Gesture Controlled Games Using OpenCV

Mohit Awachar - 191IT231
Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
mohit.191it231@nitk.edu.in

Anshul Patel - 191IT208
Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
anshulpatel.191it208@nitk.edu.in

Rakshit Kulkarni - 191IT245
Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
rakshitkulkarni.191it245@nitk.edu.in

***Abstract*—Hand gesture detection systems for virtual reality applications give users with a better engagement experience, because it blends the virtual and real world objects. Changes in Human-Computer Interaction are influenced by the growth of virtual worlds based on computer systems and the development of user interfaces. The technology allows for physically accurate interaction with the virtual world. The detection, tracking, and recognition modules comprise the hand gesture recognition system-based interface suggested and developed in this research. For this project, opencv module has been utilised to capture and process images. Also mediapipe module has been used for hand tracking and hand detection. To demonstrate the accuracy, utility and convenience of use of the proposed and realised hand gesture recognition system, comprehensive user acceptance has been evaluated. The suggested hand gesture detection system extends standard input devices capabilities for interacting with virtual worlds. The gesture-based interaction interface suggested here has a wide range of applications, including Virtual Reality, Sign Language, and Games. Despite the fact that the current article focused on gaming and educative purpose as an application area.**

Keywords— OpenCV, Gesture Recognition, PyDirectInput, Mediapipe, Flask.

I. Introduction

Human-Computer Interaction (HCI) evolution and research efforts have been immensely effective, and have fundamentally revolutionised computing. Even the phenomenal expansion of the World Wide Web is a direct outcome of HCI research: incorporating hypertext technology into browsers allows one to navigate a link globally with the click of a mouse. More than anything else, interface improvements have fueled this rapid expansion. Direct manipulation of graphical elements, as well as physical interactive devices such as a mouse and keyboard, were among the basic interactions. It appears that HCI research conducted through computer science is merely disseminated into products.Eventually, researchers working in the field of Human Computer Interaction put a common emphasis on designing and developing user interfaces capable of meeting the specified performance standards in a virtual environment. Such interfaces may be created by utilising the phenomena of intuitive communication and manipulation abilities of people, hence enhancing human-computer interaction at a higher level. As both hardware and basic software have become commodities, user interfaces are likely to contribute value to competitive advantages in the future.

As a crucial component, it uses multimedia technology to enable virtual and augmented reality. Advances in computing and semiconductor technology have made intelligent devices more inexpensive, allowing for the development of a wide range of applications. Hence Human-computer interaction (HCI) is always striving to be less expensive, easier to use, adaptable, and quick so that it might lead to a virtual environment where interaction seems natural. Currently, contact-based devices such as accelerators, data gloves, sensors/actuators, and other input devices are used to capture user movement and control item selection, manipulation, and movement in virtual settings. These devices are too expensive for the average consumer to use on a regular basis. As a result, highly qualified specialists such as surgeons and ace pilots mostly utilise them to train and conduct procedures in a virtual environment. More new HCI approaches have been focused to solve the obstacles posed by these technologies. Gesture input may be classified into several categories based on a variety of factors. The main focus of the project is on Hill Climbing Racing Game for entertainment purpose and for educative purpose a Quiz game has been included. Quiz game consists of three sections ranging from easy to hard, having five questions in each section.

The major existing works in this domain are [1] hand tracking Hand gesture based input mode. Here they have explored the ability of interacting with the application from a distance without using the traditional input devices like keyboard or mouse. Using a similar approach in [4] the authors used multi model technology to perform chess game without using markers.

A. Contribution

Along with the standard techniques used in related work,we have added the following aspects to this approach:

TABLE I: Summary of Literature Survey

Authors	Methodology	Merits	Limitations
J. Eisenstein, R. Davis [2]	multi-modal technology is used	Alphabet Recognition	Stereo Camera is used more expensive and Only one application is implemented.
Siddharth S. Rautaray and Anupam Agrawal [1]	Image processing algorithms as Camshift, Lucas Kanade, Haar to perform hand tracking	Hand gesture based input mode and ability of interacting with the application from a distance without using the traditional input devices like keyboard or mouse.	Only one Game is implemented.
Sebastien Carbini, Jean Emmanuel Viallet, Lionel Delphin-Poulat [4]	multi-modal technology is used	chess game without using markers	Only one application is implemented i,e chess game.

