MediaPipe to Recognise the Hand Gestures

LAVANYA VAISHNAVI D. A., ANIL KUMAR C., HARISH S., DIVYA M. L.

Department of ECE, R L J I T, Doddaballapur Bengaluru, INDIA

Abstract:-Human Computer Interaction (HCI) can be improved drastically using the hand gesture based recognition system. This system is designed to detect the gestures of the hands in the images captured in real time. There are certain areas of intersect in the hands that are there for the classification. The gaming devices like Xbox, PS4 and smart phones are also using this method to solve few problems. In this paper a smart method i9s developed to solve the problem. Using Python 3.9 and MediaPipe, the hand gestures are recognised in the real-time images. The background subtraction is the key method used to generate the results. The hand is detected and processed for the finding of the binary image with the fixed number of pixels. The palm position, dimension and the gesture are recognised. In this experiment, the finger count and the position of the fingers is considered as the gesture. The finger count can also be calculated after the hand is recognised. The major areas of the Image Processing and the MediaPipe in Python are covered to solve this problem.

Keywords: Mediapipe, Gesture Recognition, Python, Image processing

Received: March 19, 2021. Revised: April 15, 2022. Accepted: May 12, 2022. Published: July 2, 2022.

1. Introduction.

The gesture control applications have become widely used as the Android applications are increasing in number. We find several applications that are just based on the gesture of images. Even in Instagram we find several filters that change according to the gesture. The recognition of the gesture in the real time images is a major work of accuracy. the accuracy and the time of recognition are the most important challenges, that needs to be addressed in the real time. The main problem in the analysis of image is that the perception of human and the perception of computer on an image is not the same. Humans can easily recognise the image and its intention. Whereas the same image for a computer is just 3 dimensional matrices. The 3-dimensional Matrix does consist of the pixels. And these pixels are made out of RGB colour values. As the computer cannot analyse the intention and the image, the problem remains as a challenge.

2. Existing system

To define a sign language, It can be expressed as a sign that is made out of hands by the movement, including facial expression, posture of the body. The sign language is used by the deaf, dumbs and also sometimes in our daily lives. Sign language can also be used to communicate between two parties. In the paper [1], it is mentioned that the gesture we make has meaning to understand, separate gesture can be used for a different phrase. A simple and efficient algorithm is proposed in this paper to recognise the American Sign Language alphabets both in dynamic gesture and also in static gesture. The mentioned algorithm consists of 4 different types of techniques. The 4 key techniques to analyse the gesture are: The count of white pixels at the edge of the images. A centroid point is fixed in the image and the finger length from that point, the angles between the fingers and the various angles between various fingers or first and the last frame.

The recognition rate achieved in this experiment was 95%. It was the highest rate of recognition and clearly it has an accurate prediction of the gesture.

The paper [2], proposes a gesture control application that is used for user interface in digital machine. As in this method a low-cost motion sensor image capture techniques are used. And also, the underperforms the low lighting conditions of blurred images. This paper proposes hand image resolution enhancement technique based on multi scale decomposition and edge prevention smoothing.

This paper DT-CWT and EPS algorithms that are used to decompose into sub bands and interpolated values. Each sub band that is being decomposed, is used to enhance the images. As an experiment, a simple sign language has been recorded using Kinect camera. And the results were verified. As an experimental result, it is found that the system was accurate about a 96%.

2.1 Motivation

Apart from the implementation of the gestures in the current technology, still there is a lot of grey area where the image processing can be further uplifted to recognise the gesture efficiently. The gestures can be definitely used to convey the information and to the control the desired application or desired machine. Apart from this, the gesture can also be used to communicate or to interact with the computers. Machine learning models can be used to recognise the hand gestures, that are made in the real time. This is a paper where an attempt is made to explain and to predict the exact gesture made by the user.

2.2 Problem Statement

The two main technological boons that has been upgraded in the recent years are the interaction approaches with hand gestures based using OpenCV, mediapipe and keyboard rather than believing on mouse and pen, that are already existing. There is a lot of limitation in the usable commands that are used in these devices. For better interaction, direct usage of hands is prioritised. In this approach we tried to implement a part of Python code that users to extract the image and extract the hand from the video sequence. The segmentation of the image is done and star skeletonization and recognition is also performed using the distance signature. As an experiment, few of the basic gestures are recognised in this paper.

