International Journal of Engineering Research and Sustainable Technologies
Volume 1, No.1, September 2023, P 31 - 37
ISSN: 2584-1394

A COMPUTER VISION BASED VIRTUAL HAND GESTURE RECOGNITION SYSTEM

Purushothoman[*a]
[*a]UKhertfordshire, Hertfordshire University, London
*** Corresponding author email :**Purushoth0406@gmail.com

https://doi.org/000000/000000/

Abstract

Researchers are presently directing their efforts towards the development of devices characterized by reduced hardware dependencies. The domain of computer vision technology is experiencing a notable surge in advancement and innovation. Consequently, this article suggests the integration of hand gestures as an input modality to facilitate the creation of a virtual mouse and keyboard interface. The proposed method operates by utilizing a webcam to detect and interpret hand movements, allowing users to interact with the computer system without the need for physical hardware peripherals such as a mouse or keyboard. This virtual input system holds significant advantages, particularly in situations where physical device space is constrained. Notably, it mitigates issues related to battery consumption, which is often a concern in conventional input methods reliant on power-hungry hardware components. Furthermore, this innovative approach bears the potential to contribute to the containment of the rapid transmission of contagious agents, such as the Coronavirus, by reducing the need for physical contact with shared input devices. Additionally, this technology offers considerable utility to individuals who may face challenges when employing a traditional physical keyboard and mouse. Leveraging computer vision techniques, the virtual keyboard and mouse control system presented herein is designed to enhance user accessibility and usability, providing a more intuitive and adaptable interaction method.

Keywords: Gesture-Based Input, Virtual Mouse Control, Virtual Keyboard Control

1. Introduction

As technological advancements continue, electronic devices are progressively shrinking in size. Significant progress has been achieved in the realm of computer vision, empowering computers to adeptly recognize users through straightforward programs grounded in image processing. This technological advancement finds broad applications in everyday scenarios, ranging from facial recognition and color detection to the intricacies of autonomous driving systems. Recognizing the increasing importance of human-computer interaction (HCI), particularly in the realm of vision-based gestures and sensing, this proposal introduces a novel multimodal interactive keyboard and mouse system. This system, managing mouse tasks and keyboard functions through a webcam, relies on various image processing algorithms implemented in Python. The computer vision library OpenCV is integrated into the AI, utilizing the Media Pipe package for hand tracking and tip tracking, along with Autopy and PyAutoGUI for executing actions such as left click, right click, and scrolling. The Python-based model demonstrates exceptional accuracy, making it suitable for real-world applications on standard computer hardware without the need for a GPU. Existing attempts at virtual keyboard and mouse systems, including those accommodating eye movements, are explored through a literature review of research papers published in reputable journals.

Notable mentions in related works include Liou, Hsieh, and Lee's 2010 hand gesture recognition system with complex hand movements (Hsieh et al., 2010), and Gandhi, Dudhane, and Patil's 2013 study on hand gesture recognition with frame size limitations (Shibly et al., 2019). In the work by Pasi, Singh, and Kumari, and Shetty and colleagues, they explored cursor control using hand gestures (Kumari et al., 2016) (Islam et al., 2019), each with its unique set of advantages and limitations. In 2016, S. Shetty and colleagues employed color detection for a virtual mouse system using OpenCV (Science and Ijsrset, n.d.) (Bhuvana et al., 2017), with the system being sensitive to the background. A 2021 project by P. C. Shindhe et al. focused on mouse-free indication management through hand gestures (Gupta et al., 2021), incorporating digital camera-based motion capture and color detection methods (Rajee et al., 2023).

The proposed virtual mouse and keyboard system presented in this work rely on hand gesture patterns captured by a digital camera, with colored fingertips acting as detectable objects. This approach allows for efficient mouse operations without the need for a physical mouse and keyboard. The application scenarios include space-saving environments and mobile situations where physical mouse usage is impractical (Shibly et al., 2019) . Moreover, it addresses the challenges faced by individuals with hand-related difficulties who cannot use a traditional mouse. The system's relevance is underscored in the context of the COVID-19 situation, where physical contact with devices poses a risk of infection spread (Shriram et al., 2021) (Da Silva and Veiga, 2020). The utilization of this AI-driven virtual mouse and keyboard system (Singh et al., 2022) mitigates such risks and proves beneficial in scenarios where conventional input methods are impractical or unsafe.

