Gesture Control for Consumer Electronics

Caifeng Shan

Abstract The user interfaces of Consumer Electronics (CE) have been limited to devices such as remote control and keypad for a long time. With digital contents becoming more and more complex and interconnected, consumers are expecting more natural and powerful user interfaces. Automatic recognition of human gestures provides a promising solution for natural user interfaces. Recent years have witnessed much interest on gesture control in CE industry. In this chapter, we present a review on gesture control technologies for CE devices. We introduce different sensing technologies and then focus on camera-based gesture sensing and interpretation. Computer vision research on different steps, including face/hand detection, tracking, and gesture recognition, are discussed. We also introduce the latest developments on gesture control products and applications.

1 Introduction

The user interfaces of Consumer Electronics (CE) products (e.g., TV) have been limited to devices such as remote control and keypad for a long time. These interfaces are neither natural nor flexible for users and limit the speed of interaction. With digital contents becoming more and more complex and interconnected, consumers are expecting natural and efficient user interfaces.

Audition and vision are two important modalities for human–human interaction. The more efficient and powerful user interfaces can be achieved if the machines could "listen" and "see" as humans do. Automatic speech recognition has been well studied, and many commercial systems have been available. Voice-based interfaces have the advantage of a preestablished vocabulary (natural language). However, it may be inappropriate both for the protracted issuing of commands and for changing parameters by increments such as volume control. Moreover, in noisy environments

C. Shan (✉)
Philips Research, Eindhoven, The Netherlands
e-mail: caifeng.shan@philips.com

L. Shao et al. (eds.), *Multimedia Interaction and Intelligent User Interfaces,* Advances in Pattern Recognition,
DOI 10.1007/978-1-84996-507-1_5,

(both indoor and outdoor), it is difficult to use voice control. On the contrary, vision-based interfaces provide a promising alternative in many cases. With the advances of sensing hardware and computing power, visual sensing and interpretation of human motion has received much interest in recent years.

Human gestures are meaningful or intentional body movements, i.e., physical movements of the fingers, hands, arms, head, face, or body, for example, hand gestures, head pose and movements, facial expressions, eye movements, and body gestures. Gestures can be used as replacement for speech words or used together with speech words. As a universal body language, the gesture is one of the most natural and effective means for humans to communicate nonverbally. The ability to recognize gestures is indispensable and important for successful interpersonal social interaction. Automatic recognition of human gestures is a key component for intelligent user interfaces. Gesture recognition has been an active research area in multiple disciplines including natural language processing, computer vision, pattern recognition, and human–computer interaction [1–3]. Recent years have witnessed much interest on gesture control in CE industry [4, 5]. Gesture control has many applications, for example, virtual remote control for a TV or other home appliances, gaming, and browsing public information terminals in museums, window shops, and other public spaces. In recent Consumer Electronics Shows (CES), many companies showed prototypes or upcoming products with gesture control.

In this chapter, we present an overview on gesture recognition technologies for CE devices. The human body can express a huge variety of gestures, and hand and arm gestures have received the most attention in research community [2]. We introduce different sensors that can be used for gesture sensing and then focus on camera-based computer vision technologies. Three main components of vision-based gesture recognition, including face/hand detection, tracking, and gesture recognition, are discussed. We also introduce the latest developments on gesture control products and applications. The chapter is organized as follows. We introduce the sensing technologies in Sect. 2. Section 3 discusses the existing research on vision-based gesture recognition. We describe the gesture-control applications and products in Sect. 4. Finally, Sect. 5 concludes the paper with discussions.

2 Sensing Technologies

Different sensing technologies can be used for gesture recognition. Instrumented gloves (including exoskeleton devices mounted on the hand and fingers) can be wear to measure the position and configuration of the hand. Similarly, in some optical systems, markers are placed on the body in order to measure body motion accurately. Two types of markers, passive, such as reflective markers, and active, such as markers flashing LED lights, can be used. Although these methods can provide reliable and precise gesture data (e.g., parameters of hand and fingers), the user has to wear the expensive and cumbersome device with reduced comfort; the calibration needed can also be difficult [2]. Therefore they are too intrusive for mainstream use in CE devices. In the following, we introduce some enabling technologies that

can be considered for CE devices. These sensors can be categorized into two kinds: (1) contact-based sensors, for example, multitouch screen and accelerometer, and (2) contact-free sensors such as cameras.

