ISBN: 978-981-94-1156-6
2024 the 14th International Workshop on Computer Science and Engineering (WCSE 2024)
doi: 10.18178/wcse.2024.06.052

Hand Gesture-based Computer Operation in Multiple-user Environment

Shion Kashihara+, Eiji Kamioka [1], Chanh Minh Tran [2] Phan Xuan Tan [1]

[1] Graduate School of Engineering and Science, Shibaura Institute of Technology, Japan
[2] College of Engineering, Shibaura Institute of Technology, Japan

Abstract. Typical computer interfaces are designed for one-on-one interaction, leading to decreased efficiency in shared spaces like conference rooms where a single mouse or keyboard is shared among multiple users. Virtual mouse systems using hand gestures have been proposed to solve this issue. However, they face challenges such as low gesture recognition accuracy, lack of intuitive operation, and difficulty in recognizing authorized mouse control operation. To address these problems, this paper proposes a virtual mouse system with a novel mechanism for hand gesture control. Specifically, the concept of state transition is introduced to hand gestures recognition to improve recognition accuracy. Additionally, a method for transferring control privilege using intuitive gestures is introduced to enable accurate operation in multi-user environments. Through evaluation experiments, it is found that this proposed method achieves accurate mouse operation and can transfer control privilege more efficiently compared to existing methods.

Keywords: Gesture Recognition, Virtual Mouse, Multi-User Interface

1. Introduction

In today society, the demand for technology that supports multi-user interaction with a single device has grown, especially in collaborative settings such as meetings where participants share presentation materials. Gesture recognition technology stands as a promising avenue to meet this demand, offering a way for multiple users to interact with devices simultaneously without physical contact. However, the current state of gesture recognition faces notable challenges, such as low accuracy and the non-intuitive operation of virtual mouse, which can hinder its practical application and user experience. The purpose of this work is not just to highlight these challenges but to actively seek solutions that enhance gesture recognition accuracy and develop intuitive virtual mouse interface. In this proposed method, users acquire the initiative to control the PC through a sequence of specific gestures. Additionally, gestures inspired by actual mouse operations allow for intuitive interaction. These gestures' accuracy and intuitiveness were evaluated from both qualitative and quantitative perspectives in the evaluation experiment and got good results.

2. Related Work

Fenglin et al. [1] focused on Kinect-based hand gesture recognition, utilizing Microsoft Kinect V2 to capture human hand movements and to convert them into computer inputs. In another study, Faisal et al. [2] evaluated the effectiveness of hand gesture recognition using data gloves and deep learning. However, these methods require special sensors or devices, posing a challenge of wide adoption.

Gowra et al. [3] discussed a TensorFlow-based mouse control system, aiming to implement a virtual mouse as an alternative to traditional touchscreens and computer mouse. Additionally, Kharbanda et al. [4] and Kabid et al. [5] proposed AI-based virtual mouse systems using hand gestures to perform computer mouse functions. These studies, both utilizing cameras for virtual mouse control, remained issues such as the low recognition accuracy in certain gestures and the non-intuitive nature of gestures to replicate traditional mouse functions.

In multi-user environment, a challenge is the decrease in gesture recognition accuracy as the number of users in camera view increases. LiDAR or Wi-Fi were employed for user detection in the works in [6][7]. Dinh-Son et al. [8] introduced a real-time virtual mouse system using RGB-D imaging and fingertip detection.

+ Corresponding author. Tel.: + 81 80 4448 5865.
E-mail address: af20040@shibaura-it.ac.jp.

Enkhbat et al. [9] discussed a multi-user gesture interaction system called MOBAJES, a hybrid system combining the advantages of wearable and non-wearable technologies, where a camera-equipped mobile device worn like a pendant was used for gesture recognition. In this work, only a normal camera is used and no other specialized equipment is considered.