3. Image Processing Using MediaPipe

Image processing is a wide area where there are lot of challenges and tasks to be accomplished. There are numerous models to address the same problem. This is an approach where open CV and MediaPipe are used for your time applications of image processing. MediaPipe is a framework that is built for the performance interface over the arbitrary sensor data. With the help of Media Piper, the perception of the hand graph can be modulated. This model is used to recognise the angles between the fingers and the position of fingers. The angles and the position define what is the gesture made. A camera of more than 2 megapixel is use to capture the image. Fig one shows a different hand gestures and different positions of the fingers and the angles between them.

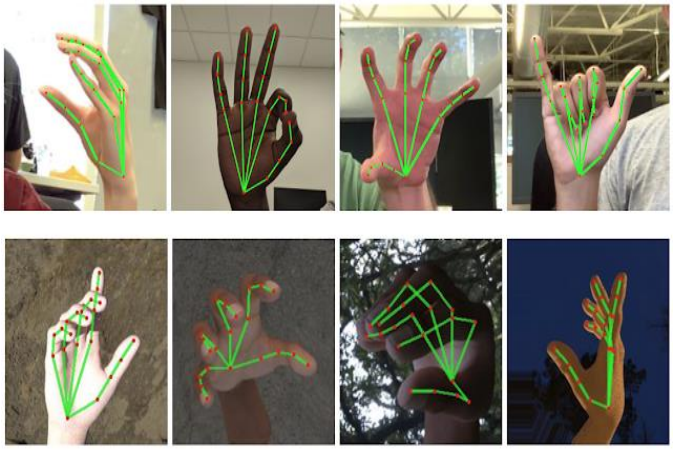


Fig 1: Image Processing Using MediaPipe

4. Methodology

A simple rule classifier is used to recognise the hand and the gestures that are done. The angles between the fingers and the position of the fingers is defined the gesture. Based on the number of finger count and the position of the fingers, the rule classifier will detect the gesture. The rule classifier is an effective and efficient algorithm that is used to detect the hand gestures.

4.1 Proposed Block Diagram & Explanation

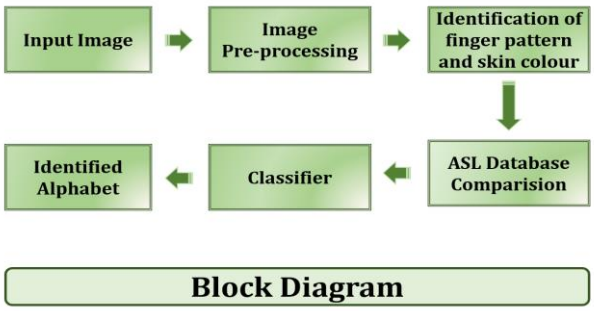


Fig 2: Block Diagram of Hand Gesture based Recognition

Fig 2 implements the block diagram of hand gesture recognition system. This system is used to identify the alphabets are the characters that is provided using the gesture. The basic steps that is involved in the conversion of image to identification is shown in the above block diagram.

Input Image:

The input images captured via the laptop camera or the web camera that is provided. The image processing takes the input of the Image. The image provided by the user consists of any of the American Sign Language alphabet. The images, taken with a 2-megapixel web camera.

Image Pre-processing:

Image pre-processing, this is a term used for the lowest level of abstraction. Before using the image in the interference or model training, this is the previous step taken to process the image. To resize, Orient and for the colour collection we make use of image preprocessing. The aim of this method is to improve the quality of the image, so that it becomes easy to analyse in a better way. Sometimes undesired distortions can also be surprised using this method.

Noise Removal:

The presence of reducing the noise from the image, all to remove the noise from the image is called as loss removal. The process removes the visible noise and smoothens the entire image area, leaving the areas that have contrast boundaries. This is one of the major steps of for getting better quality in the image processing.

Background Subtraction:

From a static camera, the moving objects and the sequences can be detected, using the method of background subtraction. The difference between the reference frame and the current frame is the major idea to implement this. This is also called as background image or background model. This image extract even the edge details using the background image.

Segmentation:

Digital images are cut into several subgroups, called image segments. Using this technique, we can reduce the complexity of the image. This also enables further

processing, easier and simpler. We can also locate the boundaries and the objects in the images.