Haptics Gestures can be sensed through haptic sensors. This is one of the commonly used gesture-sensing technologies in current CE devices, for instance, touch or multitouch screens (e.g., tablet PC and Apple iPhone). It is similar to recognizing gestures from 2D input devices such as a pen or mouse. In [6], multitouch gestural interactions were recognized using Hidden Markov Models (HMM). Haptic gesture sensing and interpretation is relatively straightforward as compared with vision-based techniques. However, it requires the availability of a flat surface or screen, and the user has to touch the surface for input. This is often too constraining, and techniques that allow the user to move around and interact in more natural ways are more compelling [2].

Handhold Sensors Another approach to gesture recognition is the use of handhold sensors. For example, in a presentation product from iMatt [7], the presenter can interact with the projector and screen using gestures, which are sensed by a handhold remote control. Similarly, Cybernet Systems [8] developed a weather map management system enabling the meteorologist to control visual effects using hand gestures that are sensed with a handhold remote control. Accelerometers and gyroscopes [9] are two types of sensors used, which measure the variation of the earth magnetic field in order to detect the motion. The Wii-mote from Nintendo uses built-in accelerometers to measure the game player's gestures. Another example is the MX Air Mouse from Logitech, which can be waved around to control programs via gestures, based on the built-in accelerometers. Since the user has to hold the sensor, this technique is often intrusive, requiring the user's cooperation.

Vision Vision-based gesture control relies on one or several cameras to capture the gesture sequences; computer vision algorithms are used to analyze and interpret captured gestures. Although, as discussed above, some vision systems require the user to wear special markers, vision-based techniques have focused on marker-free solutions. With camera sensors becoming low-cost and pervasive in CE products, vision technologies have received increasing attention, which allow unobtrusive and passive gesture sensing. Different kinds of camera sensors have been considered. Near Infrared (IR) cameras can be used to address insufficient lighting or lighting variations [10, 11]. Stereo cameras or time-of-flight cameras can deliver the depth information, which enables more straightforward and accurate gesture recognition. Vision-based gesture recognition approaches normally consist of three components: body part detection, tracking, and gesture recognition. We will discuss these in details in Sect. 3.

Ultrasound Ultrasonic sensors can also be used to detect and track gestures. For example, NaviSense [12] and EllipticLabs [13] developed ultrasound-based finger/hand gesture recognition systems (illustrated in Fig. 1). The iPoint system from

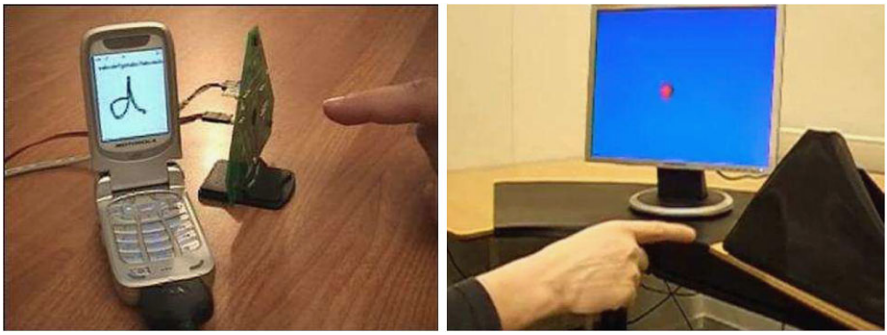

Fig. 1 Ultrasound based gesture control from NaviSense [12] (*Left*) and EllipticLabs [13] (*Right*)

NaviSense is able to track finger movements to navigate and control a cursor on the display, which can be used in mobile devices to support touchless messaging. The problems of using ultrasonic sensors were discussed in [9, 14].

Infrared Proximity Sensing Recently Microsoft [15] has developed a gesture control interface for mobile phones based on IR proximity sensors. As shown in Fig. 2, IR signal is shone outwards from the device via a series of IR LEDs embedded along each side; reflections from nearby objects (e.g., fingers) are sensed using an array of IR photodiodes. When the device is put on a flat surface (e.g., table), the user can perform single and multitouch gestures using the space around the mobile device. In the Virtual Projection Keyboard [16], an image of the full-size keyboard is projected onto a flat surface. When the user presses a key on the projected keyboard, the IR layer is interrupted; the reflections are recognized in three dimensions (Fig. 2).

Each sensing technology has its limitations, so it is promising to combine different sensors for better gesture recognition. However, the integration of multiple sensors is complex, since each technology varies along several dimensions, including accuracy, resolution, latency, range of motion, user comfort, and cost.

