Touchless interface for gaming ⊘

Bh Prashanthi; Srikar Peddi; Sai Kumar Kulai ✉; Suresh Mudavath; Uday Kiran Reddy Bommu;
Zaheer Mohammed

Check for updates

View Online   Export Citation   CrossMark

AIP Publishing

Touchless Interface for Gaming

Bh Prashanthi[1,a)], Srikar Peddi[1,b)], Sai Kumar Kulai[1,c)], Suresh Mudavath[1,d)], UdayKiran Reddy Bommu[1,e)], Zaheer Mohammed[1,f)]

[1] Department of CSE, Gokaraju Rangaraju Institute of Engineering and Technology,India

[c)] *Corresponding Author:* *kulaisaikumar0906@gmail.com*
[a)] *bhupathi.prashanthi@gmail.com*
[b)] *srikarpeddi00@gmail.com*
[d)] *sureshchowhan1999@gmail.com*
[e)] *udayreddybommu@gmail.com*
[f)] *zaheermd616@gmail.com*

Abstract: The main objective of the paper is to develop a game to enhance the gaming experience of the user cost-effectively. Some games use modern technology to give the best output possible like the PS5, VR, etc. These games are costly, and many people cannot purchase this. There are lots of games that are played on different platforms like mobile, PC, tablet where they use either the touch or the keyboard interface. We need an interface that is more interactive to the user so that user can play games more effectively. An idea to use gestures instead of touch or keypad which helps level up the gaming because user should use his gestures to interact with the computer to play which will be more realistic and enjoyable. This may not be as good as VR or any other latest technology, but we can achieve the basic enhancements required. This level of gaming is obtained using soft computing. JavaScript is used in developing the game, Deep learning is used in gesture detection. Here the camera permissions of the PC are given and using this the gestures of the user are detected based on the movements. Once the gestures are detected the result is sent to the JavaScript code present in the game, now based on the information received the decision is taken and the result is shown on the pc i.e., based on the gestures generated the movement of the object is determined. These games do not require heavy equipment just a PC with basic features is required to play. Thus, the gaming experience is enhanced cost-effectively.

INTRODUCTION

Games

People play games often to relax or enjoy their free time. Games play a crucial role in a person's mental development because based on the type of games they play and the influence it has on the person his personality changes mainly in children. So, most action games are restricted to children so that they cannot get influenced by them. Every game has the age limit of a person required to play the game. Games are designed to attract the players by using various sound effects, graphics, and UI which are comfortable to the user. Some games are played online, offline, single-player, multi-player. Online games are more popular than offline games because they are dynamic and most of these games are multi-player. We can play with our friends as allies or enemies which makes it more interesting to play. These games are encouraged by most people because we do not face the same obstacles all the time. The outcome of the game is uncertain.

PC or mobile are the major platforms for these games. Most of the games are played using the touchscreen because it gives us a lot of conveniences while playing games. The controls on the can be customized to our wills like the position and the size of the controls. This gives a lot of flexibility to the users. Users prefer the touchscreen to the keypad because it gives a good gaming experience, as users interact with the game effectively. Since online games are launched the demand for offline gaming is dropped rapidly. Offline games are mostly single player, these games are played alone where it has a disadvantage over online gaming where people play in a group. In offline games, all the obstacles are pre-programmed i.e., they can't make their own decisions so it will not be challenging to the user. So, there is something that should be added to the offline games which makes them interesting, in offline games the gaming experience should be increased to boost the demand. Interaction between user and game plays a crucial role to achieve this type of gaming. At times games are played just for fun but now people are choosing it as

a career by streaming them online. There are so many categories in games, some help to improve the personality and some are addictive. Action games are popular because it is interesting and most people like to fight and kill enemies, it has a huge audience worldwide. Some games are very helpful in improving a person's concentration, observation, and intelligence while some games are addictive and full of violence and have a negative impact on the person.

There are different ways in which a player can interact with the UI like keyboard, touch screen, etc. Some technologies use sensors, high-level gadgets to increase the level of gaming, these games are expensive to purchase. As most of the games played on a PC use keyboard and mobile use the touch screen. To enhance the gaming experience gestures can be used because as the user moves the gestures are recognized and a specific task is performed based on that this is more enjoyable and livelier than the previous methods. This paper presents a method to enhance gaming with gestures using Deep Learning cost-effectively.

