

# Human Computer Interaction – Hand Gesture Recognition

Riya Jain\*, Muskan Jain, Roopal Jain, Suman Madan

Jagan Institute of Management Studies, Sec-5, Rohini, Delhi, India



\* Corresponding Author email:

riyajain572000@gmail.com

## Article History

Received: 01 May 2021

Revised: 05 August 2021

Accepted: 07 August 2021

Published: 01 September 2021

## Student(s)

- Riya Jain
- Muskan Jain
- Roopal Jain

Academic Year: 2020-2021

Course Level: Bachelor

Course Name: BCA

Course year: 3rd Year

## Mentor(s)

- Dr. Suman Madan

## ABSTRACT

The creation of intelligent and natural interfaces between users and computer systems has received a lot of attention. Several modes of knowledge like visual, audio, and pen can be used individually or in combination have been proposed in support of this endeavour. Human communication relies heavily on the use of gestures to communicate information. Gesture recognition is a subject of science and language innovation that focuses on numerically quantifying human gestures. It is possible for people to communicate properly with machines using gesture recognition without the use of any mechanical devices. Hand gestures are a form of nonverbal communication that can be applied to several fields, including deaf-mute communication, robot control, human-computer interaction (HCI), home automation, and medical applications. Many different methods have been used in hand gesture research papers, including those focused on instrumented sensor technology and computer vision. To put it another way, the hand sign may be categorized under a variety of headings, including stance and motion, dynamic and static, or a combination of the two. This paper provides an extensive study on hand gesture methods and explores their applications.

**Keywords:** Hand gesture; Computer vision; Human-computer interaction

## 1 Introduction

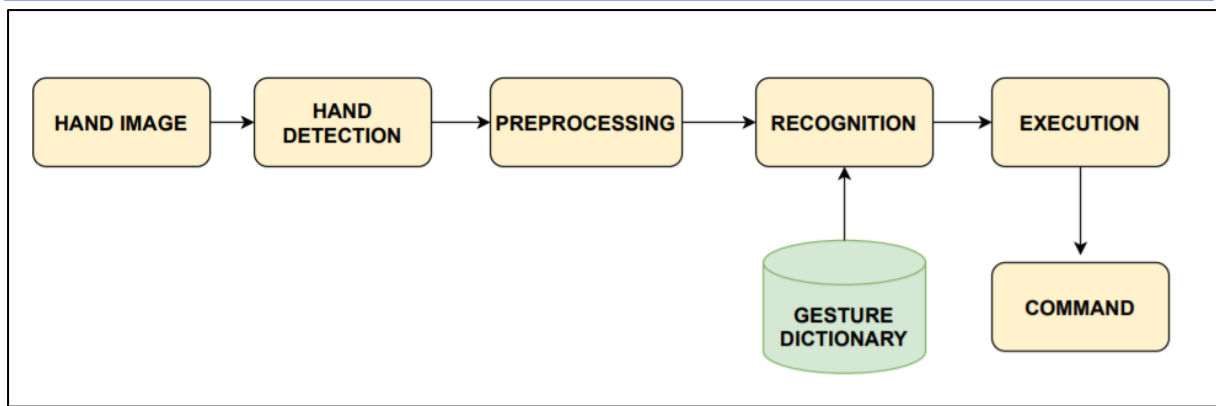
Gesture recognition is a field in computer science and language technology that interprets human gestures using computer vision techniques. There are a variety of physical motions that can cause gestures, but the most common source of gestures is the face and hands. Gesture recognition refers to the process of tracking gestures, turning them to representations, and then converting them to a useful command. Various technologies have been employed in the design and implementation of such devices, but the two most common types of technologies used for robust, accurate, and reliable hand gesture recognition systems are contact-based and vision-based technologies. Contact based devices like accelerometers, multi-touch screen, data glove etc. based on physical interaction of user who will be required to learn their usages. Whereas vision-based devices, such as cameras, must deal with a wide range of gestures.

