Hand Gesture Control System for Basic PC Features

Avay Kushwaha, K. C. Nabin Kumar, Aabhash Manandhar, Abhash Khanal, and Vani Vasudevan

Abstract Hand gesture control system is the most demanding needs in the today's world. It is one of the important means of human and computer interaction as of today. As an important part of non-verbal communication, hand gestures play an important role in our daily life. Hand gesture recognition control systems offer an innovative, natural, and user-friendly way to interact with computers. Analyzing and studying various segmentation and tracking, feature extraction, and recognition techniques, in this research paper, we have introduced an approach to recognize the hand gestures and control of different system settings and applications with the help of gestures. This paper provides the overview of systematic development of the system control using hand gesture recognition system. Hand detection model has been developed and made with the help of the computer vision library, OpenCV and Media pipe; using this hand detection system as a module, we can add specific number of gestures as different modules that will be able to control different system settings. These modules can be linked with hand detection model for controlling the system settings. The main aim of our project is to make human–computer interactions as smooth as possible such that the requirement of physical devices like keyboard, mouse, etc., as input system might not even be necessary. Therefore, overview of hand detection module and description about how system control using gestures have been implemented is described.

Keywords Hand gesture tracking · Skeleton-based recognition · Geometric attributes

A. Kushwaha · K. C. Nabin Kumar · A. Manandhar · A. Khanal · V. Vasudevan (✉)
Computer Science Engineering, Nitte Meenakshi Institute of Technology (Visvesvaraya Technological University), Bengaluru, India
e-mail: vani.v@nmit.ac.in

A. Kushwaha
e-mail: 1nt19cs042.abhay@nmit.ac.in

A. Manandhar
e-mail: 1nt19cs004.aabhash@nmit.ac.in

A. Khanal
e-mail: 1nt19cs007.abhash@nmit.ac.in

N. R. Shetty et al. (eds.), *Advances in Computing and Information*, Lecture Notes in Electrical Engineering 1104, https://doi.org/10.1007/978-981-99-7622-5_40

1 Introduction

Hand gesture recognition system refers to the movement of the hand while a gesture refers to any certain movement of the hands, arms, or any movement of anything. Hand gestures are a matter of frame language that may be conveyed by the center of the palm, the role of finger, and the form created by the hand [1]. Hand gestures may be categorized into both fixed and robust [1]. As its call implies, the static gesture refers back to the stable form of the hand as shown in determine 1, while the dynamic gesture comprises a chain of hand actions along with waving [1]. Collection of the hand gestures is done from the webcam of a PC, or any external camera is connected to a device and with different machine learning algorithms for implementation of the project in the real-life environment. Gesture recognition can be the way for the computer and robot to understand the human interaction and perform different operations as authorized by the program. It can be used to operate different machineries, robots, computers, etc. The gestures asked are converted into commands to the device through the computer vision. Hand gestures offer an exciting area of research as they facilitate communication and provide a means of natural interaction that can be used in a variety of applications [1]. The crucial aim of constructing hand gesture recognition gadget is to create an herbal interaction among human and computers in which the diagnosed gestures may be used for controlling a robot or conveying meaningful information [2]. Gestures can be static (posture or sure pose) which require less computational complexity or dynamic (series of postures) which are greater complex however appropriate for actual time surroundings.

2 Related Works

Table 1 shows the summary of the systems using skeleton-based recognition for hand gesture applications. The Intel Real depth-sensing camera has high segmentation accuracy as it is not exaggerated by light fluctuations or patterned backgrounds. However, the major problem is in the range of recognition [1]. It had an average of 84% recognition rate with maximum 91% and too expensive to work with. The Kinect V1 sensor is an outdated version of the V2 and has a fixed system that provides the feedback information from the depth sensor as metadata that provides information about the joint coordinates of the human body [1]. Kinect V1 provides skeletal joint tracking information for up to 20 joints to assist hand skeletal alignment [1]. The Kinect V2 sensor is capable to track 25 joints and up to six people concurrently with full joint tracking [1] with a detection range of 0.5–4.5 meters [1]. While the V1 has an average recognition rate of only 80%, the V2 has a rate of 95% with number of gestures taking being 14 and 26, respectively. The RGB video sequence records a video or series of pictures in fast intervals and can detect gestures broader than other models since others only work with static images. But this model has a very complex working as it is being close to that of working of a dynamic recognition