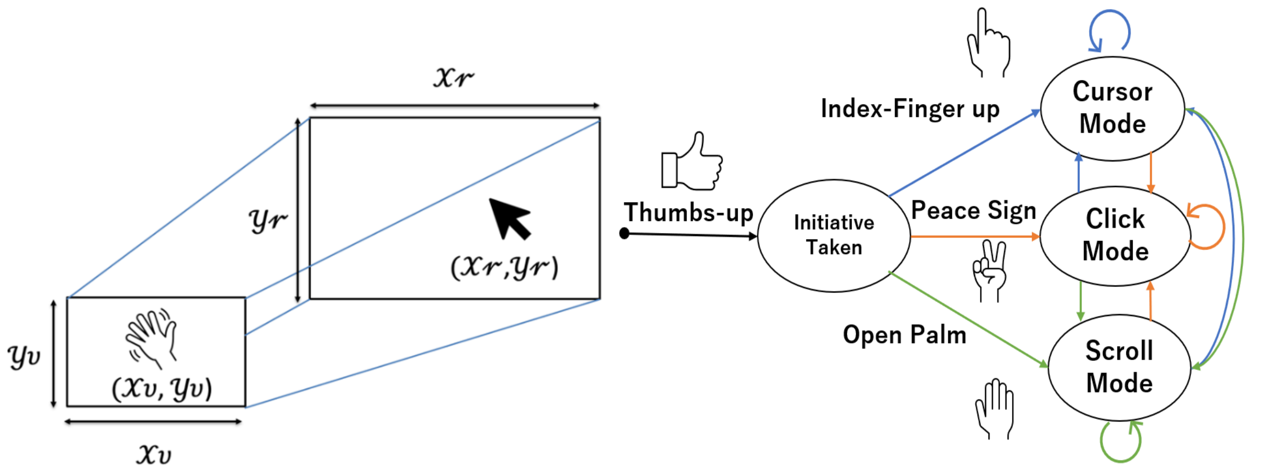
3. Proposed Hand Gesture-based Computer Operation

This study aims to improve the intuitiveness and efficiency of computer operations in environments where multiple users are present simultaneously. To realize this, a new virtual mouse system, which incorporates mechanisms for transferring mouse operation control privilege based on a specific gesture and for reducing gesture misrecognition using state transition (mode) control, is proposed. This system only requires a standard Web camera, eliminating the need for expensive specialized equipment or high-performance GPUs. This makes it a low-cost solution that has the potential for widespread adoption.

The proposed virtual mouse system utilizes a combination of OpenCV [10] MediaPipe [11], PyAutoGUI [12], and the SORT [13] algorithm to facilitate intuitive and seamless computer interactions through hand gestures. OpenCV is employed for real-time hand tracking and gesture analysis, leveraging advanced image processing to distinguish hands from other objects in the video stream from a Web camera. It ensures that gestures are recognized smoothly and accurately, enhancing system responsiveness. MediaPipe is crucial for accurate hand gesture recognition using its Hands module to detect hand landmarks and movements in real time. This technology enables the system to identify specific gestures for actions such as clicks and mode transitions. PyAutoGUI translates these recognized gestures into actual mouse operations, such as moving the cursor or clicking, by capturing user actions within the GUI. It acts as a bridge between natural hand movements and digital responses, allowing for direct interaction with the computer interface. The SORT algorithm enhances the system's tracking capabilities, assigning unique IDs to each hand gesture for continuous and accurate tracking across video frames. This supports the system's ability to handle multiple gestures simultaneously, ensuring smooth operation and minimizing errors, especially in multi-user environments.

The process begins with camera capture using OpenCV, followed by the SORT algorithm which detects how many hands are present on the screen. Simultaneously, MediaPipe captures the coordinates of 21 assigned landmarks on each hand. Subsequently, a Python script distinguishes the gestures based on predetermined thresholds. Finally, PyAutoGUI links these gestures to mouse actions.

