

# Recognition Of Hand Gesture Using CNN for American Sign Language

**HADEEB KHAN**

St. Michael's Convent School, Karachi, Pakistan

Email: [khadeeb@hotmail.com](mailto:khadeeb@hotmail.com)

**SUFYAN SIDDIUI**

The Dawood Foundation, Karachi, Pakistan

Email: [sufyansiddiqui67@gmail.com](mailto:sufyansiddiqui67@gmail.com)

**SYED MUHAMMAD BILAL MUBEEN**

Taqwa Model School, Karachi, Pakistan.

Email: [syedbilalmubeen@gmail.com](mailto:syedbilalmubeen@gmail.com)

Received: 1<sup>st</sup> July 2022; Accepted: 12<sup>th</sup> December 2022; Published: 30<sup>th</sup> December 2022

**Abstract ---** We have proposed gesture recognition system for hand gesture in this paper which would make communication easier for people who have problem of hearing impairment, and they will have more opportunities to interact and deal with the outer world because not everyone can understand the sign language so advancement in this would make a great social impact for such patients that they will easily be able to carry out their thoughts. For recognition of sign language alphabets. There are four modules on which our system works which includes hand tracking and segmentation, feature extraction, recognition of gesture, application interface. HSV (Hue Saturation Value and Camshaft method is utilized for hand tracking and segmentation. We have implemented CNN (Convolutional Neural Network) for recognition of gesture. We have proposed a gesture recognition system which is not expensive, and it is also easy to use which works for single hand gesture recognition. The system will be helpful for greater number of hearing-impaired people to be able to easily communicate with other people. The paper is divided into various section which has the introduction at first, following to it the literature review, then it contains methodology and result discussions, and lastly the conclusion with future scope.

**Keywords-** American Sign Language, hand capturing, hand segmentation, feature mining, CNN, Recognition of gesture, Single hand recognition.

## 1. Introduction

One of the most common and well-known disability which exist in the entire world is hearing impairment. About 2.17 percent of the total population of United States of America are the people who have hearing impairment as per the report presented by American hearing loss association [1]. There are various sign languages which have been built for hearing impaired people to express themselves and American sign language is one of them and our system is working on the same language, and it converts the given input into English which is spoken and understood almost globally [2]. The reason behind selecting American sign language is that it is quite evolved sign language which makes it a reliable source for the system to be trained with for hearing impaired persons. A hearing-impaired person can write to talk but writing does not fully express the emotions of the speaker and it is a frustrating and time-consuming way for them to communicate with other people. That is why sign language is prioritized because it is more expressive and fast way to communicate when a person cannot speak or has hearing impairment [3].

A very small amount of work has been carried out in this domain up till now for the development of such software, systems, and models which takes sign language as an input and converts it to voice or text form. It has been proposed to build more capable communication system for hearing impaired persons [4].

The purpose of this research is to design a system for gesture recognition which initially works for one hand either left or right. We have used the web camera of laptop for image capturing. Our main purpose is to make the system so much capable

that it completes all the modules of gesture recognition within a second. The whole system is divided into four parts: Hand capturing, segmentation, feature mining, Gesture recognition, and Interface of application [5].

## 2. Literature Review

Capturing an object can be done with various techniques which are kernel, silhouette and many more ways to capture and track an object. It has been observed during the study of this paper that camshaft is the most promising technique for obtaining good performance in real time environments amongst the different object capturing techniques [6]. 87 percent of the error is classified when implemented Kalman filter, the rotational objects can be classified by the use of SURF (speed up robust feature) whereas it flunks when it is used for high scaling different objects. Zhingeng, Lui Voila Jones, velocity weight, and cue color are few of the various hand capturing techniques which have also been observed. To achieve better accuracy and keeping the economical factor in view, it has been observed that Camshaft is well suited for our system as it can easily work with web camera [7].

