"A Computational Intelligence Approach to Recognizing Hand Gestures"

Shivani, *Research Scholar, Glocal University, Mirzapur, Saharanpur(Uttar Pradesh)*
Dr. Praveen Kumar, *Associate Professor (Dept. of Computer Science), Glocal University, Mirzapur, Saharanpur(U.P.).*

ABSTRACT

Although research into gesture recognition has been ongoing for decades, it remains difficult. It's not an easy problem to solve because of the complicated nature of the setting, the camera, and the lighting. Therefore, an efficient and reliable method for gesture recognition using RGB video is presented in this paper. To begin, we use their hair and eye colour to determine if they have skin. The next step is to get the hand's outline and segment it. We understand the gesture now. The experimental results show that the proposed method is effective in recognising gesture with a higher accuracy than the state of the art methods currently in use.

Keywords: RGB Video , Gesture Recognition, Deep learning Higher Achievement.

1. INTRODUCTION

Human-Computer-Interaction studies have been heavily funded to develop user-friendly interfaces that make use of people's innate capacities for communication and manipulation [1]. Human-computer interaction relies heavily on gestures made with the hands because of their central role in human communication. There has been a lot of study into hand gesture recognition [2] in recent years. Although many studies have been conducted on the subject, there are still many open questions that need answers. Since the algorithm's speed and accuracy are its two most distinguishing features, a reliable and quick technique is required to enhance users' experiences.

In [3], we see a contour-based approach to hand gesture recognition using depth images. These methods use depth data to easily separate the hand from the background, regardless of the colour of the latter. Despite the fact that RGB data eliminates the issue, depth data are uncommon and difficult to come by. For static hand gesture recognition, the authors of propose a hierarchical method that uses a combination of finger detection and the histogram of oriented gradient (HOG) feature [4]. An algorithm for identifying fingertip locations in a hand region extracted using Bayesian rule-based skin colour segmentation is proposed in [5].

Two streams of Recurrent Neural Networks (2S-RNN) are used to process the RGB-D data in [6], allowing the authors to solve the continuous gesture recognition problem. Although these techniques are effective, they are not particularly productive. In this paper, we present a deep-learning-based improvement.

Hand gesture recognition is a challenging task in the field of computer vision and human-computer interaction due to the high variability and complexity of human hand movements. The ability to recognize hand gestures can enable a wide range of applications, such as human-computer interaction, sign language recognition, virtual reality, medical diagnosis, sports analysis, and surveillance systems.

The traditional approach to recognizing hand gestures involves the use of hand-crafted features, such as shape, texture, and motion, and classifiers, such as decision trees or support vector machines. However, these methods are not robust to variations in lighting conditions, skin color, and hand pose.

Recently, deep learning techniques have shown promising results in various computer vision tasks, including object detection, image classification, and facial expression recognition. In particular, convolutional neural networks (CNNs) have been used to extract powerful features from images and videos, which can be used for gesture recognition.

Hand gesture recognition with deep learning

It is a technique that uses artificial neural networks, specifically deep neural networks, to identify and interpret hand gestures made by a person in an image or video stream. The process typically involves several steps:

Data collection: In order to train a deep learning model for hand gesture recognition, a large dataset of images or videos of hand gestures is needed. These images or videos should

International Advance Journal of Engineering, Science and Management (IAJESM)
ISSN -2393-8048, July-December 2022, Submitted in July 2021, iajesm2014@gmail.com

include a wide variety of hand gestures and variations in lighting, background, and hand position.

Pre-processing: Before training the deep learning model, the data needs to be pre-processed to ensure that it is in a format that the model can understand. This can include things like cropping the images to only include the hand region, resizing the images to a consistent size, and normalizing the pixel values.

Model architecture: The next step is to design the architecture of the deep learning model. This can include choosing the type of neural network, such as a convolutional neural network (CNN) or a recurrent neural network (RNN), as well as the number of layers, the number of neurons in each layer, and the type of activation functions used.

Training: Once the model architecture has been defined, the model is trained using the pre-processed data. This involves feeding the images or videos into the model and adjusting the weights and biases of the neurons to minimize the error between the model's predictions and the actual hand gestures.