Contour Extraction:

Contour can be defined as the line joining all the points in the boundary of the image. It joins the point that have same intensity. Contour extraction is an algorithm that is basically joins the lines of same intensity which are neighbor to each other. This is used to trace basic boundary lines of the image.

ASL Data Base:

American Sign Language consists of 26 6 different gestures that represents the English alphabets from A to Z. This is one of the widely used method for a deaf and dumb people to convey the information. In this prototype, the image which is considered as an import is analyse as American Sign Language.

Classifier:

In this is prototype support vector machine classifier is used to classify the American Sign Language. This is one of the mostly used supervised learning algorithm. It is used for both classification and also regression problems, in machine learning. The best boundary, which is called as hyperplane is created using the SVM algorithm. This is done via segregation of dimension spaces. The extreme points or the extreme vectors are used to create the hyperlink. The support vectors are nothing but the extreme cases in the algorithm. Hence the name support vector machine.

4.3 Flowchart & Explanation

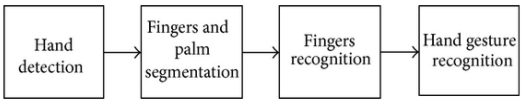


Fig 3: The Overview of the proposed Method for Hand gesture Based Recognition System.

Hand Detection:

The Fig 3 shows the methodology how the hand detection is made. The normal web camera is used to capture the image in the laptop. The hand images are considered at similar conditions for a long time. For the proper working of disc algorithm, clean background is considered and the same background is taken for all the different gesture recognition. Only in few cases for the experimental purpose, few objects were added within the frame. The background subtraction help to remove the unnecessary detailing in the images. The moving objects and the hand is differentiated by the skin colour. The HSB model is used to measure the skin colour. 315, 94, and 37 are the coordinate values of skin colour in HSV model.

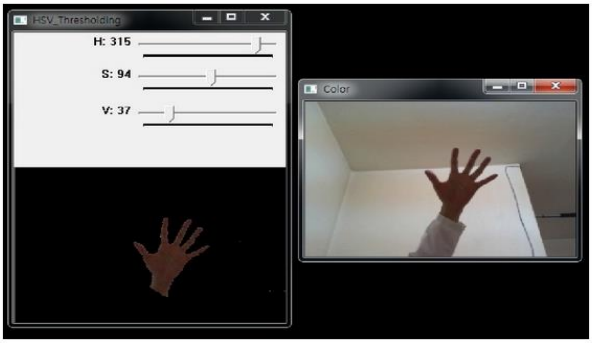


Fig 4: The Procedure of Hand Detection.

Fig 4 represent the procedure of hand detection in the algorithm. This shows the finger and palm segmentation. Initially, the value of the skin colour in HSV algorithm is shown in Fig 4. That is nothing but the pixel value of hand region. Now the same image has been converted into black and white by removing the background. The following is the procedure that shows how the finger and the palm are segmented from the binary image. And this is shown in Fig 5.

Fig 5: The Detected Hand Region.

Fingers Recognition:

The labelling algorithm is used to mark the segmentation of fingers in the images. This algorithm is used to mark fingers region on the image. The noisy region is also considered where the number of pixels are too small. And this region is discarded. Only the finger region that has enough size is regraded as finger and the remaining as the unwanted region. For each of the remaining region, the minimum boundary box is found and it is enclosed. Initially a red rectangle is used to mark the hand area where this condition is found.

The gesture recognition:

The simple rule classifier is used to detect the gesture after the fingers are detected and recognised. According to the number or and the content of finger, the hand gesture is predicted. This is showed in Fig 6. The image shows where the thumb, fore finger, middle finger, ring finger and the little fingers are recognised using the simple rule algorithm.

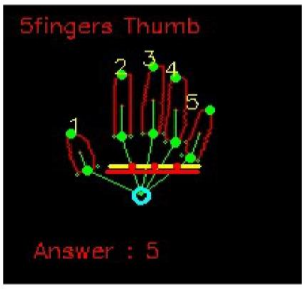


Fig 6: Recognition of the Fingers.

Fig 7: The image of hand gestures used in the experiments. From top to bottom these gestures are labeled as 0,1,2,3,4,5,6,7,8,9.

Fig 7, are the images that has been captured during the implementation of the algorithm. Different symbols and gestures were made to analyse and and has been used as an input for the algorithm. Efforts were made to recognise and to analyse the symbols.