The research done by (Ahmed et al., 2019) aims to construct a facial expression recognition system utilizing a convolutional neural network (CNN) integrated with data augmentation techniques. In the work done by (Ahmed et al., 2020), a hybrid approach combining data-driven and knowledge-driven methodologies is employed to evaluate the mental state of an individual. The integration of these approaches holds the potential to facilitate the early identification of potential suspects, allowing law enforcement agencies to preemptively address criminal activities. The performance assessment of this integrated system indicates its reliability and superiority over existing methods for facial expression assessment. In the work done by (Patil and Lambhate, 2015) a camera-based framework actively monitors the user's eyes, enabling the user to emulate mouse clicks through intentional blinks and winks. In the research endeavour done by (Gupta et al., 2019), they have crafted a digital personal assistant tailored for individuals with disabilities. That innovative system specializes in recognizing continuous Bangla voice commands, enhancing accessibility and usability for users with varying needs. A system capable of discerning the presence of the same individual across different frames of a video was introduced by (Asad et al., 2019) (Mj, 2018). In this a novel clustering algorithm, a variation of the Hierarchical Agglomerative Clustering algorithm demonstrates the ability to effectively cluster face images of humans, addressing the specific need for reliable person identification across video frames.

2. Methodology

The proposed technology facilitates the onscreen display of an associate's degree keyboard layout while leveraging the computer's digital camera. User input is accomplished through a yellow-marked finger, employed to input terms. Activation of the mouse control system is achieved by pressing the Mouse Control Module button adorned with a yellow-colored cap. Subsequently, a live video frame captures and tracks the user's hand motions, identifying mouse functions. System Development involves two fundamental stages in the color recognition process: the calibration phase and the recognition phase. The calibration phase's primary objective is to enable the system to identify the Hue Saturation Values of user-selected colors. The identified values and settings are then stored in text documents for subsequent use during the recognition phase. In the latter phase, the system captures frames and searches for color input using the values recorded during the calibration phase (Abedin et al., 2017). The depicted image Fig 1 below illustrates the phases of the virtual mouse.

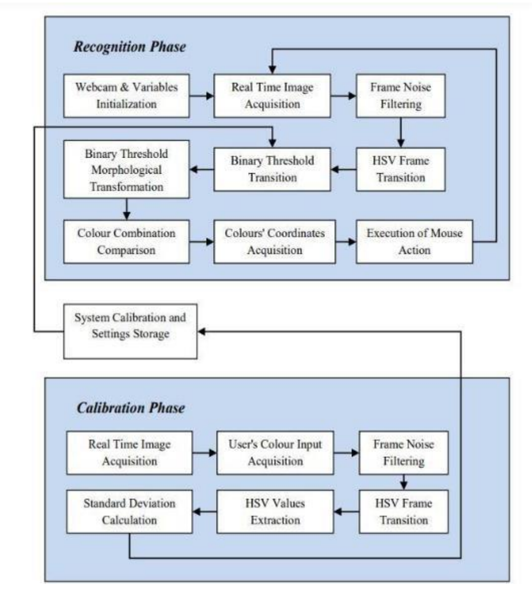

Fig. 1. System design process diagram

The program initiates by capturing live images using a webcam, awaiting human input on color. The resulting image is compressed to a manageable size to reduce processing demands. Frames, composed of user-input colors, undergo a series of calculations to derive calibrated HSV values. As frames contain noise affecting program efficacy and precision, filters are applied to eliminate unwanted noise. The Gaussian filter, a popular smoothing method, is employed for noise reduction in the frames using Gaussian Blur(Input Array src, Output Array dst, Size ksize, double sigma X, double sigmaY=0, int borderType = BORDER_DEFAULT).

The webcam of the connected laptop or desktop governs runtime actions. A Video Capture object is established for video recording, with an infinite loop ensuring continuous camera operation throughout the program. The live stream is meticulously captured, and each recorded frame is translated from RGB to HSV. OpenCV provides over 150 color-space conversion methods, but attention is focused on BGR to Grey and BGR to HSV, the most commonly used.

A mask, generated based on predetermined constraints, creates a specific region of the image. In this instance, a mask is created for an item with a red color. The HighGui's imshow() method necessitates routine invocation of waitKey to function, completing the event loop processing of imshow(). The waitKey function tracks key events for a specified delay (5 milliseconds in this case), ensuring proper HighGui event processing.

