Exploring the Effectiveness of Gesture-Based Control for YouTube: An Evaluation Study

Rayapaneni, Sai Sampath
Department of Computer Science
Missouri State University
sr852s@missouristate.edu

Kavya, Obula
Department of Computer Science
Missouri State University
ko977s@missouristate.edu

Bheemani, Venkata Vamsi Krishna
Department of Computer Science
Missouri State University
vb245s@missouristate.edu

Abstract

A project called GestureTube aims to improve YouTube's user experience by incorporating gesture-based controls. Users are constantly looking for fresh ways to engage with YouTube due to its rising popularity. Users may navigate through videos, change the volume, and carry out other actions on YouTube with the help of gesture-based controls rather than using physical buttons or voice instructions. In this research, user gestures are detected and recognized using computer vision and machine learning techniques [1]. The system uses a webcam to record photos of the user's hand movements, and computer vision techniques are then used to process the images to find and identify the gestures. The actions in YouTube that correspond to the identified movements are play, pause, skip, and rewind. Regardless of physical limitations, anyone can use the GestureTube system because it is simple to operate. Users can operate YouTube without a keyboard or mouse thanks to GestureTube, which does away with the requirement for physical buttons or voice commands. Users are able to set their own motions and actions thanks to the system's customization features. GestureTube is an original method of controlling YouTube

that has the ability to enhance the platform's user interface and make it more user-friendly and open to all users.

Index Terms

Gesture recognition, YouTube control, Human-computer interaction, Hand gesture recognition, Computer vision, Machine learning, User interface, Motion tracking, Natural user interface (NUI), Video player control.

I. Literature Review

Studies investigating the use of gestures for controlling various apps and devices have been conducted in the domain of gesture-based control for a number of years [2]. Whereas another study looked into the usage of mobile robot control via gesture recognition [3]. These studies show how gesture-based control has the potential to enhance user experience across a range of applications.

Gesture-based control has also been looked into in relation to YouTube in a number of experiments. The use of hand gestures to control YouTube video playback on a smart TV. According to the study, users were able to utilize hand gestures to accurately accomplish simple operations like play, pause, and skip. the possibilities for gesture-based control to be used on mobile devices by using a wearable gesture detection device to operate YouTube on a smartphone [4].

Gesture recognition has seen extensive usage of computer vision and machine learning techniques. the application of hand-crafted features and deep learning techniques for gesture detection. These studies offer perceptions on the various strategies for gesture recognition and can influence the GestureTube system's design [5].

In early, the world goes with interaction with complex systems that ensure a response. Gesticulation recognition systems receives attention in past and previous years because of their ability to interact with human and computer. Between the human and systems communication creates a complicated understanding about the signals. Our aim of this paper proposed the easy way to perform fast efficient of media player using gesticulation .In this concept without hardware we

can control the media using the movement of hands (Gesture).In major role of deep learning model Keras detecting the motion of hands. This video breaks down this sometimes complicated concept into easy-to-understand parts [6].

Gaze-based interaction between human and computers has opened a potential domain of effortless supervision. The eye is one of the most important organs to perceive information from our surroundings and it can be the most conspicuous way to interact with computers. Therefore, in this research, a system model has been introduced to control the multimedia player with eye gaze using an economical webcam, as well as provides higher precision and robustness. The goal is to control the play, pause, forward and backward functions of a multimedia player. The click event was triggered by detecting the user's eye blink [7].

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 VLC 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 K Nearest Neighbour algorithm. The theoretical analysis of the approach shows how to do recognition in static background. The Training Images are made by cropping the hand gesture from static background by detecting the hand motion using Lucas Kanade Pyramidical Optical Flow algorithm. 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 [8].

Controlling VLC media player withoutphysical interaction with the computer. Image is captured and verified with oursystem in which image pre-processing and other techniques are used for detectionof gesture. Image is captured and provided as the input to the application via acamera of minimum 1 MP quality for good results. Processing will take place in thesystem after providing the input as image. Here the input gesture will be recognizedon the basis of

