Hand Gesture Recognition: A Survey

**Shamama Anwar, Subham Kumar Sinha, Snehanshu Vivek
and Vishal Ashank**

Abstract A human–computer interaction is generally limited to taking input from the user using handheld devices like keyboard, mouse, or scanners. With the advancement in computers, the user interaction approaches have also advanced. Direct use of hands as an input device is an attractive method for providing natural Human–Computer Interaction. It is also helpful for people who use sign language. The chapter aims to study the existing methods for Hand Gesture Recognition and provide a comparative analysis of the same. The entire process of hand gesture recognition is divided into three phases: hand detection, hand tracking, and recognition. The chapter includes a review of the different methods used for the hand gesture recognition. The recognition phase is classified based on the way the input is received as glove based or vision based. For recognition, various methods like Feature extraction, Hidden Markov Model (HMM), Principal Component Analysis (PCA) are compared along with the reported accuracy.

Keywords Computer vision · Hand gesture · Hidden Markov model
Principal component analysis

1 Introduction

Communication among people varies from being verbal to use of body language or gestures. Gesture forms an important means of communication. People tend to involuntarily use hand movements (termed as gestures) when they normally talk or even during telephonic conversations. Hand gestures provide a separate complementary modality to speech for expressing one's ideas giving emphasis on certain points. Humans can conveniently interact with computing devices by using hand

S. Anwar (✉) · S. K. Sinha · S. Vivek · V. Ashank
Department of Computer Science and Engineering, Birla Institute
of Technology, Mesra, Ranchi 835215, India
e-mail: shamama@bitmesra.ac.in

gestures. The method is more suitable than using other input devices but the major challenge is how to make hand gestures understood by the computing devices.

For this purpose, hand gesture recognition systems have evolved. The entire hand gesture recognition process can be divided into three phases: Hand Detection, Tracking, and Recognition. In the first phase, a video input is given to the system, which is then divided into frames (images). The aim of this phase is to recognize the object of interest (i.e., the hand) from the frames. This phase may require some form of preprocessing like noise removal, background subtraction, etc. Once the hand is isolated from the frames, the tracking is done in subsequent frames to detect the motion. There are various existing models to aid this process. The chapter presents some prominently used methods.

The approaches used for hand detection can be mainly divided into "Data-Glove based" and "Vision Based" approaches based on the way the input is taken by the system. The Data-Glove-based methods require the use of sensor devices for recognizing finger and hand movement, which then needs to be represented in an appropriate form for further computations. The sensors aid the collection of hand configuration and movement data. However, the devices are quite expensive and bring much cumbersome experience to the users. In contrast, the vision-based methods acquire the input by means of a camera. This method of input is more convenient and portable as well. Any handheld or stationed device can be used for acquiring the input. These systems need to be background invariant, lighting insensitive, person and camera independent to achieve real-time performance, which is a challenge. Moreover, such systems must be optimized to meet the requirements, including accuracy and robustness. The purpose of this chapter is to present a review of Hand Gesture Recognition techniques for human–computer interaction, consolidating the various available approaches, pointing out their general advantages and disadvantages along with their reported accuracy.

1.1 Glove Based

In glove-based recognition system, a glove with sensors is provided that detects the finger and hand movement. The type of sensors used in these gloves varies from flex sensors to LED sensors. The positioning of sensors also varies across different models [1]. Some system use gloves with sensors on fingertips, while others prefer gloves with sensors at the finger joints (Fig. 1).

1.2 Vision Based

In vision-based gesture recognition, the movement of the hand is recorded by a camera. The video is decomposed into a set of images (frames). Some preprocessing may be required to isolate hand from other body parts and to eliminate the