- Generally for gesture based recognition costlier devices like Kinetic etc are required.In this project for gesture based recognition no costlier external devices are required, instead of that open source libraries like opencv and mediapipe has been used which just requires a system along with a camera. Hence the project is resource friendly.
- To provide a good and attractive user interface to the user, a user friendly web application has been built using web languages like HTML,CSS,JS etc, which follows maximum number of design guidelines. Also on a single platform, two applications, one for entertainment purpose and other for educative purpose has been clubbed and presented to the user.
- To understand the user intractability, effectiveness and smoothness of the application a proper statistical analysis has been done, taking help of tests like Anova Test and number of questions based on HCI guidelines.

The report is structured as followed. The Introduction section is followed by Problem Statement and Objectives, Literature Survey. Next section is Methodology where the work done is discussed. This is followed by Experimental Results and Analysis where Results and Statistical Analysis are subsections. This is followed by Conclusion, Individual Contribution and Timeline of the Project. The report ends with the References section.

II. Problem Statement

Hand based gesture recognition for interaction with virtual games using OpenCV and Mediapipe.

A. Objectives

- To make an effective gesture based game.
- Providing good user convenience.
- Providing interactive and effective user interface.
- To analyze time required for user in gesture based mode, manual mode and hybrid mode.

III. Literature Survey

Following are the few works and findings regarding Gesture Based Gaming.

Various image processing algorigthms like Camshift, Lucas Kanade, Haar like features etc has been employed. The gesture-based interaction interface proposed here can essentially be applied to many applications such as virtual reality, sign language and games. Here the webcam is used and front end made up of using OpenGL library [1].

Devices is hampered by numerous factors such as awkwardness, unintuitiveness, rigidity, and susceptibility to distortion from the physical environment. These devices have a cost prohibitive for their widespread use by the general user. Therefore, they are mainly used by highly qualified professionals such as surgeons and aces to train and perform their surgeries in the virtual environment. To meet the challenges arising from these technologies, more novel contemporaries of HCI have been focused. Gesture inputs can be classified into different categories depending on different characteristics [2].

One of the categories is the detect gesture, which refers to reaching out or pointing at an object. Accepting or denying an action for an event is called mimetic gestures. It is useful for the linguistic representation of gestures. An iconic gesture is a way of defining an object or its characteristics. Various researchers have proposed and implemented various pragmatic techniques for gestures as input to human-computer interfaces. Liu and Lovell [3].

Implemented a chess game without using markers. This system uses a multi-modal system, both speech and images. Regardless of the application, the system behaves like a multi-modal contact less mouse. In this system, one hand is used to point while the other hand is used to control. Using multi-modal information, the study played a chess game with

the system [4].

Other devices or multi-modal technology. Previous work using stereo cameras also made use of the YCbCr capability. In a complex background, this system found the hand and face. The thing that moves is recognized as a hand. Based on this result, they performed alphabet recognition [5].

The summary of the literature survey is presented in the Table I. There have been attempts to use image processing for hand gesture recognition. In some papers multi-modal technology is also used. All these attempts have been used to develop a gesture control based game.

IV. Methodology

OpenCV and Mediapipe are the two main python libraries which have been used for this project.

A. OpenCV :

OpenCV supports many languages like C++, java, python etc. Also it is available on different platforms like Windows, Linux, Android and iOS etc. OpenCV is a huge open source library for computer vision, machine learning, and image processing, and currently plays a major role in real-time operations, which are very important in today's systems. It can be used to process images and videos to identify objects, faces, and even human handwriting. When integrated with various libraries such as NumPy, Python can handle OpenCV array structures for analysis. Vector spaces are used to identify image patterns and their various features, and perform mathematical operations on these features.

B. MediaPipe :

MediaPipe provides cross-platform, open-source, customizable ML solutions for live and streaming media. MediaPipe Hands uses an ML pipeline consisting of several models working together: A palm detection model that works on the entire image and returns an aligned hand bounding box. A hand orientation model that operates on the clipped image area defined by the palm detector and returns high-fidelity 3D hand keypoints.

Two applications have been developed as mentioned in introduction part, one is for entertainment purpose that includes Hill Climbing Race Game and another one for educative purpose that includes Quiz Game. The user can choose any of these three modes.

- Easy.
- Medium.
- Hard.

After choosing the mode user can start quiz and he/she can see the results once it completed. The Basic requirement for the experiment is built-in camera of laptop or computer or one can also use external web camera to detect the gesture. As shown in Fig.1 once the application starts, the camera will get open in the background of the interface. Here opencv has been used to open the camera. Then image is being captured for each frame per second.

