International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 – June, 2021 ISSN: 2582-3930

# Look Based Media Player with Hand Gesture Recognition

**Saritha M [1], Soniya N [2], Sharanya [3], Skandashree H M [4], Veena B [5]**

[1] *Assistant Professor Computer Science Department, SDM Institution of Technology*
[2,3,4,5] *B.E Computer Science Student, SDM Institution of Technology*

---

**Abstract -** When you are watching a video and someone calls you have to look somewhere else or go away from PC for some time, so you miss some part of the video. Later you need to drag back the video from where you saw it. Look Based Media Player with Hand Gesture Recognition, is a media player which pauses itself when user is not looking at it. The player starts running again as soon as the user looks at it. This is done using the camera module or web camera on top of the computer. In this project we are developing an advanced media player that uses face detection to automatically play and pause the video, hand gesture recognition system to forward and backward the video. System monitors whether the user is looking at the screen or not using Haar-cascade Classifier. In case if the user is not looking at the screen or if the system couldn't detect the users face then it immediately stops the video. Controlling other functions of media player such as playing next and previous videos is done using Convolutional Neural Network. It also provides the feature of controlling functions of media players such as detection and comparing noise from environment's input to machine's output and if the input is higher, then the media player will pause.

***Key Words*: Haar-cascade Classifier, Hand-gesture Recognition, OpenCV, Face detection, Local Binary Pattern Histogram, Convolutional Neural Network.**

## 1. INTRODUCTION

Human Computer Interaction can acquire several advantages with the introduction of different natural forms of device free communication. Gestures are a natural form of actions which we often use in our daily life for interaction, therefore to use it as a communication medium with computers generates a new paradigm of interaction with computers. This paper implements computer vision and gesture recognition techniques and develops a vision based low cost input device for controlling the media player through gestures. VLC application consists of a central computational module which uses the Principal Component Analysis for gesture images and finds the feature vectors of the gesture and save it into a XML file. The Recognition of the gesture is done by Convolutional Neural Networks. The theoretical analysis of the approach shows how to do recognition in static background. This hand gesture recognition technique will not only replace the use of mouse to control the VLC player but also provide different gesture vocabulary which will be useful in controlling the application.

The face recognition is a technique to identify or verify the face from the digital images or video frame. A human can quickly identify the faces without much effort. It is an effortless task for us, but it is a difficult task for a computer. There are various complexities, such as low resolution, occlusion, illumination variations, etc. These factors highly affect the accuracy of the computer to recognize the face more effectively. First, it is necessary to understand the difference between face detection and face recognition. The face detection is generally considered as finding the faces (location and size) in an image and probably extracts them to be used by the face detection algorithm. The face recognition algorithm is used in finding features that are uniquely described in the image. The facial image is already extracted, cropped, resized, and usually converted in the grayscale.

There are various algorithms of face detection and face recognition. Here we have used face detection using the Haar-cascade Classifier.

The goal of our project is to create an advanced media player based on look and hand gestures. We have set the following objectives to achieve the goal:

A. The user interface of media player should be efficient and user friendly.
B. The media player should be accurate in terms of result.
C. The media player pause the video as soon as the users' face is not detected without much latency.
D. Not missing any part of video.
E. The hand gestures should be captured precisely and actions related to them should be performed accurately.

## 3. DESIGN OF THE PROPOSED SYSTEM

In existing systems, face detection and hand gesture recognition has been primarily conducted in a constrained environment as well as the accuracy rate is poor. Mostly existing systems use eye recognition. Due to which results aren't accurate. Face recognition and hand gestures are not implemented properly together and not even individually.

In the proposed system we are using face detection and hand gestures recognition system for controlling media player. Face recognition is used for pausing and playing the video. Various hand gestures are used for controlling other functions of media player such as playing next video, previous video and stop the video.