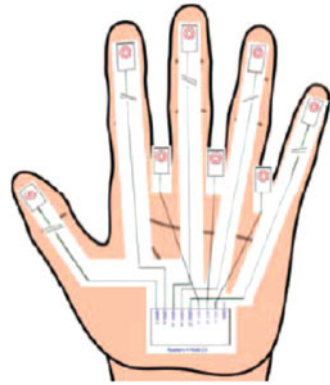**Fig. 1** Conceptual view of the gesture recognition glove consisting of sensors [1]
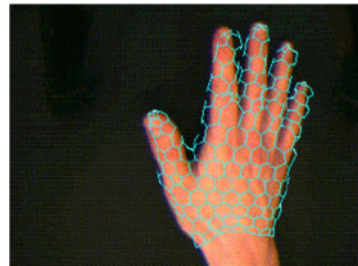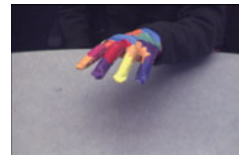

Fig. 2 Snapshot of 3D tracker [2]

Fig. 3 Real-time hand tracking using color glove [3]

background. The approaches also differ based on the background elimination techniques used. Simple background subtraction can be used if the background is static. But for real-time tracking, the background is not static. So, these implementations require a more complex background elimination technique. After the background has been eliminated, hand recognition is performed. The common approaches used for hand detection in vision-based recognition are skin color detection and 3D hand model approach [2, 3]. A description of the techniques is included in the subsequent section (Figs. 2 and 3).

2   Gesture Recognition Techniques

There are various gesture tracking techniques available. Some are glove-based recognition while others are based on vision. Some efficient algorithms exploit the advantages of both these methods. Once the hand is detected from an input frame, its movement is tracked for further recognition. There are various approaches for the same. A simple approach for recognition is the template matching technique. This method requires creating a template of predefined actions. Few researchers experimentally determined the number of templates required for a certain gesture and maintained a database of the same [4]. They also used linear regression to calculate the exact number of templates to be used on a certain gesture based on the average time the gesture was performed. The experiments were conducted on hand gestures taken on a fixed background. Hand pose recognition in the cluttered background has more applicability in real-life tracking. To achieve this, many techniques have been combined by Stenger [5] in their proposed system. The color model is initialized and updated by a frontal face detector. Hand locations and scale are hypothesized efficiently using cumulative likelihood maps, and the hand pose is estimated by normalized template matching. The system eliminated the need for background subtraction and the method was efficient enough to detect the hand in each frame independently. A drawback for template-based methods is the need to maintain a template set for the recognizable gestures.

Other than template-based matching, feature extraction-based methods have also been used. A method to recognize the unknown input gestures by using Hidden Markov Models (HMMs) was proposed by Chen et al. [6]. Since the variation of the hand gestures is usually large, the transition between states is necessary in each gesture for an effective hand tracking. The experiments in the paper recognized a single action in a stationary background. Hence, the system had a smaller search region for tracking. Addition of a new gesture required retraining the HMM for the new gesture. Repeated experiments could recognize 20 different gestures, and the recognizing rate is above 90% [6].

Another method based on feature extraction was implemented to recognize American Sign Language and Arabic numbers. The method used stereo color image sequences in HMMs. The system has three stages: preprocessing, feature extraction, and classification. In preprocessing stage, color and 3D depth maps were used to detect and track the hand. In the second stage, 3D combined features of location, orientation, and velocity with respect to Cartesian and Polar systems were used. Additionally, k-means clustering was also employed for HMM. In the final stage, the hand gesture path was recognized using Left-Right Banded topology (LRB). This system successfully recognized isolated hand gestures with 98.33% recognition rate [7]. But methods based on feature extraction are found to be computationally expensive.

Methods based on active shape model have also gained popularity. In [8], an active statistical model for hand gesture extraction and recognition is applied. After the hand contours are found out by a real-time segmenting and tracking system,

a set of feature points (landmarks) were marked out automatically and manually along the contour. Mean shape, eigenvalues, and eigenvectors are computed out and composed the active shape model. When the model parameter is adjusted continually, various shape contours are generated to match the hand edges extracted from the original images. The gesture is finally recognized after well matching.

A method using Principal Component Analysis (PCA), which used skin color detection (vision based) for hand recognition was also designed, which was tested in the controlled background and in different lightning conditions. The database collected in the ideal conditions has proved to be the most efficient database in terms of accuracy and gives 100% accuracy. When the lightning conditions are changed, the accuracy decreases as compared to the previous one. The system shows 91.43% accuracy with low brightness images [9]. But the model was not capable of working with the images containing hands of other than skin color. The proposed model does not evaluate the images clicked in other light colors, where the hand gestures have been clicked and the model works only with a static gesture. But there might be miss-recognitions in case the background has elements that resemble the human skin [10].