Table 1 Series of systems using skeleton-based recognition for hand gesture applications

	Cameras used	Methods	Extraction feature	Algorithm	Recognition rate (%)	No. of gestures
[6]	Intel Real Sense depth camera	Skeleton data	Hand skeletal joints' positions	Convolutional neural network (CNN)	91.28 84.35	14 gestures 28 gestures
[7]	Kinect camera	Laplacian-based contraction	Skeletal points cloud	Hungarian algorithm	80	12 gestures
[8]	RGB video sequence recorded	Vision-based approach and skeletal data	Hand and body skeletal features	Skeleton classification network	88.24	14 gestures
[9]	Kinect V2 camera sensor	Depth metadata	Dynamic hand gesture	SVM	95.42	10 gestures
	Proposed project	Skeletal data	Hand skeletal joints' positions	–	97.149	–

system. The accurate recognition rate for this model could not be calculated as it is being a dynamic system and it took in about 14 gestures. The proposed project uses a webcam to capture a static image and uses a skeletal recognition system to detect hand gestures. Now, according to the paper [1], recognition rate of this system, i.e., webcam using skeleton system is 95.61%. Taking the same research paper as a reference and also using accuracy score method, we roughly got the accuracy of 97.149% and furthermore is explained under result section which has greater recognition rate and accuracy than the one available in today's market.

3 Hand Gesture Methods

Figure 1 shows the types of hand gesture recognition tools with naked hand and hand with gloves.

3.1 Hand Gesture Using Instrumented Glove

In this method, sensors are attached to the gloves (Fig. 2) to receive the data. This method requires the user to be physically connected to the computer, that makes interaction between the user and the computer difficult [1]. Also, the price of these

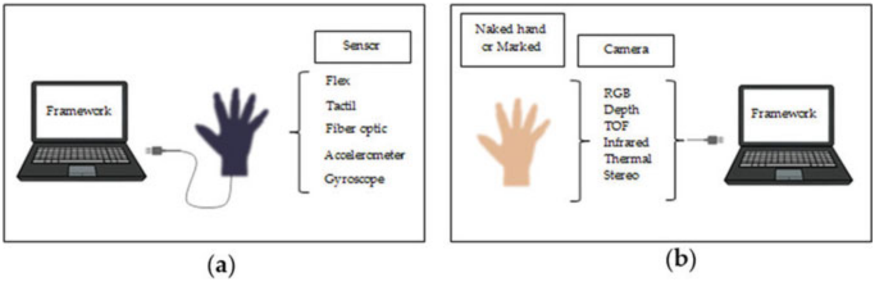

Fig. 1 Types of gesture recognition tools

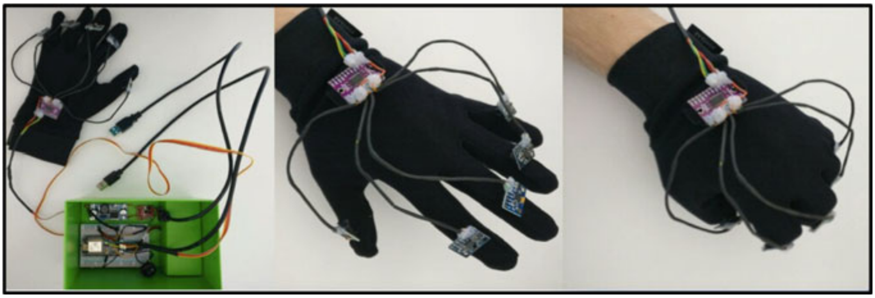

Fig. 2 Sensor-based instrumental glove

devices is extremely high. The gloves system relies on touch technology, employing input devices such as sensors, gloves, and Arduino components. In contrast, the hand gesture system aims to create a contactless interface, enhancing effective communication between humans and computers without physical touch [1].