finger count. The desired action will be performed. Errors may arisedue to invalid gestures or quick movement of hand resulting in the system notrecognizing the gesture. Invalid gestures will result in no action being performedsimilarly quick movement of hand will be ignored to avoid accidental gestures. System uses frames from an input video stream, performs morphological filteringupon it, RGB/HSV filtering is performed to detect hand glove, contours and defectsare detected and gesture is recognised and compared with number of outstretchedfingers, based on the number of outstretched fingers operations are performed inVLC media player. Communication with VLC media player is done via TCPConnection. RGB/HSV filtering techniques used have a high rate of detection [9]. controlling our media player using hand gestures with the help of OpenCV and Python. Computer applications require interaction between human and computer. This interaction needs to be unrestricted and it has made it challenging to traditional input devices such as keyboard, mouse, pen etc. Hand gesture is an important component of body languages in linguistics. Human computer interaction becomes easy with the use of the hand as a device. Use of hand gestures to operate machines would make interaction interesting. Gesture recognition has gained a lot of importance. Hand gestures are used to control various applications like windows media player, robot control, gaming etc. Use of gesture makes interaction easy, convenient and does not require any extra device. Vision and audio recognition can be used together. But audio commands may not work in noisy environments [10]. Most of all interaction tasks relevant for a general three-dimensional virtual environment can be supported by 6DOF control and grab/select input. Obviously a very efficient method is direct manipulation with bare hands, like in real environment. This paper shows the possibility to perform non-trivial tasks using only a few well-known hand gestures, so that almost no training is necessary to interact with 3D-softwares. Using this gesture interaction we have built an immersive 3D modeling system with 3D model representation based on a mesh library, which is optimized not only for real-time rendering but also accommodates for changes of both vertex positions and mesh connectivity in real-time. For performing the gesture interaction, the user s hand is marked with just four fingertipthimbles made of inexpensive material as simple as white paper. Within our scenario, the recognized hand gestures are used to select, create, manipulate and deform the meshes in a spontaneous and intuitive way. All modeling tasks are performed wirelessly

through a camera/vision tracking method for the head and hand interaction. Overall, the body of material illustrates both the viability of utilizing computer vision and machine learning techniques for gesture identification as well as the promise of gesture-based control to enhance YouTube user experience. By creating a bespoke gesture detection system especially made for operating YouTube, the GestureTube project expands on these earlier researches [11].

II. Background

As YouTube is used more often for entertainment and education, several techniques to improve user experience are being investigated. Gesture-based control is one such technique that enables users to interact with the YouTube interface without the usage of standard input devices like a keyboard or mouse. Robotics, virtual reality environments, and multimedia player control are just a few of the areas where gesture-based control has been successfully used. However, there isn't much research on how well gesture-based control works for YouTube.

The usefulness of gesture-based control for YouTube has not been thoroughly studied in previous study, which has concentrated on the usage of gestures for other applications, such as operating a media player or virtual world. This study evaluates the efficacy of gesture-based control for YouTube in an effort to fill this vacuum in the literature. The study will examine the possible advantages of gesture-based control, including elevated user involvement and less physical effort, as well as the disadvantages, including a learning curve and diminished accuracy. The analysis will take into account the user's experience utilizing gesture-based control for various YouTube operations, including finding a video to watch, playing it, pausing it, and adjusting the volume.

A user research will be done to see whether gesture-based control for YouTube is successful. With the aid of traditional input methods like a keyboard and mouse as well as gesture-based control, participants will be challenged to complete a variety of YouTube activities. A number of variables, including task completion time, accuracy, user satisfaction, and physical effort will be measured in the study. The results of this study can be used to guide the development of gesture-based control systems for YouTube and other applications that are comparable, as well as to shed light on the potential advantages and disadvantages of this kind of interaction. The

study's overall goal is to add to the increasing body of knowledge on gesture-based control and human-computer interaction.

III. Methods

A. Review of the literature

In order to inform the development of the GestureTube system, a thorough review of the literature on gesture-based control and YouTube will be done.

B. Data collection

To train and evaluate the gesture recognition system, the project entails gathering a dataset of hand movements. A webcam will be used to take pictures of users doing various hand motions for this purpose.

C. Preprocessing

To extract the pertinent characteristics for gesture identification, the acquired data will be preprocessed using computer vision techniques such image filtering, thresholding, and segmentation.

The MediaPipe Hands library's hand gesture recognition technology uses the webcam to take pictures of the hand's landmarks in order to identify various hand movements. Google created the MediaPipe Hands library, a computer vision tool that can identify the locations of the fingers, knuckles, and wrist as well as other hand landmarks. The 'Gesture' and 'Landmarks' column headers are written to the CSV file using the csv.writer() function in this method, which creates the file in write mode. The csv.writer() function is used on line 11 of the code to write the 'Gesture' and 'Landmarks' headers to the CSV file. The writer object that is created by this function makes it easier to write to a CSV file. The next step is to write a row to the file using the writer object's writerow() function. In this instance, the row has the header values "Gesture" and "Landmarks" respectively. The MediaPipe library, which is designed to recognize and track hand landmarks in real-time video frames, is used to record the gestures and movements. The VideoCapture() method of OpenCV is used in the code to capture a video stream from the

device's camera. The MediaPipe Hands module then analyzes each frame of the video stream and returns the landmarks of any hands that were found. The landmarks are then put into a list called "l" and written to the CSV file along with the label for the identified gesture. Based on the photographed hand landmarks, the recognizeHandGesture() function determines the gesture label. Using the csv.writer() function, the 'Gesture' and 'Landmarks' headers are written to the CSV file. The writer object that is created by this function makes it easier to write to a CSV file. The next step is to write a row to the file using the writer object's writerow() function. In this instance, the row has the header values "Gesture" and "Landmarks" respectively. The webcam video stream is then grabbed by the method, and the MediaPipe Hands library is used to process the acquired frames. The library is used to identify the hand's landmarks, which are then photographed and saved in a list with the suffix "l". The'recognizeHandGesture' function is then called using the 'l' list to identify the hand gesture. Using a dictionary named "gestures," the identified gesture is then mapped to the gesture name. The csv.writer() function is then used to write the gesture name and the landmarks recorded for that gesture to the CSV file. This procedure keeps going until the user closes the CSV file by pressing the ESC key. The captured landmarks can then be used to train machine learning models for hand gesture recognition.