There are multiple ways which can be used for dividing the image into many parts such as hue saturation value model which detects the random images, and it works best for objects which contains noise in it. The algorithm uses the approach of green (Y), blue (Cb), red (Cr) which detects each and every pixel of the object and find out the similar regions to pack them together [8]. After analyzing many algorithms for segmentation so it was finalized on the basis of results that hue saturation value model works properly with skin color so it can be used for dividing the image into pixel. Extracting the accurate shape from the captured and segmented image is one of the vital parts and many ways are found to do so. But GFD (generic Fourier descriptor) works best for locating or retrieval of features on basic of accurate shape of the captured hand [9].

When dealing with hand gestures so it means data dealing as well and there exist multiple techniques such as PCA (principal component analysis) and HMM (hidden markov model) which are used for feature extraction and handling ample quantity of data respectively [10]. Algorithms for gesture recognition are used to perform various operations such image classification, image segmentation, image feature extraction and so on. Approaches like ANN and GA (artificial neural network and genetic algorithm respectively) have been quite useful for recognition of random and complex pattern, but these techniques have been widely used over the past few years and it has produced promising outcomes because it works well for noisy objects as well [11]. These approaches have many benefits such as it is easily understandable and works well for random moving objects because hand gesture also means random movement with noisy environments.

The research is based on Alzheimer's disease affecting the ability of a person to perform daily life activities by targeting the memory of that person. It is divided into three stages: early, middle, and late stage. The research also indicates that in the age of 50s, 30% people are affected by Alzheimer's disease with down syndrome and the percentage gets closer to 50% when they are at the age of 60s. The hippocampal structure analysis is used to diagnose the cause of Alzheimer's disease. Image segmentation technique is used to divide an image into multiple segments to make it more understandable and easier to analyze. Medical image segmentation is done to partition the data into continuous regions to represent an object on individual anatomical object basis.[12] The technique markovian growing, which is used for segmentation, so this method did not use the shape priors instead they used the technique of growth with weak and relational geometric priors on it. These algorithms provided satisfactory outcomes having run on workstations.

In this method, watershed algorithm is used for segmentation whereas anisotropic diffusion filter was brought in use to remove the random noise. The Alzheimer's disease was detected through following brain analysis:

- Provide the image of brain.
- Bottom hat operation is applied on the brain image.
- The image is then divided into 3\*3 blocks size to extract the middle block.
- The image is converted into binary one.

- The edges are retrieved by applying the edge detection technique.
- The noise is removed, and we finally get the Hippocampus.

Once segmentation is done so then classification is carried out on grey, white, and cerebrospinal fluid features of brain. The most efficient morphological tool for image segmentation is Lucy Watershed Transformation. A marker image is created by this technique which provides even zero marker values of pixels. To achieve the efficient and reliable results from watershed algorithm, it is divided into two sections. Object identification is done through **internal markers** whereas **external markers** are used outside object classification.[13]

Image segmentation techniques is used over the wide range and in this study, radionuclide ventriculography is focused. The study provides information about left and right ventricles of the heart. Rest of the study is divided into different sections which includes basic principles of segmentation method, explanation of suggested methods, experimental results, and conclusion. It collects all the pixels which have common properties and it directly move to the division of image making each pixel a single region. Thresholding is mainly used to classify the images into different classes. Thresholding can be done in two possible ways: either bi-level thresholding or multi-thresholding. The bi-level thresholding has greater gray value whereas the other one has lower value. Canny operator is one the standard edge detection method.[14] It provided three different criteria for image edge detection:

- Good detection defined that at edge, we must have maximum signal to noise.
- Good localization provided the concept that the edges of an image must be closer to the true edge center.
- The detector must only look for single edge because there is always one edge for each object in the image. [15]

