Intell Ind Syst (2016) 2:179–191
DOI 10.1007/s40903-016-0046-9

REVIEW PAPER

A Survey on Recent Vision-Based Gesture Recognition

Haitham Badi[1,2]

Received: 26 March 2016 / Revised: 17 May 2016 / Accepted: 23 May 2016 / Published online: 30 May 2016

Abstract Considerable effort has been put toward the development of intelligent and natural interfaces between users and computer systems. In line with this endeavor, several modes of information (e.g., visual, audio, and pen) that are used either individually or in combination have been proposed. The use of gestures to convey information is an important part of human communication. Hand gesture recognition is widely used in many applications, such as in computer games, machinery control (e.g., crane), and thorough mouse replacement. Computer recognition of hand gestures may provide a natural computer interface that allows people to point at or to rotate a computer-aided design model by rotating their hands. Hand gestures can be classified in two categories: static and dynamic. The use of hand gestures as a natural interface serves as a motivating force for research on gesture taxonomy, its representation, and recognition techniques. This paper summarizes the surveys carried out in human–computer interaction (HCI) studies and focuses on different application domains that use hand gestures for efficient interaction. This exploratory survey aims to provide a progress report on static and dynamic hand gesture recognition (i.e., gesture taxonomies, representations, and recognition techniques) in HCI and to identify future directions on this topic.

Keywords Human–computer interaction · Gesture recognition · Representations · Recognition · Natural interfaces

1 Introduction

Computers have become a key element of our society since their first appearance. Surfing the web, typing a letter, playing a video game, or storing and retrieving data are few examples of tasks that involve the use of computers. Computers will increasingly influence our everyday life because of the constant decrease in the price of personal computers. The efficient use of most computer applications require more interaction. Thus, (HCI) has become an active field of research in the past few years [1]. To utilize this new phenomenon efficiently, many studies have examined computer applications and their requirement of increased interaction. Thus, human computer interaction (HCI) has been a lively field of research [2,3].

Gesture recognition and gesture-based interaction have received increasing attention as an area of HCI. The hand is extensively used for gesturing compared with other body parts because it is a natural medium for communication between humans and thus the most suitable tool for HCI (Fig. 1) [4]. Interest in gesture recognition has motivated considerable research, which has been summarized in several surveys directly or indirectly related to gesture recognition. Table 1 shows several important surveys and articles on gesture recognition. Comprehensively analyzing published surveys and articles related to hand gesture recognition may facilitate the design, development, and implementation of evolved, robust, efficient, and accurate gesture recognition systems for HCI. The key issues addressed in these research articles may assist researchers in identifying and filling research gaps to enhance the user-friendliness of HCI systems.

✉ Haitham Badi
haitham@siswa.um.edu.my

[1] College of Busniness Informatics, University of Information Technology and Communications, Baghdad, Iraq

[2] University of Malaya, 50603 Kuala Lumpur, Malaysia

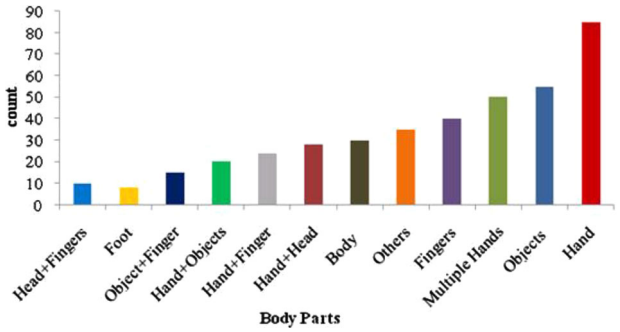

Fig. 1 The *graph* shows the different body parts or objects identified in the literature employed for gesturing [4]

2 Hand Gesture Analysis Approaches

Hand gesture analysis can be divided into three main approaches, namely, glove-based analysis, vision-based analysis, and analysis of drawing gestures [12]. The first approach employs sensors (mechanical or optical) attached to a glove that acts as transducer of finger flexion into electrical signals to determine hand posture, as shown in Fig. 2. The relative position of the hand is determined by an additional sensor. This sensor is normally a magnetic or an acoustic sensor attached to the glove. For some data-glove applications, look-up table software toolkits are provided with the glove to be used for hand posture recognition [13]. The relative position of the hand is determined by an additional sensor. This sensor is normally a mag-

Fig. 2 The cyborg glove: data glove is constructed with stretch fabric for comfort and a mesh Palm for ventilation [11]

Table 1 Analysis of some comprehensive surveys and articles

Moeslund and Granum [5]	
Scope of analysis	A survey discussing initialization, tracking, pose estimation, and recognition of motion capture systems and discusses the types of information processed
	The analysis primarily focuses on gesture recognition
	System functionality is broken down into four core processes: initialization, tracking, pose estimation, and recognition
	This survey is significant because it provides a comprehensive overview of publications on motion capture for over two decades and of the effect of machine vision on full-body motion capture and discusses the existing state of research in this area as well as future research directions
Key issues addressed	System functionality characteristics and field advancements are comprehensively discussed and evaluated. Remarkably, several features (e.g., robustness, accuracy, and speed) are required in a specific domain but not in others
	Problems predominant in a domain, such as lack of training data, considerable time required to capture gestures, lack of invariance, and robustness, are explored, and possible solutions, such as using an approach similar to speech recognition and abstracting the motion layer, are investigated
	Although these solutions have been successfully applied to a certain extent, almost all approaches are significantly limited by a lack of modularization
Derpanis [6]	
Scope of analysis	The paper reviews vision-based hand gestures for HCI
	Various aspects of vision-based gesture recognition problems related to the feature set, classification method, and underlying representation of the gesture set are discussed
Key issues addressed	Feature extraction, classification methods, and gesture representation should be investigated to facilitate HCI. The problem with most approaches is that they are based on several underlying assumptions that may be suitable for a controlled laboratory setting but not generalizable to arbitrary settings