Hand gestures are a type of body language that can be communicated by the location of the fingers, the centre of the palm, and the shape formed by the hand. Static and dynamic hand movements can be distinguished [1]. The static gesture refers to the hand's steady form, while the dynamic gesture is made up of a sequence of hand gestures such as waving [1,2]. Within a gesture, there are several different hand movements; for example, a handshake differs from one person to the next and changes with time and place [1]. The key distinction between posture and gesture is that posture emphasizes hand form while gesture emphasizes hand movement. [3].

Hand movements are a fascinating area of study because they can aid communication and provide a normal mode of interaction that can be applied to a wide range of situations. Wearable sensors mounted directly to the hand with gloves were previously used to recognize hand gestures [4]. Hand gestures or finger bending triggered these sensors to detect a physical response. The information gathered was then stored on a device that was wired to the glove. Using a sensor connected to a microcontroller, this glove-based sensor device could be made portable. With the invention of the data glove sensor, hand movements for human–computer interaction (HCI) began [5]. For a computer interface, it provided basic commands. By detecting the right coordinates of the location of the palm and fingers, the gloves were able to capture hand motion and position [1]. Curvature sensor [2], angular displacement sensor [3], optical fibre transducer [4], flex sensors [5], and accelerometer sensor [5] are examples of sensors that use the same technique depending on the angle of bending. Depending on the type of sensor, various physical concepts are used.

While the above methods have yielded promising results, they have several disadvantages that make them unsuitable for the elderly, who may encounter pain and confusion as a result of wire link issues. Furthermore, elderly people with chronic disease disorders that cause muscle weakness may be unable to put on and take off gloves, causing pain and constraining them if worn for long periods. In people with sensitive skin or those who have had burns, these sensors can cause skin damage, infection, or adverse reactions. These disadvantages prompted the production of promising and cost-effective strategies that did not necessitate the use of bulky gloves. Camera vision-based sensor technologies are the name given to these approaches. It is now simpler than ever to detect hand signals in a variety of applications, including clinical operations [1], sign language [2], robot control [2], virtual worlds [3], home automation [5], personal computer and tablet [6], and gaming [7], thanks to the evolution of open-source software libraries. The instrumented glove is effectively replaced with a camera in these techniques. To detect hands using these various types of cameras, algorithms based on computer vision methods have been developed. Hand features such as skin color, shape, motion, skeleton, depth, 3D model, deep learn detection, and more are attempted to be segmented and identified by the algorithms [8].

In the last decade, a number of studies using computer vision techniques have been published. Munir Oudah et al. [1] investigated the role and fundamental technique of HCI in terms of recognition, classification, and applications, as well as computer vision limits in various situations. Perimal et al. [2] developed a recognition system that addressed the issues of feature extraction and gesture categorization, as well as the study's application field. Sulyman et al. [3] conducted a study on hand gesture detection for mouse control applications, which included techniques and algorithms for human–machine interaction. They also gave a quick overview of the hidden Markov model (HMM). According to hand segmentation methodology, tracking, feature extraction, and recognition approaches, Prakash et al. [5] described numerous techniques and made comparisons between them, concluding that the recognition rate was a trade-off with temporal rate restricted by computational capacity. Finally, Wan,c et al. [13] examined numerous strategies for hand gesture detection, both sensor-based and vision-based, in order to increase algorithm precision by combining current techniques. As shown in Figure 1, the Hand Gesture Recognition System has several steps including Hand detection, Pre-processing of image, Recognition according to gesture dictionary and then executing its corresponding application.



**Figure 1:** Flow chart Of Hand Gesture Recognition System

Table 1 summarises a literature review of the numerous types of cameras used for different forms of hand gesture detection, as well as their features, resolution rate, number of gestures, and application areas.