3 Vision-Based Gesture Recognition

A first prototype of vision-based gesture control for CE devices can be tracked back to 1995 [17], when Freeman and Weissman developed a gesture control for TVs. As shown in Fig. 3, by exploiting the visual feedback from the TV, their system enables a user to adjust graphical controls by moving the hand. A typical interaction session is the following: (1) TV is off but searching for the trigger gesture (open hand); (2) When TV detects the trigger gesture, TV turns on, and the hand icon and graphics overlays appear; (3) The hand icon follows the user's hand movement, and a command is executed when the hand covers a control for 200 ms; (4) User closes hand to leave the control mode, and the hand icon and graphical control disappear

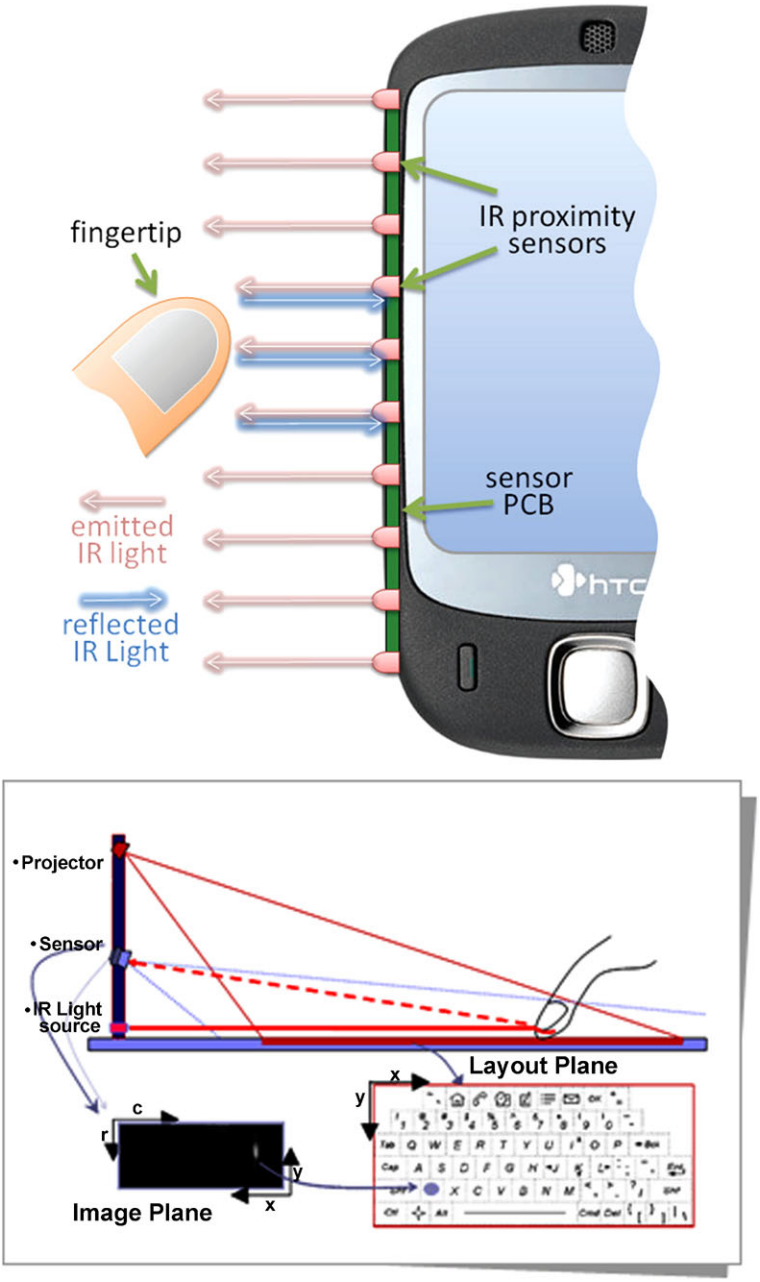

Fig. 2 IR reflection-based gesture control: (*Top*) SideSight [15]; (*Bottom*) Virtual Projection Keyboard [16]

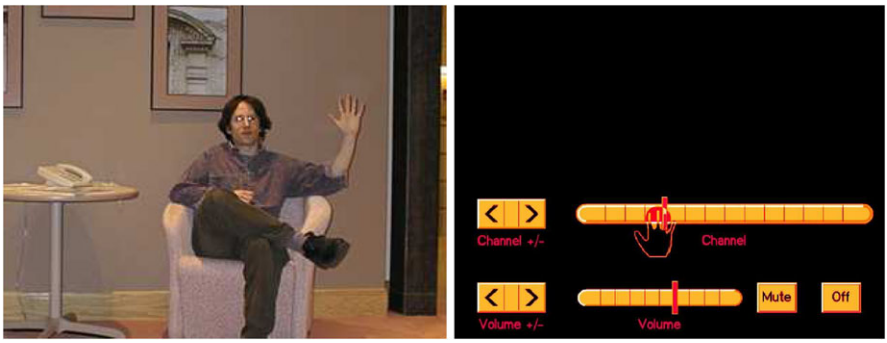

Fig. 3 Gesture control for TV [17]: the tracked hand is echoed with a hand icon on the TV

after one second. This gesture recognition system was also applied to interactive video games [18].

