AI VIRTUAL PAINTING ON VIDEO CALLS USING CV

Alet Joseph Byju , *Student, CSE, IESCE, Kerala, India*
Arjun K R , *Student, CSE, IESCE, Kerala, India*
Delvin V D , *Student, CSE, IESCE, Kerala, India*
Mridhul M L , *Student, CSE, IESCE, Kerala, India*
Guide: Ms. Neethu Prabhakaran , Faculty, CSE, IESCE, Kerala, India

ABSTRACT

The use of videoconferencing technology has increased as an alternative communication method for people to retain social and clinical contact during Covid-19 restrictions. In order to keep in contact with each other, many have taken to using video meeting software and apps, a significant amount of people experiencing videoconferencing for the very first time, and some of the stories have been incredibly heartwarming. With a huge portion of the population now finally discovering the true potential and practicality of video technology, we thought it would be a fitting time to look a little further into what current platforms offer and what to consider in the future. Our project aims to improve the existing video meeting platforms by integrating Artificial Intelligence powered hand gesture recognition and motion capture to give a virtual painting ability for the users in Real-Time. The proposed system is a A Peer-to-Peer video conferencing platform as a progressive web application[1] that recognize hand gestures and also tracks the motion of the hands that lets you virtually draw on top of the video. The proposed project enables the users to use the system as a virtual JamBoard that can be directly used on top of the video feed itself without having to open any other external links. Our proposed project idea is a novelty among other video conferencing platforms. In this post pandemic world, our idea when implemented would make life easy for everyone who wants to create or participate in online meetings.

Keywords: - *artificial intelligence, progressive web app, WebRTC, computer vision, hand motion capture, hand gesture recognition, javascript*

1. INTRODUCTION

Our project idea aims to improve the existing video meeting platforms by integrating Artificial Intelligence powered hand gesture recognition and motion capture to give a virtual painting ability for the users in Real-Time. The proposed project enables the users to use the system as a virtual JamBoard that can be directly used on top of the video feed itself without having to open any other external links.

A Peer-to-Peer video conferencing platform as a progressive web application that recognize hand gestures and also tracks the motion of the hands that lets you virtually draw on top of the video. The users of our platform does not need to switch to any other third party apps nor do they have to present their screen to show or draw anything they are struggling to convey. The app itself allows the users to do the same and more.

2. LITERATURE REVIEW

2.1 Capturing Human Hand Motion in Image Sequences

Visually capturing human hand motion requires estimating the 3D hand global pose as well as its local finger articulations. This is a challenging task that requires a search in a high dimensional space due to the high degrees of

free- dom that fingers exhibit and the self occlusions caused by global hand motion. In this paper we propose a divide and conquer approach to estimate both global and local hand motion. By looking into the palm and extra feature points provided by fingers, the hand pose is determined from the palm using Iterative Closed Point (ICP)[2] algorithm and factorization method. The hand global pose serves as the base frame for the finger motion capturing. Noticing the natural hand motion constraints, we propose an efficient tracking algorithm based on sequential Monte Carlo technique for tracking finger motion. To enhance the accuracy, pose estimations and finger articulation tracking are performed in an iterative manner. Our experiments show that our approach is accurate and robust for natural hand movements.

Hand gestures can be a more natural and articulate way for many human computer interaction applications. For instance, people can use their hands to manipulate virtual objects directly in virtual environments. But one of the main difficulties is how to capture human hand motion. As an alternative to glove-based techniques that requires users to wear a special data glove, vision-based techniques are non- invasive and more affordable. However, capturing hand and finger motion from video sequences is an extremely challenging task. Typical hand motions consist of global translation, rotation, and natural finger movements. Due to the hand anatomy, hand rotations often introduce self- occlusions that cause some of the fingers to become invisible. Since the finger motion has high degrees of freedom, many techniques of estimating finger articulations often involve a formidable search problem in a high dimensional space.