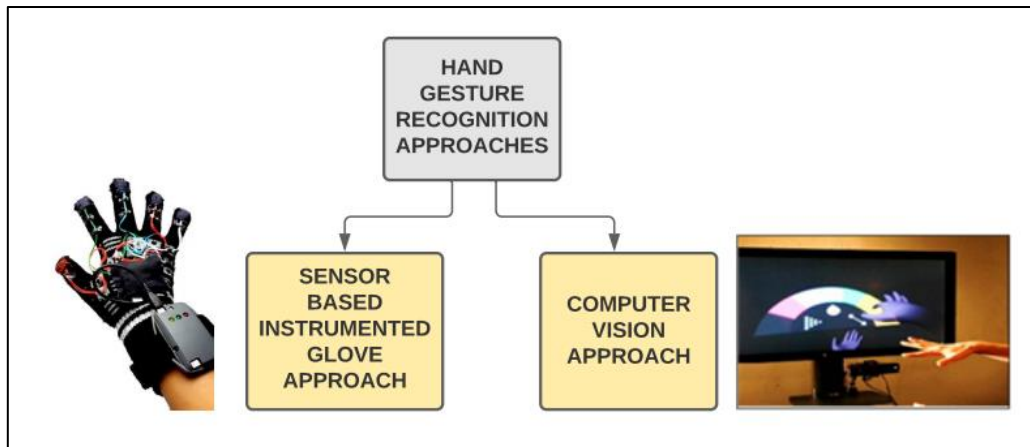
**Table 1:** Comparative study of different types of recognition of hand gesture application.

Author	Type of camera used	Feature	Resolution Rate	Number of Gestures	Area of application
Desai et al [4].	KinectV1	Hand gesture 1-5	88%	5	Electronic home appliances
Perimal et al [2].	Off-the-shelf HD webcam	Finger count	70%-100%	14	HCI
Sulyman et al [3]	Computer camera	Hand gesture (A-Z alphabet)	98%	6	Deaf people
Xi, C et al. [6]	Kinect camera depth sensor	Finger tip	-	Hand tracking	Real time hand tracking
Devineau et al. [7]	Intel real sense depth camera	Hand skeleton joints	91.28%	14	Classification methods
Konstantinidis et al.[8]	RGB video sequence recorder	Hand and body skeletal features	-	nanogesture	Sign language recognition
De Smedt et al.[9]	Sense depth camera	Hand gesture	88.25%	14	Hand gesture application
Karbasi, M et al.[10]	Kinect RGB camera	Hand blob	-	Hand gesture	Malaysian language
Ma, X et al [11].	KinectV2	Finger tip	96%	6	Natural human robot interaction
Desai, S.[4]	KinectV2	finger	90%	Fingercount	HCI

## 2 Hand Gesture Methods

The primary aim of gesture recognition research is to create a system that can identify individual human movements and use them to communicate information or command and control [1]. As a consequence, it covers not just the identification of human activity, but also the perception of that movement as meaningful commands. In most HCI applications, two methods are used to interpret gestures [2,3]. The first solution

relies on data gloves (wearable or direct contact), while the second relies on computer vision and does not require the use of any sensors [4] as shown in figure-2.



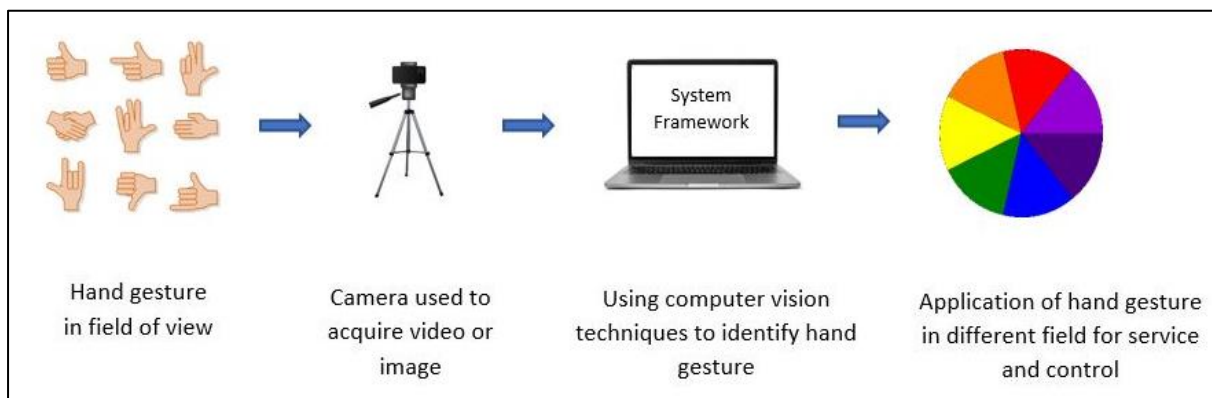
**Figure 2:** *Hand Gesture Recognition Approaches*