Table 2 Continued analysis of some comprehensive surveys and articles

Mitra and Acharya [7]	
Scope of analysis	Gesture recognition techniques, particularly those for hand and facial movements, are comprehensively surveyed
	The survey covers the use of hidden Markov models (HMMs), particle filtering and condensation, finite-state machines (FSMs), optical flow, skin color, and connection models
	Gesture recognition has manifold applications, ranging from sign language to medical rehabilitation to virtual reality
Key issues addressed	Different recognition algorithms should be developed depending on the size of the data set and the gesture made. Various combinations can be drawn out in this regard
	For example, a HMM and FSM may be hybridized to recognize a complex gesture consisting of many smaller gestures and a neural network may be used for large data sets
	Developed systems should be both flexible and expandable to maximize their efficiency, accuracy and comprehensibility
Chaudhary et al. [8]	
Scope of analysis	This paper discusses the effects of intelligent approaches, including soft computing based methods such as artificial neural networks, fuzzy logic, and genetic algorithms, on designing hand gesture recognition methods. Image preprocessing for segmentation and hand image construction is also examined
Key issues addressed	Appearance-based methods are mostly used to detect fingertips
	Soft computing enables the definition of uncertain things and uses learning models and training data with approximation. Soft computing is an effective technique where the exact position of the hand or fingers is not given

netic or an acoustic sensor attached to the glove. Look-up table software toolkits are provided with the glove for some data-glove applications for hand posture recognition [13]. The second approach, vision based analysis, is based on how humans perceive information about their surroundings. However, this approach is probably the most difficult to implement. Several different approaches have been tested thus far. One is by building a three-dimensional model of the human hand. The model is matched to images of the hand by one or more cameras (Tables 2, 3). Parameters that correspond to palm orientation and joint angles are then estimated. These parameters are then used to perform gesture classifications [13].

The third approach pertains to the analysis of drawing gestures, which usually involves the use of a stylus as an input device. The analysis of drawing gestures can also lead to the recognition of written text. Majority of hand gesture recognition work involve mechanical sensing, most often for direct manipulation of a virtual environment and occasionally for symbolic communication. However, mechanical hand posture sensing (static gesture) has a range of problems, including reliability, accuracy, and electromagnetic noise. Visual sensing has the potential to make gestural interaction more practical, but it probably poses some of the most difficult problems in machine vision [13].

2.1 Enabling Technologies for HCI

The two major types of enabling technologies for HCI are contact-and vision-based devices. Contact-based devices used in gesture recognition systems are based on the physical interaction of users with the interfacing device. That is, the user should be accustomed to using these devices; thus, these devices are unsuitable for users with low computer literacy. These devices are usually based on technologies using several detectors, such as data gloves, accelerometers, and multi-touch screens. Other devices, such as the accelerometer of Nintendo Wii, use only one detector. These contact-based devices for gesture recognition can further be classified into mechanical, haptic, ultrasonic, inertial, and magnetic devices [14]. Mechanically primed devices are a set of equipment used by end users for HCI. These devices include the IGS-190, a body suit that captures body gestures, and CyberGlove II and CyberGrasp, wireless instrumented gloves used for hand gesture recognition (Fig. 2) [11]. These devices should be paralleled with other devices for gesture recognition. For instance, the IGS-190 should be used with 18 inertial devices for motion detection. Cybergloves and magnetic trackers are also used to model trajectories for hand gesture recognition. Haptics-primed devices are commonly used, touch-based devices with hardware specially designed

Table 3 Continued analysis of some comprehensive surveys and articles

Wachs et al. [9]	
Scope of analysis	This article comprehensively discusses vision-based hand gesture applications
	The article focuses on different aspects of gesture-based interfaces using hands
	The article also provides an overview on the different challenges in vision-based gesture recognition systems.The paper also discusses different applications controllable by hand gestures
Key issues addressed	No single method for automatic hand gesture recognition is suitable for every application. Each gesture recognition algorithm depends on the cultural background, application domain, and environment of the user
	Hand gesture interaction must also address intuitiveness and gesture spotting in addition to technical obstacles, such as those related to reliability, speed, and cost
	Two-handed dynamic hand gesture interaction is a promising area for future research
Corera and Krishnarajah [10]	
Scope of analysis	This paper surveys tools and techniques used to capture hand gestures
	Different vision- and sensor-based techniques for hand gesture recognition
	The paper also discusses logic issues and design considerations for gesture recognition systems
Key issues addressed	The paper compares the merits and demerits of vision- and sensor-based techniques
	These techniques can best be advanced by integrating with modularization and scalability and essentially decentralizing the entire approach from gesture capture to recognition

for HCI, including multi- touch screen devices such as the Apple iPhone, tablet PCs, and other devices with multi- touch gestural interactions using HMMs [1]. Ultrasonic-based motion trackers are composed of sonic emitters, which emit ultrasound sonic discs that reflect ultrasound, and multiple sensors that time the return pulse. Gesture position and orientation are computed based on propagation, reflection, speed, and triangulation [14]. These devices have low resolution and precision because of ultrasonic illumination and magnetic obstacles or noise when applied to certain environments. This lack of interference makes these devices popular. Inertial-primed devices operate based on variations in the magnetic field of the earth to detect motion. Schlomer et al. [16] proposed a gesture recognition technique using a Wii controller employing an HMM independent of the target system. Bourke et al. [17] proposed recognition systems that detect normal gestures used in daily activities by using an accelerometer. Noury et al. [18] proposed a system for multimodal intuitive media browsing whereby the user can learn personalized gestures. Variations in the artificial magnetic field for motion detection are measured by magnetic primed devices, which are not preferred because of health hazards related to artificial electromagnetism. Contact-based devices are restrained by their bias toward experienced users and are thus not extensively used. Therefore, vision-based devices are used to capture inputs for gesture recognition in HCI. These devices rely on video sequences captured by one or several cameras to analyze and interpret motion [7]. Such cameras include infrared cameras that provide crisp images of gestures and can be used for night vision (Fig. 3a). Traditional monocular cameras are cheapest with variations, such as fish eye cameras for wide-angle vision and time-of-flight cameras for depth information. Stereo visionbased cameras (Fig. 3b) deliver 3D global information through embedded triangulation. Pantiltzoom cameras are used to identify details in a captured scene more precisely. Vision-based cameras also use hand markers (Fig. 3c) to detect hand motions and gestures. These hand markers can be further classified into reflective markers, which are passive in nature and shine only when strobes hit them, and light-emitting diodes, which are active in nature and flash in sequence. Each camera in these systems delivers a marker position from its view with a 2D frame that lights up with either strobe or normal lights. Preprocessing is performed to interpret the views and positions onto a 3D space.