Interfaces of Games

Every game has an interface that allows the user to interact with the game. These interfaces generate inputs when certain conditions are met. The following are different types of interfaces:

Keyboard and Mouse

Uses keyboard and mouse to take the inputs to the game like using the arrow keys to specify the direction or click the mouse to perform an action. This is used in computers and laptops.

TouchScreen

When the user touches the screen at a particular location an event is triggered, and the inputs are taken to determine the outcome. This is used in mobiles and tablets; this is best suitable to play online games.

Sensors

This uses advanced technologies like the IoT and sensor technology to generate the inputs based on the movements made by the user to play the game. This method is used in PS5, VR, etc.

Joystick

These are wire-based gadgets that are connected to the devices in which the game is played. It is a device with buttons, when clicked an input is generated and sent to the game. This is used in PS2, computers, etc.

REVIEW OF LITERATURE

Real-Time Sign Language Detection Using Human Pose Estimation, A lightweight real-time gesture recognition model is proposed in this study. The writers identified the videoconferencing for requirement of such case. They retrieved optical flow features related to human pose and showed that these features are relevant with an accuracy of 80% using a linear classifier. They notice increases of up to 91 percent accuracy when we use a recurrent model directly on the input [1].

You Only Look Once, in this paper authors introduced a new approach of detecting objects, Classifiers developed for object detection are repurposed for per-form detection. Preferably, the spatially separated bounding boxes are considered object detection to be regression issue with probabilities. A single neural network is used in detection [2].

EfficientDet: Scalable and Efficient Object Detection, the authors of this paper carefully examine the design choices for neural network architecture for object detection in an image and proposed some major enhancements to boost efficiency [3].

Continuous Gesture Segmentation, that is both successful and efficient continuous gesture recognition provided for deep structure from this study. Using the proposed dilated Res3D temporal network, continuous motion

sequences are first separated into isolated gesture instances. The recognition network is then built using a 3-D convolutional neural network [4].

Hand joints-based gesture recognition, to increase the accuracy of hand gesture identification, this research offers a nested interval unscented Kalman filter with long short-term memory network [5].

Recognition Using Bag-of-Features and Support Vector Machine Techniques, offers an overall recognition accuracy is 96.23% [6].

Sign language recognition with long short-term memory, authors explain that the Handcrafted features and Hidden Markov Models model temporal information in traditional SLR approaches. Reliable handcrafted features, on the other hand, are difficult to design and cannot adapt to the wide range of sign language. authors present an end-to-end strategy for Sign Language Recognition based on LSTM to address this challenge; Long Short-Term Memory (LSTM) can model the contextual information of sequences well [7].

Recognition using MediaPipe and Machine Learning, the main goal of this paper is to show how to use MediaPipe's opensource framework and machine learning algorithm to simplify Sign Language Recognition. The prediction model is simple to use and may be adapted to smart devices [8].

RESEARCH METHODOLOGY

Construction and Implementation Details

Steps Involved
1. Data set (Collection of Images).
2. Use of MP Holistic to highlight key points.
3. Collection of key point values for training and testing.
4. Create labels and features from data that has been preprocessed.
5. Model Generation:Create LSTM neural networks and training
6. Confusion matrix and Accuracy evaluation.
7. Run the test in real-time.
8. Converting the developed model into tensorflowJS model.
9. Developing a Game.
10. Model and Game Integration.

Data set

Some actions are predefined which are to be predicted by the model, they are left, middle, right
We have set the following constraints while data collection
- N= No_of_Sequences: 10
- M= Sequence_Length: 10

For every predefined action, for 'N' number of sequence of length 'M' we have collected images using OpenCV.Exact distance from the screen, in computer vision a big object placed far way is similar to a small object placed close to the camera, so object contour plays a major role in detection. Hence, we may conclude that an image capturing human's shoulder and hands would be perfect to detect sign language. Forexample, one meter would be an ideal distance for a camera with lens whose aperture is F/2.2.

Later, images are converted into NumPy arrays which serve as data set for training and testing.

