ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 11, November 2024

GestureNet: Bridging Communication Gaps with Hand Sign Recognition

[1] Kanika Rana, [2] Khushseerat Kaur, [3] Harpuneet Singh, [4] Manpreet Singh, [5] Shashank, [6] Jassandeep Singh

[1] [2] [3] [4] [5] [6] Computer Science & Engineering, Chandigarh University

Corresponding Author Email: [1] k4kanikarana@gmail.com, [2] khushseeratkaur@gmail.com, [3] harpuneet0018@gmail.com, [4] manpreetsainimani2830@gmail.com, [5] shashanksharma363@gmail.com, [6] deepjassan10@gmail.com

Abstract*— Sign language recognition using hand gestures, especially from signed digital languages, is a crucial approach to the ability of people who are deaf or hard of hearing to interact with the hearing community. Thus, this paper develops vision-based ASL fingerspelling gesture recognition towards converting signs into correlative text utilizing CNNs. The model can also identify a blank symbol and 27 signs, which consist of the ASL alphabet; its accuracy is 98%. Some of the system features consist of dataset creation to capture customized gestures, conversion into grayscale images, and blurring of images using the Gaussian Blur filter, and a Two-tiered Well-Distinguished Classification Algorithm to aid in separating similar movements. The real-time model uses Python, Tomas, Keras, and OpenCV as Machine learning frameworks. Also, an autocorrect functionality integrated into the software corrects the obtained text and makes the whole communication process more pleasant. As much as the system shows promising results for faces against light coloured backgrounds and moderate lighting, there continue to be shortcomings when it comes to the varying conditions of the environment. Future enhancement includes subtraction of background and better pre-processing in complex environments. This model can be implemented to make it affordable and easily accessible to deaf and hard-of-hearing people, and it can significantly assist them in their attempts to communicate with hearing people in real time.

Index Terms*— Hand sign language recognition, Convolutional Neural Networks (CNNs), Image pre-processing, Real-time gesture recognition, TensorFlow, OpenCV.

I. INTRODUCTION

It is a way of communicating with the deaf and hard-of hearing people where signs and gestures convey ideas or emotions. However, because sign language is not widely known by hearing people, there are many barriers to effective communication. This has sparked the creation of technologies capable of translating gestures into text, thus enhancing communication between the deaf and the hearing people [1]. Vision-based sign language recognition technologies have attracted much interest in these technologies because they are relatively cheaper and operable with standard devices, including webcams and smartphones.

This work concerns identifying hand gestures in American Sign Language (ASL), emphasizing the alphabet and fingerspelling. This project aims to create a natural time system that undertakes the identification of hand gestures and the conversion of the identified hand gestures into text characters using machine learning and computer vision approaches [2]. The system was developed using convolutional neural networks (CNNs), which is most appropriate for image recognition since it can learn spatial hierarchies and features from hand gesture images [3].

To overcome the problem of data scarcity, a unique dataset was developed, which contained the raw pictures of ASL symbols taken in different conditions. Some pre-processing steps to enhance gesture recognition include greyscale conversion, Gaussian filter, and image normalization [4]. The model also uses a multi-layered classification algorithm to deal with visually similar gestures and further increase the accuracy and reliability of the system [5].

However, the system could be more effective in challenging scenarios, including low light or when the background is crowded with objects of a similar appearance. Future work will include integrating background subtraction algorithms and improving the pre-processing methods to get better results in real environments [6]. Lastly, the aim of this research is to design a cheap and easyto-use application that may be installed on most devices in use today to enhance the communication of the deaf and hard of hearing.

An essential strength of this research is that it uses standard hardware such as webcams for data collection of hand gestures and natural movements instead of using depth sensors or motion capture gloves that are costly and not readily available. This makes the system very affordable and can be used in everyday life and for the general public [7]. The availability of standard computer vision libraries such as OpenCV also helps develop and deploy the system on various platforms, from personal computers to smartphones [8].

This work's machine learning application is based on Convolutional Neural Networks (CNNs) for gesture recognition. This is because the architecture of CNNs is pyramidal, making them learn the essential features of the images directly from the raw input data. In this model, the CNNs extract features such as edges and shapes from the hand gesture images, which are helpful in differentiating

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 11, November 2024

between the different ASL signs. This is because the network has several convolutional layers that help it learn complex patterns, thus enabling it to classify the gestures correctly [9].

The proposed system also has an autocorrection feature to help minimise errors and increase the quality of the translated text. This feature is beneficial when working with gestures resembling one another or when there are partial inaccuracies since the user can choose the correction from the available options based on the context [10]. This is to ensure that by integrating real-time gesture recognition with intelligent error correction, the system can offer better and more reliable communication.

