Review Paper on Computer Control using Hand and Gaze Detection

Harshal Patil , Dhairya Bhadra, Parshva Dedhia, Shubhada Labde
K.J. Somaiya Institute of Engineering and Information Technology, Sion East, Mumbai, India
harshal.sp@somaiya.edu,dhairya.bhadra@somaiya.edu,parshva.dedhia@somaiya.edu,shubhada.l@somaiya.edu

Abstract— Gestures which are the movement of a part of the body to express an idea or meaning or to act in order to convey a feeling or intention. Every day, gestures are used like waving hand without us thinking about them. Gestures are an important part of communication. The objective of this paper is to use hand and gaze detection to interact with the computer for various different purposes. Compared to existing methods, hand gesture methodology has the bonus of being simple to use. By using this method, the traditional way to use the mouse and keyboard is modified because you can instead communicate with hand gestures device. Gaze tracking is a mouse movement technique which moves the mouse according to the movement of your eye ball along with the head movement to interact with the machine or the computer system. In this paper, an attempt has been made to combine both these algorithms to achieve more accuracy.

Keywords- Gesture Recognition, Hand gestures, Gaze, CNN, System

I. INTRODUCTION

Right now computers are being used by almost every individual and the demand is still increasing. So, by Moore's law it is expected that each common people will be able to use the computer as the technology is advancing at a very high rate. Today it is necessary to have human machine interaction and keyboard, mouse and pens are not sufficient for interaction. As the technology is growing day by day, and the interactions between humans and the machines is also increasing. The use and the implementation of the gesture have been since many years. Earlier, the gestures where not efficient. The best example of it is when the Samsung that launched the "Air gesture" with its Samsung galaxy s4 in the year of 2013, which ended up in total failure because the gesture where hardware base. At first the operator had to use different hardware equipment like gloves and different sensor to read to the movement of the body. The process is too long and also difficult in real time. Recent trend uses the methods of computer vision to interpret the movements. For hand gesture identification principle component analysis is used. This required more power and memory and made the machine more slower. For head motion Hidden Markov model is used. But this requires training of head gestures. Many different models use the combination of hand and facial expression or hand and speech recognition or facial expression and speech recognition. The main aim is to develop a system where we can simultaneously control the computer system by using our hand and gaze.

II. LITERATURE REVIEW

In *COMPUTER CURSOR CONTROL MECHANISM BY USING HAND GESTURE RECOGNITION* by *Kalyani Pendke, Prasanna Khuje, Smita, Shweta Thool, Sachin Nimje*isall about the gesture control and human machine interaction. The abstract gives us some idea about the system and how we can use gestures to control computer system. The project mainly aims to mouse cursor movements and click events based on skins detection technique. It is cost effective real time working system.[1]

The paper defines the importance of computer and how it has increased to a great extent these days. [1]

It allows us to interact with the system with natural hand gestures. While operating with the system it used colour caps and if the colour of the finger and background matches the system detects an error in it.[1]

In this they only used camera to capture the hand images for a natural human computer interaction.[1]

The proposed system developed by this paper is an real time video processing that is based on real time application system. The approach is to replace the traditional input devices i.e. mouse which can be replaced with hand gesture where the user can interact with the computer more naturally. The basic architecture flow for the system is:

1. Capturing Camera View
2. Getting current frame out of it
3. Creating memory image
4. Finding pixel RGB
5. Comparing pixel color
6. Decision making

In *Eye-Control Mouse Cursor for Physically Disable Individual* authored by *Mohamed Nasor, Mujeeb Rahman K K, Maryam Mohamed Zubair, Haya Ansari, Furida Mohamed,* The paper presents an algorithm to control the movement of the cursor of a computer screen using the movement of iris. By detecting the location of the iris on the screen to a specific location, the software will allow physically disabled people to control the movement of the cursor to the left, right, up and down.The algorithm allows the person to open and close files and applications by clicking it.[2]

The methodology or the basic working of the project is by using the face recognition algorithm to detect the face on the frame that is taken by a standard webcam. After which it detects the eye from the frame. For start it only tracks one eye for fast processing time, after which the left and right corners of the eye are used as reference points to find the shift in focus.