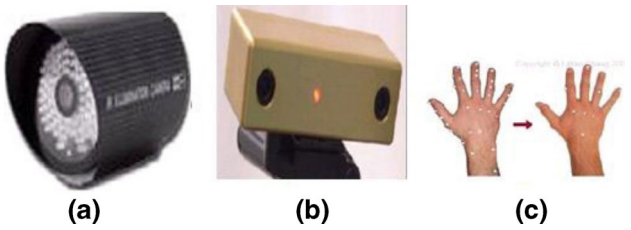

Fig. 3 infrared cameras

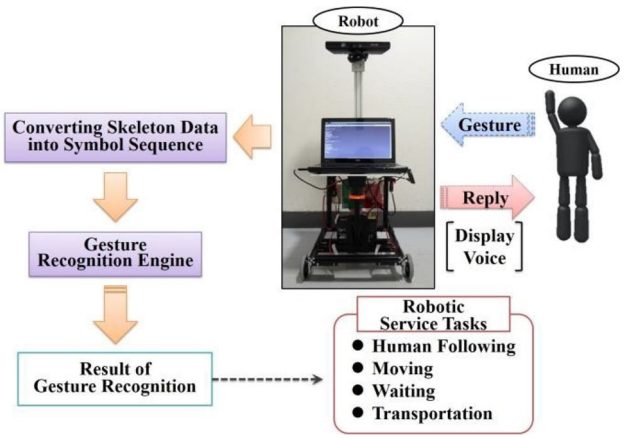

Fig. 4 Concept of the proposed service robot with gesture recognition system [19]

2.2 Service Robot with Gesture Recognition System

Figure 4 shows the concept of the proposed service robot. Firstly, the human gesture which is one of the predefined commands for the service task is given to the robot. Then, it is detected in real time and given as the position information of human arm, i.e., positions of nodes in the skeleton model, by the Kinect sensor installed in the robot. It is translated to the input signal, i.e., symbol sequence, for the recognition engine installed in the robot. After processing to recognize users command, the robot replies to the human with the display and the audio messages based on the recognition result. At the same time, the robot starts the service task ordered by the user [19].

2.3 Challenges in Vision-Based Gesture Recognition

The main challenge in vision-based gesture recognition is the large variety of existing gestures. Recognizing gestures involves handling many degrees of freedom, huge variability of 2D appearance depending on the camera viewpoint (even with the same gesture), different silhouette scales (e.g., spatial resolution), and many resolutions for the temporal dimension (e.g., variability of gesture speed). The trade-off between accuracy, performance and usefulness also requires balancing according to the type of application, cost of the solution, and several other criteria, such as real-time performance, robustness, scalability, and user-independence. In real time, the system must be able to analyze images at the frame rate of the input video to provide the user with instant feedback on the recognized gesture. Robustness significantly affects the effective recognition of different hand gestures under different lighting conditions and cluttered backgrounds. The system should also be robust against in-plane and out-of-plane image rotations. Scalability facilitates the management of a large gesture vocabulary, which may include a few primitives. Thus, this feature facilitates the users control of the composition of different gesture commands. User-independence creates an environment where the system can be controlled by different users rather than only one user and can recognize human gestures of different sizes and colors. A hand tracking mechanism was suggested to locate the hand based on rotation and zooming models. The method of hand-forearm separation was able to improve the quality of hand gesture recognition. HMMs have been used extensively in gesture recognition. For instance, HMMs were used for ASL recognition by tracking the hands based on color. An HMM consists of a set (S) of n distinct states such that S = s1, s2, s3sn, which represents a Markov stochastic process.All these enabling technologies for gesture recognition have their advantages and disadvantages. The physical contact required by contact-based devices can be uncomfortable for users, but these devices have high recognition accuracy and less complex implementation. Vision-based devices are user-friendly but suffer from configuration complexity and occlusion. The major merits and demerits of both enabling technologies are summarized in Table 4.

Table 4 Comparison between contact- and vision-based devices

Criterion	Contact-devices	Vision-devices
User cooperation	Yes	No
User intrusive	Yes	No
Precise	Yes/No	No/Yes
Flexible to configure	Yes	No
Flexible to use	No	Yes
Occlusion problem	No (Yes)	Yes
Health issues	Yes (No)	No

2.4 Hand Gestures Images Under Different Conditions

In image capture stage,as seen in Fig. 5, a digital camera Sumsung L100 with 8.2 MP and 3× optical zoom to capture