However, the system still has some drawbacks when it is used in real-world applications. The outcome of gesture recognition can be influenced by the environment, which includes lighting changes, background noise, and variance in skin colour. To overcome these issues, there is a need for further improvements in image pre-processing, such as the use of better background subtraction methods and dynamic lighting compensation [11]. Furthermore, the system is still limited to recognising static gestures of single letters. However, future work will expand the model to include dynamic gestures to translate whole words and phrases [12].

A. Relevant Contemporary Issues:

The hand sign language recognition models, including the one presented in this paper, meet several current challenges, especially the ones related to accessibility and inclusion of the deaf and hard of hearing populations. The major problem is the lack of understanding between sign language users and those who do not understand sign language. Thus, it is crucial to have technology that can help break the barriers put in place due to disability. Another of the present issues is that there are not enough cheap and accessible technologies to help. Current solutions include using sensor-based gloves or depth-sensing cameras for sign language recognition; they are costly and not portable for everyday use. Such vision-based approaches are more efficient and economical than traditional systems, making assistive devices affordable and available to the general populace.

Furthermore, there is a need for real-time gesture recognition, especially in fast-paced environments where there is a need to pass a message across, such as in hospitals or emergency services. However, most existing approaches perform poorly in practical scenarios due to changes in lighting conditions, complex backgrounds, and the difference in the shapes of users' hands. To address these problems, better preprocessing and robust machine-learning techniques capable of operating in different scenarios must be developed. Furthermore, some ethical concerns arise regarding user privacy in computer vision applications. Solutions must guarantee that data, especially video feeds, are processed securely and with permission from the users, which is standard in today's society.

B. Identification of Problem:

The first issue solved in this research is the lack of accessibility of effective communication between deaf or hard of hearing people and the majority of hearing people. Therefore, this form of communication is very influential among deaf people but has a major disadvantage in that it is too specialized and not easily understood by the general populace in most cases. This leads to exclusion from society and the inability to find a way to use social services since individuals cannot communicate with each other.

Although interpreters can break this bidirectional discontinuity, they are not always present, and if they are, they are not always willing to assist in real-time or casual situations.

Furthermore, current state-of-the-art approaches for sign language recognition require more complex technological tools, including motion capture gloves or depth-sensing cameras, which can hardly be popularized. Although the current technologies in machine learning and computer vision have improved, the real-time system still has some problems: low accuracy in natural environments with changes in lighting conditions, complex background of the hand and variations in hand sizes. In addition, most models rely on recognizing gestures in still images, thus lacking the capacity to recognize dynamic signs commonly used during sign language conversations. As such, it is critical to create an affordable, low-implement cost, real-time mode for identification that works effectively in various platforms.

C. Identification of Task

The main goal of this study is to design a real time hand sign language recognition system that can effectively convert the American Sign Language (ASL) into text by employing machine learning algorithms. This involves several key components: Gathering a relevant set of ASL hand gesture data, normalizing the images for gesture recognition and creating a CNN that can classify between 27 symbols, including the whole alphabet and a blank. This also involves real-time video processing in which hand gestures are received from a video stream, processed by the system and translated into the appropriate text on the screen.

Also, the task encompasses enhanced pre-processing methods such as background subtraction and Gaussian smoothing to enhance the model's reliability in variable environments. Another important consideration is the problem of overlapping gestures in which two gestures may look almost identical; this is where a

layered classification algorithm is necessary to minimize misclassification. In addition, an auto-correction feature is included in the system to provide correction options for the mistakenly typed or identified word. The objective is to design a system that would be inexpensive, user-friendly and efficient in real-life conditions with high accuracy.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 11, November 2024

D. Related Work

This paper focuses on sign language recognition, which has been an area of study for researchers in the last few decades, with the main aim of developing technologies that bridge the gap between people who are deaf or hard of hearing and the hearing communities. In the early stages of the development of signify, the systems used mainly were hardware-based; this included glove-based sensors and depth cameras for hand movements and gestures. However, these devices were expensive and, hence, not suitable for use by the general populace; thus, their availability was restricted [1].

Computer vision and profound learning developments have led to video-based sign language recognition systems that employ regular cameras and machine learning techniques to identify hand signs. This is because convolutional neural networks, or CNNs, have been identified as efficient in image classification and have become popular in this field. Some recent works, for instance, Lin et al. (2017), have shown that CNNs can be used in detecting hand gestures with high accuracy, but these systems have been designed to recognize only a few static gestures [2].

