IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Handy: Media Player Controller

Anikesh Yadav[1], Rushikesh Jadhav[2], Omkar Patil[3], Kalpesh Patil[4], Prof. Tejas Tambe[5]

[1]Computer Department, P K Technical Campus
[2] Computer Department, P K Technical Campus
[3] Computer Department, P K Technical Campus
[4]Computer Department, P K Technical Campus
[5]Assistant Professor, Computer Science, P K Technical Campus, Chakan, Pune

Abstract: Hand gesture recognition is considered important with development technology in industry 4.0 in Human-Computer Interactions (HCI) which gives computers the competence to capture and interpret hand gestures the executing command without touching devices physically. The MediaPipe is present as a framework built-in machine learning that has a solution for a hand gesture recognition system. In this research, we develop a simple user guide application using the MediaPipe framework. The user guide is commonly known as documentation about technical communication or a manual in a certain system to assist people. The user guide has step-by-step descriptions about handling a particular system and helps the user deal with user frustration by giving them the means to be identified, understand, and disentangle technical problems that frequently occurred by themselves. In our experiment, we captured a real-time image using camera, then trained a variety of hand gesture data, identified each hand gesture, and recognized hand gestures to convey information based on hand gestures in the system user guide application. The user can archive information user guide based on hand gestures that have been recognized. We proposed using hand gesture recognition using MediaPipe in our application to improve the convenience of utilization the user guide application and change user guide application that is still manual become a more interactive application

***Index Terms* - HCI, hand gesture recognition, VLC media player, volume control system, Mediapipe, OpenCv.**

I. INTRODUCTION

We are now in era of industry 4.0 or the Fourth Industrial Revolution which requires automation and computerized that are realized from the consolidation between various physical and digital technologies such as sensors, embedded systems, Artificial Intelligence (AI), Cloud Computing, Big Data, Adaptive Robotic, Augmented Reality, Additive Manufacturing (AM), and Internet of Things (IoT). The enhanced digital technology connectivity made technology a crucial requirement in carrying out our daily activities like doing tasks or work, shopping, communication, entertainment, and even searching for information or news. The technology works more using the machines and advances in interaction with using a broad range of gestures to recognize, communicate, or interact with each other.

The gesture is known as a form of non-verbal communication or non-vocal communication where utilize of the body's movement that can convey a particular message originating from parts of the human body, the hand or face are the most commonly adopt. Gesture-based interaction introduced by Krueger as a new type of Human-Computer Interaction (HCI) in the middle 1970s has become a magnetic area of the research. In the Human-Computer-Interaction (HCI), building interfaces of applications with managing each part of the human body to communicate naturally are the great attention to do research, especially the hands as the most effective-alternative for the interaction tool, considering their ability.

Through Human-Computer-Interaction (HCI), recognizing hand gestures could help achieve the ease and naturalness desired. When interacting with other people, hand movements have the meaning to convey something with its information. Ranging from simple hand movements to more complex ones. For example, we can use our hand to point something (object or people) or use different simple shapes of hand or hand movements expressed through manual articulations combined with their grammar and lexicon as well-known as sign languages. Hence, using hand gestures as a device then integration with computers can help people communicate more intuitively.

In this paper we present an application which is designed for human computer interaction which uses different computer vision techniques for recognizing hand gestures for controlling the VLC media player. The aim and objectives of this application is to use a natural device free interface, which recognizes the hand gestures as commands. The application uses a webcam which is used for image acquisition. To control VLC media player using defined gesture, the application focuses on some function of VLC which are used more frequently.

II. Literature Survey

Controlling Multimedia Player with Eye Gaze Using Webcam. Advantages It uses image processing for iris detection which is a faster approach for the same. Disadvantages Does not work properly in low light. Makes use of VLC media player which is heavy on resources.

Emotion Detection Utilizing Facial Expression. Advantage Several feature extraction methods have been implemented, Numerous fields such as science, medicine, & psychology within area of interest. Disadvantages Delay in the display of results, Distraction in factors like styles, facial hairs and glass wears, uses only eye for interacting with media player which may not be feasible with low resolution cameras.