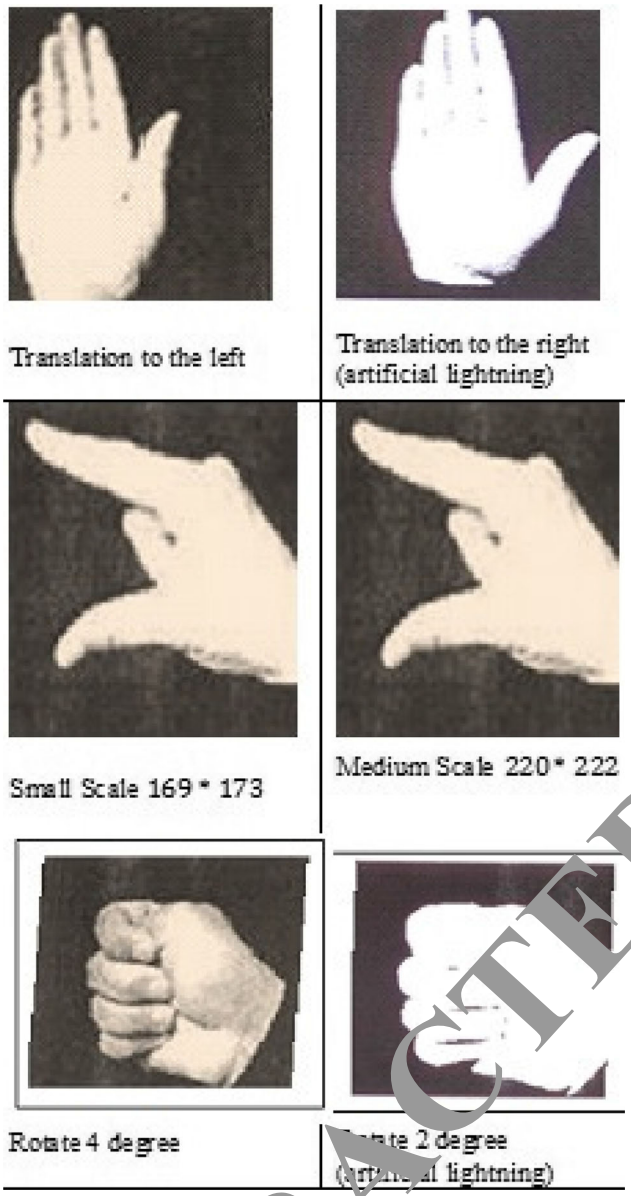

Fig. 5 Hand gestures images under different conditions

the images and each gesture is performed at various scales, translations, rotations and illuminations as follows (see Figure for some examples): 1-Translation: translation to the right and translation to the left. 2-Scaling: small scale (169173), medium scale (220222) and large scale (344348). 3-Rotation: rotate 4 degree, rotate 2 degree and rotate -3 degree. 4- Original lightning: original and artificial. Employing relatively few training images facilitates the measurement of the robustness of the proposed methods, given that the use of algorithms that require relatively modest resources either in terms of training data or computational resources is desirable [20,21]. In addition, [22] considered that using a small data set to represent each class is of practical value especially in problems where it is difficult to get a lot of examples for each class.

3 Vision-Based Gesture Taxonomies and Representations

Gesture acts a medium for nonvocal communication, sometimes in conjunction with verbal communication, to express meaningful commands. Gesture may be articulated with any of several body parts or with a combination of them. Gestures as a major part of human communication may also serve as important means for HCI. However, the meaning associated with different gestures varies with the culture, which may have an invariable or universal meaning for a single gesture. Thus, semantically interpreting gestures strictly depends on a given culture.

3.1 Vision-Based Gesture Taxonomies

Theoretically, research classifies gestures into two types, static and dynamic gestures. Static gestures refer to the orientation and position of the hand in space during an amount of time without any movement. Dynamic gestures refer to the same but with movement. Dynamic gestures include those involving body parts, such as waving the hand , whereas static gestures include single formation without movement, such as jamming the thumb and forefinger to form the OK symbol (i.e., a static pose). According to [23], 35 % of human communication consists of verbal communication, and 65 % is nonverbal gesture-based communication. Gestures can be classified into five types: emblems, affect displays, regulators, adaptors, and illustrators [24]. Emblematic, emblem, or quotable gestures are direct translations of short verbal communication, such as waving the hand for goodbye or nodding for assurance. Quotable gestures are culture-specific. Gestures conveying emotion or intention are called affect displays. Affect displays generally depend less on culture. Gestures that control interaction are called regulators. Gestures such as head shaking and quickly moving the leg to release body tension are called adaptors, which are generally habits unintentionally used during communication. Illustrator gestures emphasize key points in speech and thus inherently depend on the thought process and speech of the communicator. Illustrator gesticulations can further be classified into five sub categories: beats, deictic gestures, iconic gestures, metaphoric gestures, and cohesive gestures [24]. Beats are short, quick, rhythmic, and often repetitive gestures. Pointing to a real location, object, or person or to an abstract location or period of time is called deictic gesture. Hand movements that represent figures or actions, such as moving the hand upward with wiggling fingers to depict tree climbing, are called iconic gestures. Abstractions are depicted by metaphoric gestures. Thematically related but temporally separated gestures are called cohesive gestures. The temporal separation of these thematically related ges-

tures is due to the interruption of the communicator by another communicator.

3.2 Vision-Based Gesture Representations

Several gesture representations and models that abstract and model the movement of human body parts have been proposed and implemented. The two major categories of gesture representation are 3D modelbased and appearance-based methods. The 3-D model based gesture recognition employs different techniques for gesture representation: the 3-D textured kinematic or volumetric model, 3-D geometric model and 3-D skeleton model. Appearance-based gesture representation models include the color-based model, silhouette geometry model, deformable gabarit model, and motion-based model. The 3-D model based gesture representation defines the 3-D spatial description of a human hand for representation, with the temporal aspect being handled by automation. This automation divides the temporal characteristics of a gesture into three phases [23]: the preparation or pre-stroke phase, the nucleus or stroke phase, and the retraction or post-stroke phase. Each phase corresponds to one or more transitions of the 3-D human models spatial states. One or more cameras focus on the real target and compute parameters spatially to match the real target, and then follow its motion during the recognition process in a 3-D model. Thus, the 3-D model has an advantage in that it can update the models parameters while checking the transition consistency in the temporal model, leading to precise gesture recognition and representation. However, it becomes computationally intensive and requires dedicated hardware. Several methods [24] combine silhouette extraction with 3-D model projection fitting through the self-oriented location of a target. Three models are generally used: the 3-D textured kinematic or volumetric model provides precise details about the skeleton of the human body and information about the skin surface. 3-D textured kinematic or volumetric models are more precise than 3-D geometric models with respect to skin information, but 3-D geometric models contain essential skeleton information. Appearance-based gesture representation methods are broadly classified into two major subcategories: the 2-D static model-based methods and the motion- based methods. Each subcategory has more variants, and the commonly used 2-D models include the color-based model, which uses body markers to track the motion of the body or of a body part. Bretzner et al. [25] proposed a hand gesture recognition method employing multi-scale color features, hierarchal models, and particle filtering. Gesture tracking has a wide range of real world applications, such as augmented reality (AR), surgical navigation, ego-motion estimation for robot or machine control in industry, and in helmet-tracking systems.Recently,researchers have applied the fusion of multiple sensors to overcome the shortcomings inherent with a single sensor, andnumerous papers on sensor fusion have been published in the literature. For example, multiple object tracking has been realized by fusing acoustic sensor and visual sensor data.The visual sensor helps to overcome the inherent limitation of the acoustic sensor for simultaneous multiple object tracking, while the acoustic sensor supports the estimation when the object is occluded [26].

Silhouette geometry-based models consider several of the silhouettes geometric properties such as perimeter, convexity, surface, bounding box or ellipse elongation, rectangularity, centroid, and orientation. The geometric properties of the hand skins bounding box were used to recognize hand gestures [27]. Deformable gabarit based models are generally based on deformable active contours. Ju et al. [28] used snakes whose motion and other properties were parameterized for the analysis of gestures and actions in technical talks for video indexing. Motion-based models are used for the recognition of an object or its motion based on the motion of the object in an image sequence. A local motion histogram using an Adaboost framework was introduced by Luo et al. [29] for learning action models.