Use of MP Holistic to Highlight Key Points

Using Media Pipe, the following type of landmarks are extracted and highlighted from collected images
- Pose landmarks
- Left-hand landmarks
- Right-hand landmarks

Landmarks are drawn on image and rendered on to screen as shown below figure
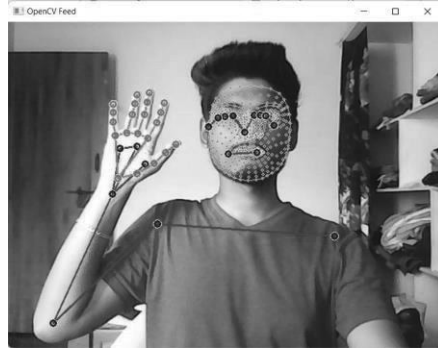
FIGURE 1. Landmarks drawn using MedaiPipe.

Collecting a List of Key Point Values for Training and Testing

The extracted key points/landmarks are stored in a NumPy array which is later used for Training and Testing. Separate directories are created for every gesture and respective NumPy arrays are stored in them.

Create Labels and Features from Data that has been Preprocessed

While preprocessing, the data is labelled using the following label map {left:0, right: 1, middle: 2}.

Model Generation: Create LSTM Neural Networks and Training

From keras models Sequential Model is used, which has linear stack of layers. With the list of layers, the sequential model is generated. We have used 3 LSTM layers and 3 dense layers.

The rectified linear activation function, or ReLU is a piece - wise linear function that will output the input directly if it is positive and will output 0 if it is negative. Because a model that utilizes it is simpler to train and generally produces higher performance, it has become the standard activation function for many neural networks. We chose the tensor board to monitor the model's performance because it clearly displays the results.
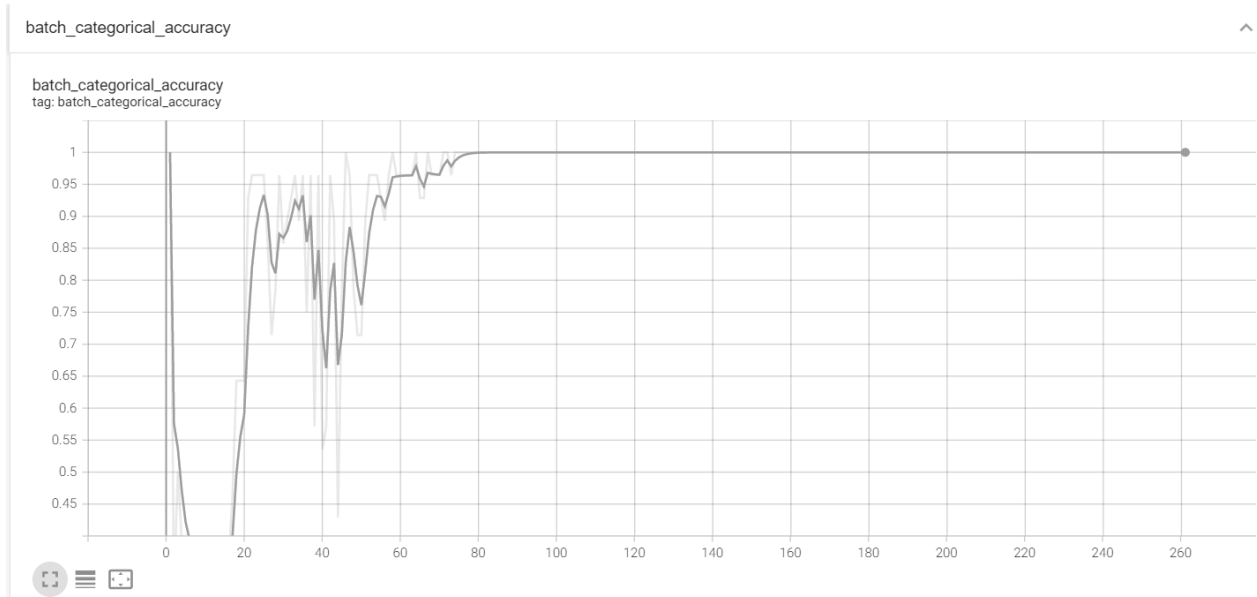