3 Comparison of the Methods

Based on the study of the different techniques, a comparison table is provided (Table 1) that lists the advantages, disadvantages, and accuracy of the different methods reviewed.

Table 1 Comparison of different techniques for gesture recognition

Method	Advantages	Disadvantages	Accuracy
Template matching	1. Simple 2. Accurate for small gesture set	1. Efficiency decreases for large posture sets	98.80% [11]
Feature extraction	1. Works well for large gesture set	1. Computationally expensive	90% [6]
Active shape model	1. More suitable for real-time recognition 2. Also efficient in posture, face recognition	1. More accurate for open hand gestures, face recognition	88% [12]
Principal component analysis	1. Better recognition accuracy	1. Depends on lighting condition and illumination 2. Needs retraining	91–100% based on lighting factor [9]

4 Conclusion

Based on the review of the different techniques involved in hand gesture recognition, it is observed that the two major ways a human–computer interaction system can take input is, either glove-based or vision-based method. The glove-based method although is more accurate but the cost of such systems is generally high due to the need of a sensor-enabled glove. An additional hardware component (glove) is needed for implementing such system. The user comfort is also compromised as these methods require a certain restriction on the hand anatomy and hence the portability of such systems is less. In contrast, the vision-based methods are portable and generally does not require any specific or special hardware for implementation. A similar study on the various recognition techniques reveal the pros and cons of the different methods used. Template matching-based methods are simple and accurate for a small set of gestures or postures. It requires maintaining a large set of databases and may not be feasible if the applicability of such systems is on a large scale.

The feature extraction-based methods and active shape model methods are more suitable for real-time recognition and are generally vision based as well. These methods along with PCA and HMM needs more training to adapt the system for more accurate recognition. The chances of misrecognition are higher in real-time HMM-based methods due to real-time moving background.

References

1. Abhishek KS, Qubeley LCF, Ho D (2016) Glove-based hand gesture recognition sign language translator using capacitive touch sensor. In: IEEE international conference on electron devices and solid-state circuits (EDSSC)
2. Garg P, Aggarwal N, Sofat S (2009) Vision based hand gesture recognition. Int J Comput Electr Autom Control Inf Eng 3(1):186–191
3. Wang RY, Popovic J (2009) Real-time hand-tracking with a color glove. ACM Trans Graph 28(3)
4. Carrera KCP, Erise APR, Abrena EMV, Colot SJS, Telentino RE (2014) Application of template matching algorithm for dynamic gesture recognition of American sign language finger spelling and hand gesture. Asia Pac J Multidiscip Res 2(4):154–158
5. Stenger B (2006) Template-based hand pose recognition using multiple cues. In: Asian conference on computer vision
6. Chen F-S, Fu C-M, Huang C-L (2003) Hand gesture recognition using a real-time tracking method and hidden Markov models. Image Vis Comput 21(8):745–758
7. Elmezain M, Al-Hamadi A, Pathan SS, Michaelis B (2009) Spatio-temporal feature extraction-based hand gesture recognition for isolated American sign language and Arabic numbers. In: 6th international symposium on image and signal processing and analysis
8. Liu N, Lovell BC (2005) Hand gesture extraction by active shape models. In: Digital image computing: techniques and applications
9. Ahuja MK, Singh A (2015) Hand gesture recognition using PCA. Int J Comput Sci Eng Technol 5(7):267–271

10. Bansal M, Saxena S, Desale D, Jadhav D (2011) Dynamic gesture recognition using hidden Markov model in static background. Int J Comput Sci Issues 8(6):391–398
11. Azad R, Azad B, Kazeroni IT (2013) Real-time and robust method for hand gesture recognition system based on cross-correlation coefficient. Adv Comput Sci Int J 2(5):121–125
12. Edwards GJ, Cootes TF, Taylor CJ (1998) Face recognition using active appearance models. In: Computer vision—ECCV, LNCS