### 2.1 Sensor Based Instrumented Glove Approach

Hand motion and location can be captured using wearable glove-based sensors. Furthermore, using sensors attached to the gloves, they can easily provide the exact coordinates of palm and finger positions, orientation, and configurations [1]. However, this technique necessitates the user being physically attached to the device, which hinders user-computer interaction. Furthermore, these devices are very expensive [4]. The new glove-based solution, on the other hand, employs touch technology, which is a more promising technology that is classified as Industrial-grade haptic technology [6]. Using microfluidic technology, the glove provides haptic feedback that allows the user to feel the form, texture, movement, and weight of a virtual object [7,8,9].

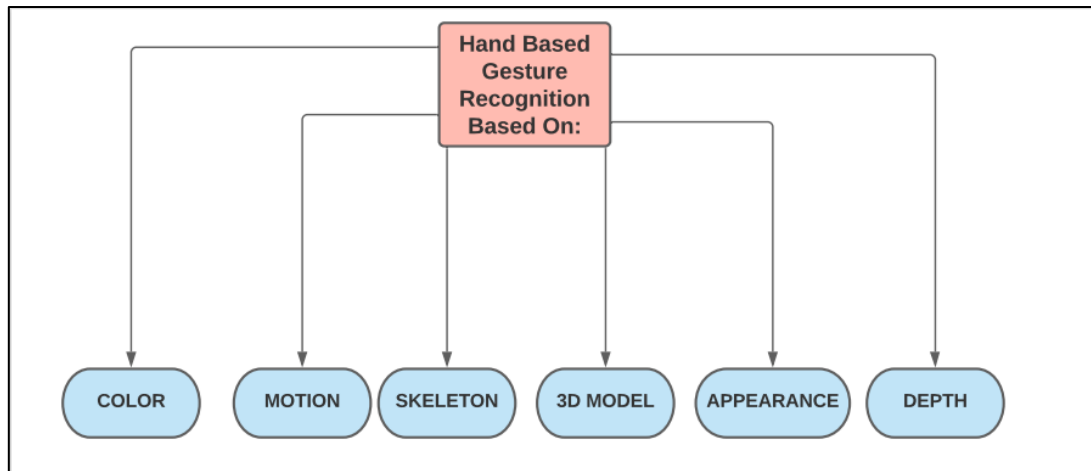
### 2.2 Computer Vision Approach

The camera vision-based sensor is a widespread, suitable, and useful approach because it allows for contactless communication between humans and computers[3]. A variety of camera setups, including monocular, fisheye, TOF, and IR, can be used[5]. Lighting changes, context problems, the influence of occlusions, a complex background, processing time vs. resolution and frame rate, and foreground or background objects of the same skin tone or otherwise resembling hands are all downsides of this method[6]. Figure 3 shows a simplified schematic of a camera vision-based sensor for extracting and detecting hand motions.



**Figure 3:** *Using computer vision techniques to identify gestures to control some possible application [1]*

As illustrated in figure 4, this part discusses computer vision-based recognition methods such as skin colour, appearance, motion, skeleton, depth, and 3D-model.