FIGURE 2. Model Performance monitored while training using tensor board

Confusion Matrix and Accuracy Evaluation

An Error matrix shown in figure 'd', is a summarized table used to evaluate a classification prediction error. The number of right and unsuccessful predictions is calculated and broken down by class using count values.
Confusion matrix of model generated is shown below is a python array
array ([[[0.97, 0],
[0.01, 0.976]],
[[0.971, 0.032],[0.02, 0.973]]], dtype=int64)
False positives and False Negatives are almost negligible for the generated model.
Accuracy of the model = 97.1%

Run the Test in Real-Time

Real-time testing is done to verify whether the generated model performs well in real-time. The same of sample shots are shown below figure. The model outcome is highlighted with a color, Here the model outcome is left and hence it is highlighted with orange color.

FIGURE 3. Real Time Testing

In real-time model performs extremely well as expected from the error matrix.

Converting the Developed Model into TensorflowJSModel

We had to convert the model to make it compatible with java script because we had to use it in a browser to control a game. TensorflowJS is a JavaScript library for using machine learning in JavaScript. We were able to convert our Keras model into a TensorflowJS model using TensorflowJS commands and successfully apply machine learning in JavaScript to control the game.

Developing a Game

A Car game is Developed using CSS and JavaScript in which the car is controlled using gestures.
Usually, games are controlled using keyboard and mouse events but since we are introducing the concept of touchless interfaces for gaming, we are controlling games using gestures.

Model and Game Integration

This is the last phase in development where we have integrated TensorflowJS model and game using JavaScript and successfully controlled the game using gestures.

RESULTS

Media pipeline performs best in collecting key points in an image and Recurrent neural network performs well in analyzing sequential data. ReLU activation function is simpler to train and produces higher performance.

LIMITATIONS

- Model performance may be reduced if the lighting of the environment is dull.
- The accuracy of the model is affected if the background environment changes continuously.

CONCLUSION

In this research, we have concluded that media pipeline performs best in collecting key points in an image to analyze pose and hand gestures when compared to the old scale-invariant feature transform in computer vision. Since the image is analyzed by media pipeline, for image processing CNN (Convolutional Neural Network) is not required.

we need to analyze the Sequential data of key points generated from the Media Pipeline using a Recurrent neural network and successfully generated a model. Now by connecting the generated model to the game, touchless interface for gaming is achieved.

FUTURE SCOPE

In future studies, many algorithms can be used to detect perfectly in a dynamic background and voice can be used to control games so that gaming is enhanced more.

REFERENCES

1. Amit, Moryossef, Ioannis, Tsochantaridis, Roee Aharoni, Sarah Ebling and Srini Narayanan, "Real-Time Sign Language Detection Using Human Pose Estimation", in *European Conference on Computer Vision* ECCV 2020: Computer Vision – ECCV 2020 Workshops , (2021),pp .237-248
2. Joseph Redmon, SantoshDivvala, RossGirshick andAliFarhadi, "You can only Look: Once: Unified, RealTime Object Detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),(IEEE,2016)*
3. Mingxing Tan, Ruoming Pang and Quoc V. Le, "EfficientDet: Scalable and Effficient Object Detection", in *Research paper from the Computer Vision Foundation ,*(IEEE,2020),pp.10781-10790
4. G. Zhu, L. Zhang, P. Shen, J. Song, S. A. A. Shah and M. Bennamoun, *IEEE Transactions on Multimedia*, **21**, 1011-1021(2019).
5. Chunyong Ma, Anni Wang, Ge Chen and Chi Xu, Springer nature paper Link journal **34**, 1053-1063(2018).
6. N. H. Dardas and N. D. Georganas, IEEE Transactions on Instrumentation and Measurement **60**, 3592-3607(2011)
7. T. Liu, W. Zhou and H. Li, "Sign language recognition with long short-term memory," *in IEEE International Conference on Image Processing (ICIP),*(IEEE, Phoenix, AZ, USA,2016),pp. 2871-2875,
8. Arpita Haldera andAkshitTayade, International Journal of Research Publication and Reviews **2,** 9-17(2021).