Centers of detected red objects are determined by averaging the maximum and minimum points of bounding boxes. The average coordinates, converted from camera to screen resolution, represent the red point in the illustration. The close gesture is executed by clicking and dragging the object, similar to the open gesture. A variable called "pinchflag" is introduced for dragging implementation. After confirming the pinchflag is set to 1 following an open gesture, the drag action is performed; otherwise, a mouse move operation is executed.

The MyFileDropTarget class is created, featuring the overridden method OnDropFiles. This method accepts dropped file paths and mouse x/y positions. Utilizing the Python socket, the program transmits the file to the recipient system module, which then creates the file after reception of the file path from the MyFileDropTarget class.

3. Experimental System

The program leverages the Media Pipe framework for hand tracking and gesture recognition, complemented by the OpenCV library for computer vision. Machine learning concepts are integrated into the program to facilitate the monitoring and differentiation of hand movements and hand tips. Media Pipe, an open-source framework developed by Google, serves as the cornerstone of the machine learning pipeline employed in the program. This framework is particularly adept at handling time series data, rendering it versatile for cross-platform programming applications.

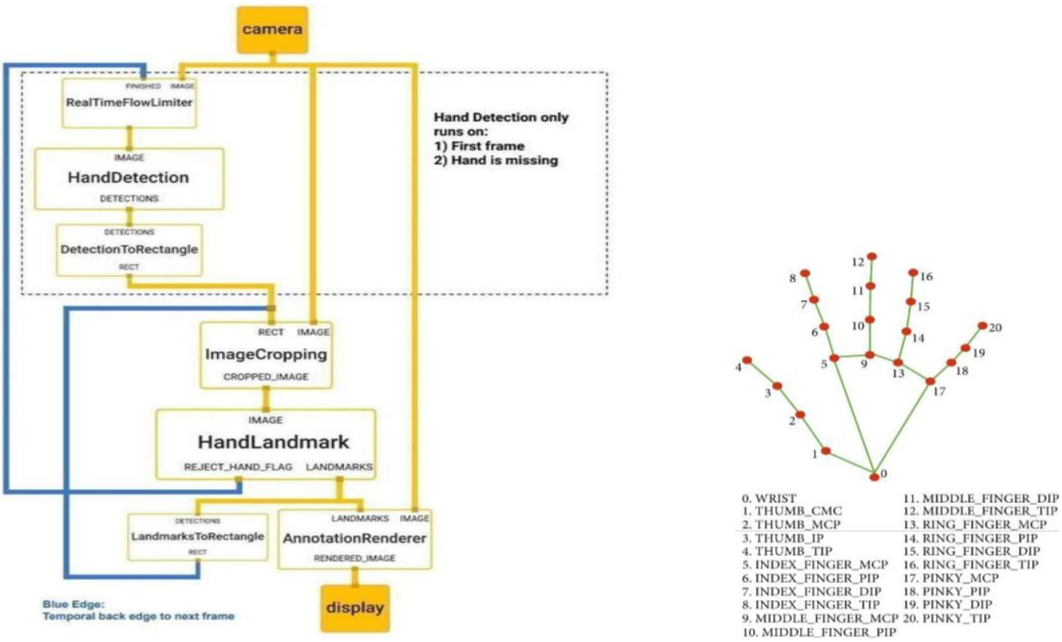

Fig. 2. Hand Recognition System

Due to the MediaPipe framework's multimodal capabilities, it is versatile enough to handle both dynamic audio and video content. The designer conceptualizes systems with practical utility, employing the MediaPipe framework for system structuring, evaluation, and graphical representation. The activities within a system utilizing MediaPipe are orchestrated within the framework's channel configuration. These channels exhibit cross-platform compatibility and scalability, catering to both desktop and mobile device applications. The foundational elements of the MediaPipe framework consist of three fundamental components: performance evaluation, a framework for reacquiring detector data, and a collection of calculable elements known as calculators.

Data packets seamlessly enter the system through channels, which are essentially graphs comprising factors termed calculators interconnected by aqueducts. In this framework, inventors retain the flexibility to define or modify custom calculators at any juncture within the graph, thereby replacing or enhancing operations as per their specific needs. The collaborative interplay between calculators and aqueducts manifests as a visual representation of data inflow, forming a graph created with MediaPipe. Each node in this graph symbolizes a calculator, and the connections between them are represented by aqueducts.