Video- Based Face Recognition using Adaptive Hidden Markov Model. Advantage Enhanced facial modelling, Improved image-based identification. Disadvantage Irrelevant result due to dynamic image recognition. Hand Gesture Recognition System to Control Slide Show Navigation. Advantage Circular profiling is controlled. Training is not required. Disadvantage High level complexity, More Time-consuming process.

III. Proposed System

Currently, many frameworks or library machine learning for hand gesture recognition have been built to make it easier for anyone to build AI (Artificial Intelligence) based applications. One of them is MediaPipe. The MediaPipe framework is present by Google for solving the problem using machine learning such as Face Detection, Face Mesh, Iris, Hands, Pose, Holistic, Hair segmentation, Object detection, Box Tracking, Instant Motion Tracking, Objection, and KIFT. MediaPipe framework helps a developer focus on the algorithm and model development on the application, then support environment application through results reproducible across different devices and platforms which it is a few advantages of using features on the MediaPipe framework

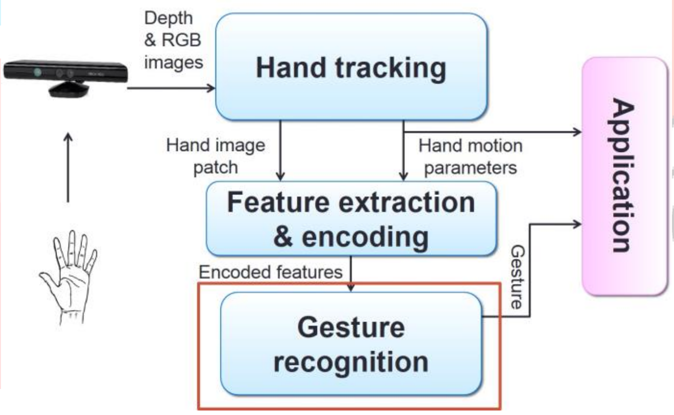
3.1 Existing System

In general, the offered frameworks apply eye acknowledgment that results in inaccurate output. Hand motions are not carried out as expected.

Drawback of existing system

- Most of the system are not able to provide real time monitoring and doesn’t respond quickly
- Accuracy issues of Machine learning based hand recognition algorithm, due to this system can’t able to identify the user hands if background and hand skin colour matches.
- Slow because use external hand gloves for detection

3.2 Proposed System



Media-Pipe Hands utilizes an ML pipeline consisting of multiple models working together: A palm detection model that operates on the full image and returns an oriented hand bounding box. A hand landmark model that operates on the cropped image region defined by the palm detector and returns high-fidelity 3D hand key points.

It employs machine learning (ML) to infer the 3D hand surface, requiring only a single camera input without the need for a dedicated depth sensor. Utilizing lightweight model architectures together with GPU acceleration throughout the pipeline, the solution delivers real-time performance critical for live experiences.

Under the hood, a lightweight statistical analysis method called Procrustes Analysis is employed to drive a robust, performant and portable logic. The analysis runs on CPU and has a minimal speed/memory footprint on top of the ML model inference.

IV. Methodology & Related Work

Today, there are many frameworks or libraries of machine learning for hand gesture recognition. One of them is MediaPipe. The MediaPipe is a framework designed to implement production-ready machine learning that must build pipelines to perform inference over arbitrary sensory data, has published code accompanying research work, and build technology prototypes. In MediaPipe, graph modular components come from a perception pipeline along with the function of inference model function, media processing model, and data transformations. Graph of operations are used in others machine learning such as Tensor flow, MXNet, PyTorch, CNTK, OpenCV 4.0.

Using MediaPipe for hand gesture recognition has been researched by Zhang before, using a single RGB camera for AR/VR application in a real-time system that predicts a hand skeleton of the human. We can develop a combined MediaPipe using other devices

4.1 MediaPipe Framework

The MediaPipe implements pipeline in Figure 1. consists of two models for hand gesture recognition as follows:

1. A palm detector model processes the captured image and turns the image with an oriented bounding box of the hand.
2. A hand landmark model processes on cropped bounding box image and returns 3D hand key points on hand.
3. A gesture recognizer that classifies 3D hand key points then configuration them into a discrete set of gestures.