The general approach to vision-based gesture recognition consists of three steps: body part detection, tracking, and gesture recognition. The first step is to automatically find the body part of interest (e.g., face, hand, etc.) in the input image. Initialized by the detection, a visual tracking method is normally adopted to track the body part over time. Based on the tracked (or detected) body part, gesture recognition is thereafter performed, which can be static posture recognition in the single frame or dynamic gesture recognition in the sequence. In this section, we review research in each of these steps.

3.1 Body Part Detection

The face and hands are the major body parts involved in gesture interaction, with the ability of expressing a huge number of gestures. Most of gesture recognition systems developed so far target recognizing hand gestures and face/head gestures. In these systems, face detection and/or hand detection are required as the first step. In the following, we introduce existing work in these topics.

Face detection plays a crucial role in face-related vision applications. Due to its practical importance, face detection has attracted a great amount of interest, and numerous techniques have been investigated over the years (see [19] for a survey). In these methods, facial features, such as edge, intensity, shape, texture, and color, are extracted to locate the faces using statistical or geometric models. The face detection scheme proposed by Viola and Jones [20, 21] is arguably the most commonly employed frontal face detector, which consists of a cascade of classifiers trained by AdaBoost employing Haar-wavelet features. AdaBoost [22, 23] provides a simple yet effective approach for stagewise learning of a nonlinear classification function. Later their approach was extended with rotated Haar-like features and different boosting algorithms [24]. In [25], by incorporating Floating Search into AdaBoost, FloatBoost was proposed for improved performance on multiview face detection.

Many other machine learning techniques, such as Neural Network and Support Vector Machine (SVM), have also been introduced for face detection. In [26], the Bayes classifier was adopted with discriminating feature analysis for frontal face detection. The input image, its 1D Haar-wavelet representation, and its amplitude projections are combined to derive a discriminating feature vector. Later the features were extended and combined with an SVM-based classifier [27]. To improve the detection efficiency, Garcia and Delakis [28] designed a convolutional neural network for face detection, which performs simple convolutional and subsampling operations. More recently, the approach in [26], Viola and Jones's approach [20, 21], and the approach in [28] are modified and combined for a fast and robust face detector in [29]. Overall, face detection technique is fairly mature, and a number of reliable face detectors have been built based on existing approaches.

Compared to face detection, less work has been done on finding hands in images [30]. Most earlier attempts to hand detection make assumptions or place restrictions on the environment. For example, in the prototype developed in [17], a hand template was used for hand detection and tracking based on normalized correlation of local orientations. Their approach works with clean background and could fail in case of cluttered background. Skin color is one of the distinctive features of hands. Zhu et al. [31] presented a skin color-based hand detector. Skin color can be modeled in different color spaces using nonparametric (e.g., color histograms) or parametric (e.g., Gaussian Mixture Models) methods. Skin color-based methods may fail if skin-colored objects exist in background. Furthermore, lighting conditions (e.g., insufficient lighting) could also make them less reliable. With a skin color prior, Bretzner et al. [32] used multiscale blob detection of color features to detect an open hand with possibly some of the fingers extended. Kölsch and Turk [33] presented an approach to finding hands in grey-level images based on their appearance and texture. They studied view-specific hand detection following the Viola and Jones' method [20]. To address the high computational cost in training, a frequency analysis-based method was introduced for instantaneous estimation of class separability, without the need for training. In [34], a hand detector was built based on boosted classifiers, which achieves compelling results for view- and posture-independent hand detection.

Considering gradient features could better encode relevant hand structures, Zondag et al. [35] recently investigated Histogram of Oriented Gradients (HOG) features for real-time hand detector. Cluttered background and variable illumination were considered in their data (shown in Fig. 4). Toshiba has developed a gesture control system for displays [36, 37]. The system initially performs face detection. Once a face is detected, the user is prompted to show an open hand gesture within the area below the face (as shown in Fig. 5), which works for multiple users. The scale of the detected face is used to define the size of the interaction area, and the areas are ordered according to scale, giving easier access to users who are closer to the camera. The first detection of an open hand triggers the gesture tracking and recognition. Face recognition is also triggered by hand detection, and the content and functionality can be customized according to the user's profile.