Up and down motions were highly inconsistent. A prolonged blink was used as a signal for clicking the pattern. [2]

In *Vision-based Multimodal Human-Computer Interaction using Hand and Head Gestures* authored by *Anupam Agrawal, Rohit Raj and Shubha Porwal,* the project was implemented for controlling applications such as video and image viewers. Hand and Faces were detected real-time using Haar classifiers based on Viola Jones algorithm. The training was done on static background and cannot be used with other backgrounds. The method of Optical Flow is used to analyze the movement of head and a finite state automata is used to classify the movements into gestures. Hand gestures are classified using ANN by feeding it the biggest contour found by background image subtraction is used to detect the gestures camshift algorithm is used to track the movements of hand. Users were given option to choose an action to corresponding gesture.[3]

In *Accurate Hand Keypoint Localization on Mobile Devices* by *Filippos Gouidis, Paschalis Panteleris, IasonOikonomidis, AntonisArgyros,* in this paper they have used a 2D key-point position was used for standard RGB input for hand. Actually, with the traditional RGB image representing a hand, its main aim is to find a series of predefined keypoints / marks, such as the centroid of the handwrist and the finger joints.[4]

They have applied CNN to detect 2D pose of a hand. Then the pixel wise classification of each joint is done. This classification is used to learn the feature maps better. There have been a significant effort used to develop methods for hand estimation.[4]

The current hand pose estimation approaches can be classified into three categories, called generative, discriminative and hybrid. Common ways that they use to learn the mapping include the use of Random Forests and CNNs. The optimization algorithm is then used to find for the hand position that best matches the visual input.[4]

They have used a neural network that is pre-trained on the Image dataset.[4]

The input has standard color images while the output is K+1 heatmaps, one for each hand mark and the other for the background.[4]

For training purposes, they have used only their right hand. During the test they made use of mirror images of left hand before transferring it to the network. The training images are cut around the appropriate hand center. And the images are moved from $-30^{\circ}$to 30°, mapped up to 30 pixels and scaled by a factor that ranged from 0.8 to 1.5.[4]

In *ENHANCED CURSOR CONTROL USING EYE MOUSE by MARGI SADASHIV NARAYAN, WARANG PRAVIN RAGHOJI,* this paper presents controlling the computer using iris tracking with the help of single web-camera. And also showed that it can be detected easily for gaze estimation. The calibrated points that get generated on the screen are used for measuring the accuracy and precision. The process followed is 1)detect the face position, 2) track the eye, and 3) map the gaze to the screen. If the detection of any step fails then start tracking from first. For face detection they have used viola-jones algorithm to detect face in a live video. For gaze detection, the exact location of eye region needs to be tracked for necessary features. This can be done by Haar-like object detectors to give more accurate pixels. Once this is done the cursor can be moved from current point to where the user gazes it. And clicking event can also be done through it. [5]

In Infotainment Devies Control by Eye Gaze and Gesture Recognition Fusion by Tabassam Nawaz, Muhammad Saleem Mian, the main idea of this paper is to control the consumer devices using eye gaze and gestures. This recognition consists tracking of human movement and interpreting it into meaningful commands. Interaction of gaze with computer is a natural process and its direction can be determined by the position of human head. Face detection is the initial stage to extract the frames. It has much variability in size, colour and shape. These features act as a classifier and are cascaded to give efficient result. After this detection of eye is necessary within the frame for gaze estimation. Once the iris is tracked the mapping of the coordinates of eyes with the cursor is matched and then the cursor is moved in the direction of gaze. [6]

They have used camera to detect the eye and gesture. This system can be used in home and office. Patients can also use this as they are not able to use the devices in regular way. For example different option to play video, watch movies. [6]

In Hand Gesture Recognition Using Deep Learning by Soeb Hussain, Rupal Saxena, Xie Han, Jameel Ahmed Khan, Prof. Hyunchul Shin proposed a new way to interact with the computer system. They have detected hand shape using transfer learning. Transfer learning is a machine learning technique that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. The classifier is trained using transfer learning over a pretrained model which is trained using CNN. The CNN model uses VGG16 architecture for the pretrained model. It consists of total 13 convolution which are full connected by 3 layers. The model consist of 6 static and rest are dynamic gesture. Static gesture are predefined gesture which are set by using pretrained model. The pretrained model uses 70 percent of dataset for training and rest for testing. For dynamic hand gesture, tracing of detected hand is used. In this they have used the transfer learning to take the new input of the gesture and keep on learning every time the model is trained again. The protype was tested with different person without training the model and the accuracy was reported to be 93.09 percent. The model showed accurate results in different background and in different light conditions.[7]