**Figure 4:** Hand Gesture Recognition Methods Based on Computer Vision approach.

### 2.2.1 Recognition Based on color:

#### i. Using colored Glove Marker

This method involves monitoring the movement of the hand with a camera while wearing a glove with various coloured markings [1]. This method has been used to communicate with 3D objects, allowing for zooming, shifting, drawing, and writing with a virtual keyboard [2]. The colours on the glove enable the camera sensor to monitor and detect the position of the palm and fingers, allowing for the extraction of a geometric model of the hand's form [3]. The advantages of this technology over the sensor data glove are its ease of use and low cost [5]. Colored gloves are still required, and the amount of natural and spontaneous contact with the HCI is reduced [7,8,11].

#### ii. Using Skin Color

One of the most widely used methods for hand segmentation is skin color detection, which is used in a variety of applications including object classification, degraded photograph recovery, individual movement monitoring [2], video observation [3], HCI applications [5], facial recognition [6], hand segmentation [7], and gesture identification [8]. Two techniques were used to detect the color of the skin. The first approach is pixel-based skin detection, which classifies each pixel in an image as skin or not based on its neighbors [1]. The second technique is area skin detection, which involves spatially processing skin pixels based on details including strength and texture [2]. Color space is a mathematical model that can be used to reflect image color details. Various color spaces can be used depending on the type of application, such as digital graphics, image processing applications, TV transmission, and computer vision techniques [7,12].

For skin segmentation, a variety of color space formats are available, as described below:

R–G–B and RGB-normalized red, green, and blue

H–S–V, H–S–I, and H–S–L (hue, saturation, and lightness)

YIQ, Y–Cb–Cr, and YUV are all examples of luminance.

Color space properties such as hue/saturation family and brightness family perform well under varying lighting conditions. When there is a lot of variances in colour value, the conversion from RGB to HSI or HSV takes a long time (hue and saturation). As a result, a pixel within a given intensity range is chosen. Because of the conversion from Cartesian to polar coordinates, the RGB to HSV translation may take some time. As a result, HSV space is excellent for simple picture detection. When compared to the HSV colour family in terms of skin colour detection and segmentation, transforming and dividing channels in the Y–Cb–Cr colour space is simple.[1]

### **2.2.2 Recognition based on Appearance**

This approach works by extracting image features and comparing them to features extracted from the input image frames to model visual appearances like hands [1]. Without a previous segmentation process, the features are determined directly from the pixel intensities [2]. Due to the ease with which 2D image features can be removed, the method runs in real-time and is considered simpler to implement than the 3D model method [7]. This approach can also detect different skin tones. The AdaBoost learning algorithm can solve the occlusion problem by maintaining fixed features such as key points for a portion of the hand [1,5].

### **2.2.3 Recognition based on motion**

For identification purposes, motion-based recognition may be used; it extracts the object from a sequence of image frames [1]. To recognize the gesture, the AdaBoost algorithm used for object detection, characterization, movement modelling, and pattern recognition is needed [3]. The most common problem with motion recognition is that it happens when another gesture is involved during the recognition process, and the complex context has a negative impact [4]. Furthermore, occlusion among tracked hand gestures or an error in region extraction from tracked gestures can result in gesture loss, with long-distance effects on region appearance [12].

### **2.2.4 Recognition based on skeletons**

Skeleton-based recognition defines model parameters that can help detect complex features more accurately [1]. Where the different representations of skeleton data for the hand model can be used for classification, it defines geometric properties and constraints and easily converts data characteristics and similarities, to concentrate on geometric and statistical features [8]. The joint orientation, space between joints, skeletal joint position, degree of angle between joints, and trajectories and curvature of the joints are the most common features used [13].