Figure 1 illustrates the proposed virtual mouse system. A virtual monitor in the virtual space, which is located between the real physical monitor and the user, translates the user's hand movements into cursor movements naturally and intuitively. This process is supported by a coordinate transformation algorithm between the actual and virtual monitors, providing high compatibility with different screen sizes and resolutions. Additionally, a customizable sensitivity adjustment feature allows users to finely tune the cursor's responsiveness according to their preferences. Figure 1(a) shows the correspondence of coordinates between the virtual monitor (x_v, y_v) and the real physical monitor (x_r, y_r)



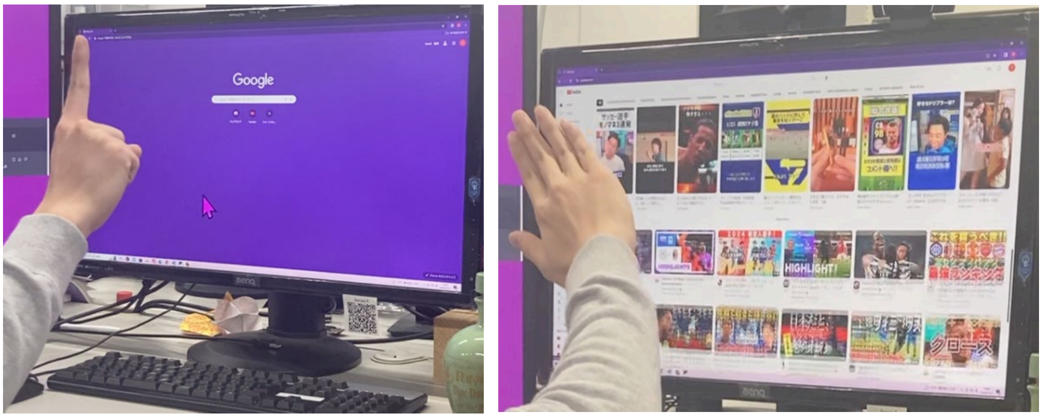
(a) Correspondence of monitor coordinates (b) Initiative acquisition and state transition (mode) control

Fig. 1: Proposed hand gesture-based computer operation

In the idle state of the proposed system, all hands of multiple users captured by the Web camera are detected. At this point, any user can take the mouse control initiative. The initiative can be obtained by conducting a "Thumbs-up" gesture. Even though a user has the initiative, another user can get the initiative by conducting a “Thumbs-up” gesture. Through this gesture, the initiative can be transferred to other users, enabling a robust multi-user operation.

The proposed virtual mouse system distinctly differentiates among various operations such as cursor movement, left and right clicks, and scrolling through the mode-switching function. Users can easily switch among these modes using specific gestures, executing only the operations allowed in each mode. More concretely, when a user acquires the mouse control initiative, the proposed system transitions to “Cursor Mode”, “Click Mode”, or “Scroll Mode”, depending on a specific pre-determined hand gesture for each mode (e.g., “Index-Finger up”, “Peace Sign”, “Open Palm”, and so on) conducted by the user. This determines the current operation mode, allowing only the limited hand gestures designated for each mode to be detected thereafter, reducing misdetection and malfunctions. Figure 1(b) shows the state transition (mode) control in the proposed system.

Figures 2(a) and (b) depict the actual usage of the system. In (a), the mouse cursor is controlled by a gesture of raising only the index finger, while in (b), scrolling through the screen is performed with an open palm gesture.



(a) Moving the mouse cursor (b) scrolling

Fig. 2: Usage Scenarios of the System

4. Performance Evaluation

4.1. Evaluation Experiments

Evaluation experiments conducted in this study aimed to thoroughly test the proposed system in real-world scenarios to assess its practicality, reliability, and applicability. This experiment was designed to evaluate how accurately, efficiently, and intuitively the system can perform mouse operations. Twenty university students underwent two main experiments. The first experiment was conducted to measure the accuracy of basic mouse operations and the effectiveness of transferring the initiative for computer operations using the proposed system. This process involved a practice session for participants to familiarize themselves with the system's operation principles and specific actions, such as how to execute clicks. After the practice session, the participant continued to the exercise session, where they had to repeat designated mouse operations. In this session, the operation recognition accuracy was assessed to evaluate the system performance. The mouse movement test involved navigating the mouse cursor through black circles located at each screen corner. For click operations, the exact number of clicks was measured using a click counter. The naturalness and accuracy of scrolling were evaluated by navigating through actual Internet pages during scroll operations.