Nose, Eyes, and Ears: Head Pose Estimation By Locating facial Keypoints by Aryaman Gupta, Kalpit Thakkar, Vineet, Gandhi

and P J Narayanan proposed the idea behind this paper is to estimate the head pose by 3D orientation of head with camera using digital images. They have used computer vision for computing the human pose directly. They have used RGB inputs to detect the head. For estimation of head they employed a Multi Layer Perceptron using the five keypoints. They assumed that the keypoints estimated are accurate in face image. As the face is defected in 3D, then the heatmap images are converted to 2D to reduce the gap between real and synthetic data. Through this they have directly trained the convolutional regression.So by using CNN they have used to train and test the model to run on single Nvidia.[8]

The Iris Feature Point Averaging Method in Student Eye Gaze Tracking by Fangfang Yang, Yaping Dai, Lei Wang, Zhiyang Jai. The paper states that the location of iris plays an important role gaze tracking, biometric identification. It also takes the images through the camera and then detects the face and eye and gets to the center of iris. The face recognition algorithm requires high pixels with Haar features. The Haar features are divided into various categories to change size, location and type. The classifier is used to train the weak series to strong ones. Once the iris is located then through hough transform the circle of image is recognized. The extraction of eye region of human to capture the region and then detect the edge by using opencv. The central coordinates of iris are detected by Hough of left and right eye to center coordinate. Although the noise is been reduced but there are some incomplete images that leads to failure in locating the center of iris. So, Average Feature Point was used to solve the problem. So by AFP the center of iris was easily detected.[9]

III. METHODOLOGY

A. *Existing Model*

The existing system made for human machine interaction along for hand gesture where mainly based on RGB 3D model base pose recognition of image sequences of a monocular silhouette. The method's idea is to search for the hand position that best suits a profile in a picture of potential candidates generated from the 3-D hand model. Gaze tracking a video-based eye gaze tracking system basically consists of one or more digital cameras, near infra-red (NIR) LEDs and a monitor with a screen displaying a user interface where the user gaze is recorded. Existing system for gaze tracking a mainly hara cascade which use the color of eye balls to track the mouse pointer where ever the user is looking. It is not as accurate as required for tracking a mouse but it can track the eyes which enables it to mouse the pointer. Hand pose detection using convolution neural network is another existing algorithm used for tracking the hand for gesture control. It uses a predefined dataset which includes the color image of hand along with different gestures image. The RGB images make the processing take more time than normally required. Another system for hand detection is hand pose detecting using skin color segmentation. The process for which is the model is given the dataset which tracks the skins of the hand to convert the image into a outline of the hand. The new image has an white outline of the hand which is filled with the same white color. The rest of the part of the image is consider as the background which is colored in black so that it can be ignored by the model.