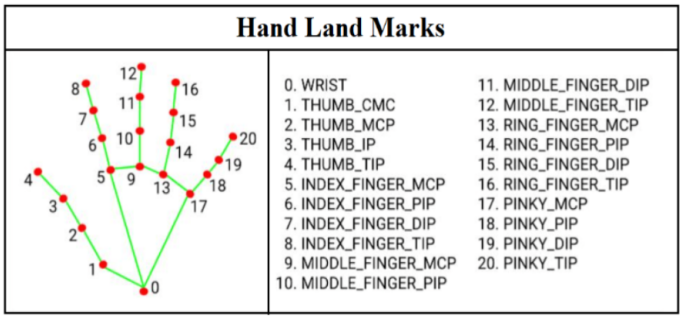

Fig. 1: Hand Land Marks

The captured image is being converted into *RGB* format as various mediapipe's library requires *RGB* format of image for its further processing. The processed image is being then used for hand detection using mediapipe's library. Then further all the 21 land mark positions have been marked on the hand. In Fig.1 HandLand Marks one can see the land marks of 21 key points where starting id is 0 and last id is 20. For each id a particular name is given. Along with the information of id,position of the landmarks are being stored.

The above information is being used to find the proper action for the gesture of the hand. Here for required event a predefined gesture will be mapped with a predefined action according to the need so that when a user performs any action with the hand then the gesture is being recognized and appropriate action will be done. In this way modelling of the gesture will be get done. Then at last required action will get performed.The entire process is being shown as a Virtual game architecture in Fig 2.

The Work Flow : This paper included two games in this paper one is Hill Climbing Race Game and another one is Quiz Game. It includes Hill Climbing Race Game and Quiz game using HTML,CSS and JavaScript. To make it user friendly it has a home page which includes a navigation bar, image slider, footer and the general information of the work. The navigation bar includes buttons for Hill Climbing Game, another for Quiz Game and one more for home page. On clicking the following buttons it will get redirected to the respective applications as per the name of the button suggests.

1) Hill Climbing Race: After clicking on its button in home page, it will get redirected to this gaming user interface. After being get opened, camera will get on for image capturing so that gesture recognition can be done. Here user is supposed to close the palm to accelerate the car and open the palm to reduce its speed i.e. to apply breaks.

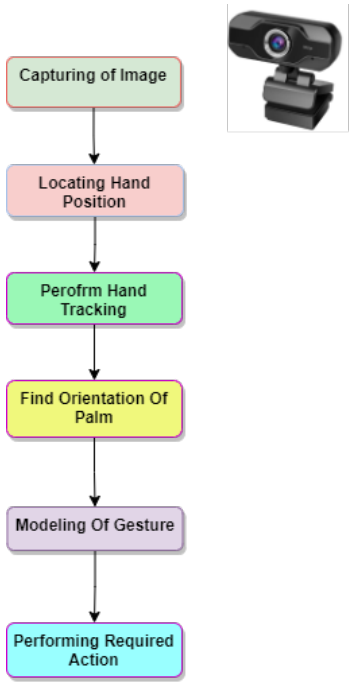

Fig. 2: Virtual game architecture

Basically hand gesture is being recorded as per the id's of the landmark given to a hand. If ids 20,16,12 and 8 drops down below to 18,14,10 and 6 respectively and also if id 4 is right to the id 8,7,6 or 5 then this action will be considered as a closed palm i.e. here used for acceleration instead of right key of the keyboard which will be clicked by the python package called *PyDirectInput*. Similarly if vice versa action is done then it will be considered as open palm used for slowing down the car or as brakes for it instead of using left key from the keyboard.

2) Quiz Game: After clicking on its button in home page,it will get redirected to a *start quiz* page it consist of a button,after clicking on this it will get redirected to the instruction page. In this page instruction about the game will be given to the user. Also level of the game ranging from easy to hard will be given to the user. As per the difficulty, different scoring system is consider,for easy ,medium and hard level respectively one , three and five points will be awarded to the user.