3.2 Hand Gesture Based on Computer Vision

Sensors based on video cameras are a common and applicable method as they enable zero contact communication between human beings and machine, i.e., computers [1]. A variety of camera configurations are available, including monocular, fisheye, TOF, and IR. However, there are many problems with this method: lighting changes, background issues, complex backgrounds, or background objects with the same skin color. Figure 3 shows the hand gesture recognition by computer vision (CV).

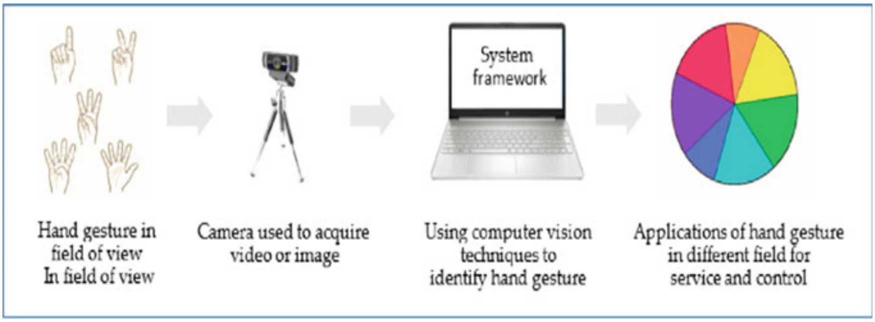

Fig. 3 Recognition by computer vision

3.2.1 Color-Based Recognition

Color Glove

This technique uses camera that detects the movements of the glove-fitted human hand which is divided into different parts and colored differently. These different colors of the glove (Fig. 4) allow the camera to track and determine the position of the palm and fingers, thus extracting a geometric model of the hand [1]. This method was used to interact with 3D models, and actions like zoom-in, zoom-out, movement detection, sketching, and content writing can be done flexibly using a virtual keyboard [1]. The computer imaginative and prescient device for monitoring and identifying the hand positions that manage the menus is based totally on a combination of multi-scale coloration function detection; view-based hierarchical hand fashions and particle filtering [3]. Hand poses or states are represented as a hierarchy of multi-scale silhouette features at different scales using qualitative correlations regarding scale, position, and orientation. Multi-scale shading feature detection is performed on each image. Hand poses are then simultaneously detected and tracked using particle filtering using an extension of layered sampling called hierarchical layered sampling.

Based on Skin Color

Skin coloration detection is a popular method for hand segmentation. Skin color determination may be executed the usage of techniques. The primary technique is pixel primarily based skin detection, where every pixel in a photo is differentiated from neighboring pixels by categories if whether or not it is skin [1]. The other technique is skin region identification, which spatially processes skin pixels primarily based on data which includes depth and texture [1]. It is used in an extensive range of programs along with item category, broken photos recovery, human movement

Fig. 4 Colored gloves

tracking, facial recognition, video surveillance, hand segmentation, HCI package, and gesture recognition [1].

3.2.2 Skeleton-Based Recognition

Skeleton-based recognition system specifies model parameters that can improve recognition of multiplex features [1]. When different representations of skeletal data for hand models are available for classification, easily transform data features and correlations to describe geometrical properties and constraints and focus on geometrical and statistical features. The most commonly used features are the orientation of joints.

The distance between joints, the location of skeletal joints, and the angles between joints and trajectories and curvatures of joints within the skeletal structure identifies key patterns vital for accurate hand gesture recognition allowing recognition systems to precisely classify gestures. Gesture recognition based on hand skeletal tracking is becoming the primary interaction method for next-generation virtual reality (VR) and mixed reality (MR) devices such as the Oculus Quest and Microsoft HoloLens enabling unique and advanced solutions [4]. You can use the Kinect camera's depth sensor to do hand segmentation, then use 3D connection for fingertip position, Euclidean distance, and skeletal hand pixel geodesic distance for better accuracy [1]. There are further different methods under this skeleton-based approach. Some of them use Intel Real Sense depth camera, Kinect camera, RGB video sequence recorded, and Kinect v2 camera sensor.