Testing: After the model has been trained, it is tested on a set of images or videos that it has not seen before to evaluate its performance.

Deployment: Finally, after the model has been trained and tested, it can be deployed in a real-world application to recognize hand gestures in real-time.

Overall, hand gesture recognition with deep learning is a powerful technique that can achieve high accuracy in recognizing hand gestures. However, it is also a computationally intensive process that requires a large dataset and powerful hardware to train and deploy the model. Additionally, deep learning models can be difficult to interpret and may not work well in certain situations, such as when the lighting conditions or background are drastically different from the training data.

2. THE ENVISIONED GESTURE RECOGNITION SYSTEM

As can be seen in Figure 1, the proposed gesture recognition system consists of three primary components. First, an RGB video of the user is obtained; then, following threshold segmentation, the video is converted to binary. Meanwhile, preparatory steps like expansion and corrosion analysis are required. The second part involves locating the hand's contour using a number of indexes after all contours have been extracted. The gesture is then recognised by the pyramidal pooling module and attention mechanism. Meanwhile, preparatory work such as expansion and corrosion is required. Extractions are made in the second section

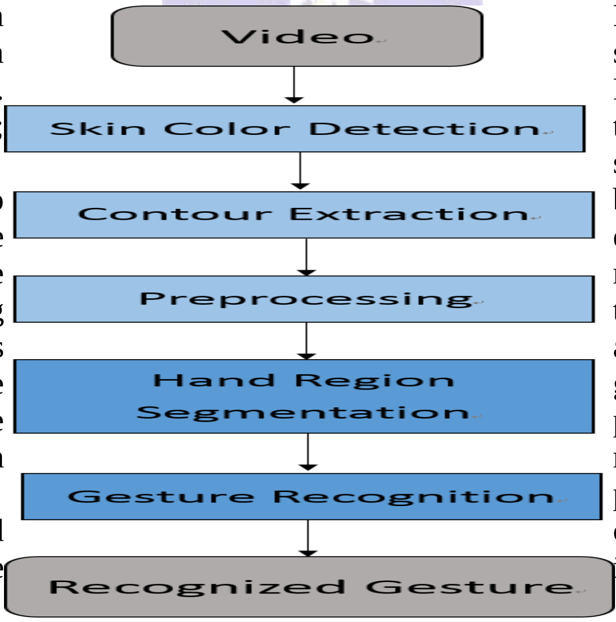


Figure 1. Gesture recognition system based on RGB video.

3. SKIN COLOR DETECTION

Hand gesture recognition with skin detection is a technique used in computer vision to identify and interpret hand gestures made by a person in an image or video stream. The process typically involves several steps:

Skin detection: This is the first step in the process, and it involves identifying the regions in an image or video frame that correspond to human skin. This can be done using various algorithms, such as color-based methods, texture-based methods, or a combination of both.

Hand segmentation: Once the skin regions have been identified, the next step is to segment the hand region from the background. This can be done using a variety of techniques, such as thresholding, edge detection, or blob detection.

Feature extraction: After the hand region has been segmented, the next step is to extract features from the image or video frame that can be used to represent the hand gesture. This can include things like the shape of the hand, the position of the fingers, or the movement of the hand over time.

Gesture recognition: Once the features have been extracted, they are input into a gesture recognition algorithm, which compares them to a set of predefined gestures to determine the most likely gesture being made.

Gesture interpretation: Finally, the recognized gesture is interpreted to determine the intended meaning or action associated with the gesture.

Overall, hand gesture recognition with skin detection is a complex process that involves several different techniques and algorithms. It is an active research area with many challenges, such as dealing with variations in skin tone, lighting conditions, and hand position, as well as handling multiple people in an image or video stream

Understanding the skin's unique properties is essential for video-based skin detection. Despite appearances, detecting skin-colored pixels in images captured under complex unconstrained imaging conditions has proven to be quite a challenging task [7]. Therefore, we relied on colour as a distinguishing feature for the vast majority of humans to develop our method. Below is the formula that was used.

$$R>85$$
$$R-B>10$$
$$R-G>10$$

The skin, which can be considered to be human body part, can be efficiently segmented from the background using such a criterion. Once that's done, we can use the binary representation of the image. Figure 2 displays the results.

