# **Hand Gesture Recognition Based on Computer Vision**

# Niranjan Maurya

Masters of Computer Applications

Thakur Institute of Management Studies, Career,

Development & Research (TIMSCDR)

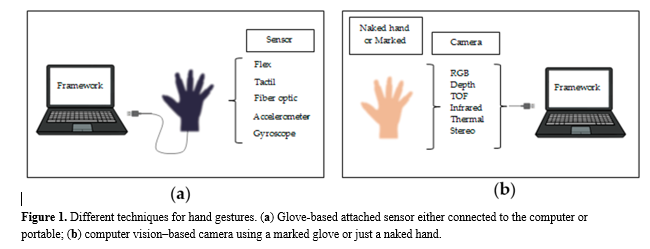
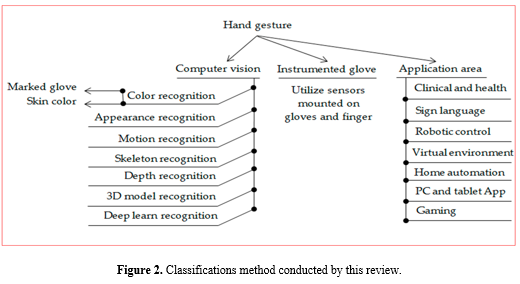
Mumbai, [niranjanmaurya03@gmail.com](mailto:niranjanmaurya03@gmail.com)

**Abstract:** Hand gestures are a form of nonverbal communication that can be used in several fields such as communication between deaf-mute people, robot control, human–computer interaction (HCI), home automation and medical applications. Research papers based on hand gestures have adopted many different techniques, including those based on instrumented sensor technology and computer vision. In other words, the hand sign can be classified under many headings, such as posture and gesture, as well as dynamic and static, or a hybrid of the two. This paper focuses on a review of the literature on hand gesture techniques and introduces their merits and limitations under different circumstances. In addition, it tabulates the performance of these methods, focusing on computer vision techniques that deal with the similarity and difference points, technique of hand segmentation used, classification algorithms and drawbacks, number and types of gestures, dataset used, detection range (distance) and type of camera used. This paper is a thorough general overview of hand gesture methods with a brief discussion of some possible applications.

**Keywords:** hand gesture; hand posture; computer vision; human–computer interaction (HCI).

## **1. Introduction**

Hand gestures are an aspect of body language that can be conveyed through the center of the palm, the finger position and the shape constructed by the hand. Hand gestures can be classified into static and dynamic. As its name implies, the static gesture refers to the stable shape of the hand, whereas the dynamic gesture comprises a series of hand movements such as waving. There are a variety of hand movements within a gesture; for example, a handshake varies from one person to another and changes according to time and place. The main difference between posture and gesture is that posture focuses more on the shape of the hand whereas gesture focuses on the hand movement. The main approaches to hand gesture research can be classified into the wearable glove-based sensor approach and the camera vision-based sensor approach.

Hand gestures offer an inspiring field of research because they can facilitate communication and provide a natural means of interaction that can be used across a variety of applications. Previously, hand gesture recognition was achieved with wearable sensors attached directly to the hand with gloves. These sensors detected a physical response according to hand movements or finger bending. The data collected were then processed using a computer connected to the glove with wire. This system of glove-based sensor could be made portable by using a sensor attached to a microcontroller.

As illustrated in Figure 1, hand gestures for human–computer interaction (HCI) started with the invention of the data glove sensor. It offered simple commands for a computer interface. The gloves used different sensor types to capture hand motion and position by detecting the correct coordinates of the location of the palm and fingers. Various sensors using the same technique based on the angle of bending were the curvature sensor, angular displacement sensor, optical fiber transducer, flex sensors and accelerometer sensor. These sensors exploit different physical principles according to their type.

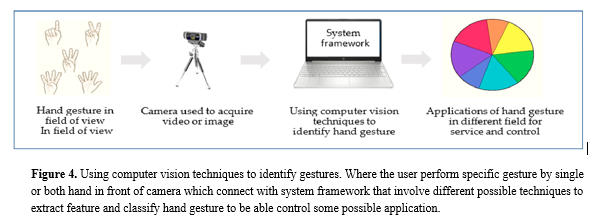
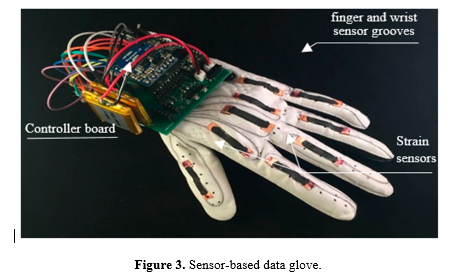
These drawbacks led to the development of promising and cost-effective techniques that did not require cumbersome gloves to be worn. These techniques are called camera vision-based sensor technologies. With the evolution of open-source software libraries, it is easier than ever to detect hand gestures that can be used under a wide range of applications like clinical operations, sign language, robot control, virtual environments, home automation, personal computer and tablet, gaming. These techniques essentially involve replacement of the instrumented glove with a camera. Different types of camera are used for this purpose, such as RGB camera, time of flight (TOF) camera, thermal cameras or night vision cameras.