Several gesture representations and models that abstract and model the movement of human body parts have been proposed and implemented. The two major categories of gesture representation are 3D modelbased and appearance-based methods. The three-dimensional (3D) model-based gesture recognition has different techniques for gesture representation, namely, 3D-textured volumetric, 3D geometric model, and 3D skeleton model. Appearance-based gesture representation include color-based model, silhouette geometry model, deformable gabarit model, and motion-based model. The 3D model-based gesture representation defines a 3D spatial description of a human hand for temporal representation via automation. This automation divides the temporal characteristics of gesture into three phases [24], namely, the preparation or prestroke phase, the nucleus or stroke phase, and the retraction or poststroke phase. Each phase corresponds to one or more transitions of the spatial states of the 3D human model. In the 3D model, one or more cameras focus on the real target, compute parameters that spatially match this target, and follow the motion of the target during the recognition process. Thus, the 3D model has an advantage because it updates the model parameters while checking the transition matches in the temporal model. This feature leads to precise gesture recognition and representation, although making it computationally intensive requires dedicated hardware. Many methods [25] combine silhouette extraction with 3D model projection fitting by finding a self-oriented target.

Three kinds of model are generally used. 3D-textured kinematic/volumetric model contains highly detailed information on the human skeleton and skin surface. 3D geometric models are less precise than 3D-textured kine-

matic/volumetric models with regard to skin information but contain essential skeleton information. Appearance-based gesture representation methods are broadly classified into two major subcategories: 2D static model-based methods and motion-based methods. Each subcategory has several variants. The commonly used 2D models include the following:

- based model uses body markers to track the motion of a body or a body part. Bretzner et al. [27] proposed hand gesture recognition that involves multiscale color features, hierarchical models, and particle filtering.
- Silhouette geometry-based models include several geometric properties of the silhouette, such as perimeter, convexity, surface, bounding box/ellipse, elongation, rectangularity, centroid, and orientation. The geometric properties of the bounding box of the hand skin are used to recognize hand gestures [28].
- Deformable gabarit-based models are generally rooted in deformable active contours (i.e., snake parameterized with motion and their variants). Ju et al. [29] used snakes to analyze gestures and actions in technical talks for video indexing.
- Motion-based models are used to recognize an object or its motion based on the motion of the object in an image sequence. Luo et al. [30] introduced the local motion histogram that uses an Adaboost framework for learning action models.

4 Vision Based Gesture Recognition Techniques

Several common techniques used for static and dynamic gesture recognition are described as follows:

- K-means [31]: This classification searches for statistically similar groups in multi-spectral space. The algorithm starts by randomly creating k clusters in spectral space.
- K-nearest neighbors (K-NN) [32]: This is a method for classifying objects according to the closest training examples in the feature space. K-NN is a type of instance-based or lazy learning where function is only locally approximated and all computations are deferred until classification.
- Mean shift clustering [33]: The mean shift algorithm is a non-parametric clustering technique that requires no prior knowledge of the number of clusters and does not constrain cluster shape. The main idea behind mean shift is to treat the points in the d-dimensional feature space as an empirical probability density function where dense regions in the feature space correspond to the local maxima or modes of the underlying distribution.
- Support vector machine (SVM) [34]: SVM is a nonlinear classifier that produces classification results superior to those of other methods. The idea behind the method is to nonlinearly map input data to some high dimensional space, where the data can be linearly separated, and thus provide desired classification or regression results.
- Hidden Markov Model (HMM) [35]: is a joint statistical model for an ordered sequence of variables. HMM is the result of stochastically perturbing variables in a Markov chain (the original variables are thus "hidden").
- Dynamic time warping (DTW) [36]: DTW has long been used to find the optimal alignment of two signals. The DTW algorithm calculates the distance between each possible pair of points out of two signals according to their feature values. DTW uses such distances to calculate a cumulative distance matrix and finds the least expensive path through this matrix.
- Time delay neural networks (TDNNs) [37]: TDNNs are special artificial neural networks (ANNs) that work with continuous data to adapt the architecture to online networks and are thus advantageous to real-time applications. Theoretically, TDNNs are an extension of multilayer perceptrons. TDNNs are based on time delays that enable individual neurons to store the history of their input signals.
- Finite state machine (FSM) [38]: An FSM is a machine with a limited or finite number of possible states (an infinite state machine can be conceived but is impracticable). An FSM can be used both as a development tool for approaching and solving problems and as a formal way of describing solutions for later developers and system maintainers.
- Artificial neural networks (ANNs) [39]: An ANN is an information processing paradigm based on the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the structure of the information processing system. An ANN is composed of many highly interconnected processing elements (neurons) working in unison to solve specific problems. Similar to humans, ANNs learn by example [24]. A neural network consists of interconnected processing units that operate in parallel. Each unit receives inputs from other units, sums them up, and then calculates the output to be sent to other units connected to the unit.
- Template matching [40]: One of the simplest and earliest approaches to pattern recognition is based on template matching. Matching is a generic operation I-pattern recognition used to determine similarities between two entities (points, cures, or shapes) of the same type. In template matching, a template (typically a 2D shape) or prototype of the pattern to be recognized is available and is matched against the stored template while considering

Table 5 Analysis of major literature related to vision static and dynamic hand gesture recognition

Author/References	3 Model	Appearance	Static	Dynamic	Application
Cheng et al. [43]		Yes		Yes	Virtual reality application
Sangineto and Cupelli [44]		Yes	Yes		Desktop application
Wang et al. [45]		Yes	Yes		Sign language
Radkowski and Stritzke [46]	Yes			Yes	Augmented reality
Tran and Trivedi [47]	Yes			Yes	Gaming application
Rautaray and Agrawal [48]		Yes	Yes		Desktop application
Reale et al. [49]		Yes	Yes		Desktop application
Vrkonyi-Kczy and Tusor [50]	Yes		Yes	Yes	Smart environment application
Gorce et al. [51]	Yes			Yes	Desktop application
Sajjawiso and Kanongchaiyos [52]	Yes		Yes		Desktop application
Henia and Bouakaz [53]	Yes			Yes	Virtual reality application
Ionescu et al. [54]		Yes		Yes	Gaming application
Yang et al. [55]		Yes	Yes		Desktop application
Huang et al. [56]		Yes	Yes		Desktop application
Ionescu et al. [57]		Yes	Yes	Yes	Entertainment application
Bergh and Gool [58]		Yes	Yes		Entertainment application
Bao et al. [59]		Yes		Yes	Virtual reality application
Bellarbi et al. [60]		Yes		Yes	Desktop application
Hackenberg et al. [61]			Yes	Yes	Virtual reality application
Du et al. [62]	Yes		Yes		Virtual reality application
Rautaray and Agrawal [63]		Yes	Yes		Entertainment application
Visser and Hopf [64]		Yes	Yes		Desktop application
He et al. [65]		Yes	Yes		Gaming application
Tan et al. [66]		Yes	Yes		Gaming application
Ho et al. [67]	Yes		Yes		Desktop application
Huang et al. [68]		Yes	Yes	Yes	Desktop application
Hsieh et al. [69]		Yes		Yes	Desktop application

all allowable poses (translation and rotation) and scale changes.