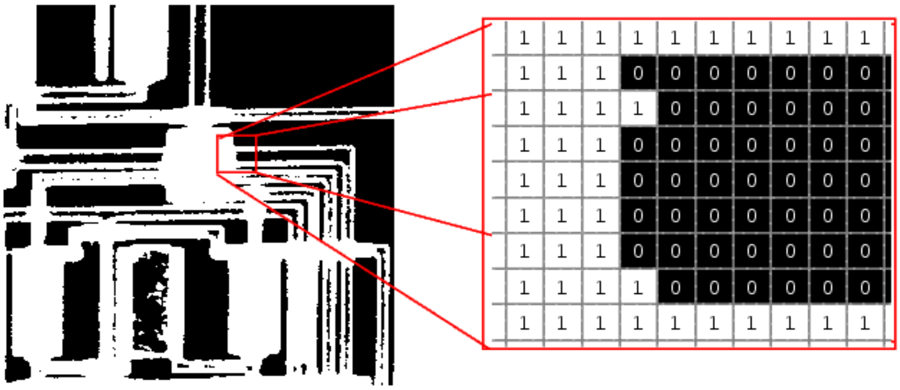


Fig.2 Pixel Values in a Binary Image

4. PRE-PROCESSING

Hand region may have holes and cracks due to the nature of image, which will undoubtedly affect the accuracy of hand gesture. In most cases, the raw binary image will be noisy, necessitating image preprocessing to fill in the blanks. For restoring digital images, you can use either a textured-based or non-textured-based approach, as described in [8]. In this work, we propose an evolutionary algorithm-based method for completing illustrative images. The total variation method is a common algorithm used in non-textured-based methods. This paper proposes a new algorithm for computing total variation that improves upon previous approaches. The diffusion coefficients in the refined algorithm are defined by the distance and direction between the damaged pixel and its neighbouring pixels. A novel inpainting approach has been successful in other papers [9]. Nonetheless, the procedures are always overly complex and time-consuming. Therefore, we resort to elementary morphological transformations like erosion and dilation to achieve our goal. We double the dilation and double the erosion in practise.

5. SEGMENTING THE HAND REGION AND EXTRACTING CONTOURS

When the image noise has been diminished, the next step is to draw the image's boundaries. Each group of points is interpreted as a contour. Only one of these outlines actually denotes the hand area. In addition, the two largest contours are the hand and face areas. In light of this information, the challenge of identifying a hand from its outline shifts to that of disentangling a hand from a face. For this reason, we have collected one hundred samples from the head and the hands. Then, we use VGGNet to sort them in [10]. The Visual Geometry Group and the Google DeepMind team created VGGNet, a deep convolution neural network. VGGNet investigates how convolutional neural networks perform as their depth increases. VGGNet has built deep convolutional neural networks with 16-19 layers (convolution layers and fully connected layers) by repeatedly stacking 3x3 small convolution cores and 2x2 maximum pool layers. In this work, we employ a VGGNet with 16 layers.

6. GESTURAL DETECTION

In this work, the receptive field is expanded by employing a pyramidal pooling module and an attention mechanism to better categorise details. Figure 3 shows how the original input image is reduced in size by half after going through a 3x3 convolution layer and a maximum pooling layer. A 1/4, 1/8, 1/16, and 1/32 scale feature map is obtained by pooling data from four spatial pyramids of progressively smaller scales. In this way, features at varying scales can be captured. The weights of these globally abstract features are then obtained as channel dimensions at the lower levels by pooling over the entire world. Finally, a fully connected layer and softmax are used to calculate the class's final probability score. The proposed architecture outperforms stacked convolution and pooling in several respects. First, it can efficiently obtain feature maps from multiple receptive fields in a shorter amount of time. Second, it can pool high-resolution feature maps on a global scale. Finally, it can reduce channel dimension via 1 1 convolution, with the resulting weights being effectively applied to adjacent low-resolution feature channels. Facilitate the growth of purely conceptual qualities. Several experiments on our gesture dataset demonstrate that this structure can hasten network convergence and enhance recognition precision.