The second experiment tested whether multiple users could effectively transfer the initiative of the virtual mouse among themselves. Using a thumbs-up gesture, participants sequentially acquired and then passed on control of the system. This process was repeated 30 times to measure the success rate of transferring initiative.

Following the above experiments, a usability survey was conducted, asking participants to provide feedback on a series of questions regarding the naturalness of cursor movement, the intuitiveness of operations, and the overall user experience. They rated these aspects on a five-point scale and also free-writing answer. This survey aimed to identify the system's ease of use from the user's perspective and any areas requiring improvement.

4.2. Evaluation Results

Figure 2 shows the evaluation results obtained by the proposed method in this study, comparing with the methods by Kabid et al. [5] and Guha et al. [14]. In Fig. 2, the vertical axis represents the accuracy of gesture recognition, while the horizontal axis represents the types of gestures. Kabid et al. employed colored fingertip caps to recognize hands. The proposed system utilizes a webcam, either externally connected or integrated within a laptop, to capture video frames in real-time. Through the instantiation of a video capture object, the system is capable of acquiring video using a primary camera (identified by a device index of "0"), with the capability to integrate additional cameras by assigning subsequent indices (1, 2, etc.). These captured frames are sequentially passed to the system for further processing. In an ongoing process, the webcam captures each frame within an infinite loop, continuing until the termination of the program. Subsequently, the captured frames undergo a conversion from the Blue-Green-Red (BGR) color space to the Hue-Saturation-Value (HSV) color space. This transformation is pivotal for enhancing the efficiency of the subsequent color detection and masking phases. The system's approach to color detection involves identifying the color pixels of fingertips, which are marked with color caps, against the background of the captured frames. This initial step is crucial for the system's functionality, resulting in a grayscale image where the color caps are prominently highlighted against the rest of the frame. To track these color caps, rectangle bounding boxes (masks) are generated around them. The system then detects gestures based on the tracking of these color caps, facilitating an innovative method for interpreting user commands.

Guha et al. employed the similar method to the proposed method in this work, using OpenCV, MediaPipe, and Autopy (an older version of PyAutoGUI.) The AI virtual mouse system employs the instructive algorithm to transform the coordinates of the fingertip from the camera screen to the full screen of the PC window. When a hand is detected, and it is unclear which finger is raised to perform a specific mouse function, a rectangular box is drawn around the PC window within the camera area, allowing the mouse pointer to be moved within the window. In this system, the AI mouse utilizes Media-Pipe to identify which specific finger is raised and locates its precise coordinates. Based on the part of the finger that is raised, the actual mouse function is executed. After all, the rectangle matches the actual monitor, and moving the hand within the rectangle realizes mouse functions.

From this result, in all types of gestures, the proposed method in this work achieved higher accuracy than the one by the Kabid et al., and the same accuracy or slightly better than the one by Guha et al. It is believed that the low accuracy in the method by Kabid et al. is because they did not utilize machine leaning.

Figure 4 shows the success rate of initiative transfer in several multi-user scenarios (from two to six groups). For comparison, the research results by Dinh-Son et al. [8] are also included. In the fig. 3, the vertical axis represents the success rate of recognition of the user with initiative, while the horizontal axis represents the number of users. Dinh-Son et al. evaluated the success rate of fingertip tracking for multiple users, which can be considered equivalent to the initiative transfer success rate in this study. From this result, in the method by Dinh-Son et al., the success rate was highest when there were two users, at 89.78%. Additionally, it can be observed that the accuracy gradually decreases as the number of users increases from two to six. Meanwhile, the proposed method superiorly achieved 100% accuracy in all cases. Additionally, as an extra experiment, this study examined how many hands could be recognized simultaneously and whether initiative transfer could be achieved with them. As a result, it was found that initiative transfer could be conducted with up to eight users simultaneously. Although experiments with more than eight users were not conducted due to camera angle limitations, considering the obtained results, it can be expected that this functionality would be effective for even more participants.

Figure 5 represents an example of the results from a survey conducted among participants, covering two out of the total nine questions. Looking at the pie chart, it is evident that over 90% of the responses regarding the issue of clicking—a problem in related research—were rated as either excellent or good. Furthermore, regarding the handover of initiative, which is central to this study, everyone rated it as either excellent or good. From this, it can be inferred that the proposed method in this study is superior as a user interface.