The outdoor images are classified into two categories. One is unstructured which consist of ground, sky etc. And the second one is structured objects which includes people, vehicles etc. Segmentation works on intensity level similarity and dissimilarity. Similarity divides the image in regions which are similar whereas dissimilarity defines the sudden changes which occur in the intensity of an image. Global threshold technique is used for gray images and local threshold technique is used for colored images. Clustering method is also used in this paper to make the pixel values of images of small clusters for similar value pixels. Another technique which is used in this paper is graph cut which is applied over greater diversity of pixel value. The last technique used in this research is graph cut based on color which identifies the foreground and makes the background darker. All of the above-mentioned techniques are used to separate the background and foreground of an image.[16] The graph cut technique is used to show the difference of the changes which occur between objects and highlights those changes by drawing a boundary on it. One has to find out by looking at the image that which portion is background, and which is foreground. The background is then turned black whereas the foreground is kept with the colors. Graph cut based on color information method performs task within single computation and produces good result. Clustering is the method of grouping pixels into similar and dissimilar category. There are various techniques of clustering, but k-mean clustering is widely used because it is considered compatible and standard one. [17] In k-mean clustering, a color is given to every cluster to differentiate between each other.

An image processing technique which was recently designed for segmentation of IUVS images. It collects information through the frequencies of range 20 to 45 MHz which are generated by using dissimilar transducer operator for the borders of luminal in IVUS image detection. All the challenges and weaknesses which occur during its process are discussed in this paper. Different detection techniques are used for border detection which includes gradient based, edge tracking, active contour based, statistical-probabilistic based and multiscale expansion-based techniques. Edge detection technique is used by generating IVUS signal which provides distinctive edge patterns. For refinement of the edges and reduction of noise different hybrid algorithms have been developed to collect more pleasant edge results. [18]

MRI images of brain are segmented through morphological operations. It deals with the concept of set theory of image processing and analysis. Nonlinear transformations are part of morphological filters which are used to modify the images geometric features. Morphological operations are performed on 0 and 1 like only on binary images. It produces another image through original image of size which is required for structuring element. This process is also called mathematical morphology. MATLAB tool is used for morphological implementation of MR images and for comparison EM and FCM segmentation methods were brought in use. It has been analyzed that there is no specific technique for brain MRI

segmentation because it is very difficult to decide a proper method for any image. [19] Through the study, it was seen that morphological operations are pretty fast and effective for image detection dataset.

### **3. Methodology**

The web cam is used to take the input of hand gesture. Tracking of hand would be done using the CNN technique. HSV color model is used to carry out segmentation of hands. Certain features are kept in mind to represent the segmented image. The features gathered are utilized for gesture recognition by Gesture recognition Algorithm to attain accurate and optimized outcome.[20] The outcome which has been obtained so it is turned into voice or text form. It is classified into four parts:

- Hand capturing.
- Segmentation of hand.
- Feature mining.
- Recognition of gesture.

CNN seems to be one of the best techniques for manipulating the random behavior of natural selection and gesture recognition. It is widely used as it works best for the recognition of images and it belongs to the class of deep learning.

#### **3.1. Hand Capturing**

The general camshaft algorithm which has been used for hand tracking from captured images is defined below that how it is done:

- Select the hand area from the captured image.
- Make the hand region colorful and draw its histogram.
- Now use the histogram to make frame of probability distribution.
- Process the frame which has been made through histogram.

#### **3.2. Segmentation**

We used HSV color model to segment the hand from the background after the hand is tracked.

#### **3.3. Mining Features**

In mining process, all the features are extracted with their generality and relationship which is found between classes and features. It is one of the most important parts because it allows us to extract each and every factor which is important in the recognition of gestures.

#### **3.4. Recognition of Gesture**

After the process of featuring mining. All the extracted features are used to recognize the gesture based on the features with having similarity to that which has been extracted and stored. This is the final part in the whole process of recognition system.