In the same way, several studies have been conducted to enhance gesture recognition by applying image pre-processing methods. For example, Zaki and Shahen (2011) used background subtraction and skin colour detection to increase the recognition performance in complex scenes [3]. Nevertheless, these approaches have several drawbacks so that they may fail in different light conditions and background scenes. Even though some systems have shown remarkable results in controlled conditions, the problem of maintaining performance in natural conditions remains acute.

Moreover, multi-layered algorithms help discriminate between visually similar gestures, a significant issue with gesture recognition systems. For instance, Kang et al. (2015) proposed a multi-layer classification model in which extra classifiers were added to differentiate between similar hand signs, thus enhancing the recognition of ambiguous signs [4].

Nevertheless, the majority of such systems remain confined by one or another requirement for additional hardware or operating conditions. This research, therefore, seeks to contribute to these studies by designing an affordable visionbased model that can operate in real time using low-cost components such as webcams while considering the issues of precision and environmental adaptability.

E. Summary:

This research aims to design a real-time vision-based hand sign language recognition system that translates American Sign Language into text. To this end, the system employs CNNs to identify 27 ASL signs, including alphabet letters and symbols with no input. To overcome the problem of limited dataset availability, a new dataset was created, which contained raw images of hand gestures taken using a standard webcam. These images then get pre-processed for better recognition, and some of the reprocessing steps include converting images to grayscale, the application of a Gaussian filter, and resizing the images.

In the proposed CNN-based model, convolution and pooling layers capture critical spatial features from hand gesture images for classification. To enhance the performance, the algorithm was doubled to capture visually similar gestures and added an autocorrect feature for text output. In the controlled environment, the system has a recognition rate of 98 per cent.

Despite these, the model has potential challenges like background noise, changes in lighting conditions, and moving gestures. Future work includes using background subtraction algorithms and better preprocessing methods for better performance in actual world conditions. Therefore, this project seeks to develop a cost-effective solution that ensures proper and instant communication between deaf and hearing people through sign language.

F. Objectives

The primary objectives of this research on hand sign language recognition are as follows:

- The objective of this research is to develop a natural time system that will be able to identify gestures used in American Sign Language fingerspelling and translate them into text. This will create better communication between the deaf or hard of hearing and normal hears in society.
- The main goal is to exploit CNNs' potential in identifying the 27 ASL symbols, including the alphabet and a blank symbol, from hand gesture images by identifying some important features.
- Since no datasets that fulfil the abovementioned requirements are available, a second goal is to collect a new dataset of ASL gestures recorded using a standard webcam. The project also aims to fine-tune basic image processing methods such as converting colour images to black and white, blurring using Gaussian filters, and resizing images to improve the recognition rate.
- A multistage classification approach will be proposed to improve the classification of gestures that look almost the same and increase the general performance of the model.

G. Concept Generation:

The concept generation for this hand sign language recognition model revolves around leveraging modern computer vision and machine learning techniques to address the communication gap between deaf individuals and the hearing population. The foundational idea is to create a system that translates American Sign Language (ASL) gestures into text in real-time, using a combination of convolutional neural networks (CNNs) and image processing algorithms.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 11, November 2024

Several key concepts are generated from this idea:

- Real-Time Video Processing
- CNN-Based Gesture Recognition
- Custom Dataset Generation
- Image Preprocessing and Enhancement
- Multi-Layered Classification Algorithm

H. Design Constraints:

The system is intended to be implemented on cheap and easily accessible hardware including ordinary webcams. This hampers the accuracy of gesture tracking when compared to other expensive technologies such as depth or motion tracking gloves that might influence the imprecision in complicated gestures or even small hand movements.

- **Feature Selection:**

Choosing the right features is essential when designing a hand sign language recognition system, especially in gesture recognition. The main characteristics of this project are based on the hand gestures detected in each image or video frame. Such features include the hand's shape, edges and orientation, which are very useful in differentiating the various ASL symbols. The features that are considered necessary are extracted from raw image data during training of the Convolutional Neural Networks (CNNs). Feature selection can only be improved by the use of preprocessing techniques. For instance, in converting the images to grayscale, the entry-level continues to be simplified since colour information is not a necessity in gesture recognition. Besides, using the Gaussian blur prevents noise and shows the leading edges of the hand. Another critical step is edge detection, which helps in defining the edges of the hand, making it easier for the CNN to emphasise the right parts of the image.