4. Results and Discussion

The proposed AI virtual mouse system introduces the concept of utilizing computer vision to enhance human-computer interaction. However, assessing the performance of the system is challenging due to the limited availability of datasets for comparison. Specifically, the aspects under scrutiny include finger-tip detection and the tracking of hand motions. These pivotal elements collectively contribute to the system's ability to interpret and respond to user inputs, thereby shaping the overall effectiveness of the proposed AI virtual mouse system.

The identification of relevant fingertip positions is crucial for effective hand gesture tracking in the AI virtual mouse system. This involves associating specific tip IDs with corresponding fingers: Tip ID 0 for the thumb finger, Tip ID 1 for the index finger, Tip ID 2 for the middle finger, Tip ID 3 for the ring finger, and Tip ID 4 for the little finger. The evaluation of hand gesture tracking and fingertip detection has been conducted at various distances from the webcam.

A comprehensive experimental test was executed to gather data, as outlined in Table 1. This involved the participation of four individuals, each performing the test 25 times, resulting in a total of 600 hand motions for meticulous hand labeling. To ensure robust evaluation, each participant tested the AI virtual mouse technology under diverse lighting conditions and varied viewing distances. The camera underwent testing under different scenarios: ten times under strong lighting conditions, five times in low light, five times in close proximity, and five times at a distance of five feet. The experimental results, encapsulating the outcomes of these rigorous tests, are meticulously presented in Table 1. This extensive testing protocol aimed to comprehensively assess the system's performance across a spectrum of real-world conditions, thereby providing valuable insights into its robustness and adaptability.

Table 1
Proposed AI Virtual Mouse System

Hand tip gesture	Mouse function performed	Success	Failure	Accuracy(%)
Tip ID 1 or both tip IDs 1 and 2 are up	Mouse movement	100	0	100
Tip Ids 0 and1 are up the distance between the fingers is<30	Left button click	99	1	99
Tip IDs 1 and 2 are up and the distance between the fingers is<40	Right button click	95	5	95
Tip IDs 1 and 2 are up and the distance between the fingers is>40 and both fingers are moved up the page	Scroll up function	100	0	100
Tip IDs 1 and 2 are up and the distance between the fingers is>40 and both fingers are moved down the page	Scroll down function	100	0	100
All five tip IDs 0, 1, 2, 3, and 4 are up	No action performed	100	0	100
Result		594	6	99

Table 1 clearly demonstrates that the proposed AI virtual mouse system has achieved an impressive level of accuracy, hovering around 99%. This high degree of accuracy underscores the system's efficacy in fingertip detection and hand gesture tracking across various distances from the webcam. The robust performance, as indicated by the accuracy rate, affirms the system's capability to consistently and precisely identify relevant fingertip positions associated with distinct fingers. Such accuracy is particularly notable given the diverse testing conditions, including variations in lighting and viewing distances. This outcome speaks to the system's reliability and proficiency in real-world scenarios, reinforcing its potential as an effective tool for human-computer interaction.

The results obtained from a variety of mouse movements and distinct keyboard strokes are visually illustrated in Figure 3 below. Evidently, the system demonstrates a notable capability to accurately recognize and distinguish between different mouse gestures and keyboard inputs. This high level of accuracy underscores the effectiveness of the system in interpreting and responding to diverse user interactions, affirming its reliability and precision across a range of input modalities.

