. The consequences of the testing were remarkable concerning the framework's ease of use and learning influence. This undertaking was finished with the help and meeting of gesture-based communication specialists. Despite its average accuracy, our system outperforms existing systems because it can perform recognition at the given accuracy with a larger vocabulary and without the use of gloves or hand markings.

REFERENCES

[1] https://www.kaggle.com/datasets/datamunge/sign-language-mnist

[2] https://edu.authorcafe.com/academies/6813/sign-language-recognition

[3] https://www.irjet.net/archives/V6/i3/IRJET-V6I3639.pdf

[4] https://en.wikipedia.org/wiki/Sign_language

[5] https://scikit-learn.org/stable/

[6] https://www.geeksforgeeks.org/python-grayscaling-of-images-using-opencv/

[7] F. R. Session, A. Pacific, and S. Africa, Senat’2117, 2014, pp. 1–3.

[8] L. G. Zhang, Y. Chen, G. Fang, X. Chen, and W. Gao, “A vision-based sign language recognition system using tied-mixture density HMM,” in Proc. the 6th International Conference on Multimodal Interfaces, 2004, pp. 198-204.

[9] Q. Wang, X. Chen, L. G. Zhang, C. Wang, and W. Gao, “Viewpoint invariant sign language recognition,” Computer Vision and Image Understanding, vol. 108, no. 1-2, pp. 87–97, 2007.

[10] T. Starner, J. Weaver, and A. Pentland, “Real-time American sign language recognition using desk and wearable computer based video,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371-1375, 1998.

[11] C. Vogler and D. Metaxas, “Handshapes and movements: Multiple-channel American sign language handshapes and movements: Multiple-channel ASL recognition,” in Proc. International Gesture Workshop, 2003, pp. 247-258.

[12] T. Chouhan, A. Panse, A. K. Voona, and S. M. Sameer, "Smart glove with gesture recognition ability for the hearing and speech impaired," in Proc. 2014 IEEE Global Humanitarian Technology Conference-South Asia Satellite (GHTC-SAS), 2014, pp. 105-110.

[13] M. Süzgün, H. Özdemir, N. Camgöz, A. Kındıroğlu, D. Başaran, C. Togay, and L. Akarun, "HospiSign: An interactive sign language platform for hearing impaired," Journal of Naval Sciences and Engineering, vol. 11, no. 3, pp. 75-92, 2015.

[14] J. A. Deja, P. Arceo, D. G. David, P. Lawrence, and R. C. Roque, "MyoSL: A Framework for measuring the usability of two-arm gestural electromyography for sign language," in Proc. International Conference on Universal Access in Human-Computer Interaction, 2018, pp. 146-159.

[15] C. Ong, I. Lim, J. Lu, C. Ng, and T. Ong, "Sign-language recognition through Gesture & Movement Analysis (SIGMA)," Mechatronics and Machine Vision in Practice, vol. 3, pp. 232-245, 2018.

[16] I. N. Sandjaja and N. Marcos, "Sign language number recognition," in Proc. 2009 Fifth International Joint Conference on INC, IMS and IDC, 2009, pp. 1503-1508.

[17] E. P. Cabalfin, L. B. Martinez, R. C. L. Guevara, and P. C. Naval, "Filipino sign language recognition using manifold learning," in Proc. TENCON 2012 IEEE Region 10 Conference, 2012, pp. 1-5.

[18] P. Mekala, Y. Gao, J. Fan, and A. Davari, "Real-time sign language recognition based on neural network architecture," in Proc. 2011 IEEE 43rd Southeastern Symposium on System Theory, 2011, pp. 195–199.

[19] J. P. Rivera and C. Ong, "Recognizing non-manual signals in Filipino sign language," in Proc. Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018, pp. 1-8.

[20] J. P. Rivera and C. Ong, "Facial expression recognition in Filipino sign language: Classification using 3D Animation units," in Proc. the 18th Philippine Computing Science Congress (PCSC 2018), 2018, pp. 1-8.

[21] J. Bukhari, M. Rehman, S. I. Malik, A. M. Kamboh, and A. Salman, "American sign language translation through sensory glove; SignSpeak," International Journal of u-and e-Service, Science and Technology., vol. 8, no. 1, pp. 131–142, 2015.