Back-propagation learning algorithm Example: Basically the error back-propagation process consists of two passes through the different layers of the network a forward pass and a backward pass [41,42]. The algorithm is as follows [42]:

(1) Step 0. Initialize weights. (Set to small random values)
(2) Step 1. While stopping condition is false do steps 2:9
(3) Step 2. For each training pair do steps 3:8 Feed-forword:
(4) Step 3. Each input unit (X i , i = 1,…, n) receives input signal X i And broadcasts this signal to all units in the layer above (the Hidden units).
(5) Step 4. Each hidden unit (Z j , j = 1,…, p) sums its weighted input Signals, Zin = Voj + X i Vij i =1,V0 j Bias on hidden unit j.Vij Weight between input unit and hidden unit. Applies its activation function to compute its output signal
(6) Step 5. Each output unit (Yk , k = 1,…, m) sums its weighted input signals, yink = Wok + Z j W jk Back propagation of error:
(7) Step 6. Each output unit (Yk , k = 1,…, m) receives a target pattern corresponding to the input training pattern, computers its error information term, computers its error information term,
(8) Step 7. Each hidden unit (z j , j = 1,…, p) sums its delta inputs (from unit in the layer above),
(9) Step 8. Each output unit (Yk , k = 1,…, m) updates its bias and weights (j =0,…, p).

5 Analysis of Existing Literature

Research on hand gesture recognition has significantly evolved with the increased usage of computing devices in daily life. This section surveys studies on HCI to classify them according to the gesture representations used for man

Table 6 List of several commercial products and software

Name	Applications	Technology	Interface	Product
SoftKinetic IISU SDK,	Build gesture recognition application compatible with all 3D cameras and platforms	Real-time 3D gesture recognition software platform	PC, consoles, Smart TVs, Set Top Boxes	Camera and software
Hand GKET,	Toolkit facilitates integration of hand gesture control with games and VR applications	This middleware recognizes user's hand gestures and generates keyboard or mouse event to control applications in computer using computer vision techniques	PC based interface	Camera and software
Mgestyk,	Interaction with computer tooperate games and application	Software for hand-gesture processing and 3D camera	PC based interface	Camera and software
Wii Nintendo,	Wireless and motion sensitive remote with game console	Game with any TV, computer etc.	Interface in the screen	Game consoleand remotecontrol
HandVu,	Interaction with computer to operate games and application	Real time gesture recognition using computer vision techniques	PC based interface	Camera and software

and machine interaction. Table 5 classifies previous hand gesture interaction research based on gesture representations used, such as 3D modelbased or appearance-based gesture representations and the techniques used in proposed systems from 2005 to 2012. An exhaustive list of these studies is given in Table 5. The list has been made as exhaustive as possible. Detailed analysis of the table reveals interesting facts about ongoing research on HCI in general and vision-based hand gesture recognition in particular. Literature presents various interesting facts that compare and contrast the two object representation techniques: 3D model- and appearance-based. The 3D modelbased representation technique is based on computer-aided design through a wired model of the object, whereas the appearance-based technique segments the potential region with an object of interest from the given input sequence. Although the 3D model allows for real-time object representation along with minimal computing effort, the major difficulty with this approach is that the system can handle only a limited number of shapes. The appearance-based model uses global and local feature extraction approaches. Local feature extraction has high precision with respect to the accuracy of shapes and format. Appearance-based methods use templates to correlate gestures to a predefined set of template gestures and thus simplify parameter computations. However, the lack of precise spatial information impairs the suitability of the method for manipulative postures or gestures as well as their analysis. Appearance-based models are sensitive to viewpoint changes and thus cannot provide precise spatial information. This constraint makes these models less preferred for more interactive and manipulative applications.

6 Commercial Products and Softw

Hand gestures are considered a promising research focus for designing natural and intuitive methods for HCI for myriad computing domains and applications. This section shows several commercially available products and software based on vision-based hand gesture recognition technology for interaction with varied applications. However, these commercial products are still in the initial phases of acceptance but may still be made robust with user requirements and feedback. The criteria considered for designing and developing such products and technological constraints limit the capabilities of these products, which have to be supported by research and development in the associated technological areas. These products should be modified in terms of cost-effectiveness, robustness under different real-life and real-time application environments, effectivity, and end-user acceptability. Table 6 lists several vision-based hand gesture recognition commercial products and software available for interacting with the computing world.

7 Conclusion

Numerous methods for gestures, taxonomies, and representations have been evaluated for core technologies proposed in gesture recognition systems. However, these evaluations do not depend on standard methods in some organized format but have been conducted based on increasing usage in gesture recognition systems. Thus, an analysis of the surveys presented in the paper indicates that appearance- based gesture representations are preferred over 3D-based gesture representations in hand gesture recognition systems. Despite the considerable information and research publications on both techniques, the complexity of implementation of 3D modelbased representations makes them less preferred. The existing state of the applications of gesture recognition systems indicates that desktop applications are the most implemented applications for gesture recognition systems. Future research in gesture recognition systems will provide an opportunity for researchers to create efficient systems that overcome the disadvantages associated with core technologies in the existing state of enabling technologies for gesture representations and recognition systems as a whole. Industrial applications also require specific advances in man-to-machine and machine-to-machine interactions.