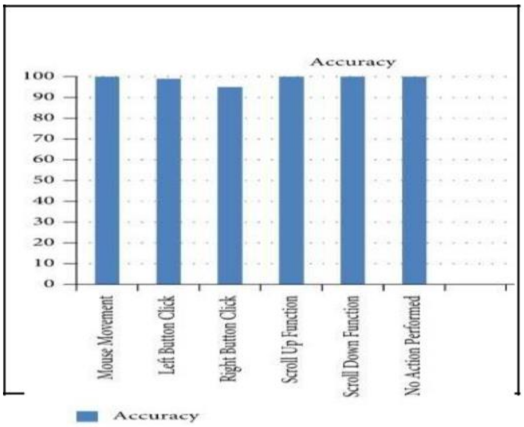

Fig. 3 (a). Accuracy of Mouse Movements

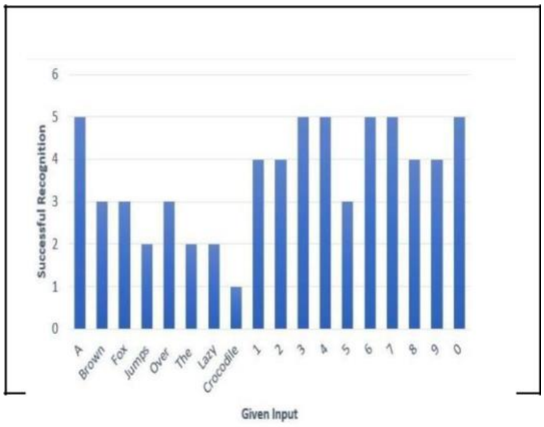

Fig. 3(b). Accuracy of Keyboard Strokes

The experimental variables in our study are encapsulated by the alphanumeric phrase "A Brown Fox Jumps Over the Slothful Crocodile" and the numerical sequence from 1 to 0. This comprehensive set of characters was subjected to a series of meticulous tests, totaling five administrations and encompassing a total of 44 characters, excluding spaces. The experiment was also done with a 52-year-old individual who had suffered a stroke, exhibiting notable limitations in left-side motor control. Despite these challenges, the participant engaged in the tests, contributing valuable insights to the experiment. The culmination of the study is visually represented in Figure 4 below, which graphically illustrates the system's success rate in accurately recognizing each word and digit within the specified character set. This visual depiction serves as a succinct yet comprehensive summary of the experiment's findings, shedding light on the system's performance under the unique conditions presented by the stroke patient's limitations.digit.

The Hand Tracking Modules effectively identify the anatomical features of the hand, as demonstrated by the display of a circle outlining its location. Within this depiction, 26 distinct points delineate key landmarks on the hand. In the figure 4 below, the operational dynamics of the mouse are illustrated. The system captures the nuances of mouse movements, including actions such as moving, holding, stopping, and zooming. This graphical representation provides a visual insight into the seamless interaction between the hand tracking modules and the corresponding functionalities of the virtual mouse system.

Fig 4: Hand Tracking, Mouse Movement & Mouse Zooming Recognition

Fig.5 (a). Virtual Keyboard

Fig.5 (b) Virtual Keyboard Typing

Figure 5 above depicts the output of the virtual keyboard, showcasing the result of typing a word. Subsequent to the input, the typed word is visibly displayed beneath the screen. This functionality extends to various applications within the system, enabling the use of words in software such as Word, Notepad, and other platforms where textual input is required. The virtual keyboard seamlessly integrates with different software applications, providing a versatile solution for word input across a spectrum of contexts.

5. Conclusion

The keyboard and mouse serve as integral components of a computer system. This system, designed for individuals with physical limitations, provides a means for paralyzed individuals to access and operate computers. By leveraging a virtual system, users can interact with the computer without the need for a physical keyboard and mouse. This innovation has the potential to redefine the landscape of human-computer interaction, eliminating the necessity for direct physical contact with computers. The incorporation of OpenCV into the system has proven to be highly successful, demonstrating precise control over mouse and keyboard functionalities. This advancement holds particular promise for individuals who lack control over their limbs. Impressively, the system achieves an accuracy rate of approximately 99%, showcasing its reliability. The Virtual Mouse module within the system emulates mouse functions by interpreting hand gestures, including counting cutlet figures, enabling six distinct actions such as left and right clicks, as well as movements in different directions. Additionally, a virtual keyboard facilitates word codification, providing a text input solution where a physical keyboard is not viable. The trials for the virtual keyboard involved consistent room arrangement and lighting conditions, ensuring consistency in performance. This innovative system, which allows for computer operation without a physical keyboard or mouse, stands as a testament to the potential of technology to enhance accessibility and inclusivity for individuals facing physical challenges.