Algorithms have been developed based on computer vision methods to detect hands using these different types of cameras. The algorithms attempt to segment and detect hand features such as skin color, appearance, motion, skeleton, depth, 3D model, deep learn detection and more.

## **2. Hand Gesture Methods**

The primary goal in studying gesture recognition is to introduce a system that can detect specific human gestures and use them to convey information or for command and control purposes. Therefore, it includes not only tracking of human movement, but also the interpretation of that movement as significant commands. Two approaches are generally used to interpret gestures for HCI applications. The first approach is based on data gloves (wearable or direct contact) and the second approach is based on computer vision without the need to wear any sensors.

### 2.1. Hand Gestures Based on Instrumented Glove Approach

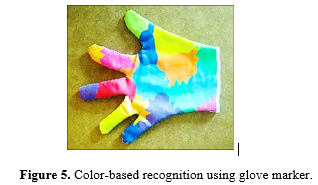
The wearable glove-based sensors can be used to capture hand motion and position. In addition, they can easily provide the exact coordinates of palm and finger locations, orientation and configurations by using sensors attached to the gloves. However, this approach requires the user to be connected to the computer physically, which blocks the ease of interaction between user and computer. In addition, the price of these devices is quite high. However, the modern glove based approach uses the technology of touch, which more promising technology and it is considered Industrial-grade haptic technology. Where the glove gives haptic feedback that makes user sense the shape, texture, movement and weight of a virtual object by using microfluidic technology. Figure 3 shows an example of a sensor glove used in sign language. [](https://physicsworld.com/a/smart-glovetranslates-sign-language-into-digital-text/)

### 2.2. Hand Gestures Based on Computer Vision Approach

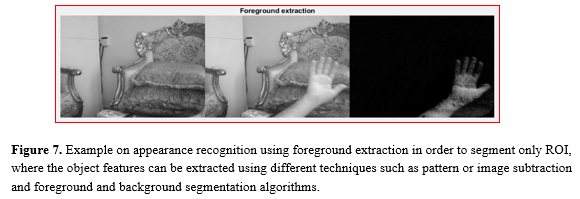
The camera vision based sensor is a common, suitable and applicable technique because it provides contactless communication between humans and computers. Different configurations of cameras can be utilized, such as monocular, fisheye, TOF and IR. However, this technique involves several challenges, including lighting variation, background issues, the effect of occlusions, complex background, processing time traded against resolution and frame rate and foreground or background objects presenting the same skin color tone or otherwise appearing as hands. These challenges will be discussed in the following sections. A simple diagram of the camera vision-based sensor for extracting and identifying hand gestures is presented in Figure 4.

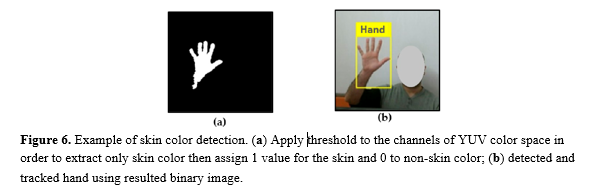
**2.2.1. Color-Based Recognition:**

Color-Based Recognition Using Glove Marker

This method uses a camera to track the movement of the hand using a glove with different color marks. This method has been used for interaction with 3D models, permitting some processing, such as zooming, moving, drawing and writing using a virtual keyboard with good flexibility. The colors on the glove enable the camera sensor to track and detect the location of the palm and fingers, which allows for the extraction of geometric model of the shape of the hand. The advantages of this method are its simplicity of use and low price compared with the sensor data glove. However, it still requires the wearing of colored gloves and limits the degree of natural and spontaneous interaction with the HCI. The color-based glove marker is shown in Figure 5. 