5. Implementation

The implementation of this algorithm had a lot of challenges in the initial time. Python was selected as the best language to implement the algorithm. Support vector machine is one of the major algorithms that is used to. Train the input data and to make the algorithm intelligent to understand what gesture has been displayed. Support vector Machine place an important role in analysing the Gesture. This is one of the mostly used supervised learning algorithm. It is used for both classification and also regression problems, in machine learning. The best boundary, which is called as hyperplane is created using the SVM algorithm. This is done via segregation of dimension spaces. The extreme points or the extreme vectors are used to create the hyperlink. The support vectors are nothing but the extreme cases in the algorithm. Hence the name support vector machine.

The first step is to take the input image. The images, captured using 2 megapixel camera in the laptop which is also called as web camera. The images taken are reduced to resolution of 256 x 256. This image is further fed for image preprocessing. In the preprocessing stage, the image is resized. In the next stage, the noise is removed Using the algorithms. Once the noise has been removed, a better quality image has been produced. In the next step, the background subtraction is done. Only the hand image is extracted. In the next step, segmentation is carried out. Where the image is differentiated into subgroups and hence it will be simpler for the further processing. Next is an important step of extracting the hand image. And this is called Contour extraction. In this Do points that are having the same intensity are joined and made the structure of hand. To perform all the above operation, different functions and different libraries are used in the Python. Open CV and MediaPipe are the 2 important libraries that are being used here. Several functions that are used to accumulate the data and to process the data for the analysis purpose.

The number of frames has been counted using a counter. An infinite loop is made to read the frames from the web camera continuously. The aspect ratio of the frame is maintained properly. The octane image is flipped to get a mirror view of the image which resembles the original image. In the next step, the region of interest is being analysed using Numpy slicing. Then the image is converted to grayscale and also minimises the high frequency component. Up to 30 frames are collected and 30 frames are run. An average update to the model. These 30 frames are analysed using the SVM. And finally, SPM recognises what gesture has been made and that is displayed on the screen.

5.1 Software Interface

The following is the software interface that is used to execute the Prototype. Few of the basic requirements are mentioned in the software interface.

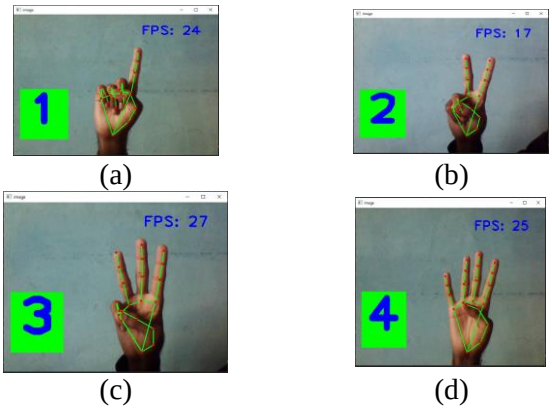
- Operating system- Microsoft Windows 7 SP 1 or above
- Microsoft Visual Studio 2010
- MinGW and Visual C++ compilers (for Windows)
- Supporting Webcam Drivers
- Anaconda – Spyder

5.2 Hardware Interface:

All the physical equipment's i.e., input devices, processor, and output device & inter connecting processor of the computer s called as hardware.

- Hard Disk minimum of 40 GB.
- RAM minimum of 2 GB.
- Dual Core and up ,15" Monitor.
- Integrated webcam or external webcam (15 -20 fps).

6. Results and Outputs



(a) (b) (c) (d)