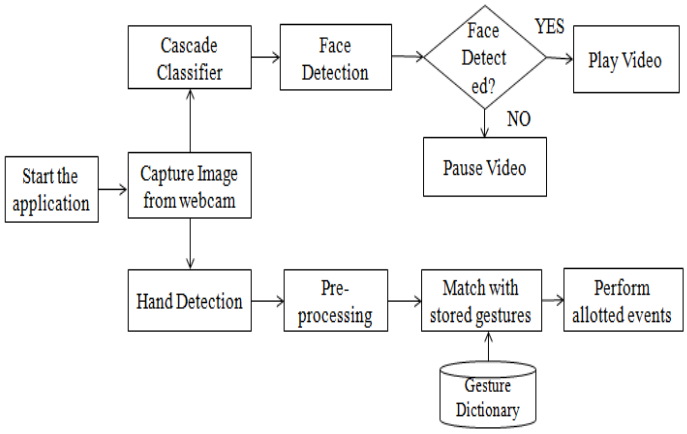



Figure: Architecture of the proposed system

When user runs the application, the web-camera starts capturing the frames. It is divided into three conditions as follows:

A. User watching the video: When the user is looking the screen or when the face is detected, the video continues without any interruption.

B. User not watching the video: When the user stops watching video or when the face is not detected, the video pauses automatically until the face is detected again.

C. User showing hand gestures: When the user shows some hand gestures such as any numeric values. If the user shows one, the control goes to next video or if the user shows two, the control goes to previous video in the play track or if the user shows three, the media player stops completely.

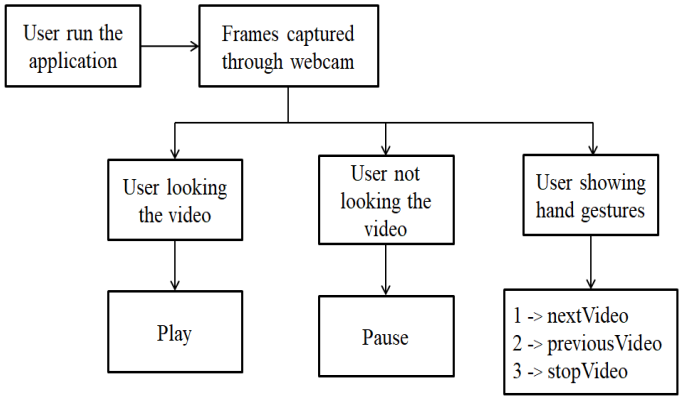



Figure: User Interface of the proposed system

## 4. METHODOLOGY

A. OpenCV is a Python open-source library, which is used for computer vision in Artificial intelligence, Machine Learning, face recognition, etc. In OpenCV, the CV is an abbreviation form of a computer vision, which is defined as a field of study that helps computers to understand the content of the digital images such as photographs and videos. The purpose of computer vision is to understand the content of the images. It extracts the description from the pictures, which may be an object, a text description, and three-dimension model, and so on. For example, cars can be facilitated with computer vision, which will be able to identify and different objects around the road, such as traffic lights, pedestrians, traffic signs, and so on, and acts accordingly.

Computer vision allows the computer to perform the same kind of tasks as humans with the same efficiency. There are a two main task which are defined below:

- Objects Classification - In the object classification, we train a model on a dataset of particular objects, and the model classifies new objects as belonging to one or more of your training categories.
- Objects Identification - In the object identification, our model will identify a particular instance of an object.

The picture intensity at the particular location is represented by the numbers. In the above image, we have shown the pixel values for a grayscale image consist of only one value, the intensity of the black color at that location. There are two common ways to identify the images:

- Grayscale images are those images which contain only two colors black and white. The contrast measurement of intensity is black treated as the weakest intensity, and white as the strongest intensity. When we use the gray-scale image, the computer assigns each pixel value based on its level of darkness.
- An RGB is a combination of the red, green, blue color which together makes a new color. The computer retrieves that value from each pixel and puts the results in an array to be interpreted.