Color-Based Recognition of Skin Color

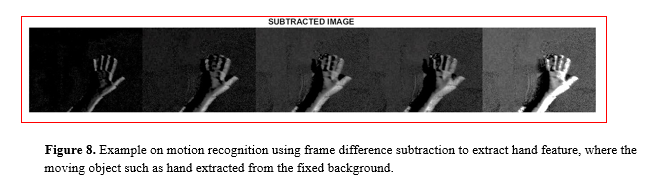
Skin color detection is one of the most popular methods for hand segmentation and is used in a wide range of applications, such as object classification, degraded photograph recovery, person movement tracking, video observation, HCI applications, facial recognition, hand segmentation and gesture identification. Skin color detection has been achieved using two methods. The first method is pixel based skin detection, in which each pixel in an image is classified into skin or not, individually from its neighbor. The second method is region skin detection, in which the skin pixels are spatially processed based on information such as intensity and texture.

Color space can be used as a mathematical model to represent image color information. Several color spaces can be used according to the application type such as digital graphics, image process applications, TV transmission and application of computer vision techniques. Figure 6 shows an example of skin color detection using YUV color space.

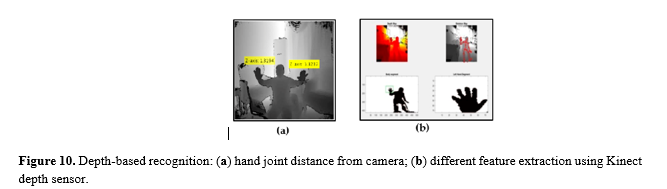
**2.2.2. Appearance-Based Recognition**

This method depends on extracting the image features in order to model visual appearance such as hand and comparing these parameters with feature extracted from the input image frames. Where the features are directly calculated by the pixel intensities without a previous segmentation process. The method is executed in real time due to the easy 2D image features extracted and is considered easier to implement than the 3D model method. In addition, this method can detect various skin tones. Utilizing the Ada Boost learning algorithm, which maintains fixed feature such as key points for a portion of a hand, which can solve the occlusion issue, it can separate into two models: a motion model and a 2D static model. Figure 7 below show simple example on appearance recognition.

**2.2.3. Motion-Based Recognition**

Motion-based recognition can be utilized for detection purposes; it can be extracts the object through a series of image frames. The AdaBoost algorithm utilized for object detection, characterization, movement modeling, and pattern recognition is needed to recognize the gesture. The main issue encounter motion recognition is this is an occasion if one more gesture is active at the recognition process and also dynamic background has a negative effect. In addition, the loss of gesture may be caused by occlusion among tracked hand gesture or error in region extraction from tracked gesture and effect long-distance on the region appearance. Figure 8 gives simple example on hand motion recognition.

**2.2.4. Skeleton-Based Recognition**

The skeleton-based recognition specifies model parameters which can improve the detection of complex features. Where the various representations of skeleton data for the hand model can be used for classification, it describes geometric attributes and constraint and easy translates features and correlations of data, in order to focus on geometric and statistic features. The most common feature used is the joint orientation, the space between joints, the skeletal joint location and degree of angle between joints and trajectories and curvature of the joints.

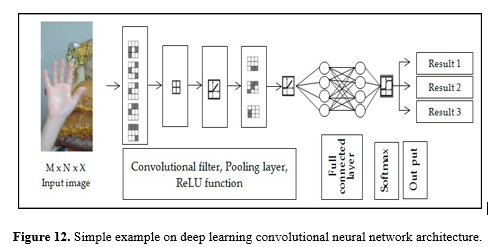
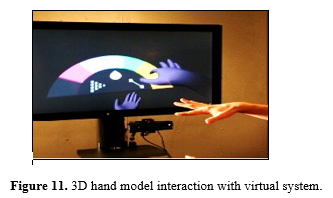
Hand segmentation using the depth sensor of the Kinect camera, followed by location of the fingertips using 3D connections, Euclidean distance, and geodesic distance over hand skeleton pixels to provide increased accuracy. Figure 9 show an example on skeleton recognition.

**2.2.5. Depth-Based Recognition**

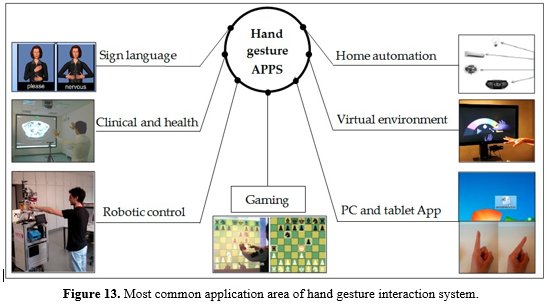
Approaches have proposed for solving hand gesture recognition using different types of cameras. A depth camera provides 3D geometric information about the object. Previously, both major approximations were utilized: TOF precepts and light coding. The 3D data from a depth camera directly reflects the depth field if compared with a color image which contains only a projection. Using this approach, the lighting, shade, and color did not affect the result image. However, the cost, size and availability of the depth camera will limit its use.