4 Proposed System

Hand recognition system is one of the most demanding concepts in the growing industry and human computer interaction field. Many projects have used motion detecting sensor gloves, color-based recognition using glove marker, or even infrared cameras as their main interface to communicate and input the gesture commands to the computer. In this project, we are trying to make the system more interactive by feeding the input through direct real-time image capturing, detecting the hand gestures, and then assigning those hand gestures to certain system specific functions which also works in many applications using just the web camera of a typical laptop. We have used the OpenCV and the media pipe to prepare the hand recognition module. As the first phase of our project, we have developed this module that tracks the hand movements, and based on this module, the application recognizes the palm and fingers accordingly. OpenCV is the computer vision library which is used in the real-time computer vision. On the other hand, media pipe is the hand and the finger tracking library or solution where it employs 21 3D landmarks coordinates in the hand (Fig. 5). Using these two libraries, we can construct the hand detection module.

On a successful construction of the hand recognition module, this module now can link with the control modules which we have provided into the application as system controlling module that uses gestures for functionalities. These modules can link with the different system settings accordingly to control the basic features of the operating system. So far, the application can control system settings such as changing the volume of the system. We have done so such that there is also an extra function where we can give an additional gesture that can lock the functionality. In this case, we can have an extra gesture that sets the volume in a certain value so that it is not constantly changed as we move our hand away from the view of camera. The hand detection/recognition module is taken as the class, then imported to control modules for the further linking of control gestures. The volume module traces the distance between the hand landmark 4 and 8 in a frame and increases the volume accordingly and the volume set works in such way that if the hand landmark 20 overlaps or is near to the 17, it locks the volume.

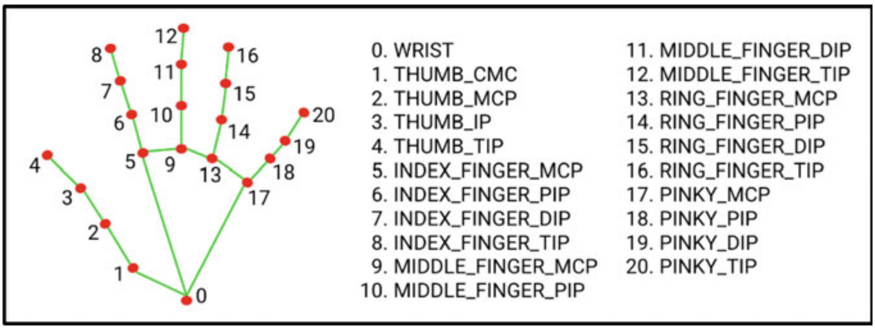

Fig. 5 Hand landmarks using media pipe

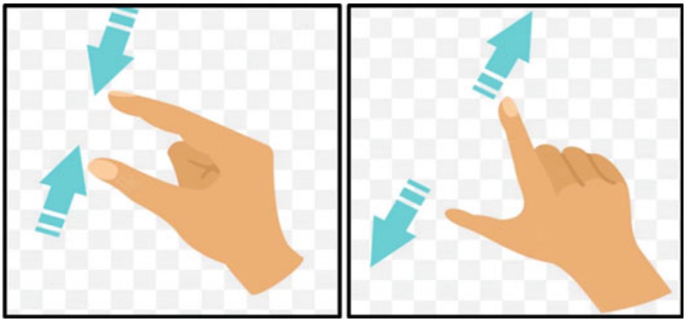

Fig. 6 Gesture to control the volume of the PC

For controlling the volume of the system (Fig. 6), we have used the Pycha library which is an inbuilt function. With the help of this library, we can control the volume function of the system by using gesture commands. Importing other basic modules to control the basic functions of operating system with the help of different gestures is very much possible in the future additions to this application. Hand detection system is the main module through which we can add in different other modules to control the system features accordingly.