After selecting difficulty level user will be get redirected to the game. In the game user will be provided with five questions and for each question user will be given a total of twenty seconds of time to read the question. Then user is supposed to select the option by showing a gesture related to option number i.e. if user wants to select option one then user is suppose to show one finger open and other being closed in front of camera. Similarly it can be done for other options also as shown in the Fig.3. After showing the gesture,immediately, chosen and the right option will shown in front of user. Then user is supposed to show the closed palm in front of camera to get redirected to next question.At the end of all the questions,user will be shown its score.After that user will be given two choices,one is to replay the quiz and other to quit the quiz to return to the home page.

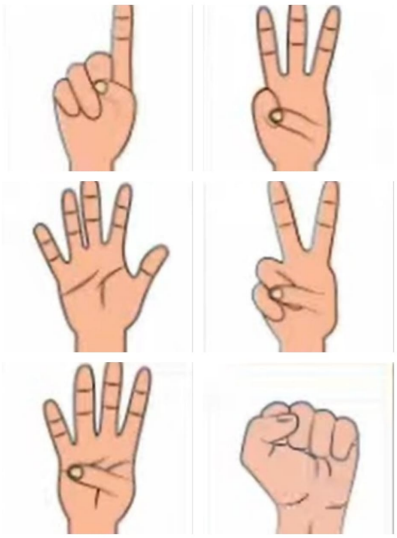

Fig. 3: Actions

V. HUMAN COMPUTER INTERACTION PRINCIPLES USED

The work of the following project has been done under consideration of various Human Computer Interaction guidelines. The interface follows Ben Shneiderman's Eight Golden Rules for User Interface Design. Each of these aspects are explained below with proper justification for the same.

1) **Strive for Consistency** :Users should not have to guess whether different words, contexts, or actions indicate the same thing, according to the document. Here also interface exhibits consistent quality across screens/applications both visually as well as behaviorally.
2) **Cater to Wide Range Type of Users** : This guideline states that the interface should accommodate to all types of users, from novices to professionals, according to this guideline. Here also user has been provided an easy to use, user interface along with gesture based controlability.
3) **Offer Informative Feedback** : This guideline states that interface should be communicative and should provide user feedback if they are proceeding in the right direction. As per survey results in Figure 17 this guideline is followed.
4) **Design Dialogue to Yield Closure** : This rule indicates that when pop-ups and dialogue boxes are required, the interface should give them. The following websites provides same to the user, whenever needed.

5) **Prevent Errors** : The guideline states that interface should minimize human errors. Being a simple and well instructed website, it shrinks down the mannualy done error.
6) **Permit Easy Reversal of Actions** : The guideline states that interface should be designed in such a way to allow easy retrieval of actions. Navigation amongst various pages in the application should be smooth and easy. Here a proper navigation bar has been supported so that easy redirection can be done.
7) **Support Internal Locus of Control** : This guideline states that interface should be interactive enough to make user feel he is in control. The website provides easy and simple to use web interface in front of user so that user can never loose its control.
8) **Reduce Short Term Memory Load** : This guideline states that the interface should not expect user to remember a sequences to reach a particular page. This guideline is followed in the interface by providing proper navigation and well instruction whenever required.

VI. Experimental Results and Analysis

A. *Results :*

In brief this paper includes various user interfaces of the project work. It has home page for this project as shown in the Fig.4 and consisting of general information,a page for instruction of the quiz game as shown in the Fig.5, user interface for quiz game as shown in the Fig.6 also the result of the quiz game shown in Fig.7 and user interface for hill climbing race game as shown in the Fig.8