### **4. Results and Discussions**

Convolutional Neural Network (CNN). This is used most of the time to make evaluation for visible photographs and this belongs to the class of deep neural networks. CNN has widespread application and be discovered widely in the domain of classification of pictures, photo evaluation, recognition of photographs, picture prediction, advice system, textual content prognosis, recognition of video material, processing of herbal language and a lot of more such things. Additionally, it is a regularized model which works on the concept of multilayer perceptron but in this model all the neurons are connected with each other which makes the model totally linked however, it goes by different approach for regularization of data. Convolutional neural network utilizes comprehensive and minor structures whose shape is in hierarchical form of statistics hence, this makes CNN to be less connected as compared to multilayer approach. CNN is the concept which is used to be

pushed from organic system which offers the affinity amongst various neurons to different components in the model. Convolutional neural network makes less use of pre-processing approach in comparison with other algorithms which perform classification of pictures thus this makes CNN adopt the matters greater correctly other than different usual algorithm approaches that utilizes much of the time to process the facts and data. One of the subtle benefits of CNN is, the human efforts which are required in CNN are impartial as compared to other algorithms. CNN is additionally recognized as area invariant synthetic neural networks (SIANN). CNN technique is implemented because it is considered to be the most precise and best working technique for recognition of hand gesture because it works best for the random processes by generating the best results.[21] It works in the following process.

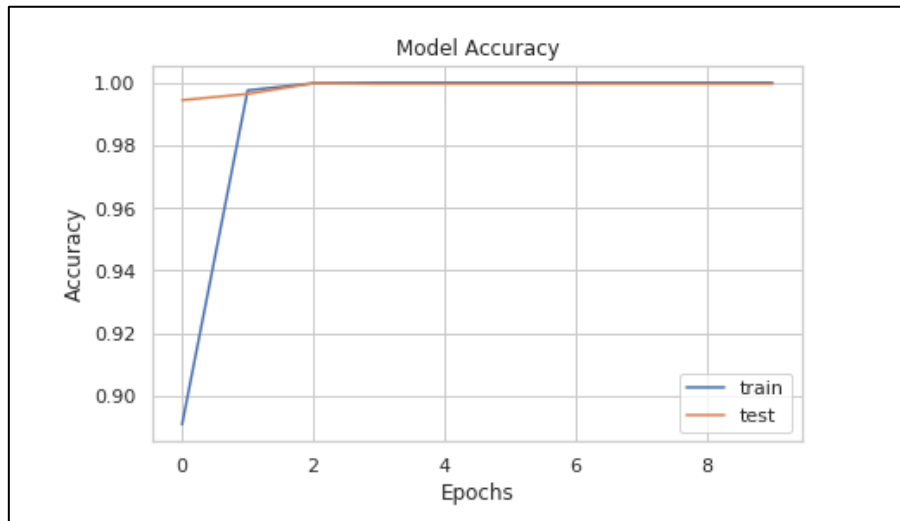


Fig.1. Model Accuracy

It can be clearly seen in figure 1 that we achieved accuracy of about 99% through the use of CNN because as we know that sign language patterns may vary from hand to hand, so CNN works best for such random behavior, and we have been able to attain 99% accuracy through it. The dataset was taken from the Kaggle to train and test the model. The overall dataset was divided into 3:7 ratio for testing and training respectively. The data was divided into this ratio because it allowed to train the model over more tuples of data so that accurate results can be generated when testing the model.