5 Results and Implications

5.1 Hand Recognition Module

The camera was able to detect the hand and the movements (Fig. 7) with ease. Along with hand detection, very specific palm detection was also successful along with the identification of hand landmarks. Now, the system can be confirmed to be accurate because the code uses the library of Google's media pipe which is trained with more than 3800 + images for hand detection, and by the literature survey above, we can conclude that the skeleton system used in this project has higher accuracy and recognition rate compared to the other system until and unless it uses super expensive external hardware.

Also, to make sure that we do not just rely on the previously used libraries, we use an accurate-score method to calculate the accuracy where we have taken 77 samples of hands with seven different people's hands. Out of 77 hands, we got 73 true-positive, 1 false-positive, 1 false-negative, and 2 true-negative, resulting in the accuracy rate of approximately 97.149% by the formulae, i.e., accuracy = (TP + TN)/TP + TN + FN + FP. This is just a rough calculation done using bunch of samples and compared with the results of the paper [1] which turns out to be similar in result. Also, the Google's media pipe claims to have achieved 95.7% of accuracy

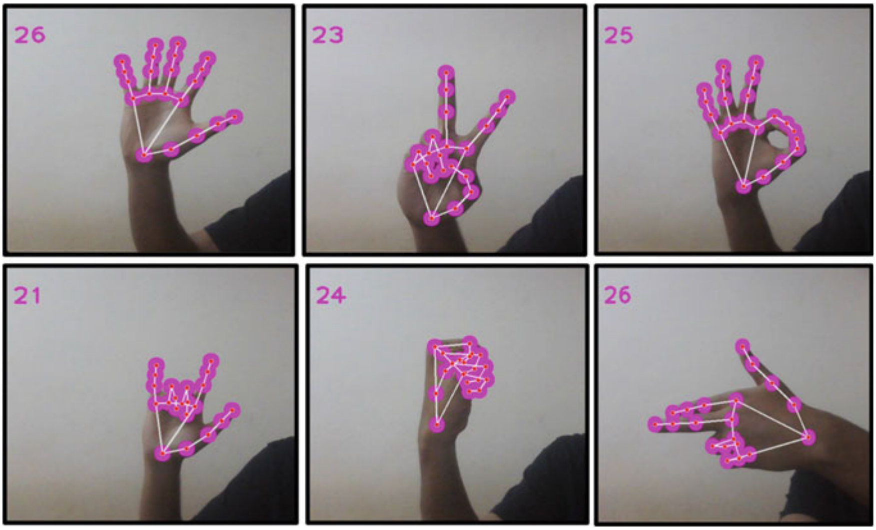

Fig. 7 Hand recognition module

with hand recognition system which rounds off to recognition rate of 95% + at the best-case scenarios, i.e., proper lighting, and clear, sharp camera lens.

5.2 Volume Control Module

The recognition of gestures by detecting and using the hand movements has been successful with a few modules. Modules like controlling the volume (Fig. 8) of the system and locking the functioning have been added and successfully implemented. We are able to link the control modules with the main hand recognition module, which simply means that we will be able to link more control systems as per the requirements. The volume control module had an issue of fluctuation in the volume when the hands were moved after adjusting the volume. This issue led us to integrate an extra function of locking the volume using a new gesture, i.e., adjusting a volume using thumb and forefinger, then using pinky finger to lock the volume at that adjustment.