2.2 Hand Gesture Recognition Using Deep Learning

In order to offer new possibilities to interact with machine and to design more natural and more intuitive interactions with computing machines, our research aims at the automatic interpretation of gestures based on computer vision. In this paper, we propose a technique which commands computer using six static and eight dynamic hand gestures. The three main steps are: hand shape recognition, tracing of detected hand (if dynamic), and converting the data into the required command. Experiments show 93.09% accuracy.

Gesture recognition is the mathematical interpretation of a human motion by a computing device. Modern research of the control of computers changes from standard peripheral devices to remotely commanding computers through speech, emotions and body gestures. Our application belongs to the domain of hand gesture recognition which is generally divided into two categories i.e. contact-based and vision-based approaches. The second type is simpler and intuitive as it employs video image processing and pattern recognition.

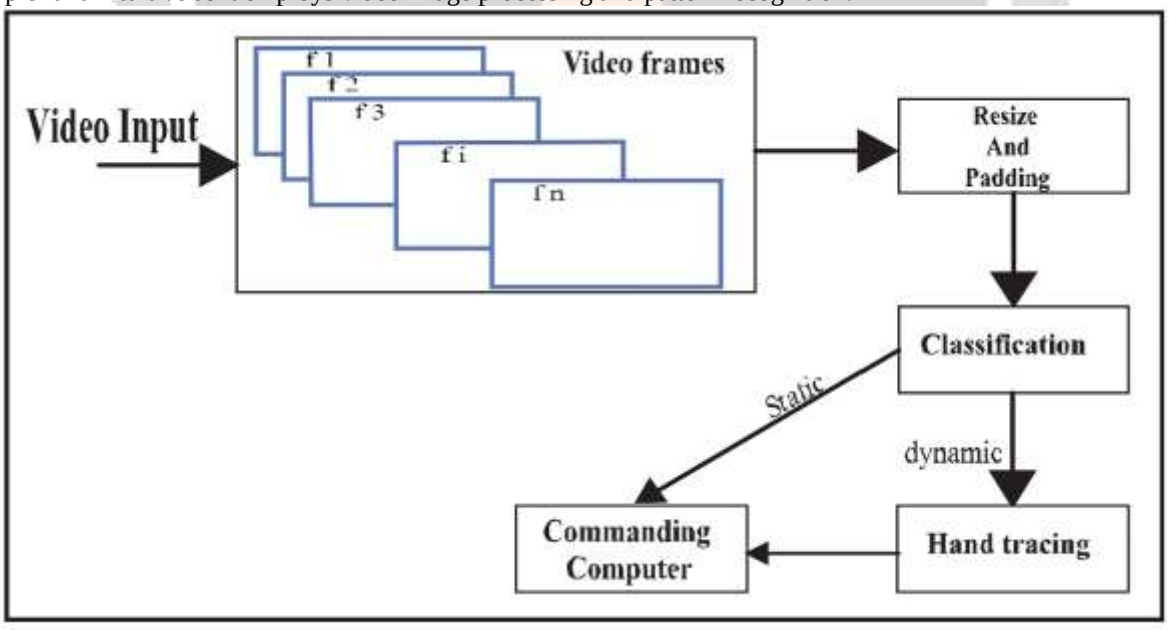

Fig.- 1. Hand gesture recognition block diagram

2.3 Hand Tracking and Motion Detection from the Sequence of Stereo Color Image Frames

This paper presents an approach for hand tracking which is based on color and motion. In particular, the YcbCr color system and the difference of the two consecutive frames of images are considered for detection of a moving hand. Since the luminance component is concentrated in Y, while the chrominance components are in Cb and Cr, the image region in which the luminance varies significantly is able to be effectively selected by the chrominance components only. All the objects and background that are not in motion are removed from the image by the difference operation. The head and the hand that may slightly move are also narrowed down by the difference operation. Thus, the region of the moving hand is extracted from a composite image. This approach is useful in detecting moving object of narrow banded color. The locations of the centers of hands or objects are obtained in a sequence of stereo images. The movement of a hand or an object is estimated with an equation that operates the set of locations obtained. A complete tracking of a hand or an object is described and presented.