B. Haar-cascade Classifier is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images. Here we will work with face detection. Initially, the algorithm needs a lot of positive images and negative images to train the classifier. Then we need to extract features from it. For this, Haar features shown in the below Figure are used. They are just like our convolutional kernel. Each feature is a single value obtained by subtracting sum of pixels under white rectangle from sum of pixels under black rectangle.

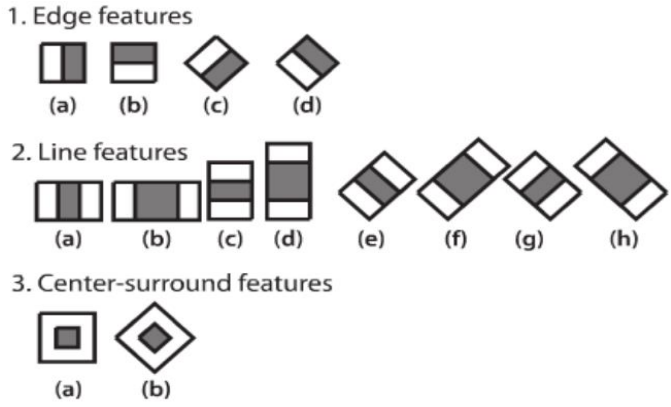



Figure: Haar features

C. Convolution Neural Network is used in finding features that are uniquely described in the image. The hand image is already extracted, cropped, resized, and usually converted in the grayscale. By performing predefined gestures users can make use of various media player functions. In this system various gestures are assigned to move to the next video and the previous video. Steps for hand gesture recognition are as follows:

- Input image - OpenCV allows a straightforward interface to capture live stream with the camera (webcam). It converts video into grayscale and display it.
- HSV - Hue Saturation Value scale provides a numerical readout of the captured image.
- Binary image - The image consists of only two color, black and white in this case.
- Convex Hull - The convex hull of binary image is the set of pixels included in the smallest convex polygon that surround all white pixels in the input image.
- Convexity Defects – For finding the convexity defects of a contour.

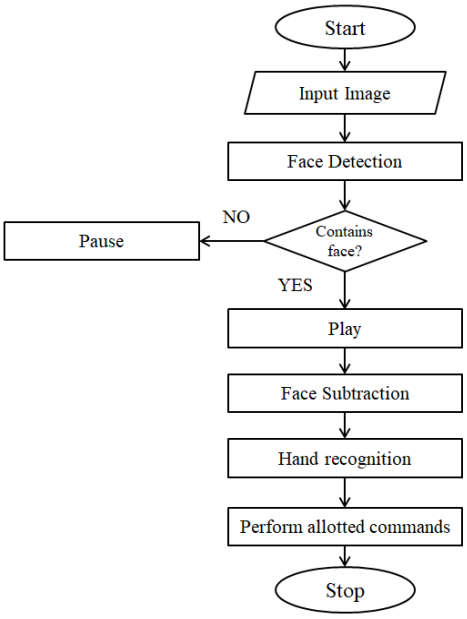



Figure: Flowchart of the proposed system.

## 5. RESULTS AND DISCUSSION

The terms used as validation metrics for verifying quality of a segmented image. In a scenario where you want to compare a segmented image with ground truth, then taking the ground truth image as base of comparison you can make assumption of taking foreground as "white" pixels and background as "black" pixels in ground-truth. The metrics calculated using TP (True Positive), FP (False Positive), FN (False Negative), FP (False Positive) are mentioned below:

- Sensitivity: The sensitivity tells us how likely the test is come back positive in someone who has the characteristic. This is calculated as TP/ (TP+FN).
- Specificity: The specificity tells us how likely the test is to come back negative in someone who does not have the characteristic. This is calculated as TN/ (TN+FP).
- Accuracy: (TP+TN)/(TP+FP+TN+FN)