### **2.2.5 Recognition based on depth**

Various methods for solving hand gesture recognition using various types of cameras have been proposed. A depth camera captures 3D geometric data about an object [7]. Previously, both TOF precepts and light coding were used as approximations [8]. When compared to a color image that only contains a projection, the 3D data from a depth camera directly reflects the depth field [1,2]. The resulting image was unaffected by lighting, shade, or color when using this method [3]. To work with Kinect depth data, we require a thorough understanding of its characteristics. Kinect's depth stream is made up of pixels that store distance data from the camera plane to the nearest object. The distance is measured in millimetres. This feature is useful when we need to remove background scene from image. The depth image is available in three resolutions: 640x480 (default), 320x240, and 80x60 pixels. However, the depth camera's cost, size, and availability will limit its use [6].

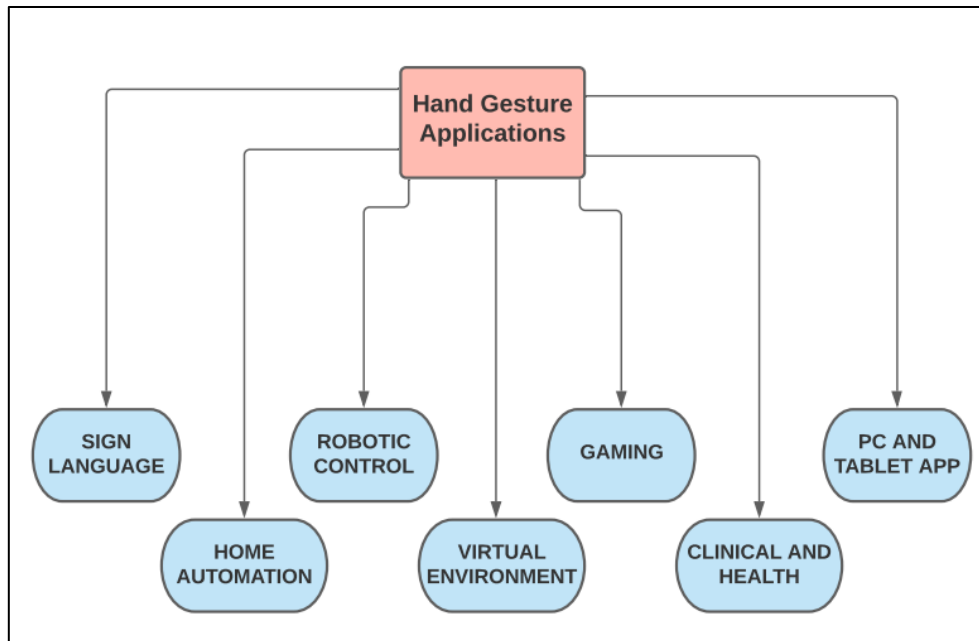
### **2.2.6 Recognition based on 3D models**

The 3D model is based on a large-degree-of-freedom 3D Kinematic hand model, with hand parameter estimation obtained by comparing the input image with the two-dimensional appearance projected by the three-dimensional hand model [1]. Furthermore, by forming a volumetric, skeletal, or 3D model that is identical to the user's hand, the 3D model introduces human hand features such as pose estimation [3]. During the matching process, the 3D model parameter is updated. To improve accuracy, the depth parameter is added to the model [5].

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

Hand gesture research has become a fascinating and important area because it allows for natural contact while also lowering the cost of using sensors in data gloves [1]. Traditional interactive approaches depend on a variety of items, including a mouse, keyboard, touch screen, joystick for gaming, and consoles for

computer control [2,3]. The parts that follow go into some of the most popular uses for hand gestures [3,4]. The most popular application field for hand gesture recognition techniques is shown in Figure 5.



**Figure 5:** Application areas of Hand Gesture

### 3.1 Clinical and Health

A surgeon can require information about the patient's entire body structure or a comprehensive organ model during clinical operations to minimize operating time or improve the accuracy of the outcome [1]. This is accomplished by using a medical imaging device, such as an MRI, CT, or X-ray system [1], which gathers data from the patient's body and displays it as a detailed image on the screen [2]. Using a computer vision technique, the surgeon can encourage contact with the displayed images by making hand movements in front of the camera [5]. These movements can be used to perform tasks like zooming, flipping, cropping images, and moving to the next or previous slide without the use of a mouse, keyboard, or touch screen [6]. Sterilization is needed for any additional devices, which can be difficult in the case of keyboards and touch screens. Hand movements can also be used for assistive purposes, such as guiding a wheelchair [7].