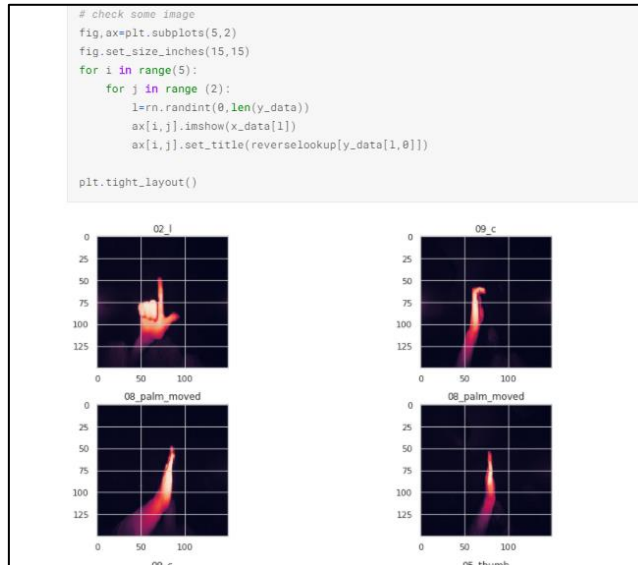


Fig.2. Hand Gesture Images

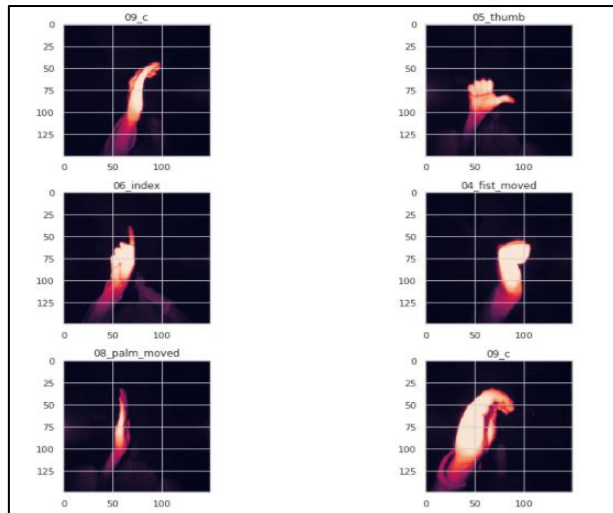


Fig.3. Hand Gesture Images

It is prominent in figure 2 and 3 that how the movement of hand is captured in the system which is then segmented and features are extracted from the image to detect that what is being said. Webcamera is used to capture these images and these are few of the alphabets which are part of the sign language. In figure 2 and 3, each captured image depicts the different alphabet of sign language.

## 5. Conclusion and Future Work

In this paper, we have worked on American Sign Language recognition through the use different algorithms to attain the best accuracy which could be implemented in real time systems. It can handle various hand gesture and by taking the hand gesture input, it is going to provide the meaning of that gesture as an output. It can work for recognition of single hand gesture.

The system is trained to recognize the hand movement which is not covered with any cloth or something. It is not expensive at all that is why it is going to facilitate hearing impaired persons to a great extent. American sign language is effectively

recognized by the proposed system. In future, the system will be trained for gesture recognition of both of the hands in real time environment. It can be expanded in a way that the same system can be enhanced to translate speech into sign language.

## Acknowledgement

The authors will take this opportunity to show utmost appreciation to our supervisor who helped us in finalizing this paper. We are also grateful to the institution which provided us with all the technical support which was needed for completion of the paper. We would also thank our friends and family for supporting us morally and offering deep insight into the study.