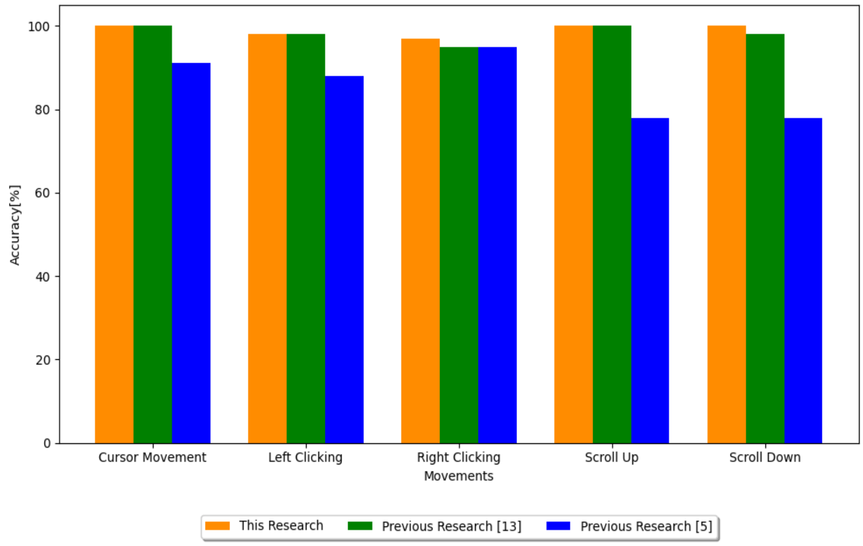


Fig. 3: Comparison of Gesture Recognition Accuracy

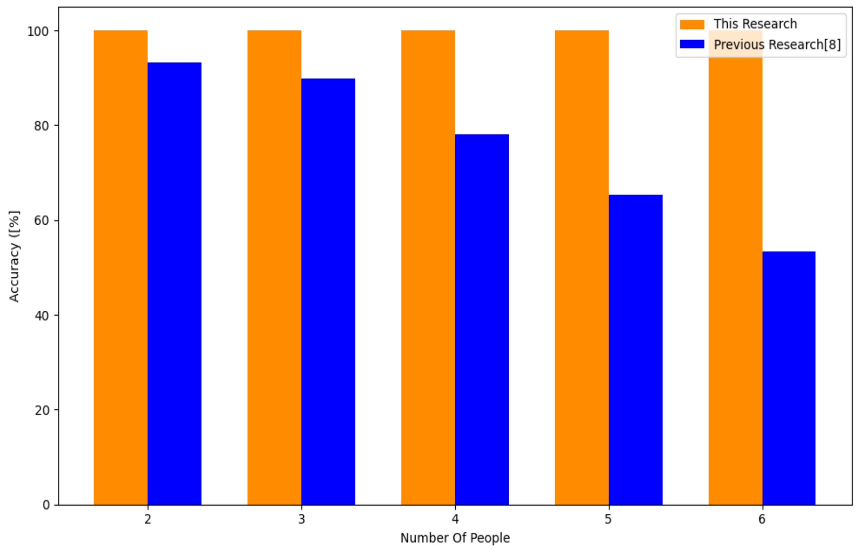


Fig. 4: Comparison of initiative transfer success rate

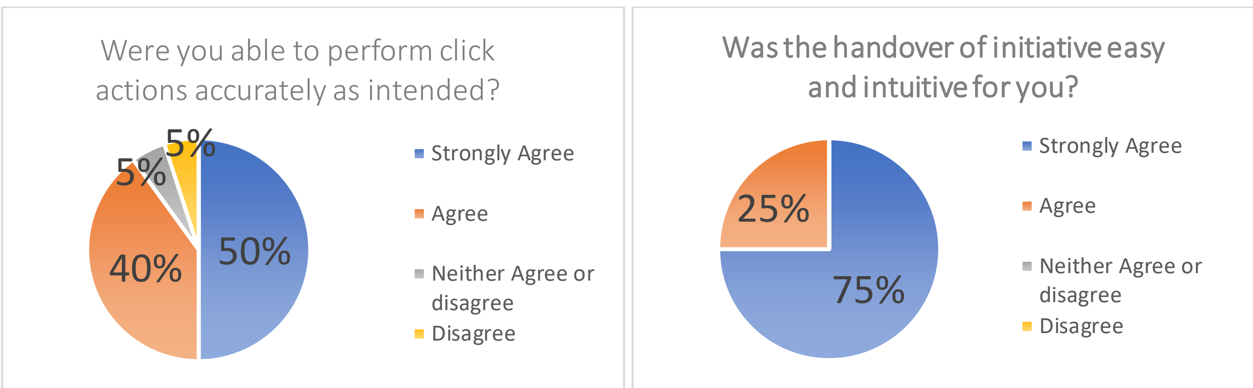


Fig. 5: Result Examples of User Interviews

5. Conclusion

In this paper, a method for realizing a multi-user compatible virtual mouse without using special equipment or expensive devices was proposed. Specifically, by leveraging a standard Web camera and utilizing OpenCV, MediaPipe, PyAutoGUI, and the SORT algorithm, gesture-based mouse operations and initiative transfer among users were achieved. Performance evaluation experiments showed that the success rate of the basic functions of the computer mouse exceeded that of existing methods. Furthermore, experiments on initiative transfer in a multi-user environment achieved a 100% success rate even with eight users simultaneously, which is a significantly superior result compared to existing research. In conclusion, this system is cost-effective to implement, and excels in accuracy of mouse operation and intuitive usability in a multi-user environment compared to existing research. However, it is necessary to address environmental factors such as background image processing, as the recognition accuracy of hand gestures may deteriorate depending on the background color seen from the camera.

6. References

[1] Fenglin et al., "Kinect-based hand gesture recognition using trajectory information, hand motion dynamics and neural networks," Artificial Intelligence Review, 2019, Vol. 52, Issue 1, pp. 563-583.

[2] Faisal et al., "Exploiting domain transformation and deep learning for hand gesture recognition using a low-cost dataglove," Scientific Report, 12, 21446, 2022.