Since we do not have any solid dataset and records about this particular module, we had about 250 hand samples to test if the accuracy rate of the volume module is as good as the hand recognition module. Fifty samples each of five different person's hand were taken and tested if we get the similar output and optimized results. Both the worst-case and best-case scenarios were tested. Out of 250 samples, we could successfully adjust the volume perfectly in 235 samples, whereas the other 15 samples were unable to do so due to the anti-light and hands that were out of the reach of camera. And 220 out of 250 samples were able to pull out the volume set functionality

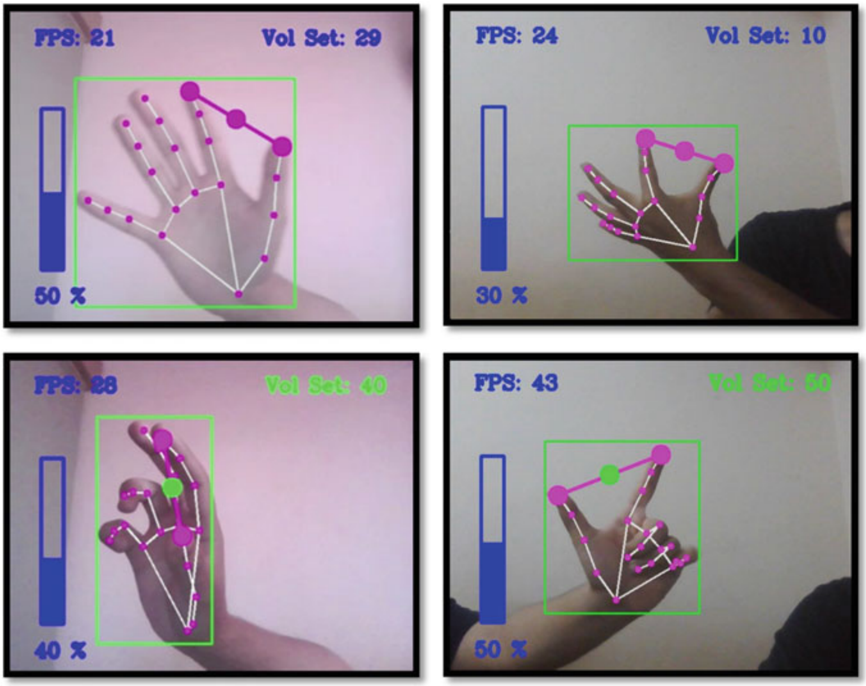

Fig. 8 Volume control module

perfectly, while the other 30 samples could not do so because of the improper gesture movement by the user which provides us with 94 and 88% of accuracy, respectively. Figure 8 shows the example of the samples taken for the test cases.

6 Gap Analysis

With lot of different approaches and models for hand recognition, there comes separate challenges with each model. For example, using an instrumented glove recognition system, the model is built in such a way that it must be physically connected to the system for it to work. Also, the parts used cannot be cost-effective. In colored based models, especially using colored gloves or skin-recognized models, it has its own set of challenges such as the system not detecting the instruments properly. Analyzing these kinds of challenges, we concluded that the skeleton-based model tends to be the better model for hand recognition with high accuracy and effectiveness. Skeleton-based models take the physical skeletal models of the hand which describe geometric attributes and constraints, transform feature data correlations, and focus on geometric and statistical features [1]. Also, this model uses a webcam as its input device unlike other models which require external hardware such as Kinect camera, sensors, gloves which might cost more, in process nullifying the cost-effectiveness of the project [5]. The hand gesture system which we use is more efficient than the other skeleton-based hand gesture recognition system in the market as the gesture recognition system using the other recognition system has the efficiency rate less than 95% and our hand gesture recognition system has the efficiency rate more than 95% which is also tested and provided with the library of media pipe [1]. Now, the major drawback of the systems available in the market is that they are way more expensive than what we are building and also does not have an ecosystem that holds all the different functionalities in the same system. The systems that use external hardware and super expensive production cost might in some case perform better than what we have built so far, but they do not provide such cost-effective system and also do not have any product that is specifically made for controlling the operating systems of PC. The system which we have developed is uniquely designed only for the PC and laptop which is compatible for only Windows OS. The main goal of our project is to control the media and features of the PC. Our solo goal was to focus on building the application that is compatible for the laptop that have Windows OS which provide the contactless input interaction between human and the computer. So the meaning of the system is uniquely designed for PC only is it is compatible for PC only which consist of Windows OS and only control the features of the PC media and other features such as volume, cursor movement etc.