References

1. Just, A.: Two-handed gestures for human–computer interaction. PhD Thesis (2006)
2. Hasan, H., Abdul-Kareem, S.: Fingerprint image enhancement and recognition algorithms: a survey. Neural Comput. Appl. (2012). doi:10.1007/s00521-012-1113-0
3. Hasan, H., Abdul-Kareem, S.: Static hand gesture recognition using neural networks. Artif. Intell. Rev. (2012). doi:10.1007/s10462-011-9303-1
4. Karam, M.: A framework for research and design of gesture-based human computer interactions. PhD Thesis, University of Southampton (2006)
5. Moeslund, T., Granum, E.: A survey of computer vision based human motion capture. Comput. Vis. Image Underst. **81**, 231–268 (2001)
6. Derpanism, K.G.: A Review of vision-based hand gestures. http://cvr.yorku.ca/members/gradstudents/kosta/publications/file (2004). Gesture review
7. Mitra, S., Acharya, T.: Gesture recognition: a survey. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **37**(3), 311–324 (2007)
8. Chaudhary, A., Raheja, J.L., Das, K., Raheja, S.: Intelligent approaches to interact with machines using hand gesture recognition in natural way: a survey. Int. J. Comput. Sci. Eng. Surv. **2**(1), 122–133 (2011)
9. Wachs, J.P., Kolsch, M., Stern, H., Edan, Y.: Vision-based hand-gesture applications. Commun. ACM **54**, 60–71 (2011)
10. Corera, S., Krishnarajah, N.: Capturing hand gesture movement: a survey on tools techniques and logical considerations. In: Proceedings of Chi Sparks 2011 HCI Research, Innovation and Implementation, Arnhem, Netherlands. http://proceedings.chi-sparks.nl/documents/Education-Gestures/FP-35-AC-EG (2011)

11. Kevin, N.Y.Y., Ranganath, S., Ghosh, D.: Trajectory modeling in gesture recognition using CyberGloves. In: TENCON 2004, IEEE Region 10 Conference (2004)
12. Ionescu, B.: Dynamic hand gesture recognition using the skeleton of the hand. EURASIP J. **2005**, 1–9 (2005)
13. Symeonidis, K.: Hand gesture recognition using neural networks. Neural Netw. **13**, 1–5 (1996)
14. Kanniche, M.B.: Gesture recognition from video sequences. PhD Thesis, University of Nice, 2009 (2009)
15. Webel, S., Keil, J., Zoellner, M.: Multi-touch gestural interaction in X3D using hidden markov models. In: VRST 08 Proceedings of the 2008 ACM symposium on Virtual reality software and technology, pp. 263–264. ACM, New York, NY, USA (2008)
16. Schlomer, T., Poppinga, B., Henze, N., Boll, S.: Gesture recognition with a wii controller. In: TEI 08 Proceedings of the 2nd International Conference on Tangible and Embedded Interaction, pp. 11–14. ACM, New York, NY, USA (2008)
17. Bourke, A., Brien, J.O., Lyons, G.: Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. Gait Posture **26**(2),194–199 (2007). http://www.sciencedirect.com/science/article/B6T6Y-4MBCJHV-1/2/f87e4f1c82f3f93a3a5692357e3fe00c
18. Noury, N., Barralon, P., Virone, G., Boissy, P., Hamel, M., Rumeau, P.: A smart sensor based on rules and its evaluation in daily routines. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2003, vol. 4, pp. 3286–3289 (2003)
19. Tatsuya, F., Jae, H.: Gesture recognition system for human–robot interaction and its application to robotic service task. In: Proceedings of the International Multi Conference of Engineers and Computer Scientists, voI 1, IMECS, pp. 63 (2014)
20. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. Comput. Vis. Image Underst. **106**(1), 59–70 (2007)
21. Kanan, C., Cottrell, G.: Robust classification of objects, [illegible] and flowers using natural image statistics. In: Paper Presented [illegible] the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2010)
22. Guodong, G., Dyer, C.R.: Learning from examples in the small sample case: face expression recognition. IEEE Trans. Syst. Man Cybern. Part B Cybern. **35**(3), 477–488 (2005). doi:10.1109/TSMCB.2005.846658
23. Hall, E.T.: The Silent Language. Anchor Books. ISBN: 13 978-0385055499 (1973)
24. McNeill, D.: Hand and Mind: What Gestures Reveal About Thought. University Of Chicago Press. ISBN: 9780226561325,1992-99
25. Boulay, B.: Human posture recognition for behavior understanding. PhD thesis, Universite de Nice-Sophia Antipolis (2007)
26. Zhou, S., Fei, F., Li, G., Mai, J.D.: 2D human gesture tracking and recognition by the fusion of MEMS inertial and vision sensors. IEEE Sens. J. **14**(4), 1160–1170 (2014)
27. Bretzner, L., Laptev, I., Lindeberg, T.: Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. In: Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 405–410. doi:10.1109/AFGR.2002.1004190 (2002)
28. Binh, A., Hassanpour, R.: Region based hand gesture recognition. In: 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, pp. 17 (2008)
29. Ju, S.X., Black, M.J., Minneman, S., Kimber, D.: Analysis of gesture and action in technical talks for video indexing. In: Technical report, American Association for Artificial Intelligence. AAAI Technical Report SS-97-03 (1997)
30. Luo, Q., Kong, X., Zeng, G., Fan, J.: Human action detection via boosted local motion histograms. Mach. Vis. Appl. (2008). doi:10.1007/s00138-008-0168-5
31. Lindsay, J.: K-Means Classifier Tutorial. http://www.uoguelph.ca/hydrogeo/Whitebox/Help/kMeansClass.html (2009)
32. Thirumuruganathan, S.: A Detailed Introduction to K-Nearest Neighbor (KNN) Algorithm. http://saravananthirumuruganathan.wordpress.com/2010/05/17/a-detailed-introduction-to-k-nearest-neighbor-knnalgorithm/ (2010)
33. Derpanis, K.G.: Mean Shift Clustering. In: Lecture Notes. http://www.cse.yorku.ca/kosta/CompVisNotes/meanshift.pdf (2005)
34. Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. Kluwer Academic Publishers, Boston (1998)
35. Ramage, D.: Hidden Markov Models Fundamentals. In: Lecture Notes. http://cs229.stanford.edu/section/cs229-hmm (2007)
36. Senin, P.: Dynamic Time Warping Algorithm Review. In: Technical Report. http://csdl.ics.hawaii.edu/techreports/08-04/08-04 (2008)
37. Wohler, C., Anlauf, J.K.: An adaptable time-delay neural-network algorithm for image sequence analysis. IEEE Trans. Neural Netw. **10**(6), 1531–1536 (1999)
38. Holzmann, G.J.: Finite State Machine Ebook. http://www.spinroot.com/spin/Doc/Book91_PDF/F1
39. Stergiou, C., Siganos, D.: Neural networks. http://www.doc.ic.ac.uk/-nd/surprise_96/journal/vol4/cs11 (1996)
40. Jain, A.K., Duin, R.P.W., Jianchang, M.: Statistical pattern recognition: a review. IEEE Trans. Pattern Anal. Mach. Intell. **22**(1), 4–37 (2000)
41. Haykin, S.S.: Neural Networks: A Comprehensive Foundation. Prentice Hall, Englewood Cliffs (2007)
42. Kinnebrock, W.: Neural Network, Fundamentals, Applications, Examples. Galotia Publications, Delhi (1995)
43. Cheng, J., Xie, X., Bian, W., Tao, D.: Feature fusion for 3D hand gesture ecognition by learning a shared hidden space. Pattern Recognit. Lett. **33**, 476–484 (2012)
44. Sangineto, E., Cupelli, M.: Real -time viewpoint-invariant hand localization with cluttered backgrounds. Image Vis. Comput. **30**, 26–37 (2012)
45. Wang, G.W., Zhang, C., Zhuang, J.: An application of classifier combination methods in hand gesture recognition. Math. Probl. Eng. **2012**, 1–17 (2012). doi:10.1155/2012/346951
46. Radkowski, R., Stritzke, C.: Interactive hand gesture-based assembly for augmented reality applications. In: ACHI: The Fifth International Conference on Advances in Computer–Human Interactions, IARIA, pp. 303–308 (2012)
47. Tran, C., Trivedi, M.M.: 3-D posture and gesture recognition for interactivity in smart spaces. IEEE Trans. Ind. Inform. **8**(1), 178–187 (2012)
48. Rautaray, S.S., Agrawal, A.: Real time hand gesture recognition system for dynamic applications. Int. J. UbiComp **3**(1), 21–31 (2012)
49. Reale, M.J., Canavan, S., Yin, L., Hu, K., Hung, T.: A multi-gesture interaction system using a 3-D iris disk model for gaze estimation and an active appearance model for 3-D hand pointing. IEEE Trans. Multimed. **13**(3), 474–486 (2011)
50. Vrkonyi-Kczy, A.R., Tusor, B.: Human–computer interaction for smart environment applications using Fuzzy hand posture and gesture models. IEEE Trans. Instrum. Meas. **60**(5), 1505–1514 (2011)
51. Gorce, M.D.L., Fleet, D.J., Paragios, N.: Model-based 3D hand pose estimation from monocular video. IEEE Trans. Pattern Anal. Mach. Intell. **33**(9), 1793–1805 (2011)