D. Recognition of gestures

Using the preprocessed data, a machine learning model will be trained to identify various hand gestures. Given their effectiveness in picture identification tasks, deep learning methods like convolutional neural networks (CNNs) will be investigated for this purpose.

Gestures will be translated into specific YouTube actions, such as play, pause, skip, or rewind, once they have been identified. The user will define the actions, and they can be altered to fit their tastes.

E. Evaluation

The accuracy, responsiveness, and user satisfaction of the GestureTube system will be assessed. User studies and surveys will be used in the evaluation to get input on the system's effectiveness and usability. To examine the information gathered during hand gesture recognition, use the

CSV file. The following are some possible analyses: Analysis of gesture frequency: Count how many times in the dataset each gesture was made. The gestures that are used more frequently than others can be determined using this. Analyze the length of time that each gesture was held by computing its duration. This makes it easier to tell which gestures were made rapidly and which ones were held for a longer period of time. Analyze the locations of landmarks for each gesture to find any trends or connections between particular landmarks and particular movements. Examine how frequently and in what order various gestures were executed to spot any recurring transitional patterns. Overall, the CSV file can offer helpful insights on how humans make hand gestures and how computer vision techniques might be used to identify them.

IV. Task division

A. *Rayapaneni, Sai Sampath*

conducting a thorough literature review on YouTube and gesture-based control and Finding and comparing several deep learning gesture recognition systems and Python and TensorFlow are used in this project to put the gesture recognition model into practice and refine it. Using a gathered dataset, assess and confirm the accuracy of the gesture recognition model.

B. *Kavya Obula*

A dataset of hand movements should be gathered and prepared for gesture recognition and Combine the YouTube platform with the gesture recognition model. Create and use computer vision preprocessing methods on the dataset to improve the gesture recognition model, working with Person 1. In YouTube, associate recognized gestures with particular activities.

C. *Bheemani, Venkata Vamsi Krishna*

Doing user research and surveys to assess the system's usability and efficiency. Evaluating user comments and make system suggestions. With Person 2, adapt the actions in response to user feedback and Check the system's responsiveness, confirm it, and suggest improvements.

REFERENCES

[1] S. Song, D. Yan, and Y. Xie, "Design of control system based on hand gesture recognition," in *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, 2018, pp. 1–4.

[2] P. N. Arathi, S. Arthika, S. Ponmithra, K. Srinivasan, and V. Rukkumani, "Gesture based home automation system," in *2017 International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2)*, 2017, pp. 198–201.

[3] C. Hu, X. Wang, M. Mandal, M. Meng, and D. Li, "Efficient face and gesture recognition techniques for robot control," in *CCECE 2003 - Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology (Cat. No.03CH37436)*, vol. 3, 2003, pp. 1757–1762 vol.3.

[4] Y. Gao, Y. Jin, S. Choi, J. Li, J. Pan, L. Shu, C. Zhou, and Z. Jin, "Sonicface: Tracking facial expressions using a commodity microphone array," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 5, no. 4, dec 2022. [Online]. Available: https://doi.org/10.1145/3494988

[5] T. T. M. Nguyen, N. H. Pham, V. T. Dong, V. S. Nguyen, and T. T. H. Tran, "A fully automatic hand gesture recognition system for human-robot interaction," ser. SoICT '11. New York, NY, USA: Association for Computing Machinery, 2011. [Online]. Available: https://doi.org/10.1145/2069216.2069241

[6] K. B.S, Y. Yakkeshwar, J. R, and M. Priya, "Gesticulation based media player controller using deep learning," *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, vol. 07, 02 2023.

[7] F. Ahmed, M. Rayhan, S. Rahman, N. Benazir, A. Chowdhury, and M. Imran, "Controlling multimedia player with eye gaze using webcam," 02 2019.

[8] S. Rautaray and A. Agrawal, "A vision based hand gesture interface for controlling vlc media player," *International Journal of Computer Applications*, vol. 10, 11 2010.

[9] I. Vadgama, Y. Khot, Y. Thaker, P. Jouras, and Y. Mane, *Gesture Based Audio/Video Player*, 09 2017, pp. 369–378.

[10] D. Varshika, "Media-player controlling by hand gestures," *International Journal for Research in Applied Science and Engineering Technology*, vol. 9, pp. 2022–2030, 06 2021.

[11] H. Kim, G. Albuquerque, S. Havemann, and D. W. Fellner, "Tangible 3D: Hand Gesture Interaction for Immersive 3D Modeling," in *Eurographics Symposium on Virtual Environments*, E. Kjems and R. Blach, Eds. The Eurographics Association, 2005.