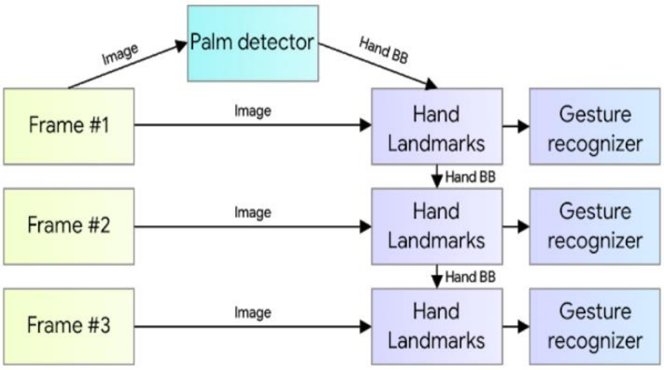


Figure 1

4.1.1 Palm Detector Model

MediaPipe framework has built detect initial palm detector called BlazePalm. Detecting the hand is a complex task. Step one is to train the palm instead of the hand detector, then using the non-maximum suppression algorithm on the palm, where it is modeled using square bounding boxes to avoid other aspect ratios and reducing the number of anchors by a factor of 3-5. Next, encoder-decoder of feature extraction that is used for bigger scene context-awareness even small objects, lastly, minimize the focal loss during training with support a large number of anchors resulting from the high scale variance

4.2.2 Hand Landmark

Achieves precise key point localization of 21 key points with a 3D hand-knuckle coordinate which is conducted inside the detected hand regions through regression which will produce the coordinate prediction directly which is a model of the hand landmark in MediaPipe, see in Figure 2

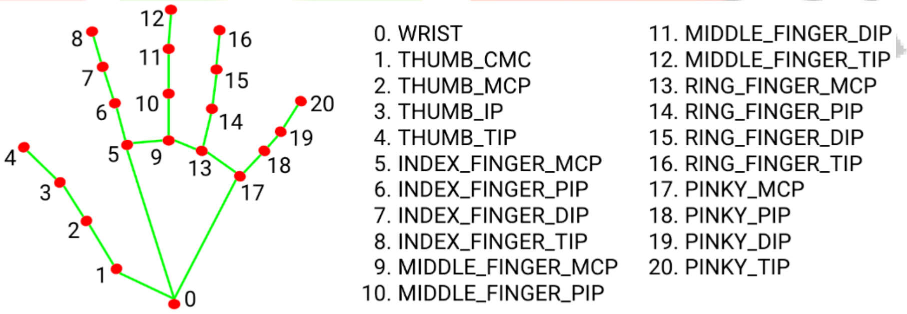


Figure 2

Each hand-knuckle of the landmark has coordinate is composed of x, y, and z where x and y are normalized to [0.0, 1.0] by image width and height, while z representation the depth of landmark. The depth of landmark that can be found at the wrist being the ancestor. The closed the landmark to the camera, the value becomes smaller.

V. Research Methods

In this research, the VLC media application controller allows user to display steps taken by the system by identifying hand gestures as a certain command

We developed an application that implements hand gestures recognition using camera to capture hand pose and then recognize it for running the application. We are using the MediaPipe framework and python programming language to develop an application. For detail, can see Figure 3 below.