[3] Gowra et al., "Virtual Mouse using Machine Learning and GUI Automation," Proceedings of the 8th International Conference on Advanced Computing and Communication Systems 2022 (ICACCS 2022), pp. 1112-1117.

[4] Kharbanda et al., "GESTURE CONTROLLED VIRTUAL MOUSE USING ARTIFICIAL INTELLIGENCE," International Research Journal of Modernization in Engineering Technology And Science, Vol. 05, Issue 01, January, 2023, e-ISSN 2582-5208.

[5] Kabid et al., "Design and Development of Hand Gesture Based Virtual Mouse," Proceedings of the 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019, pp. 1-5.

[6] Alexandros et al., "Multi-User Macro Gesture Recognition using mmWave Technology," Proceedings of the 18th European Radar Conference (EuRAD), 2022, pp. 37-40.

[7] Jaya et al. "Real-Time Hand Gesture Recognition Using Fine-Tuned Convolutional Neural Network," Sensors, Vol. 22, No. 3, 706, 2022, pp. 1-14.

[8] Dinh-Son et al., "Real-time virtual mouse system using RGB-D images and fingertip detection," Multimedia Tools and Applications, Vol. 80, 2021, pp.10473-10490.

[9] Enkhbat et al., "MOBAJES: Multi-user Gesture Interaction System with Wearable Mobile Device," Proceedings of the 15th International Conference, HCI International 2013, pp. 196-204.

[10] Open CV, https://opencv.org/

[11] MediaPIpe, https://developers.google.com/mediapipe

[12] PyAutoGUI, https://pyautogui.readthedocs.io/en/latest/

[13] SORT https://jp.mathworks.com/help/fusion/ug/implement-simple-online-and-realtime-tracking.html

[14] Guha et al., "AI Virtual Mouse Using Hand Gesture Recognition," International Journal for Science Technology and Engineering, Vol. 10, Issue 4, pp. 3070-3076, 2022