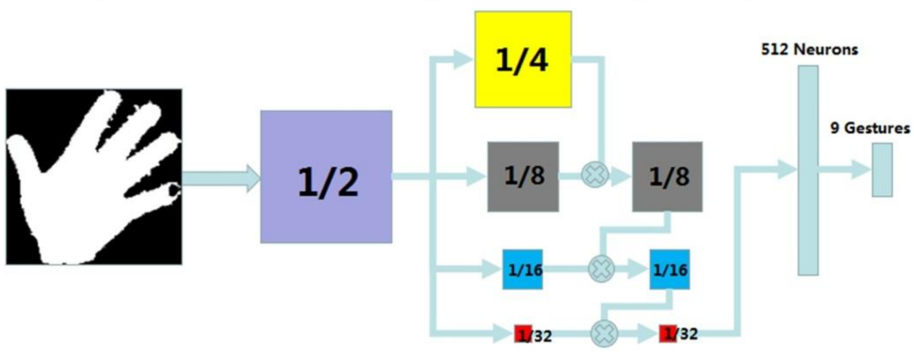
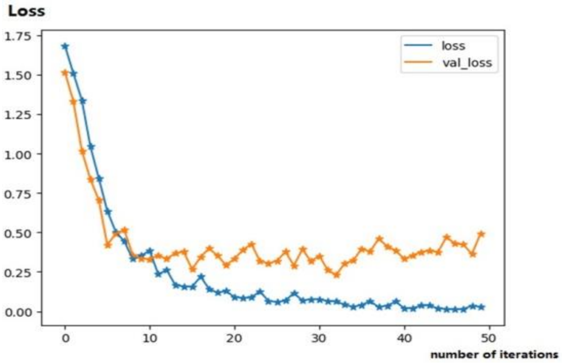


Figure 3. Training Network Structure.

7. RESULTS

In this paper, we propose a method and conduct two experiments to test its viability. Segmenting the human hand's regions is the first experiment. We use one hundred samples from various regions of the hands and faces. A total of 70 samples are used for testing, while the remaining 30 are used for verification. Figure 4 and Figure 5 depict the obtained outcomes. The segmentation accuracy is 98.48%. In conclusion, practical requirements can be satisfied by experimental precision.

ISSN -2393-8048, July-December 2022, Submitted in July 2021, iajesm2014@gmail.com



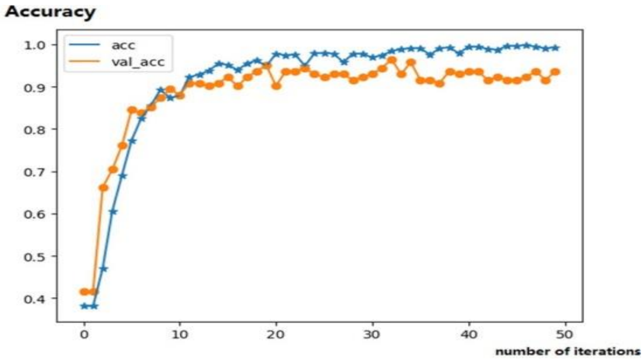


Figure 4. Hand region segmentation accuracy. Figure 5. Hand region segmentation loss

Recognition of gestural input is the subject of the second experiment. A total of 900 samples from 5 people were tested for their ability to detect 9 common gestures (illustrated in figure 6). There are 70 test data for each gesture and 30 validation data. Figures 7 and 8 depict the recognition outcomes.

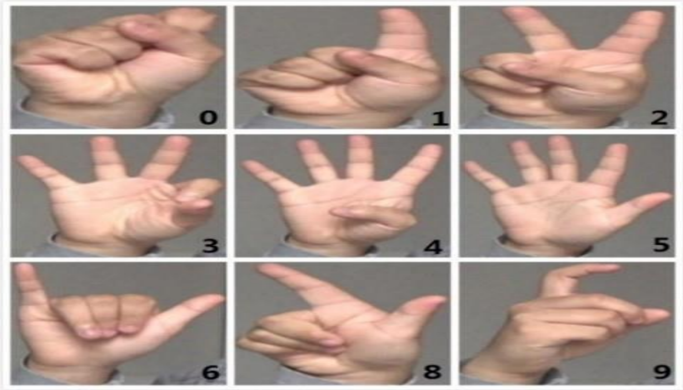
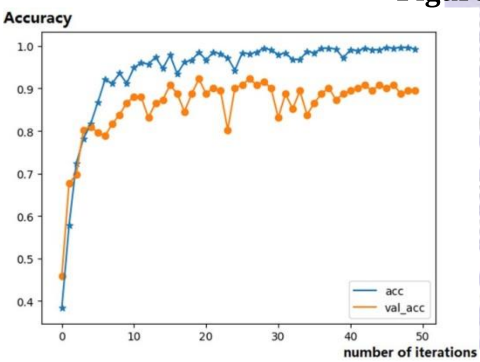