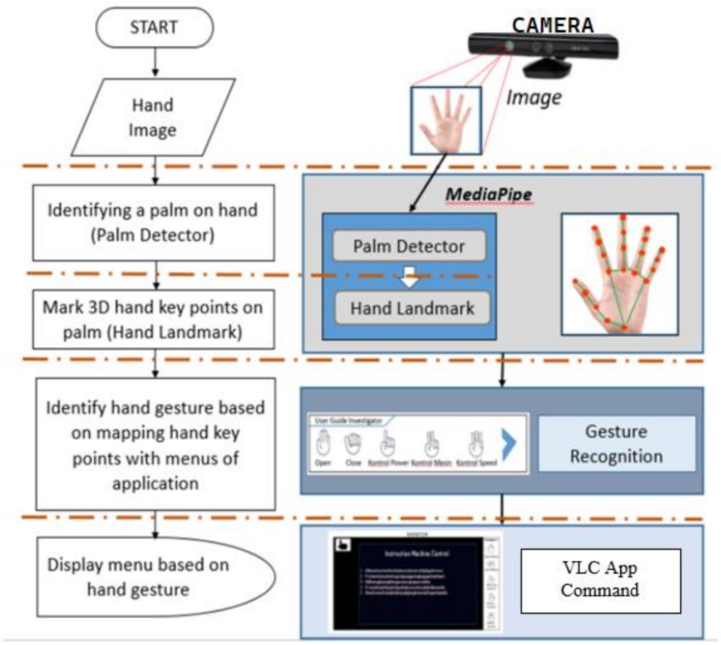
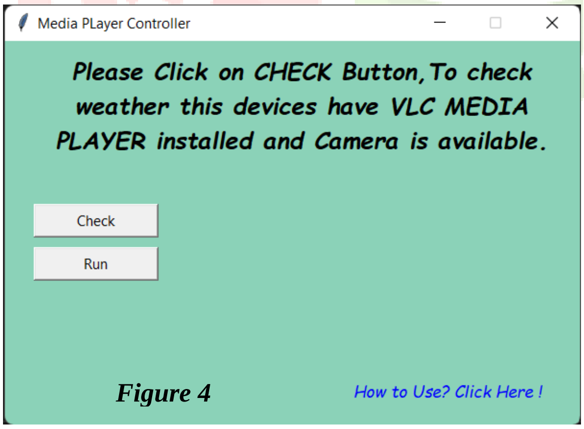

START
Hand Image
CAMERA
Image
Identifying a palm on hand (Palm Detector)
MediaPipe
Palm Detector
Hand Landmark
Mark 3D hand key points on palm (Hand Landmark)
Identify hand gesture based on mapping hand key points with menus of application
Gesture Recognition
Display menu based on hand gesture
VLC App Command

Figure 3

We use the camera x360 with RGB camera with a resolution of 640x240 pixel for capturing a real-time image for process using MediaPipe framework. MediaPipe will read the image that received from camera, then on an image will do palm detection and make hand landmark that made return 3D hand key points and joint it make up like skeleton. 3D key points in the palm that has been marked in the image will be computed and initialized as a tool for reading pose hand and recognition based that will be conveyed information based on hand pose had been initialized before.

5.1 Handy: Media Player Application

To develop handy application, we prepared a window to check all the requirements to run our application. As show in figure 4. The system will capture an image of a hand pose with all fingers had stated open, identifying it then do command to display mock-up as seen in Figure 5 is shown a application that contains hand pose is initialized with a tracking of 21 key points with a 3D hand-knuckle coordinate If the user performs defined gestures than application preform that defined function of VLC media player without any delay. The right corner on the mock-up shows hand getting tracked continuously.



Media PLayer Controller

Please Click on CHECK Button,To check weather this devices have VLC MEDIA PLAYER installed and Camera is available.

Figure 4

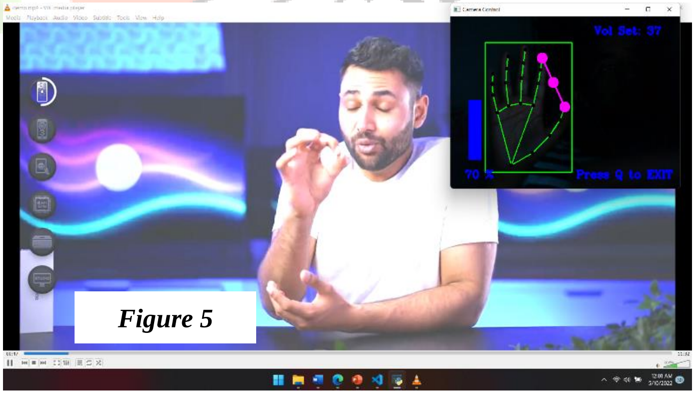
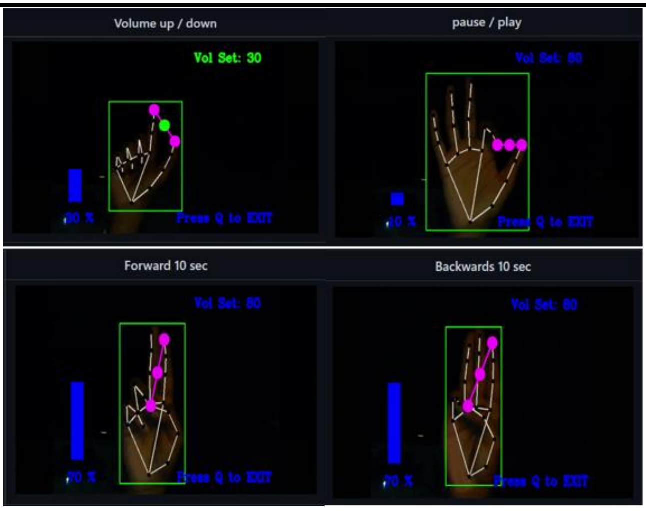
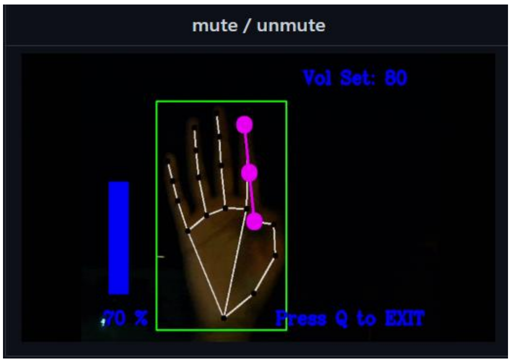