B. *Proposed System*

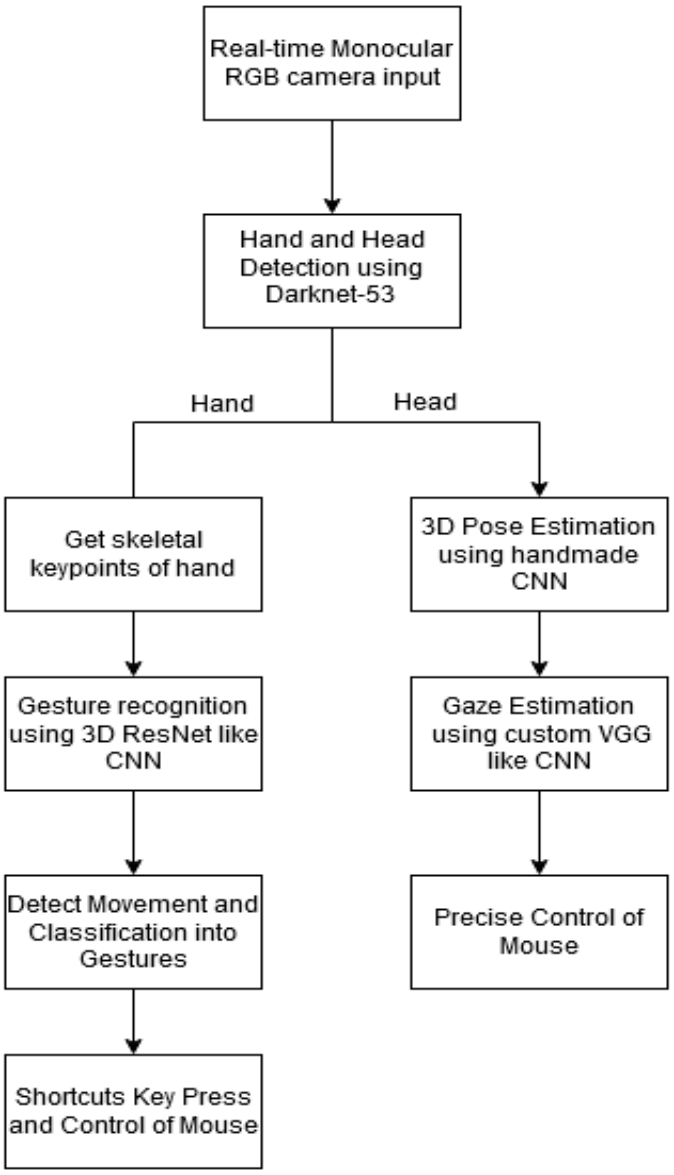

Fig 1 Proposed system flowchart

Hand and Gaze are well known human machine interaction techniques. There are many available system for individual system, but our system purposes to use both the technology together in a single system. We aim to combine to different algorithm into one, which can be used for controlling the computer which basic day to day gestures shown in fig 1.

In the first module we have completed the initial step of designing an Iris tracking system.

This refers to the new head-free gaze tracking systems. This encourages a greater user of eye tracking technology in consumer electronics. The gaze tracking contained in the

proposed method is done using a single web-cam without any infrared camera and sensors. The findings are positive with best results for iris detection and gaze prediction. The result of this approach is that a simple web camera can be used to build an efficient and accurate gaze tracking system. Future work will include carrying out iris monitoring on mobile phones. The device shall be optimized without the use of reference points to improve the precision. After mounting the device on a moving vehicle, friction allows the picture shot to become distorted. This in effect reduces the pace of follow-up. We need to tackle these issues. After which the mouse cursor movements are to be embedded with this tracking algorithm along with gesture.

The Second module which consists of hand tracking will be implemented using image processing along with gesture implementation. The gesture will be trained to be efficient which will require a large number of data sets. Gestures like right click, left click, double click, scroll will be implemented in another module.

The Openvino tool enables us to combine these different algorithms into a single system. The system will detect two main things, first one being Gaze detection which is the detection of the head movements for efficient mouse control. Gaze detection is accurate for mouse movement and is easy to use for many users. The process for which is, the user will start with the system. The system will detecting the face of the user and it will follow the movements of the face to move the mouse, for example if the user moves the face left the mouse pointer or the mouse cursor will move towards left. This can be achieved using openvino tool kit which provides us with custom CNN. The second part of out system is hand gesture detection which includes to detect the hand and its movement, for example, we use fast forward in different video streaming application for which we use right and left arrow, using hand movement to replace the use of keyboard can be done by this system. The system will consist of a pre-trained data model. The model will be trained using black and white or grey scale images of the hand to increase the accuracy as the black and white image provides us with better accuracy and efficient training dataset because of less noise in the background. Using RGB images will make the system less efficient. The parameters for training will consist of the length of the hand, width of the hand, width of the palm, and figure tips. The model also can be trained by openvino tool and the using grey level images will eliminate the problem of the skin tones of varying from user to user. So the working of the system will be the user uses the hand as the input which will be captured by the camera. The system will detect the gesture and analyze it and will give the skeletal keypoints of the hand. Once detected the gesture will perform the required task.

C. Head GestureWe have used a gaze estimation model which requires left and right eye images along with head position angles. To get the eye images we have used a custom CNN for facial landmark detection which is used for getting coordinates of eyes which takes face image as input. For head pose we have another trained CNN which also takes face image as input. The face image is given as input to above networks using another CNN based on MobileNet with depth-wise convolutions to reduce the amount of computation in convolution.

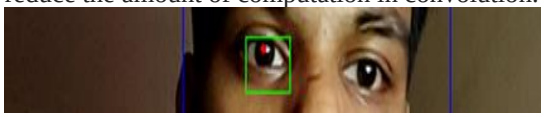

Fig. a

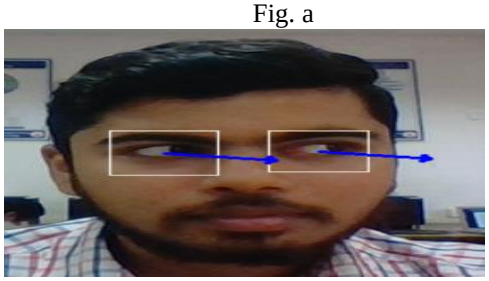

Fig. b

The above figure is the output of two of the models we developed. The fig a is the old model which is based on haarcascades to detect the head. Once detected it captures the coordinates of the iris and connects with the cursor of the mouse. It uses PythonautoGUI library for mouse control.