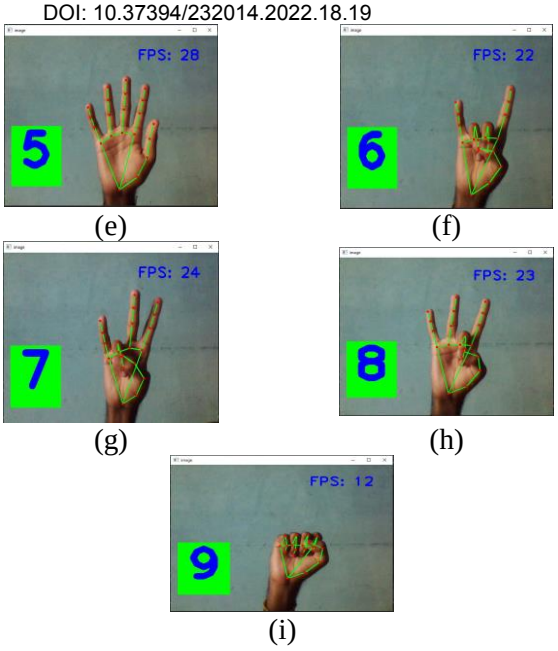


Fig 8: Identifications of the Handgestures. (a) Gesture of Number "One", (b) Gesture of Number "Two", (c) Gesture of Number "Three", (d) Gesture of Number "Four", (e) Gesture of Number "Five", (f) Gesture of Number "Six", (g) Gesture of Number "Seven", (h) Gesture of Number "Eight", (i) Gesture of Number "Nine"

The algorithm is executed on the previously mentioned computer with the specified hardware. Using the web camera, the images were captured and the same have been taken as the input for the algorithm. The Fig 8, shows the different gestures, and the results displayed on the computer screen recognising the gesture as 1,2,3 and others.

6.1 Advantages and Applications of Proposed System

Hand gesture is one of the basic recruitments in communication, even a for a normal human being. Implementing the hand gesture monitoring system into the computer might reduce several tasks and peripherals. The flexibility of usage of the machine can also be improved by using the hand gestures. The usage of mouse and keyboard can also be avoided if this technique is implemented in a smart way.

Gesture recognition can also be used in several other purposes also. Several machines can be automize just by the gesture recognition. For instance, we can just make a gesture to open the door instead of using IR sensors that often opens for a simple obstacle. Nodding head can also be implemented to turn on a machinery or a device. Like this each and every gesture of human being can be analysed and can be automated using these machine learning algorithms.

The touchless and contactless system can also be developed using this. This can also have an application in virtual environment to control the robots remotely. Or to develop a music just by waving hands. And also, to translate the sign language that is being used by the freedom people to the normal human being.

6.2 Challenges of Proposed System

In the previous section, the implementation is being discussed. In each and every stage of the implementation, there were few challenges that has to overcome to get a better result. First challenge was the image size reduction of the image size was a tedious task at the initial stages. As the images captured directly through the web camera. Handling such big image also reduced and hence increase the processing speed. This segmented image was then directly passed to the feature extraction stage. Initially features of only Statistical parameters and Orientation Histogram were extracted. Also, from the images it can be seen that the orientation data due to the wrist of the hand, gets added to the actual original information and dilute the information content. These are the few problems and challenges that were faced during the implementation.

6.4 Future Scope

The recognition rate can also be improved by adding other features in the feature extraction techniques. This can make the system more robust and accurate. Support vector machine is the algorithm used to recognise The gesture. Other algorithms can be used and made experiment to analyse the robustness also. This complete exercise was done taking the background plane. So fuzzy background can be used to recognise the same. We have implemented only the American Sign Language and the numbers. There are several other gestures present, so each of them a can be used to analyse and check. The algorithm that is developed is on Windows platform. The same can also be extended towards Android OS. If it is implemented on Android, it will be more user friendly and more commonly used by the user.

7. Conclusion

The gesture recognition is successfully implemented using the algorithm. By the improvement of the human machine, interaction can be improved robustly. The complete exercise gave the expected result with a good speed. Different numbers from 0 to 9 was recognised using the gestures. Have it was implemented using support vector machine algorithm. The model was tested and a trend for the accurate result. The plain background was considered to examine the prototype. The images were captured using the web camera and the results was displayed on the monitor screen of the laptop. Robust and accurate system was developed by using MediaPipe and the Python coding.

References

[1] Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur, Dr. Dayananda P, 2020, Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCAIT – 2020 (Volume 8 – Issue 15),

[2] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer

Vision," 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141.

[3] A. Thongtawee, O. Pinsanoh and Y. Kitjaidure, "A Novel Feature Extraction for American Sign Language Recognition Using Webcam," 2018 11th Biomedical Engineering International Conference (BMEiCON), 2018, pp. 1-5, doi: 10.1109/BMEiCON.2018.8609933.

[4] Pradeep kumar B P Resolution Enhancement of American sign language Image Using DT-CWT and EPS algorithm August 2019IEIE Transactions on Smart Processing and Computing 8(4):265-271 DOI:10.5573/IEIESPC.2019.8.4.265

[5] M. Z. Iqbal, A. Ghafoor and A. M. Siddiqui, "Satellite Image Resolution Enhancement Using Dual-Tree Complex Wavelet Transform and Nonlocal Means," in IEEE Geoscience and Remote Sensing Letters, vol. 10, no. 3, pp. 451-455, May 2013, doi: 10.1109/LGRS.2012.2208616.