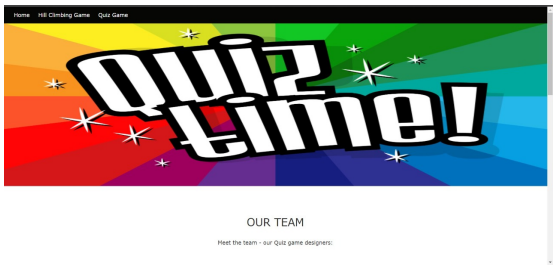

Fig. 4: Home Page

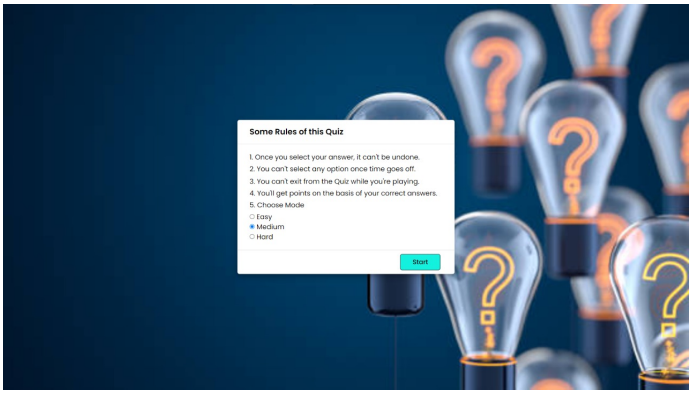

Fig. 5: Quiz Instruction Page

Fig. 6: Quiz Questions Page

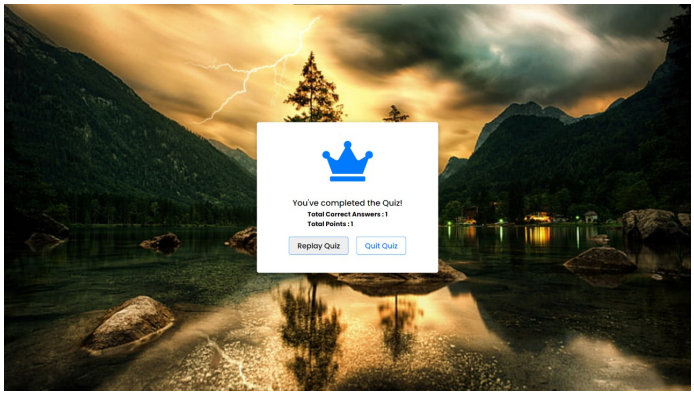

Fig. 7: Quiz Result Page

Fig. 8: Hill Climbing Race Game

B. *Statistical Analysis :*

1) Anova Test: Anova Test also known as Analysis of Variance is a statistical test to determine if the two or more groups are similar or not. The test has a null hypothesis which considers that means of all the groups is the same. The alternative hypothesis states that mean of all the groups is not the same. The null hypothesis is accepted if the p-value is less than 0.05 otherwise the alternative hypothesis is accepted.

The anova test is used here to analyze different times which user takes to complete the quiz game. The game is played in three modes Gesture based Mode, Manual Mode(using mouse and keyboard) and Hybrid mode(Camera, mouse and keyboard). In the Gesture based mode, the user makes use of gestures to answer the questions. In the Manual mode, the user makes use of keyboard and mouse to answer the questions. In hybrid mode the user uses both to answer the questions. The data collected is the time taken to complete the quiz from 15 different users for all the above modes. The users are categorized into three categories namely novice who have no or very few experience with computers, intermediate who have some experience with computers and experts who are experts with computers. 5 users from each category to are used to collect data. The null hypothesis is that mean time of all modes should be similar. The time required in seconds to complete the quiz is shown in the below(Figure 9).

		Time(sec)	
User type	Gesture based	Manual	Hybrid
Novice	219	270	241
Novice	220	281	242
Novice	214	263	247
Novice	216	273	237
Novice	219	255	246
Intermediate	206	240	217
Intermediate	213	233	229
Intermediate	210	237	220
Intermediate	205	228	219
Intermediate	213	225	211
Expert	209	200	201
Expert	200	207	205
Expert	207	210	195
Expert	202	205	201
Expert	203	212	203

Fig. 9: Time Required to complete quiz

The analysis of data in shown in the Figure 10. As one can observe the mean time for gesture based mode is the least followed by hybrid mode and the manual mode. The variance is also least in the gesture based mode followed by hybrid mode and manual mode.

Data Summary				
Groups	N	Mean	Std. Dev.	Std. Error
Group 1	15	210.4	6.5115	1.6813
Group 2	15	235.9333	26.9933	6.9696
Group 3	15	220.9333	18.2031	4.7

Fig. 10: Analysis of Data

The results of the anova test are shown in Figure 11. The F-value is 6.72 and p-value is 0.0029 which is less than alpha(0.05). So the null hypothesis is rejected.

ANOVA Summary					
Source	Degrees of Freedom DF	Sum of Squares SS	Mean Square MS	F-Stat	P-Value
Between Groups	2	4939.4991	2469.7495	6.7211	0.0029
Within Groups	42	15433.4702	367.4636		
Total:	44	20372.9693			

Fig. 11: Anova test results

From the Anova test it is concluded that Gesture based mode is improvement in terms of time when compared to other 2 modes. It is also observed that time taken is almost similar for all types of user when using the gesture based mode. In the manual mode, the time is dependent on the user and the experts can finish the quiz quickly and novice users take time. The story is similar in the hybrid mode. Figure 12 shows the average+standard deviation for all three modes.