Figure 5

Gesture you can use in this applications are fast-forward for 10 sec, Pause, Play, Mute, Unmute, fast-backward for 10 sec & Volume up/down as show in figure 6





VI. Result

Using MediaPipe for implementing machine learning on hand gesture recognition in a handy application achieved good performance. We can also see that a few hand gestures could make a false prediction of hand gestures. Results in validation accuracy of 95% for hand gestures to be recognized. A false prediction of hand gestures was getting the percentage accuracy performance to become descend. This is caused by lighting, a distance between camera as a picture catcher and user while using the application, and degrees of angle from camera is placed. Figure 4 shows the result of images of deploying the handy application using hand gestures as a command for display information

VII. Future Scope

The present application is robust in recognition phase. If somehow, we manage to add it as a dedicated feature for certain devices like controlling whole television and music set with hand gestures of entire room. It will more convenient to use. We can also implement this idea with VR technologies or Augmented Reality technologies.

VIII. Conclusions

The Hand gesture recognition system has become an important role in building efficient human-machine interaction. Implementation using hand gesture recognition promises wide-ranging in technology industry. The MediaPipe as one framework based on machine learning plays an effective role in developing this application using hand gesture recognition, with the result has shown an accuracy performance of 95%. We would like to extend our system further to develop collaboration with other devices and other human body parts and experiment with both static and dynamic hand gesture recognition systems**.**

References

[1] R. Jain, M. Jain, R. Jain, and S. Madan, “Human Computer Interaction – Hand Gesture Recognition”, Adv. J. Grad. Res., vol. 11, no. 1, pp. 1–9, Sep. 2021.

[2] Ge, L.; Ren, Z.; Li, Y.; Xue, Z.; Wang, Y.; Cai, J.; Yuan, J. “3d hand shape and pose estimation from a single rgb image”. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, pp. 10833–10842, 15–20 June 2019.

[3] Perimal, M.; Basah, S.N.; Safar, M.J.A.; Yazid, “Hand-Gesture Recognition-Algorithm based on Finger Counting”, J. Telecommun. Electron. Comput. Eng.,10, 19–24, 2018.

[4] https://doi.org/10.21467/ajgr.11.1.1-9

[5] OpenCV Tutorial (tutorialspoint.com)

[6] Hands - mediapipe (google.github.io)

[7] International Journal of Computer Applications (0975 – 8887) Volume 10– No.7, November 2018-19

[8] Liuhao Ge, Hui Liang, Junsong Yuan, and Daniel Thalmann. Robust 3d hand pose estimation in single depth images: from single-view cnn to multi-view cnns. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3593–3601, 2016

[9] MediaPipe Hands: On-device Real-time Hand Tracking Fan Zhang Valentin Bazarevsky Andrey Vakunov Andrei Tkachenka George Sung Chuo-Ling Chang Matthias Grundmann.