[6] J. J. M. Ople, D. S. Tan, A. Azcarraga, C. -L. Yang and K. -L. Hua, "Super-Resolution by Image Enhancement Using Texture Transfer," 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 953-957, doi: 10.1109/ICIP40778.2020.9190844.

[7] Y. Yang et al., "Deep Networks with Detail Enhancement for Infrared Image Super-Resolution," in IEEE Access, vol. 8, pp. 158690-158701, 2020, doi: 10.1109/ACCESS.2020.3017819.

[8] Á. Makra, W. Bost, I. Kalló, A. Horváth, M. Fournelle and M. Gyöngy, "Enhancement of Acoustic Microscopy Lateral Resolution: A Comparison Between Deep Learning and Two Deconvolution Methods," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 67, no. 1, pp. 136-145, Jan. 2020, Doi: 10.1109/TUFFC.2019.2940003.

[9] M. Rashid, B. Ram, R. S. Batth, N. Ahmad, H. M. Elhassan Ibrahim Dafallaa and M. Burhanur Rehman, "Novel Image Processing Technique for Feature Detection of Wheat Crops using Python OpenCV," 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 2019, pp. 559-563 Doi: 10.1109/ICCIKE47802.2019.9004432.

[10] X. Liu and H. Li, "An Electrolytic-Capacitor-Free Single-Phase High-Power Fuel Cell Converter with Direct Double-Frequency Ripple Current Control," in IEEE Transactions on Industry Applications, vol. 51, no. 1, pp. 297-308, Jan.-Feb. 2015, Doi: 10.1109/TIA.2014.2326085.

[11] R. Harshitha, I. A. Syed, and S. Srivasthava, "Hci using hand gesture recognition for digital sand model," in Proceedings of the 2nd IEEE International Conference on Image Information Processing (ICIIP '13), pp. 453–457, 2013.

[12] M. R. Malgireddy, J. J. Corso, S. Setlur, V. Govindaraju, and D. Mandalapu, "A framework for hand gesture recognition and spotting using sub-gesture modeling," in Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10), pp. 3780–3783, August 2010.

[13] M. Elmezain, A. Al-Hamadi, and B. Michaelis, "A robust method for hand gesture segmentation and recognition using forward spotting scheme in conditional random fields," in Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10), pp. 3850–3853, August 2010.

[14] A. D. Bagdanov, A. Del Bimbo, L. Seidenari, and L. Usai, "Real-time hand status recognition from RGB-D imagery," in Proceedings of the 21st International Conference on Pattern Recognition (ICPR '12), pp. 2456–2459, November 2012.

[15] A. Traisuwan, P. Tandayya and T. Limna, "Workflow translation and dynamic invocation for Image Processing based on OpenCV," 2015 12th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2015, pp. 319-324, Doi: 10.1109/JCSSE.2015.7219817.

[16] J. Bai, Y. Li, L. Lin and L. Chen, "Mobile Terminal Implementation of Image Filtering and Edge Detection Based on OpenCV," 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), 2020, pp. 214-218, Doi: 10.1109/AEECA49918.2020.9213537.

[17] K. Hu, S. Canavan, and L. Yin, "Hand pointing estimation for human computer interaction based on two orthogonal-views," in Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10), pp. 3760–3763, August 2010.

[18] G. Dewaele, F. Devernay, and R. Horaud, "Hand motion from 3d point trajectories and a smooth surface model," in Computer Vision—ECCV 2004, vol. 3021 of Lecture Notes in Computer Science, pp. 495–507, Springer, 2012.

[19] C. L. NEHANIV. K J DAUTENHAHN M KUBACKI M. HAEGELEC. PARLITZ R. ALAMI "A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction", 371-377 2014.

[20] JC. MANRESARVARONAR.MASF. PERALES "Hand tracking and gesture recognition for human-computer interaction",2012.

[21] H. HASAN S. ABDUL-KAREEM "Static hand gesture recognition using OpenCV", 2014.

[22] D DIAS R MADEO T. ROCHA H. BISCARO S. PERES "2009. Hand movement recognition for American sign language: a study using distance-based OpenCV.,2009