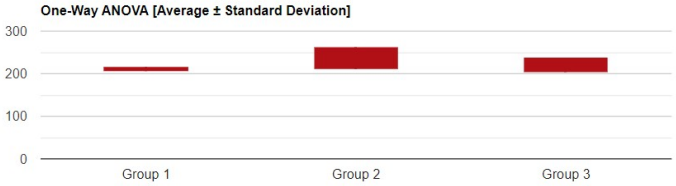

Fig. 12: Average + Standard Deviation

2) Analysis of feedback form: To get fair data regarding the fact whether the work follows design guidelines or not , proper information has been collected, compared and visualized. Each of these aspects are explained below along with the cause for it.

1) **Consistency of interface** : The feedback has been taken against a query,"Rate the consistency of this application across different screens sizes". The poll result is shown in Figure 13.This shows that major of the user has given positive response to it with 57.8 percent rated 5 out of 5 points.

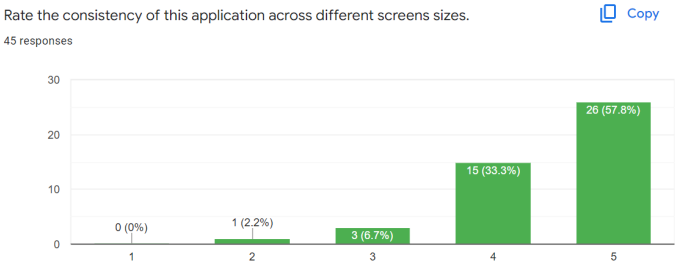

Fig. 13: Consistency of Interface

2) **Cater to Wide Range Type of Users** : Here general information of of the user has bee recorded. Fig 14. shows data as per the "Age group",Fig 15. shows data corresponding to "user background" and Fig 16. records data as per the familiarity of user for the computer.

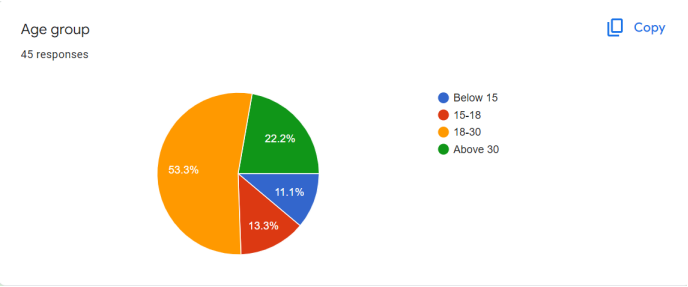

Fig. 14: Age groups Analysis

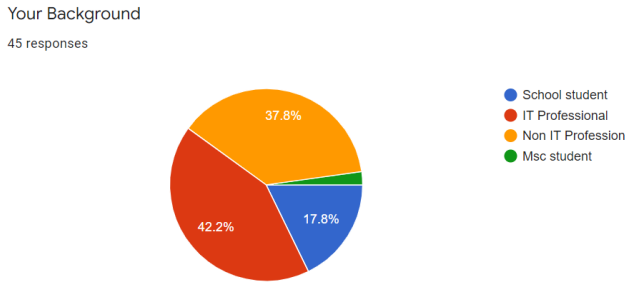

Fig. 15: User Analysis

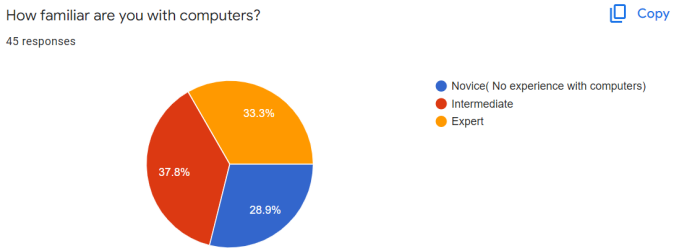

Fig. 16: Analysis 3

3) **Mode of usage** : Fig 17. shows the data corresponding to mode of the usage of the application i.e. whether gesture based has been done or manually used.The percentage for camera and keyboard was 70 and 30 percent respectively.