The finger earth mover’s distance (FEMD) approach was evaluated in terms of speed and precision, and then compared with the shape-matching algorithm using the depth map and color image acquired by the Kinect camera. Figure 10 shows an example of segmentation using Kinect depth sensor.

**2.2.6. 3D Model-Based Recognition**

The 3D model essentially depends on 3D Kinematic hand model which has a large degree of freedom, where hand parameter estimation obtained by comparing the input image with the two-dimensional appearance projected by three-dimensional hand model. In addition, the 3D model introduces human hand feature as pose estimation by forming volumetric or skeletal or 3D model that identical to the user’s hand. Where the 3D model parameter updated through the matching process. Where the depth parameter is added to the model to increase accuracy. Figure 11 shows an example of a 3D hand model interaction with virtual system.

There are some reported limitations, such as 3D hand required a large dataset of images to formulate the characteristic shapes of the hand in case multi-view. Moreover, the matching process considers time consumption, also computation costly and less ability to treat unclear views

**2.2.7. Deep-Learning Based Recognition**

The artificial intelligence offers a good and reliable technique used in a wide range of modern applications because of using a learning role principle. The deep learning used multilayers for learning data and gives a good prediction out result. The most challenges facing this technique is required dataset to learn algorithm which may affect time processing.

Figure 12 below shown example on deep learn convolution neural network.

**3. Application Areas of Hand Gesture Recognition Systems**

Research into hand gestures has become an exciting and relevant field; it offers a means of natural interaction and reduces the cost of using sensors in terms of data gloves. Conventional interactive methods depend on different devices such as a mouse, keyboard, touch screen, joystick for gaming and consoles for machine controls. The following sections describe some popular applications of hand gestures. Figure 13 shows the most common application area deal with hand gesture recognition techniques.

**3.1. Clinical and Health**

During clinical operations, a surgeon may need details about the patient’s entire body structure or a detailed organ model in order to shorten the operating time or increase the accuracy of the result. This is achieved by using a medical imaging system such as MRI, CT or X-ray system, which collects data from the patient’s body and displays them on the screen as a detailed image. The surgeon can facilitate interaction with the viewed images by performing hand gestures in front of the camera using a computer vision technique. These gestures can enable some operations such as zooming, rotating, image cropping and going to the next or previous slide without using any peripheral device such as a mouse, keyboard or touch screen. Any additional equipment requires sterilization, which can be difficult in the case of keyboards and touch screen. In addition, hand gestures can be used for assistive purpose such as wheelchair control.

### 3.2. Sign Language Recognition

Sign language is an alternative method used by people who are unable to communicate with others by speech. It consists of a set of gestures wherein every gesture represents one letter, number or expression. Many research papers have proposed recognition of sign language for deaf-mute people, using a glove-attached sensor worn on the hand that gives responses according to hand movement. Alternatively, it may involve uncovered hand interaction with the camera, using computer vision techniques to identify the gesture. For both approaches mentioned above, the dataset used for classification of gestures matches a real-time gesture made by the user.

### 3.3. Robot Control

Robot technology is used in many application fields such as industry, assistive services, stores, sports and entertainment. Robotic control systems use machine learning techniques, artificial intelligence and complex algorithms to execute a specific task, which lets the robotic system, interact naturally with the environment and make an independent decision. Some research proposes computer vision technology with a robot to build assistive systems for elderly people. Other research uses computer vision to enable a robot to ask a human for a proper path inside a specific building.

### 3.4. Virtual Environment

Virtual environments are based on a 3D model that needs a 3D gesture recognition system in order to interact in real time as a HCI. These gestures may be used for modification and viewing or for recreational purposes, such as playing a virtual piano. The gesture recognition system utilizes a dataset to match it with an acquired gesture in real time.

### 3.5. Home Automation

Hand gestures can be used efficiently for home automation. Shaking a hand or performing some gesture can easily enable control of lighting, fans, television, radio, etc. They can be used to improve older people’s quality of life.

### 3.6. Personal Computer and Tablet

Hand gestures can be used as an alternative input device that enables interaction with a computer without a mouse or keyboard, such as dragging, dropping and moving files through the desktop environment, as well as cut and paste operations. Moreover, they can be used to control slide show presentations. In addition, they are used with a tablet to permit deaf-mute people to interact with other people by moving their hand in front of tablet’s camera. This requires the installation of an application that translates sign language to text, which is displayed on the screen. This is analogous to the conversion of acquired voice to text.

### 3.7. Gestures for Gaming

The best example of gesture interaction for gaming purposes is the Microsoft Kinect Xbox, which has a camera placed over the screen and connects with the Xbox device through the cable port. The user can interact with the game by using hand motions and body movements that are tracked by the Kinect camera sensor.