### 3.2 Sign Language Recognition

People who are unable to communicate with others using words use sign language as an alternative. It is made up of a series of movements, each of which represents a letter, number, or word [6]. Many research papers have suggested using a glove-attached sensor worn on the hand that responds to hand movement to recognize sign language for deaf-mute people [9]. Alternatively, exposed hand contact with the camera may be used, with computer vision techniques used to recognize the gesture [12]. The dataset used for gesture classification in each of the methods described above matches a real-time gesture made by the user [15].

### 3.3 Robot Control

Industry, assistive services [1], stores, sports, and entertainment are only a few of the areas where robot technology is used [5]. Machine learning methods, artificial intelligence, and complex algorithms are used by robotic control systems to carry out a particular mission, enabling the robotic system to communicate naturally with the environment and make autonomous decisions [8]. According to some studies, combining computer vision technology with a robot can be used to create assistive devices for the elderly [10]. Another study [2] uses computer vision to allow a robot to ask a human for a proper path within a house.

### 3.4 Virtual Environment

Virtual worlds are built on a three-dimensional model that requires a three-dimensional gesture recognition system to interact within real-time as an HCI. These movements can be used for editing and viewing, as well as for entertainment, such as playing a virtual piano. A dataset is used by the gesture recognition system to align it with an acquired gesture in real-time [1,2,13].

### 3.5 Home Automation

Hand gestures are an effective way to automate your home. Lighting, fans, television, radio, and other devices may all be controlled simply by shaking a hand or making a gesture. They can be used to improve the quality of life of older individuals [13].

### 3.6 Personal Computer and Tablet

Hand movements can be used to communicate with a computer without using a mouse or keyboard, such as dragging, falling, and transferring files across the desktop environment, as well as cut and paste operations [12]. Furthermore, they can be used to monitor slide shows [14]. They are often used in conjunction with a tablet to enable deaf-mute people to communicate with others by placing their hands in front of the tablet's camera [13]. This necessitates the installation of a program that transforms sign language into text and shows it on the computer. This is close to translating acquired voice to text [15].

### 3.7 Gestures for Gaming

The Microsoft Kinect Xbox, which has a camera positioned over the screen and connects to the Xbox console through the cable port, is the best example of gesture interaction for gaming purposes [13]. The Kinect camera sensor tracks the user's hand gestures and body movements, which can be used to interact with the game [11].

## 4 Conclusion

Hand gesture recognition is used to overcome a flaw in interaction methods. Controlling things by hand is more natural, simpler, more versatile, and less expensive, and there is no need to address problems produced by hardware devices because none are required. From the previous sections, it was evident that investing a lot of work into designing accurate and stable algorithms with the help of a camera sensor with a certain characteristic to encounter common challenges and produce a dependable result was necessary. Each of the strategies outlined above, on the other hand, has its own set of advantages and disadvantages, and performs admirably in some situations while failing miserably in others.

## 5 Declarations

### 5.1 Acknowledgements

The authors would like to express their gratitude to the Jagan Institute of Management Studies for providing us with the opportunity to produce this research paper and for encouraging us to do so.

### 5.2 Competing Interests

There are no conflicts of interest that are relevant to this study for the authors of this manuscript.

### How to Cite this article

R. Jain, M. Jain, R. Jain, and S. Madan, "Human Computer Interaction – Hand Gesture Recognition", *Adv. J. Grad. Res.*, vol. 11, no. 1, pp. 1–9, Sep. 2021. <https://doi.org/10.21467/ajgr.11.1.1-9>

### References

- [1] Munir Oudah., Ali Al-Naji., Javaan Chahl., "Hand gesture recognition based on computer vision: A review on technique" *Journal of Imaging*, 6 , 73, 2020.
- [2] Perimal, M.; Basah, S.N.; Safar, M.J.A.; Yazid, "Hand-Gesture Recognition-Algorithm based on Finger Counting", *J. Telecommun. Electron. Comput. Eng.*,10, 19–24, 2018.