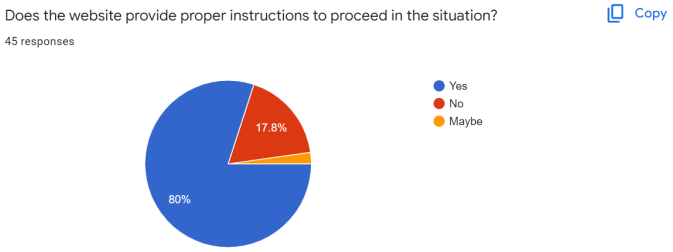

Fig. 17: Mode of usage

4) **Well-instructed** :The data has been recorded against the query,"Does the website provide proper instructions to proceed in the situation?".Fig 18. shows result for the same.

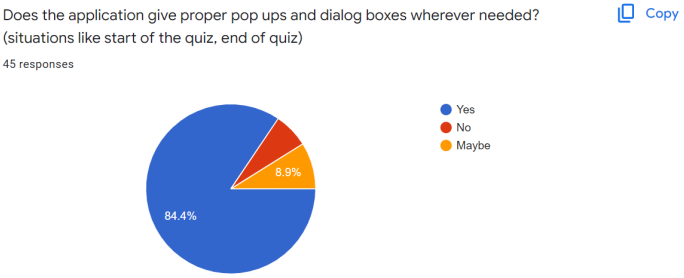

Fig. 18: Well-instructed

5) **Prevent Errors** :The data has been recorded against the query,"Rate how well does the application handle human errors". As per survey, results has been shown in Figure 19 .

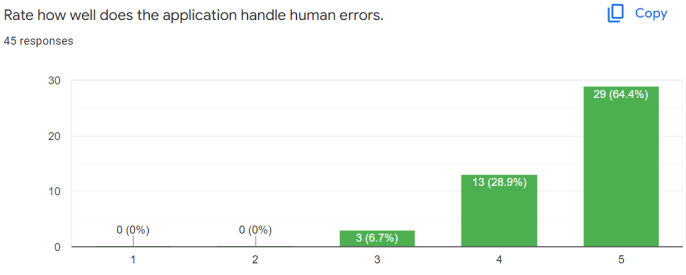

Fig. 19: Handling Human Errors

6) **Navigation amongst pages** : The data has been recorded against the query,"Rate navigation amongst pages inside the application". As per survey, results has been shown in Figure 20.

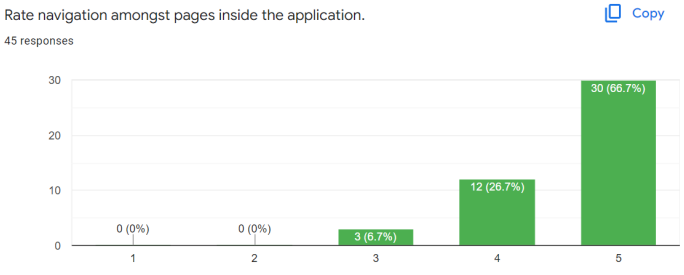

Fig. 20: Analysis of Navigation amongst pages

7) **Interactive Interface** : The data has been recorded against the the interaction with the interface. Survey, results are shown in Fig. 21.

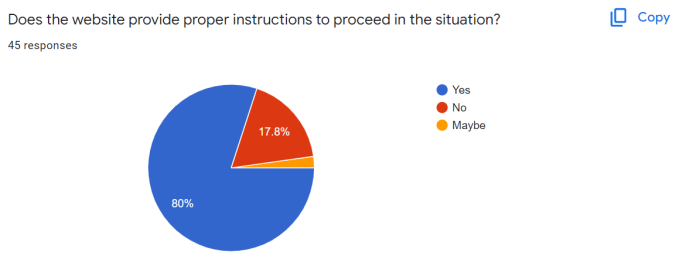

Fig. 21: Interactive Interface

VII. Conclusion

From Anova Test results, it is observed that experts users are not affected much by Gesture Based interface but Novice and Intermediate users show increased productivity in terms of time required when playing the quiz game. This project has applications for users where they can control the user interface by just doing actions with their hands i.e. gesture recognition. The application's user interface in the form of website(not hosted yet). The technologies used for the design are HTML, CSS, Javascript and the back-end mainly consists of Flask and Python. For image capturing and its further processing this mainly used two modules i.e. Mediapipe and Opencv. The core idea of the project was to provide gesture based user interface to the user so that actions done by the hands can be recognized by the system instead of using keyboard or mouse for the same action required. Therefore this concept can be used in two applications, first is hill climbing race game used for fun and second is for a quiz game i.e. for educative purpose. It has a lot of fun working on the project and got to learn the important libraries for image capturing , its detection and tracking. The knowledge acquired by us based on gesture based recognition also got increased during course of this project.