## **4. Research Gaps and Challenges**

From the previous sections, it is easy to identify the research gap, since most research studies focus on computer applications, sign language and interaction with a 3D object through a virtual environment. However, many research papers deal with enhancing frameworks for hand gesture recognition or developing new algorithms rather than executing a practical application with regard to health care. The biggest challenge encountered by the researcher is in designing a robust framework that overcomes the most common issues with fewer limitations and gives an accurate and reliable result. Most proposed hand gesture systems can be divided into two categories of computer vision techniques. First, a simple approach is to use image processing techniques via Open-NI library or OpenCV library and possibly other tools to provide interaction in real time, which considers time consumption because of real-time processing. This has some limitations, such as background issues, illumination variation, distance limit and multi-object or multi-gesture problems. A second approach uses dataset gestures to match against the input gesture, where considerably more complex patterns require complex algorithm. Deep learning technique and artificial intelligence techniques to match the interaction gesture in real time with dataset gestures already containing specific postures or gestures.

Although this approach can identify a large number of gestures, it has some drawbacks in some cases, such as missing some gestures because of the classification algorithms accuracy contrast. In addition, it takes time more than first approach because of the matching dataset in case of using a large number of the dataset. In addition, the dataset of gestures cannot be used by other frameworks.

## **5. Conclusions**

Hand gesture recognition addresses a fault in interaction systems. Controlling things by hand is more natural, easier, more flexible and cheaper, and there is no need to fix problems caused by hardware devices, since none is required. From previous sections, it was clear to need to put much effort into developing reliable and robust algorithms with the help of using a camera sensor has a certain characteristic to encounter common issues and achieve a reliable result. Each technique mentioned above, however, has its advantages and disadvantages and may perform well in some challenges while being inferior in others.

# **6. References**

1. Gupta, H.P.; Chudgar, H.S.; Mukherjee, S.; Dutta, T.; Sharma, K. A continuous hand gestures recognition technique for human-machine interaction using accelerometer and gyroscope sensors. *IEEE Sens. J.* **2016**, *16*, 6425–6432.
2. [Munir Oudah](https://sciprofiles.com/profile/1401383).; [Ali Al-Naji](https://sciprofiles.com/profile/706896).; [Javaan Chahl](https://sciprofiles.com/profile/1071415).; Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. [J. Imaging 2020](https://doi.org/10.3390/jimaging6080073.), 6(8), 73
3. Wachs, J.P.; Kölsch, M.; Stern, H.; Edan, Y. Vision-based hand-gesture applications. *Commun. ACM* **2011**, *54*, 60–71.
4. Nguyen, X.S.; Brun, L.; Lézoray, O.; Bougleux, S. A neural network based on SPD manifold learning for skeleton-based hand gesture recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12036–12045.
5. Malik, J.; Elhayek, A.; Stricker, D. Structure-aware 3D hand pose regression from a single depth image. In Proceedings of the International Conference on Virtual Reality and Augmented Reality, London, UK, 22–23 October 2018; pp. 3–17.
6. Taylor, J.; Bordeaux, L.; Cashman, T.; Corish, B.; Keskin, C.; Sharp, T.; Soto, E.; Sweeney, D.; Valentin, J.; Luff, B. Efficient and precise interactive hand tracking through joint, continuous optimization of pose and correspondences. *ACM Trans. Graph.*
7. Tekin, B.; Bogo, F.; Pollefeys, M. H+ O: Unified egocentric recognition of 3D hand-object poses and interactions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4511–4520.
8. Li, Y. Hand gesture recognition using Kinect. In Proceedings of the 2012 IEEE International Conference on Computer Science and Automation Engineering, Beijing, China, 22–24 June 2012; pp. 196–199.
9. Ma, X.; Peng, J. Kinect sensor-based long-distance hand gesture recognition and fingertip detection with depth information. *J. Sens.* **2018**, *2018*, 5809769.
10. Kim, M.-S.; Lee, C.H. Hand Gesture Recognition for Kinect v2 Sensor in the Near Distance Where Depth Data Are Not Provided. *Int. J. Softw. Eng. Its Appl.* **2016**, *10*, 407–418.
11. De Smedt, Q.; Wannous, H.; Vandeborre, J.-P.; Guerry, J.; Saux, B.L.; Filliat, D. 3D hand gesture recognition using a depth and skeletal dataset: SHREC’17 track. In Proceedings of the Workshop on 3D Object Retrieval, Lyon, France, 23–24 April 2017; pp. 33–38.