52. Sajjawiso, T., Kanongchaiyos, P.: 3D hand pose modeling from uncalibrate monocular images. In: Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 177–181 (2011)
53. Henia, O.B., Bouakaz, S.: 3D hand model animation with a new data-driven method. In: Workshop on Digital Media and Digital Content Management, IEEE, pp. 72–76 (2011)
54. Ionescu, D., Ionescu, B., Gadea, C., Islam, S.: A multimodal interaction method that combines gestures and physical game controllers. In: Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN), IEEE, pp. 1–6 (2011)
55. Yang, J., Xu, J., Li, M., Zhang, D., Wang, C.: A real-time command system based on hand gesture recognition. In: Seventh International Conference on Natural Computation, pp. 1588–1592 (2011)
56. Huang, D., Tang, W., Ding, Y., Wan, T., Wu, X., Chen, Y.: Motion capture of hand movements using stereo vision for minimally invasive vascular interventions. In: Sixth International Conference on Image and Graphics, pp. 737–742 (2011)
57. Ionescu, D., Ionescu, B., Gadea, C., Islam, S.: An intelligent gesture interface for controlling TV sets and set-top boxes, In: 6th IEEE International Symposium on Applied Computational Intelligence and Informatics, pp. 159–164 (2011)
58. Bergh, M., Gool, L.: Combining RGB and ToF cameras for real-time 3D hand gesture interaction. In: Workshop on Applications of Computer Vision (WACV), IEEE, pp. 66–72 (2011)
59. Bao, J., Song, A., Guo, Y., Tang, H.: Dynamic hand gesture recognition based on SURF tracking. In: International Conference on Electric Information and Control Engineering (ICEICE), pp. 338–341 (2011)
60. Bellarbi, A., Benbelkacem, S., Zenati-Henda, N., Belhocine, M.: Hand gesture interaction using color-based method for tabletop interfaces. In: IEEE 7th International Symposium on Intelligent Signal Processing (WISP), pp. 16 (2011)
61. Hackenberg, G., McCall, R., Broll, W.: Lightweight palm and finger tracking for real-time 3D gesture control. In: IEEE Virtual Reality Conference (VR), pp. 19–26 (2011)
62. Du, H., Xiong, W., Wang, Z.: Modeling and interaction of virtual hand based on virtools. In: International Conference on Multimedia Technology (ICMT), pp. 416–419 (2011)
63. Rautaray, S.S., Agrawal, A.: A novel human computer interface based on hand gesture recognition using computer vision techniques. In: International Conference on Intelligent Interactive Technologies and Mult imedia (IITM-2011), pp. 292–296 (2011)
64. Visser, M., Hopf, V.: Near and far distance gesture tracking for 3D applications. In: 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), pp. 1–4 (2011)
65. He, G.F., Kang, S.K., Song, W.C., Jung, S.T.: Real-time gesture recognition using 3D depth camera. In: 2nd International Conference on Software Engineering and Service Science (ICSESS), pp. 187–190 (2011)
66. Tan, T., De, Geo, Z.M.: Research of hand positioning and gesture recognition based on binocular vision. In: EEE International Symposium on Virtual Reality Innovation 2011, pp. 311–315 (2011)
67. Ho, M.F., Tung, C.Y., Chen, C.C., Huang, C.L.: A multi-view vision-based hand motion capturing system. Pattern Recognit. **44**, 443–453 (2011)
68. Huang, D.Y., Hu, W.C., Chang, S.H.: Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination. Expert Syst. Appl. **38**(5), 60316042 (2011)
69. Hsieh, C.C., Liou, D.H., Lee, D.: A real time hand gesture recognition system using motion history image. In: 2nd International Conference on Signal Processing Systems (ICSPS), pp. 394–398 (2010)