## References

- [1] A. El-Sawah, C. Joslin, N. D. Georganas, and E. M. Petriu, "A Framework for 3D Hand Tracking and Gesture Recognition using Elements of Genetic Programming," in *Fourth Canadian Conference on Computer and Robot Vision (CRV '07)*, Montreal, QC, Canada, May 2007.
- [2] S. DiPaola and O. N. Yalcin, "A multi-layer artificial intelligence and sensing based affective conversational embodied agent," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, Cambridge, United Kingdom, Sep. 2019.
- [3] P. Kumar, H. Gauba, P. Pratim Roy, and D. Prosad Dogra, "A multimodal framework for sensor-based sign language recognition," *Neurocomputing*, vol. 259, pp. 21–38, Oct. 2017.
- [4] L. Haldurai, T. Madhubala, and R. Rajalakshmi, "A Study on CNN and its Applications," *International Journal of Computer Sciences and Engineering*, vol. 4, p. 6, 2016.
- [5] R. Bryll, R. T. Rose, and F. Quek, "Agent-Based Gesture Tracking," *IEEE Trans. Syst., Man, Cybern. A* vol. 35, no. 6, pp. 795–810, Nov. 2005.
- [6] F. Ullah, "American Sign Language recognition system for hearing impaired people using Cartesian Genetic Programming," in *the 5th International Conference on Automation, Robotics and Applications*, Wellington, New Zealand, Dec. 2011.
- [7] S. Rajaganapathy, B. Aravind, B. Keerthana, and M. Sivagami, "Conversation of Sign Language to Speech with Human Gestures," *Procedia Computer Science*, vol. 50, pp. 10–15, 2015.
- [8] M. Salem, S. Kopp, I. Wachsmuth, and F. Joublin, "Generating robot gesture using a virtual agent framework," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Oct. 2010.
- [9] G. Kosta and M. Benoit, "Group Behavior Recognition for Gesture Analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 2, pp. 211–222, Feb. 2008.
- [10] A. S. Ghotkar, R. Khatal, S. Khupase, S. Asati, and M. Hadap, "Hand gesture recognition for Indian Sign Language," in *2012 International Conference on Computer Communication and Informatics*, Coimbatore, India, Jan. 2012.
- [11] "Hand Gesture Recognition using Ant Colony Optimization Technique," *International Journal of Engineering Research*, vol. 3, no. 19, p. 3, 2015.
- [12] Meng Lei, Teng Lin. Karhunen loeve transform-based online handwritten signature scheme via dynamic bit allocation method and naive bayes for biometric key generation[J]. *Journal of Information Hiding and Multimedia Signal Processing*, v 9, n 4, p 896-903, July 2018.
- [13] Jing Yu, Hang Li, Shoulin Yin. Dynamic Gesture Recognition Based on Deep Learning in Human-to-Computer Interfaces [J]. *Journal of Applied Science and Engineering*, vol. 23, no. 1, pp.31-38, 2020.

- [14] Ligu Wang, Yin Shoulin, Hashem Alyami, et al. A novel deep learning-based single shot multibox detector model for object detection in optical remote sensing images [J]. *Geoscience Data Journal*, 2022
- [15] R. Anitha, Prakash, and S. Jyothi, “A segmentation technique to detect the Alzheimer’s disease using image processing,” in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, Chennai, India, Mar. 2016, pp. 3800–3801.
- [16] Y. Ben Fadhel, S. Ktata, and T. Kraiem, “Cardiac scintigraphic images segmentation techniques,” in *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, Monastir, Tunisia, Mar. 2016, pp. 364–369.
- [17] P. Bhosale, A. Gokhale, and Y. Motey, “Image segmentation using graph cut technique for outdoor scene images,” in *2013 International Conference on Communication and Signal Processing*, Melmaruvathur, India, Apr. 2013, pp. 280–282.
- [18] A. Swarnalatha and M. Manikandan, “Review of segmentation techniques for intravascular ultrasound (IVUS) images,” in *2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*, Chennai, India, Mar. 2017, pp. 1–4.
- [19] T. D. Vishnumurthy, H. S. Mohana, and V. A. Meshram, “Automatic segmentation of brain MRI images and tumor detection using morphological techniques,” in *2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)*, Mysuru, India, Dec. 2016, pp. 6–11.
- [20] Kang-Hyun Jo, Sung-Eun Kim, and Kyung Sup Park, “Interaction between agents from recognition of face direction and hand gestures in the virtual space,” in *Proceedings KORUS 2000. The 4th Korea-Russia International Symposium on Science and Technology*, Ulsan, South Korea, 2000.
- [21] A. A. Hosain, P. S. Santhalingam, P. Pathak, J. Kosecka, and H. Rangwala, “Sign Language Recognition Analysis using Multimodal Data,” *arXiv:1909.11232 [cs, stat]*, Sep. 2019, Accessed: Aug. 24, 2020.