

Real Time Human Gesture Recognition: Methods, Datasets and Strategies

Rameez Shamalik¹ and Dr.Sanjay Koli²

¹*Research Scholar, G. H. Raisoni College of Engineering and Management, Pune, India*

²*Professor, Department of E&TC, Dr.D.Y. Patil School Of Engineering, Pune, India*

Abstract. Gestures are universal means of communication without any language barrier. Detecting gestures and recognition of its meaning are key steps for researchers in computer vision. Majority of the work is done in sign language already. Sign language datasets are compared with respect to their usability and diversity in terms of various signs. This paper highlights the available datasets from three dimensional body scans to hand action gestures. Their usability and strategies used to achieve the desired results are also discussed. Major neural networks are evaluated in terms of varied parameters and features. A Methodology for effective gesture recognition in real is proposed. Lastly Results achieved through an Open CV in combination with Sci-kit learn library based technique for gesture recognition are presented and analyzed in terms of efficacy and efficiency.

Keywords. Gesture Recognition, Datasets, Open CV, Hand Gestures, Sci-Kit Learn

1. Introduction

Gestures facilitate a user to interact with given environment. Gestures not only allow front end interaction with others with no language barrier but also remote interactions with a smart screen, virtual-reality and augmented-reality objects. Digital computing is made possible thanks to gesture-based human-computer interfaces[1] from everyday life physical objects like lights, mirrors, doorknobs, notebooks to the specific and focused tasks like driving a car or a medical surgery. Gesture study helps to further understand semantic meaning with contextual information about a person's behaviour and reactions. Among a plethora of various gestures which include hand, body and head gestures, hand gestures stand class apart as they simply carry more information due to multiple combinations than the other types of gestures.

Gestures play a vital role to help people with speaking and hearing disabilities as it replaces voice as the primary means of communication[2]. Dynamic gesture recognition through video processing is tedious due to the complications in the gesture background, ambiguity of video devices and the non-uniformity in a collected data. Enhanced requirements for computing devices especially in the recent lockdowns increased the volumes and necessity of easy to use computer interfaces.

2. Related Theory

There is a rapid transition in video processing techniques based on requirements of various application[3]. Three Dimensional(3D) videos are gaining increasing importance amongst various techniques[4]. Image based gesture recognition is the most basic techniques[5], also Haar cascade is repeatedly used for static hand and face recognition[6].Adaboost algorithm based fingure tracking and contours detection is also recently achieved[7]. Intelligent cars are being designed having a simple RGB camera in front connected to wifi for real time gesture control[8].

Computer vision plays an important role not only in gesture recognition but also in facial identification[9], Crop disease detection[10] and even in accident detection in various weather situations[11].

Datasets of particular project are the core of the research topic. Efficient and accurate data with respect to the application needs to be selected. When it comes to gestures, sign language datasets emerge first but a lot of work is already done upon them with greater accuracy. Some of the examples include American sign language dataset and Brazilian sign language dataset[12].

MPI-FAUST[13] dataset includes a complete three dimensional (3D) scans of hundreds of subjects highlighting their gestures. **DensePose**[14] is another dataset made by Facebook research team, it maps all human pixels of an RGB image to the 3D surface of the human body. **DensePose- COCO**[15] is an upgraded version of the prior, to work on 3D estimation of pose in wild. When it comes to Hand gestures, the most effective dataset is found to be **EgoGesture**[16] developed by Chinese Academy of sciences. It has both images and videos of 26 different signs made by hands with depth. **NVGesture**[17] is a very similar dataset for Dynamic Hand Gesture Dataset is a resource for researchers in the field of dynamic hand gesture recognition. **Jester**[18] is another dataset which is a collection of labelled video clips that show humans performing pre-defined hand gestures. **MSR 3D** dataset[19] is a sequence of 100 images captured from 8 cameras showing the breakdancing and ballet scene developed by Microsoft. **DHG** i.e. Dynamic Hand Gesture dataset[20] contains sequences of 14 handgestures performed in two ways: using one finger and the whole hand.

All of the above mentioned datasets are of size ranging from 1 GB to 30 GB in size as they contain image as well as video sequences to be worked upon. Also an extra depth feature is also involved to achieve greater accuracy. While working with such a huge amount of data, Annotations become difficult for each frame, hence pre-processing is limited to negligible. Deep Neural Networks are the valid solution to achieve better results in such scenario. Neural networks starting from simple Convolutional Neural Networks (CNNs) to **VGG-16**, **AlexNet**, **ResNet** etc. are used to process the data for higher accuracy. Basic comparison is highlighted in table 1.

Table 1. Comparison between Major neural networks with key features

| Neural Network | Special feature | Parameters | Accuracy |
|----------------|-------------------|------------|----------|
| VGG Net | Static kernels | 138M | 92.3% |
| AlexNet | Deep structure | 60M | 84.7% |
| ResNet | Short connections | 62M | 60.3% |

3. Methodology

First part in gesture recognition is to find the favorable pixels in the given video frame or image. Gestures especially hand ones has lot to do about edges. The pixel density is always high upon region of interest(ROI).The Canny algorithm performs brilliantly when it comes to a function that finds edges in the input image and marks them in the output map edges. Edge detection helps in finding both depth and geometry of a frame. It involves noise reduction and finding intensity gradient of the image.

The next step especially in hand gesture recognition is to find the hand region from all the edges available. It is important to eliminate unwanted regions in the video sequence. It is done with help of Background Subtraction which uses concept of running averages. 30 frames are looked upon over a given video sequence to calculate it. A difference image is thus obtained by calculation of absolute difference between the background model and the current frame to highlight foreground portion[21].