- [3] Sulyman, A.B.D.A.; Sharef, Z.T.; Faraj, K.H.A.; Aljawaryy, Z.A.; Malallah, F.L. , “Real-time numerical 0-5 counting based on hand-finger gestures recognition.”, *J. Theor. Appl. Inf. Technol.* 2017, 95, 3105–3115.
- [4] Desai, S.; Desai, A., “Human Computer Interaction through hand gestures for home automation using Microsoft Kinect.” *In Proceedings of the International Conference on Communication and Networks, Xi'an, China*, pp. 19–29, 10–12 October 2017.
- [5] Prakash, J.; Gautam, U.K. , “Hand Gesture Recognition.”, *Int. J. Recent Technol. Eng.* 2019, 7, 54–59.
- [6] Xi, C.; Chen, J.; Zhao, C.; Pei, Q.; Liu, L., “Real-time Hand Tracking Using Kinect.”, *In Proceedings of the 2nd International Conference on Digital Signal Processing, Tokyo, Japan*, pp. 37–42, 25–27 February 2018.
- [7] Devineau, G.; Moutarde, F.; Xi, W.; Yang, J., “Deep learning for hand gesture recognition on skeletal data.”, *In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China*, pp. 106–113, 15–19 May 2018.
- [8] Konstantinidis, D.; Dimitropoulos, K.; Daras, P. , “Sign language recognition based on hand and body skeletal data.”, *In Proceedings of the 2018-3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), Helsinki, Finland*, pp. 1–4, 3–5 June 2018.
- [9] 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*, pp. 33–38, 23–24 April 2017.
- [10] Karbasi, M.; Muhammad, Z.; Waqas, A.; Bhatti, Z.; Shah, A.; Koondhar, M.Y.; Brohi, I.A. “A Hybrid Method Using Kinect Depth and Color Data Stream for Hand Blobs Segmentation”; *Science International: Lahore, Pakistan*, Volume 29, pp. 515–519, 2017.
- [11] Ma, X.; Peng, “J. Kinect sensor-based long-distance hand gesture recognition and fingertip detection with depth information.”, *J. Sens.* 5809769, 2018.
- [12] Desai, “S. Segmentation and Recognition of Fingers Using Microsoft Kinect”, *In Proceedings of the International Conference on Communication and Networks, Paris, France*, pp. 45–53, 21–25 May 2017.
- [13] Wan, C.; Probst, T.; Van Gool, L.; Yao, A., “Self-supervised 3d hand pose estimation through training by fitting.”, *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA*, pp. 10853–10862, 15–20 June 2019.
- [14] 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*, pp. 4511–4520, 15–20 June 2019.
- [15] Ge, L.; Ren, Z.; Li, Y.; Xue, Z.; Wang, Y.; Cai, J.; Yuan, J. “3d hand shape and pose estimation from a single rgb image”. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA*, pp. 10833–10842, 15–20 June 2019.

**Publish your books with AIJR publisher-**

- ✓ Publish with ISBN and DOI.
- ✓ Publish Thesis/Dissertation as Monograph.
- ✓ Publish Book Monograph.
- ✓ Publish Edited Volume/ Book.
- ✓ Publish Conference Proceedings
- ✓ Retain full copyright of your books.

Submit your manuscript at [books.aijr.org](http://books.aijr.org)

**Publish your research article in AIJR journals-**

- ✓ Online Submission and Tracking
- ✓ Peer-Reviewed
- ✓ Rapid decision
- ✓ Immediate Publication after acceptance
- ✓ Articles freely available online
- ✓ Retain full copyright of your article.

Submit your article at [journals.aijr.org](http://journals.aijr.org)