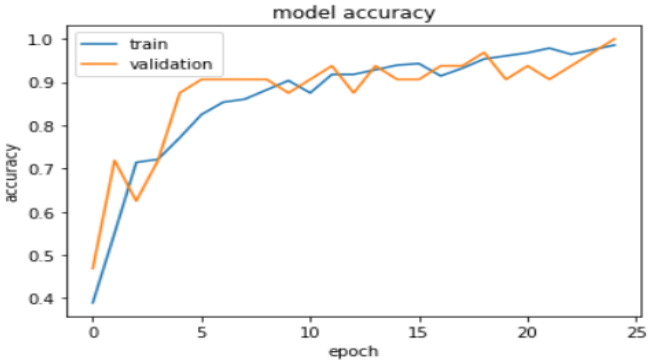



Figure: Model accuracy

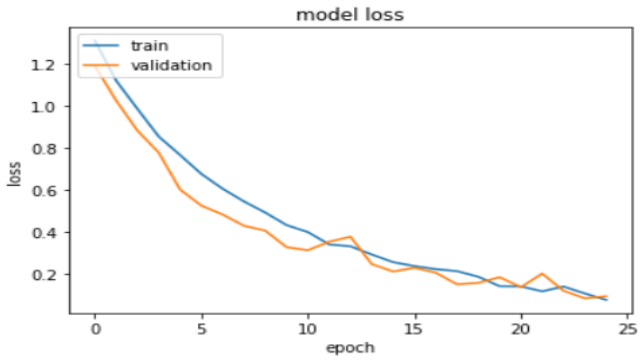



Figure: Model loss

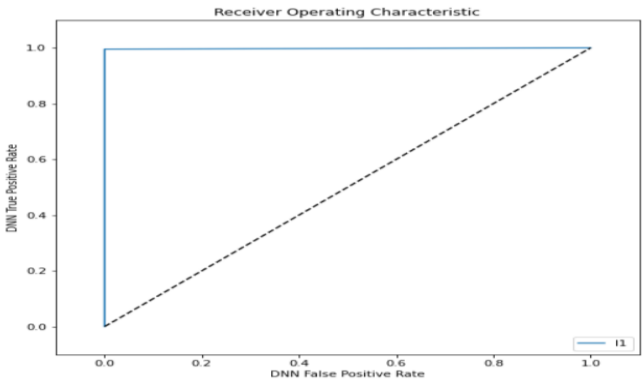


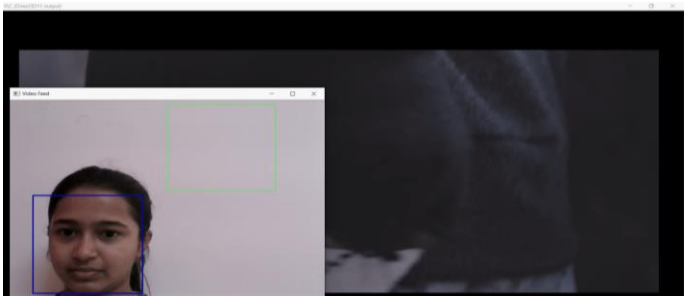

Figure: ROC curve

A ROC curve (receiver operating characteristic curve) is a graph showing the performance of the model. This curve plots two parameters: True Positive Rate (Sensitivity) and False Positive Rate (Specificity).

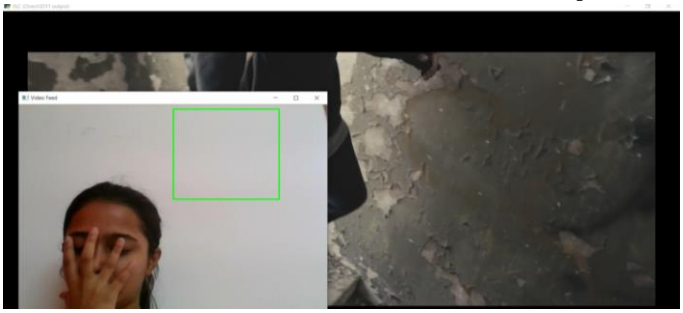
| SL.No | Performance metrics | Convolutional Neural Network (CNN) | Haar cascade classifier |
|---|---|---|---|
| 1 | accuracy | 99 | - |
| 2 | loss | 0.06 | - |
| 3 | function | Hand gesture recognition | Face detection |

Table: Performance metrics of CNN and Haar-cascade classifier

Case 1: When the users' face is detected, video is playing.