7 Future Work

The proposed system is so far able to recognize the hand and control and set the volume of the PC. Since this gesture control system is integrated within the operating system and the PC's default mechanism, no matter what applications are being used it is able to control the applications as well. As said above, we can link more control modules with different gestures assigned to them, and we will be working on to add the basic and most useful control commands in an operating system, i.e., mouse, screenshot, and brightness control. In successful linking of these control modules with the main module, we will be linking the whole project with a Windows app that will have an option to activate and deactivate the modules as per the needs. The app will be kept as a startup app which will run automatically when the system is booted up. This will be done so that the user does not have to run the app every time when they turn on the PC and also to overcome the disadvantage of the high startup time requirement. If the proposed work is successfully implemented with no lags and issue, the optimization will be done to decrease the compile time.

8 Conclusion

The developing technologies have always been about the easier, more flexible, cheaper, and scalable products. Working for this product made us learn so much about how the technologies are developing every single day but with also increment in the both production cost, high market price, and use of high-tech systems even after the availability of the cheaper, easier, and reliable systems. This project not only looks after the entertainment sector but also covers education, health cares, and ease in use of technology. The proposed work has both its advantages and disadvantages and might perform outstanding in best case and sometimes not as expected in worst case. Since the computers are everywhere nowadays, we are trying to build something that spreads throughout the computer world making it lot easier for users to know what future looks like and generate idea on how can this be applied in more applications by making it far better than what it is today.

References

1. Oudah M, Al-Naji A, Chahl J (2020) Hand gesture recognition based on computer vision: a review of techniques. J Imaging 6(8):73. https://doi.org/10.3390/jimaging6080073
2. Zaman Khan R (2012) Hand gesture recognition: a literature review. Int J Artif Intell Appl 3(4):161–174. https://doi.org/10.5121/ijaia.2012.3412
3. Mhatre A, Arutla KD, Desai A, Amoda N (2018) Hand gesture technology using colour markers for interactive HCI. Int J Eng Res Technol 2(4). https://doi.org/10.17577/IJERTCONV2IS04037
4. Caputo A, Giachetti A, Soso S et al (2021) SHREC 2021: skeleton-based hand gesture recognition in the wild. Comput Graph 99(0097–8493):201–211. https://doi.org/10.1016/j.cag.2021.07.007
5. Priyanka R, Sriram P, Jayasree LN, Gladston A (2021) Shape-based features for optimized hand gesture recognition. Int J Artif Intell Mach Learn 11(1):23–38. https://doi.org/10.4018/ijaiml.2021010103
6. Devineau G, Moutarde F, Xi W, Yang J (2018) Deep learning for hand gesture recognition on skeletal data. In: 2018 13th IEEE International conference on automatic face & gesture recognition (FG 2018), pp 106–113. https://doi.org/10.1109/FG.2018.00025
7. Jiang F, Wu S, Yang G, Zhao D, Kung SY (2014) Viewpoint-independent hand gesture recognition with Kinect. Signal Image Video Process 8(S1):163–172. https://doi.org/10.1007/s11760-014-0668-x
8. Konstantinidis D, Dimitropoulos K, Daras P (2018) Sign language recognition based on hand and body skeletal data. In: 2018—3DTV-conference: the true vision—capture, transmission and display of 3D video (3DTV-CON), pp 1–4. https://doi.org/10.1109/3DTV.2018.8478467
9. Chen Y, Luo B, Chen Y-L, Liang G, Wu X (2015) A real-time dynamic hand gesture recognition system using kinect sensor. In: 2015 IEEE International conference on robotics and biomimetics (ROBIO), pp 2026–2030. https://doi.org/10.1109/ROBIO.2015.7419071