Motion Detection and Thresholding play a vital role in featuring the visibility of hand in question in order to get rid of rest of the portion which is colored as black and hand particularly becomes white. This exercise can also help understand the depth in the frames available as logically hand gestures should be in the foreground.

Contours help in understanding the size and the shape of the hands by making a boundary around the particular gesture made by hand. Bounding box over the contour helps in segmentation of the ROI from the frame.

4. Results and Discussion

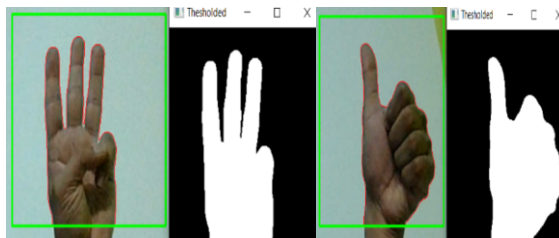


Figure 1. Various Hand gesture capture with threshold frame

Figure 1 shows three steps happening together which includes background subtraction happening in the bounding box shown with green borders. Bounding box helps with segmentation as whole video frame need not to be processed but just a specific area. Finding exact contours is the key to proper gesture detection as it forms the exact shape of the gesture. Proper thresholding is important for highlighting the gestures in a separate window where the background is shown as black and foreground i.e. a gesture as white in a specified bounding box only. Note that system collects 20 frames per second hence active only for first 20 seconds, later it will continue to display the captured contour shape.

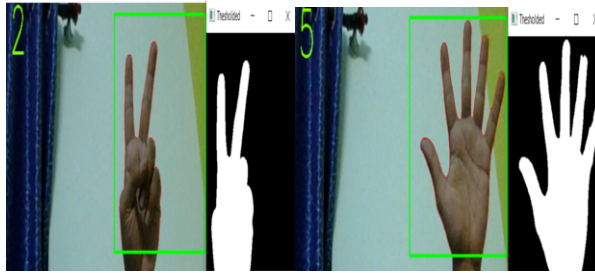


Figure 2. Detection of count of fingers

Capturing gesture is just not enough but understanding them adds value. The shapes are finalized with the help of contours and thresholding but recognition of the gestures require a separate approach. After successfully thresholding the hand gesture we could now count the number indicated by the fingers in the hand gesture. A convex hull is formed around the contour in the Region of interest (ROI). Basically it tracks the movement of hands to form the convex hull. Next step is to find the exact distance between the fingers for which Euclidean distance is a standard solution in computer vision. It is measured with the help of Sci-kit learn library in built function which performs bitwise AND operation to display the numbers as shown in figure 2. This model conveniently segments the background, no matter how complex it is.

5. Conclusion

There are various datasets and techniques available for evaluation of Gestures. Majority of work done is on the sign language detection. Only capturing the gesture is not enough but to gain understanding of the same is more important. Depth and geometry provide insight in the real time gesture detection thus adding extra variables to be worked upon. Accurate depth and geometry of the frame is obtained by using canny edge detection. In Python programming libraries like Open CV, Numpy and Sci-Kit Learn are used to recognize hand gestures and numbers associated with it. A specific technique needs to be opted for effective evaluation of gestures which will suit the applications of the design. The developed framework needs to be high efficiency, economical Model for Human Gesture Detection and can find its application in Human-Machine Interaction, Virtual Reality, Augmented Reality, Sport Videos, Airport Security and other public transportation hubs.

References

- [1] M. P. Pacaux-Lemoine et. Al, "Emulated haptic shared control for brain-computer interfaces improves human-robot cooperation", *IEEE International Conference on Human-Machine Systems (ICHMS)*, Rome, Italy, 2020.
- [2] R. Nair et. Al, "Hand Gesture Recognition system for physically challenged people using IoT", *6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2020.
- [3] Sanjay Koli, Rameez Shamalik, "Transformation of Video Signal Processing Techniques from 2D to 3D: A Survey", *Proceedings of the 2nd International Conference on Communications and Cyber Physical Engineering, springer lecture notes in electrical engineering(LNEE)*, volume 570, pages 63-70, ICCCE 2019.

- [4] Rameez Shamalik, Sanjay Koli, "Emergence and Functionality of 3D Videos", *International Journal of Engineering and Advanced Technology (IJEAT)*, ISSN: 2249 – 8958, Volume-9 Issue-3, February, 2020
- [5] O. Saman and L. Stanciu, "Image Processing Algorithm for Appearance-Based Gestur Recognition", *23rd International Conference on System Theory, Control and Computing (ICSTCC)*, 2019, pp. 681-684, doi: 10.1109/ICSTCC.2019.8885888.
- [6] S. Sharma, S. Jain and Khushboo, "A Static Hand Gesture and Face Recognition System for Blind People", *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*, 2019, pp. 534-539, doi: 10.1109/SPIN.2019.8711706.
- [7] R. M. Gurav and P. K. Kadbe, "Real time finger tracking and contour detection for gesture recognition using OpenCV", *International Conference on Industrial Instrumentation and Control (ICIC)*, 2015, pp. 974-977, doi: 10.1109/IIC.2015.7150886.
- [8] G. Dai and P. Wang, "Design of intelligent car based on WiFi video capture and OpenCV gesture control", *2017 Chinese Automation Congress (CAC)*, 2017, pp. 4103-4107, doi: 10.1109/CAC.2017.8243499.
- [9] Manisha Talware, Sanjay Koli, "Video-Based Person Re-Identification: Methods, Datasets, and Deep Learning", *International Journal of Engineering and Advanced Technology (IJEAT)* ISSN: 2249 – 8958, Volume-9 Issue-3, February, 2020.