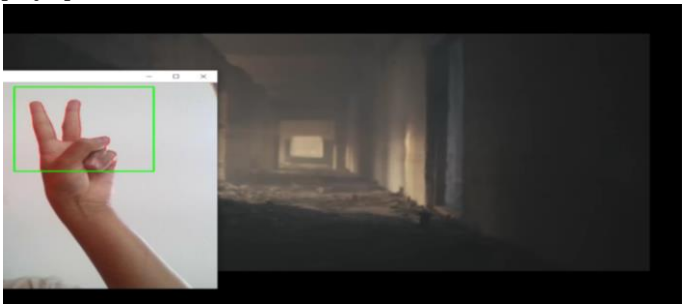


Case 2: When the user's face is not detected, video is paused.



Case 3: When user shows hand gesture one, media player plays next video.


Case 4: When user shows hand gesture two, media player plays previous video.



## 6. CONCLUSION

The user gets better experience of using advanced media player. Hand gesture recognition and face detection is used for controlling the functions of the media player, helps the user to watch keyboard and mouse free. Playing and pausing the video based on user watching the video through face detection is done using Haar-cascade classifier. Controlling other functions of media player such as playing next and previous video is done through hand gesture recognition system.

This project can be further enhanced to detect the facial expression and control the media player according to the expression. There are several studies that should be carried on in future work. This system can be implemented in various technologies like devices and applications that do not have traditional mouse and keyboard interface.

## 7. REFERENCES

[1] Sidharth Rautaray et Anupam Agrawal, "A Vision Based Hand Gesture Interface for Controlling VLC Media Player", Journ. International Journal Of Computer Applications (0975-8887), vol. 10-No.7, November 2010.

[2] D Jadhav et Prof L.M.R.J Lobo "Hand Gesture Recognition System To Control Slide Show Navigation", Journ. International Journal Of Applications or Innovation in Engineering & Management (IJAIEM)", vol. 3, PP-283-286, January 2014.

[3] N. Krishna Chaitanya et R. Janardan Rao "Controlling of Windows Media Player Using Hand Recognition System", Journ. The International Journal Of Engineering And Science (IJES), vol. 3, PP 01-04, 2014.

[4] Xiaoming Liu et Tsuhan Chen "Video-Based Face Recognition using Adaptive Hidden Markov Model", Electrical and Computer Engineering, Caregie Mellon University, Pittsburgh, PA, 15213, U.S.A.

[5] Davis A, Nordholm S, Togneri R: Statistical voice activity detection using low-variance spectrum estimation and an adaptive threshold. Audio, Speech, Lang. Proc. IEEE Trans 2006,14(2):412–424.

[6] Jyoti Rani et Kanwal Garg, "Emotion Detection Using Facial Expression", Journ. International Journal Of Advanced Research In Computer Science & Software Engineering, Vol. 4, PP-465467, April 2014.

[7] Ghosh P, Tsiartas A, Narayanan S: Robust voice activity detection using long-term signal variability. Audio, Speech, and Lang. Proc. IEEE Trans 2011,19.
[8] Prasad RV, Sangwan A, Jamadagni HS, Chiranth MC, Sah R, Gaurav V: Comparison of voice activity detection algorithms for VoIP. In Proceedings of the Seventh International Symposium on Computers and Communications (ISCC'02), Washington, Piscataway: IEEE; 2002:530–535.
[9] Harsha Jadhav, Sabiha Pathan, Neha Rokade et Uma Annamalai "Controlling Multimedia Applications Using Hand Gesture Recognition", Journ, International Research Journal Of Engineering and Technology (IRJET), vol. 02, PP 1200-1203, August-2014.
[10] Rabiner LR, Sambur MR: An algorithm for determining the end points of isolated utterances. Bell Syst. Techn. J, 1975.
[11] Paul Viola and Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001.