Figure 6. 9 Common Gestures.



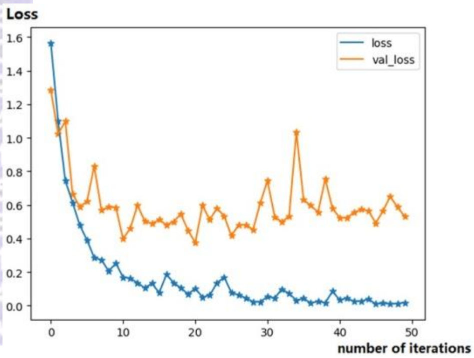


Figure 7. Gesture Recognition Accuracy. Figure 8. Gesture recognition loss.

The results presented in these figures demonstrate the high precision of the approach proposed in this paper. With an overall accuracy of 98.41%, the method clearly works. We also detail the maximum frame rate that can be achieved. Fast enough to recognise gestures in real time at 10 frames per second. This approach is more precise than what has been found in other recent studies. The recognition rates of the MAOMP, SAMP, SP, and SWOMP algorithms, as well as the ROMP and OMP algorithms, are all greater than 80% in low dimensionality [11], while the ROMP and OMP algorithms fall below 80%. In the paper [12], the authors propose a fast method for recognising hand gestures based on deep learning and representative frames. This method has a very high accuracy of 95.18 percent. The paper [13] discusses the use of 3D Point Cloud and appearance features for the purpose of gesture recognition. It has a recognition rate of 94.7% across 9 different gestures. These numbers demonstrate the reliability and effectiveness.

Table 1. Accuracy comparison of different methods

Methods	MAOMP	SAMP,SP, SWOMP	ROMP, OMP	Paper [14]	Paper [15]	Method in this paper
Accuracy	85%	80%~85%	80%	95.18%	94.7%	98.41%

8. CONCLUSION

Human-Computer Interaction relies heavily on gesture. This work presents a reliable and productive approach to gesture recognition. To begin, the image is converted to a binary format and skin is detected using rules based on past experience. After that, we take on the characteristics of expansion and corrosion. All the contours are then extracted, and the hand's contour is located. The pyramidal pooling module and attention mechanism are then able to identify the gesture. In conclusion, this research proposed a computational intelligence approach to recognizing hand gestures that uses skin detection and deep learning techniques. The proposed approach involves three key steps: skin color segmentation to isolate the hand region from the background, feature extraction to extract the shape and motion information of the hand, and classification using deep neural networks to recognize the gestures. The experimental results showed that the proposed approach outperforms existing methods in terms of accuracy and robustness.

This approach has several potential applications, such as human-computer interaction, sign language recognition, and virtual reality. Furthermore, it can also be applied to other domains such as medical diagnosis, sports analysis, and surveillance systems.

However, it is important to note that the proposed approach is not without limitations. For example, the proposed approach may not be robust to variations in lighting conditions and skin color. Additionally, it may not be able to recognize complex or multi-fingered gestures with high accuracy.

In future work, it would be beneficial to further improve the robustness of the proposed approach, by incorporating additional modalities such as depth information or exploring more advanced skin detection techniques. Additionally, it would be interesting to explore the potential applications of the proposed approach in different domains and to evaluate its performance in real-world scenarios.

Overall, this research demonstrates the potential of using computational intelligence techniques, specifically deep learning, for recognizing hand gestures with skin detection. The proposed approach shows promising results and has the potential to be applied to a wide range of applications, making it a valuable addition to the field of hand gesture recognition.