The future work includes hosting this website online and making it available to the interested users. Also planning to add more games to it so that user can get more choices. Further planning to shrink down the manual interaction as its least i.e, support of gesture for main home website also.

Individual Contribution and Timeline

The Time-line for the project is shown in Figure 23. The individual contributions are mentioned in Figure 24.

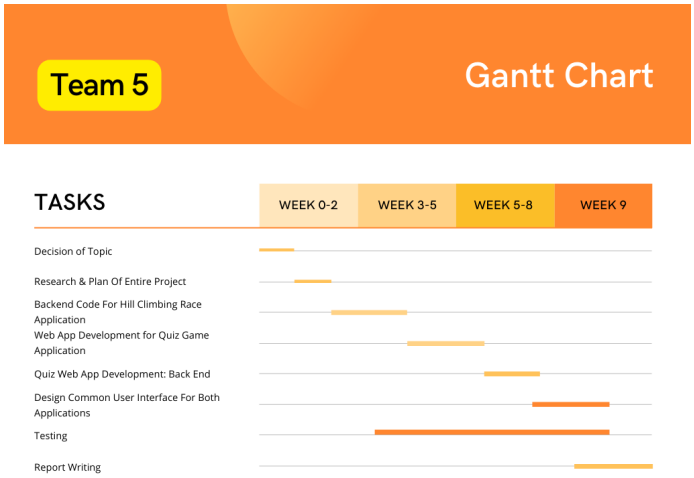

Fig. 22: Time-Line of the Project

Fig. 23: Individual Contribution

Implemented/Base Paper

The paper called *Interaction with Virtual Game through Hand Gesture Recognition* is the base paper for our project. The authors are Siddharth S. Rautaray and Anupam Agrawal, 2011 International Conference on Multimedia, Signal Processing and Communication Technologies.

In this base paper, the authors talk about hand gesture recognition to control the application, this application uses Camshift and Lucas Kanade Optical Flow technique to track the hand. The following step has been implemented to achieve robust hand motion tracking in the video sequence. This is done using the CAMSHIFT tracking algorithm.The noise and lighting variations also hit the pixel dimensions throughout the characteristic area that could be counteracted. The hand is tracked using background subtraction and motion detection.

The noises generated during use are reduced by eroding and dilating.

REFERENCES

[1] Siddharth S. Rautaray and Anupam Agrawal,"Interaction with Virtual Game through Hand Gesture Recognition",2011 International Conference on Multimedia, Signal Processing and Communication Technologies.

[2] J. Eisenstein, R. Davis, Visual and linguistic information in gesture classification, In Proceedings of 6th International conference on Multimodal interfaces. ACM Press, 113-120, 2004

[3] N. Liu, B. Lovell, Mmx-accelerated realtime hand tracking system. In Proceedings of IVCNZ, 26–28, November 2001.

[4] Sebastien Carbini, Jean Emmanuel Viallet, Lionel Delphin-Poulat, "Context dependent interpretation of multimodal speech-pointing gesture interface", Saarbr¨ucken 2007.

[5] Mahmoud Elmezain, Ayoub Al-Hamadi, Gerald Krell, Sherif El-Etriby, Bernd Michaelis, "Gesture Recognition for Alphabets from Hand Motion Trajectory Using Hidden - 1096 - Markov Models", Institute for Electronics, Signal Processing and Communications (IESK).

[6] Andrea Fusiello, Emanuele Trucco and Alessandro Verri, "A compact algorithm for rectification of stereo pairs", Machine Vision and Applications (2000) 12: 16—22.

[7] Matthew A.Turk and Alex P. Pentland, "Face Recognition Using Eigenfaces," Proc. of CVPR, pp.586-591, 1991 [10] Sebastien Carbini, Jean Emmanuel Viallet, Lionel Delphin-Poulat, "Context dependent interpretation of multimodal speech-pointing gesture interface", ICMI2005.

[8] Q. Chen, N.D. Georganas, E.M. Petriu, Realtime Vision-based Hand Gesture Recognition Using Haar-like Features, In Proceedings of IEEE Instrument and Measurement Technology Conference, 1-6, 2007.

[9] Viola and Jones, Rapid object detection using boosted cascade of simple features", In Proceedings of Computer Vision and Pattern Recognition, I-511 - I-518, 2001.