Fig b is a custom model created in openvino tool kit using its model optimizer. The application performs inference on auxiliary models to obtain head pose angles and images of eyes regions serving as an input for gaze estimation model. The model generates output using inference results with auxiliary model shown in fig 2.

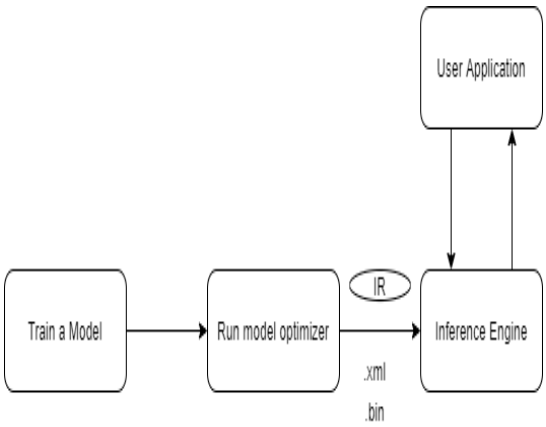

Fig 2 Iris Detection

IV. Conclusion

We propose this Gaze and Hand tracking using the openvino tool kit. The method is robust and the gaze tracking is more accurate than our first model. The Gaze can track the eye movement along side with face and head movement which enables us to accurately track the eye movements. The next part of paper which is the hand detection can detect the hand points.

Thus detection of eyes and gaze estimation has been implemented, the output of which will be the input for the

alagorithm controlling the mouse pointer. Using Openvino model optimizer we are able to achieve real-time, accurate gaze estimation on CPU without high resource consumption which is important if the system is to be used widely. This input method along with hand gestures will be used to improve interactions with computer. The hand gestures will be used to provide intuitive controls to users. Once the system is complete, it would be revolutionary in HCI as we move towards a contactless input method.

REFERNECE

[1] Kalyani Pendke, Prasanna Khuje, Smita Narnaware, Shweta Thool, "Computer Cursor Control Mechanism By Using Hand Gesture Recognition"

[2] Mohamed Nasor, Mujeeb Rahman K K, Maryam Mohamed Zubair, Haya Ansari, Furida Mohamed "In Eye-Control Mouse Cursor for Physically Disable Individual"

[3] Anupam Agrawal, Rohit Raj, Shubha Porwal, "Vision-based Multimodal Human-Computer Interaction using Hand and Head Gestures".

[4] Filippos Gouidis, Paschalis Panteleris, Iason Oikonomidis, Antonis Argyros " Accurate Hand keypoint Localization on Mobile Devices", International Conference on Machine Vision Applications (MVA) National Olympics Memorial Youth Center, Tokyo, Japan, May 27-31, 2019.

[5] MARGI SADASHIV NARAYAN, WARANG PRAVIN RAGHOJI "ENHANCED CURSOR CONTROL USING EYE MOUSE" Volume-3, Issue-12, Dec.-2016

[6] Tabassam Nawaz, Muhammad Saleem Mian, "Infotainment Devies Control by Eye Gaze and Gesture Recognition Fusion.

[7] Soeb Hussain, Rupal Saxena, Xie Han, "Hand gesture recognition using deep learning", 2017 Interational SoC Design Conference (ISOCC) .pp. 48-49

[8] Aryaman Gupta, Kalpit Thakkar, Vineet, Gandhi and P J Narayanan "Nose, Eyes, and Ears: Head Pose Estimation By Locating facial Keypoints", 2018.

[9] Fangfang Yang, Yaping Dai, Lei Wang, "The Iris Feature Point Averaging Method in Student Eye Gaze Tracking", 37th Chinese Control Conference (CCC), 2018, pp.5520-5524.

[10] Alper Aksac,Tansel Ozyer, "Real – Time multi-objective hand posture/gesture recognition by using distance classifiers and finite state machine for virtual operations", 7th International Conference on Electrical and Electronics Engineering(ELECO), 2016, pp: 457-461

[11] Jing-Hao Sun, Ting-Ting Ji, Shu-Bin Zhang, Jia-Kai Yang, "Research on the Hand Gesture Recognition Based on Deep Learning" 12th Internal Symposium on Antennas, Propagation and EM Theory (ISAPE), 2018,

[12] Aleksei Bukhalov, Viktoriia Chafonova, "An eye tracking algorithm based on hough transform", 2018 International Symposium on Consumer Technologies (ISCT), 2018., pp. 49-50.