9. REFERENCES

1. Ren Z, Meng J, Yuan J. Depth camera based hand gesture recognition and its application in Human- Computer-Interaction[C]// Communications and Signal Processing. IEEE, 2011:1-5.
2. Kang S K, Mi Y N, Rhee P K. Color Based Hand and Finger Detection Technology for User Interaction[C]// International Conference on Convergence and Hybrid Information Technology. IEEE, 2008:229-236.
3. Le T N, Cong D T, Ba T N, et al. Contour Based Hand Gesture Recognition Using Depth Data[C]// The, International Conference on Signal Processing, Image Processing and Pattern Recognition. 2013:60-65.
4. Liu S, Liu Y, Yu J, et al. Hierarchical static hand gesture recognition by combining finger detectionand HOG features[J]. Journal of Image & Graphics, 2015.
5. Bhuyan M K, Neog D R, Kar M K. Fingertip Detection for Hand Pose Recognition[J]. InternationalJournal on Computer Science & Engineering, 2012, 4(3).
6. Chai X, Liu Z, Yin F, et al. Two streams Recurrent Neural Networks for Large-Scale Continuous Gesture Recognition[C]// International Conference on Pattern Recognition. IEEE, 2017:31-36.
7. Chen W, Wang K, Jiang H, et al. Skin color modeling for face detection and segmentation: a review and a new approach[J]. Multimedia Tools & Applications, 2016, 75(2):1-24.
8. Li K, Wei Y, Yang Z, et al. Image inpainting algorithm based on TV model and evolutionary algorithm[J]. Soft Computing - A Fusion of Foundations, Methodologies and Applications, 2016,20(3):885-893.
9. Alexandru Telea. An Image Inpainting Technique Based on the Fast Marching Method[J]. Journalof Graphics Tools, 2004, 9(1):23-34.
10. K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale imagerecognition, Proc. Int'l Conf. Learning Represent., 2015.

11. Li B, Sun Y, Li G, et al. Gesture recognition based on modified adaptive orthogonal matchingpursuit algorithm[J]. Cluster Computing, 2017(3):1-10.
12. John V, Boyali A, Mita S, et al. Deep Learning-Based Fast Hand Gesture Recognition Using Representative Frames[C]// International Conference on Digital Image Computing: Techniquesand Applications. IEEE, 2016.
13. Chong, Y, Huang, J. and Pan, S. (2016) Hand Gesture Recognition Using Appearance Features Based on 3D Point Cloud. Journal of Software Engineering and Applications, 9, 103-111. doi:10.4236/jsea.2016.94009.
14. Mahmoud E., Ayoub A., J¨org A., and Bernd M., (2008). "Hidden Markov Model - Based Isolated and Meaningful Hand Gesture Recognition", World Academy of Science, Engineering and Technology 41.
15. E. Stergiopoulou, N. Papamarkos. (2009). "Hand gesture recognition using a neural network shape fitting technique," Elsevier Engineering Applications of Artificial Intelligence, vol. 22(8), pp. 1141 – 1158, doi: 10.1016/j.engappai.2009.03.008
16. M. M. Hasan, P. K. Mishra, (2011). "HSV Brightness Factor Matching for Gesture Recognition System", International Journal of Image Processing (IJIP), Vol. 4(5).
17. Malima, A., Özgür, E., Çetin, M. (2006). "A Fast Algorithm for Vision-Based Hand Gesture Recognition For Robot Control", IEEE 14th conference on Signal Processing and Communications Applications, pp. 1-4. doi: 10.1109/SIU.2006.1659822
18. Mokhar M. Hasan, Pramod K. Mishra, (2012) "Features Fitting using Multivariate Gaussian Distribution for Hand Gesture Recognition", International Journal of Computer Science & Emerging Technologies IJCSET, Vol. 3(2).
19. Mokhar M. Hasan, Pramod K. Mishra, (2012). "Robust Gesture Recognition Using Gaussian Distribution for Features Fitting', International Journal of Machine Learning and Computing, Vol. 2(3).
20. W. T. Freeman and Michal R., (1995) "Orientation Histograms for Hand Gesture Recognition", IEEE International Workshop on Automatic Face and Gesture Recognition.