References

1. Abedin, Md.Z., Nath, A.C., Dhar, P., Deb, K., Hossain, M.S., 2017. License plate recognition system based on contour properties and deep learning model, in: 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). Presented at the 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), pp. 590–593. https://doi.org/10.1109/R10-HTC.2017.8289029
2. Ahmed, T.U., Hossain, S., Hossain, M.S., Ul Islam, R., Andersson, K., 2019. Facial Expression Recognition using Convolutional Neural Network with Data Augmentation, in: 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR). Presented at the 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), IEEE, Spokane, WA, USA, pp. 336–341. https://doi.org/10.1109/ICIEV.2019.8858529
3. Ahmed, T.U., Jamil, M.N., Hossain, Mohammad Shahadat, Andersson, K., Hossain, Mohammed Sazzad, 2020. An Integrated Real-Time Deep Learning and Belief Rule Base Intelligent System to Assess Facial Expression Under Uncertainty, in: 2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR). Presented at the 2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR), IEEE, Kitakyushu, Japan, pp. 1–6. https://doi.org/10.1109/ICIEVicIVPR48672.2020.9306622
4. Asad, M.-U., Mustafa, R., Hossain, M., 2019. An Efficient Strategy for Face Clustering use in Video Surveillance System. https://doi.org/10.13140/RG.2.2.31972.42882
5. Bhuvana, D.S., Ashwin, E., Boopathi, R., Victor, A.D., 2017. Virtual Keyboard Interaction with System Based on Webcam. Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol. 2, 445–449. https://doi.org/10.32628/CSEIT1722157

6. Da Silva, R.A., Veiga, A.C.P., 2020. Algorithm for decoding visual gestures for an assistive virtual keyboard. IEEE Lat. Am. Trans. 18, 1909–1916. https://doi.org/10.1109/TLA.2020.9398632
7. Gupta, D., Hossain, E., Hossain, Mohammed Sazzad, Hossain, Mohammad Shahadat, Andersson, K., 2021. An Interactive Computer System with Gesture-Based Mouse and Keyboard, in: Vasant, P., Zelinka, I., Weber, G.-W. (Eds.), Intelligent Computing and Optimization, Advances in Intelligent Systems and Computing. Springer International Publishing, Cham, pp. 894–906. https://doi.org/10.1007/978-3-030-68154-8_76
8. Gupta, D., Hossain, E., Hossain, M.S., Andersson, K., Hossain, S., 2019. A Digital Personal Assistant using Bangla Voice Command Recognition and Face Detection, in: 2019 IEEE International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON). Presented at the 2019 IEEE International Conference on Robotics, Automation, Artificial-intelligence and Internet-of-Things (RAAICON), IEEE, Dhaka, Bangladesh, pp. 116–121. https://doi.org/10.1109/RAAICON48939.2019.47
9. Hsieh, C.-C., Liou, D.-H., Lee, D., 2010. A real time hand gesture recognition system using motion history image, in: 2010 2nd International Conference on Signal Processing Systems. Presented at the 2010 2nd International Conference on Signal Processing Systems, pp. V2-394-V2-398. https://doi.org/10.1109/ICSPS.2010.5555462
10. Islam, Md.Z., Hossain, M.S., Ul Islam, R., Andersson, K., 2019. Static Hand Gesture Recognition using Convolutional Neural Network with Data Augmentation, in: 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR). Presented at the 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), IEEE, Spokane, WA, USA, pp. 324–329. https://doi.org/10.1109/ICIEV.2019.8858563
11. Kumari, P., Singh, S., Pasi, V.K., 2016. Cursor Control using Hand Gestures. IJCA Proc. Recent Trends Future Prospect. Eng. Manag. Technol. RTFEM 2016, 25–29.
12. Mj, J., 2018. Finger recognition and gesture based augmented keyboard.
13. Patil, M.I.D., Lambhate, P.P.D., 2015. Virtual Keyboard Interaction Using Eye Gaze and Eye Blink. Int. J. Recent Innov. Trends Comput. Commun. 3, 4849–4852. https://doi.org/10.17762/ijritcc.v3i7.4750
14. Shibly, K., Dey, S., Islam, Md.A., Showrav, S., 2019. Design and Development of Hand Gesture Based Virtual Mouse. https://doi.org/10.1109/ICASERT.2019.8934612
15. Shriram, S., Nagaraj, B., Jaya, J., Shankar, S., Ajay, P., 2021. Deep Learning-Based Real-Time AI Virtual Mouse System Using Computer Vision to Avoid COVID-19 Spread. J. Healthc. Eng. 2021, e8133076. https://doi.org/10.1155/2021/8133076
16. Singh, J., Goel, Y., Jain, S., Yadav, S., 2022. Virtual Mouse And Assistant: A Technological Revolution Of Artificial Intelligence. J. Pharm. Negat. Results 13.