Real-time hand tracking in vision systems can be applied to control different functions of computers, robots, or TV . The location, motion and pose of the hands could contain the information necessary for the purpose of interaction between the human and machine. This information can be analyzed by the vision system and converted to meaningful commands which could substitute the signals of the existing tedious media.

2.4 A WebRTC-Based Video Conferencing System with Screen Sharing

The popularity of mobile devices makes it possible to promote communication and collaboration among different devices. In this paper, we propose a video conferencing system with enhanced screen sharing feature. In order to make the screen sharing across platforms, we proposed a scheme based on the WebRTC[3] technology under the Browser/Server framework. Both the system architecture and its components are described in detail in the paper. Compared with the traditional screen sharing systems, this proposed WebRTC based scheme not only brings a cross platform, cross-device and multifunctional user experience, but also insures good quality even in the low bit rate communication networks.

Screen sharing aims at sending the compressed screen images from one computer to another one through networks, which could support a lot of applications such as wireless display and remote assistance. In particular, this feature plays an essential role in the video conferencing applications for multi-party remote collaborations. When developing an end-to-end solution for interactive screen sharing, we need to make sure that the system has high visual quality, low delay and bandwidth cost, as well as supporting various clients.

There are some commercial systems supporting screen sharing, e.g., AirPlay of Apple Inc. and Chromecast of Google Inc., which can share the computer screen to other lightweight devices. Besides, some recent research on cloud computing has proposed a solution of sharing the computer screen with other computers and mobile devices through the cloud. In particular, the screen needs to be virtualized and rendered on the cloud side first, and then be sent to the clients through Internet.

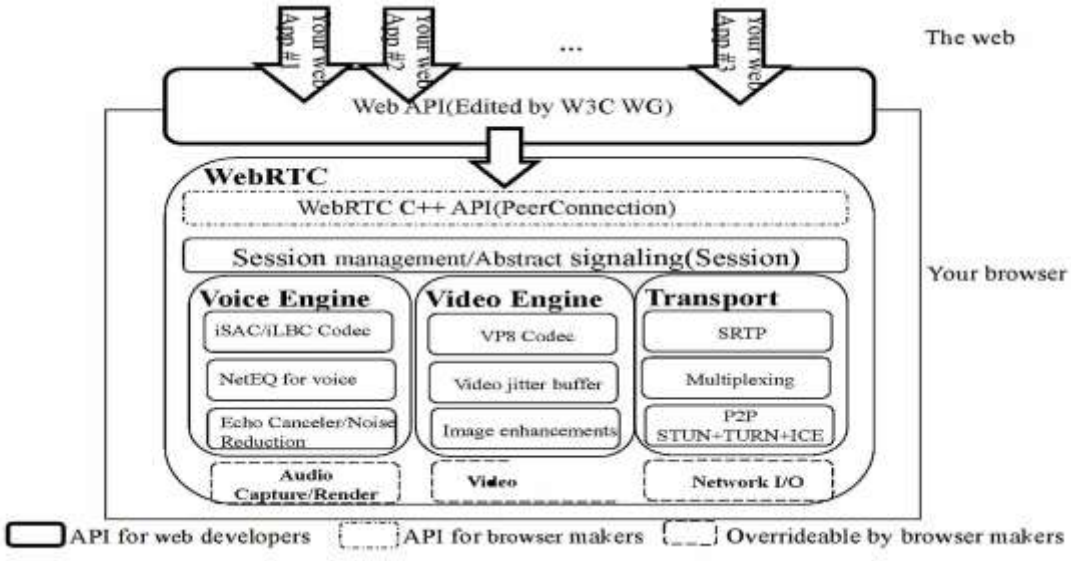

Fig.- 2. Overview of WebRTC[4]