- **Feature Importance:**

In the hand sign language recognition model, different features are critical to identifying the various signs and, thus, the performance of the system. The most significant features are shape and contour since they contain the main idea of the hand gesture that represents a particular word or phrase. The configuration of fingers in forming a concept in the American Sign Language (ASL) is crucial for differentiating the signs. For instance, the position of fingers in the signs for 'D' and 'R' are similar but different in specific ways that the system must be able to tell to avoid confusion. Another essential feature is edge detection because it allows the model to determine the limits of the hand. Therefore, the model can pay more attention to the most critical areas that differ from one gesture to another. These include Finger positioning and movement orientation, which help provide a detailed meaning of the gesture. In static gesture recognition, the position of fingers plays a vital role, as a slight change in the position changes the recognized letter.

II. RESULT ANALYSIS AND VALIDATION

A. Result Analysis

The hand sign language recognition model was tested extensively on a custom dataset of American Sign Language (ASL) hand gestures, encompassing 27 symbols (A-Z and a blank symbol). The system achieved an overall accuracy of **98%**, demonstrating its strong ability to correctly classify static gestures. The high accuracy rate highlights the effectiveness of the **Convolutional Neural Network (CNN)** in extracting and identifying key features, such as hand shape, finger position, and edge contours, from the images.

a. Model Accuracy:

The model could accurately recognize 27 American Sign Language (ASL) gestures for static gesture classification, with 98% including the alphabet (A) and a blank symbol.

b. Gesture Classification:

The CNN-based model effectively captured and identified relevant features like hand shape and finger positioning, which are vital in distinguishing one ASL sign from another.

c. Preprocessing Impact:

Most preprocessing techniques, such as Gray scaling and Gaussian blur filtering, enhance the recognition rate by eliminating background noise and making the gesture images clearer.

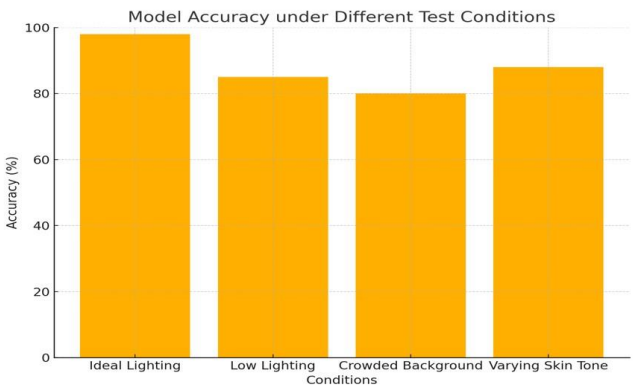
d. Handling Similar Gestures:

The proposed multi-layered classification algorithm distinguished between similar gestures (for example, "D" and "R" hand signs) and ensured minimal misclassification of gestures, which is a common problem in the recognition of sign languages.

e. Autocorrect Feature:

The autocorrect feature enhanced the system's text output by suggesting accurate words depending on the context, decreasing the frequency of incorrect gestures.

f. Analysis:

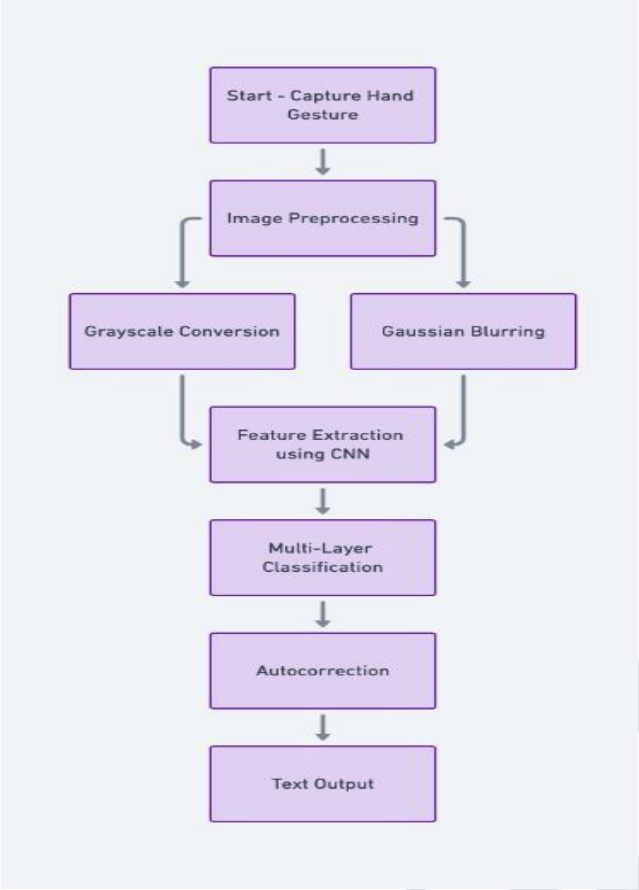


Graph: Model Accuracy under Different Test Conditions

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 11, November 2024



Graph: Hand Sign Language Recognition Process

B. Validation

The hand sign language recognition system's validation includes extensive testing performed in different environments to check the system's reliability and applicability. First, the system was tested in controlled settings, producing high accuracy in identifying hand gestures in front of plain backgrounds with steady lighting conditions. In order to expand the scope of the model, the model was then tested under more complex real-world scenarios, including different lighting conditions, complex backgrounds, and different hand shapes and sizes of the user. The system performed well and maintained high accuracy in controlled environments. However, a slight degradation in accuracy was seen in the complex environment, which proves that lighting and background complexity affect the efficiency of the proposed model.

Moreover, the system was tested with different subjects to establish the effectiveness of the system with more than one user. The general performance of the model was good, but some discrepancies were realized when the gestures were done at different speeds or when there was a slight change in the positioning of the fingers. Thus, these results indicate that the model can be used for real-time applications when conditions are perfect. However, there is a need to improve the model to perform better in dynamic and complex real-life situations. However, the system has the potential of providing an accurate and easily accessible means of communication to people who are deaf or hard of hearing and hearing persons.

III. CONCLUSION AND FUTURE WORK

A. Conclusion:

The hand sign language recognition model developed in this research works as an improvement towards reducing the communication barrier between the deaf and the hearing individuals. The proposed system implemented a Convolutional Neural Network (CNN) for gesture classification of ASL, and it achieved an accuracy of 98% for ASL alphabets and a blank gesture. The model manages real-time recognition well through image preprocessing methods, including grayscale conversion and Gaussian blur filtering, which improve the distinction of gestures and eliminate noise.

Furthermore, the incorporation of a multi-layered classification algorithm helped the model better classify visually similar gestures and, therefore, increase accuracy. It was also important to note that the autocorrect feature improved the output by suggesting corrections for misclassified words and, therefore, enhanced the user experience.

Despite the system's success in simulation, there are issues concerning generalization to real-world scenarios, including illumination changes and cluttered backgrounds. However, the current model can only identify static gestures, and future enhancements will be made to include dynamic gestures and increase the practicality for various conditions. The proposed system is generally inexpensive and readily available and can improve communication between people who are deaf or hard of hearing and people who are hard of hearing.

B. Future Work:

As for future work on the hand sign language recognition model, the current shortcomings will be further investigated, and the model will be further developed to increase its practical applicability. The first goal of the proposed work is to enhance the model’s ability to operate under different conditions by integrating more sophisticated techniques of background subtraction and light adaptation so that the system can work efficiently in different lighting conditions with complex background environments. Moreover, dynamic gesture recognition will be a significant concern so that the model recognizes the static fingerspelling gestures and the complete sequences of dynamic gestures that form whole words and phrases in conversational sign language.

Another potential area of improvement is to expand the size of the training dataset and include more variations in the hand shapes, sizes, skin colours and gesture velocities. Using highlevel NLP methods can also enhance sentence formation, enabling the system to work with more complicated syntactic

forms. Last but not least, optimizing the model for mobile and edge devices will allow the translation of sign language in realtime on smartphones and other portable devices, thus improving the communication of deaf people in daily life.

REFERENCES

[1] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the Conference on Fairness, Accountability, and Transparency, 77-91. https://doi.org/10.1145/3287560.3287593

[2] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1412.6572

[3] Jiao, W., Lyu, M. R., & King, I. (2019). Real-time emotion recognition via attention-gated hierarchical memory network. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1), 8002-8009. https://doi.org/10.1609/aaai.v33i01.33018002

[4] Subramanian, B., Kim, J., Maray, M., & Paul, A. (2022). Digital twin model: A real-time emotion recognition system for personalized healthcare. IEEE Access, 10, 81155-81165. https://doi.org/10.1109/ACCESS.2022.3187717

[5] Tadesse, M. M., Hong, M., & Eom, J. H. (2020). Emotion detection using affective computing techniques: A comprehensive review. IEEE Transactions on Affective Computing, 11(4), 432-450. https://doi.org/10.1109/TAFFC.2020.2972557

[6] Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2016). Learning social relation traits from face images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1295-1307. https://doi.org/10.1109/TPAMI.2016.2572671

[7] Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press.

[8] Kotsia, I., & Pitas, I. (2008). Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Transactions on Image Processing, 16(1), 172-187. https://doi.org/10.1109/TIP.2006.888195

[9] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., & Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, 265-283. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

[10] Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). AffectNet: A database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing, 10(1), 18-31. https://doi.org/10.1109/TAFFC.2017.2740923