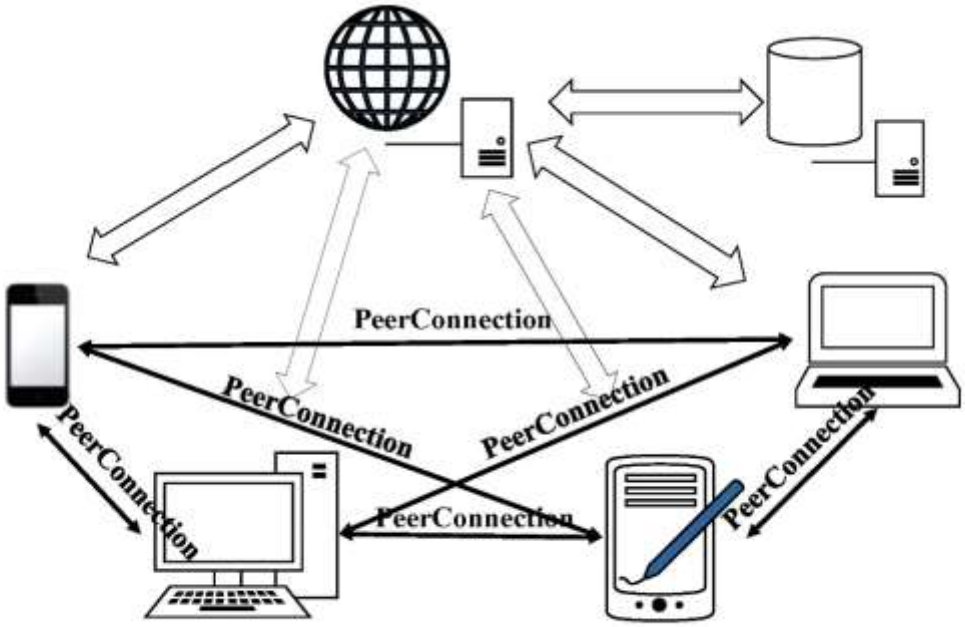

Figure 3. The structure of the WebRTC screen sharing system.

3. PROS & CONS

3.1 Advantages

- A new "smart" AI powered video chat app that is smart enough to convey your imagination more effectively & efficiently.
- No need to switch between 3rd party apps/services to share your thoughts.
- Users doesn't have to present their screen to doodle.
- End-to-End encryption.

3.1 Disadvantages

- Possibility of slight delay/lag for slow internet connections.
- Won't be an ideal User Experience for handicapped individuals.

4. CONCLUSIONS

A WebRTC based AI powered video conferencing system has been proposed and presented in this paper. The combined method of color and motion works sufficiently accurate to track moving hands and objects. Due to simplified approach the method could be applied for real-time tracking intended for human-computer interaction. The system after careful analysis has been identified that every user is a peer. All the participants in the conference make use of peer-to-peer communication for better ping. The users have the control to switch their camera & mic ON/OFF. Our proposed project idea is a novelty among other video conferencing platforms. In this post pandemic world, our idea when implemented would make life easy for everyone who wants to create or participate in online meetings.

5. ACKNOWLEDGMENT

The authors would like to thank the reviewers for their constructive comments. We would also like to thank IES COLLEGE OF ENGINEERING THRISSUR of Dr. APJ Abdul Kalam Technological University for supporting tools and environment for this research.

6. REFERENCES

[1]. https://developer.mozilla.org/en-US/docs/Web/Progressive_web_apps
[2]. https://en.wikipedia.org/wiki/Iterative_closest_point
[3]. https://developer.mozilla.org/en-US/docs/Web/API/WebRTC_API
[4]. WebRTC Architecture [S/OL] .
https://www.webrtc.org/architecture

- https://ieeexplore.ieee.org/document/1182220
- https://ieeexplore.ieee.org/document/4608702
- https://ieeexplore.ieee.org/document/8368821
- https://ieeexplore.ieee.org/document/7924748
- https://towardsdatascience.com/handtrackjs-677c29c1d585