Fig. 4 Positive and negative examples for hand detection [35]

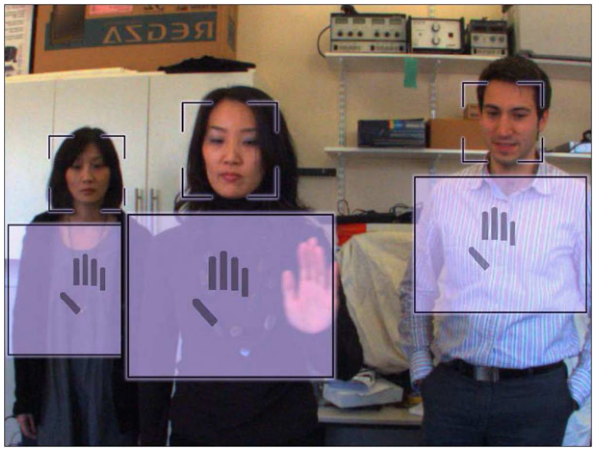

Fig. 5 Toshiba's gesture control is initialized by face detection and hand detection [36, 37]

3.2 Gesture Tracking

After detecting the body part of interest (e.g., face or hand), a tracking method is usually needed to track the gesture over time. Visual tracking in complex environments, a challenging issue in computer vision, has been intensively studied in the last decades (see [38] for a survey). Here we review relevant work on gesture tracking, mainly hand tracking and face/head tracking.

Hand tracking, aiming to estimate continuous hand motion in image sequences, is a difficult but essential step for hand gesture recognition. A hand can be represented by contours [39, 40], fingertips [41], color [42], texture, and so on. The edge feature-based hand tracker in [17] works when the hand moves slowly but tends to be unstable when motion blur occurs. Isard and Blake [40] adopted parameterized B-spline curves to model hand contours and tracked hands by tracking the deformed curves. However, the contour-based trackers usually constrain the viewpoint [39] and assume that hands keep several predefined shapes. Oka et al. [41] exploited

fingertips for hand tracking. Many color-based trackers have been utilized to track hand motion based on skin color cues [42, 43].

In order to overcome limitations of each individual feature, many approaches have considered multiple cues for hand tracking [44–48]. In [45], the contour-based hand tracker was augmented by skin-colored blob tracking. Huang and Reid [44] developed a Joint Bayes Filter for tracking and recognition of the articulated hand motion, where particle filtering [49] was adopted for color-region tracking to assist HMM in analyzing hand shape variations. Kölsch and Turk [47] presented a multi-cue tracker that combines color and short tracks of local features under "flocking" constraints; the color model is automatically initialized from hand detection. However, this approach struggles with rapid hand motion and skin-colored background objects. In [50], we combined particle filtering and mean shift [51, 52] for real-time hand tracking in dynamic environments, where skin color and motion were utilized for hand representation. In [36], normalized cross-correlation (NCC) was adopted for frontal fist tracking, which works for slow hand motion. In case of failure, a second tracker using color and motion (CM) was used. NCC tracker and CM tracker were switched online, and a Kalman filter was used to combine the estimates with a constant-velocity dynamic model.

Another kind of approaches to hand tracking is based on 3D model [53–57]. These methods have the ability to cope with occlusion and self-occlusion and can potentially obtain detailed and accurate gesture data. Usually, the state of a hand is estimated by projecting the prestored 3D hand model to the image plane and comparing it with image features. Lu et al. [53] presented a model-based approach to integrate multiple cues, including edges, optical flow, and shading information, for articulated hand motion tracking. In [58], the eigen-dynamics was introduced to model the dynamics of natural hand motion. Hand motion was modeled as a high-order stochastic linear dynamic system (LDS) consisting of five low-order LDSs, each of which corresponds to one eigen-dynamics. Sudderth et al. [55] adopted nonparametric belief propagation for visual tracking of a geometric hand model. 3D hand tracking can also base on 3D data obtained by stereo cameras or scanners [59]. In [60], a 3D search volume was set for efficient palm tracking using two cameras.

Face/head tracking has been widely studied in the literature because of its practical importance. Reliable head tracking is difficult due to appearance variations caused by the nonrigid structure, occlusions, and environmental changes (e.g., illumination). 3D head models have been utilized to analyze head movements. Basu et al. [61] adopted a 3D rigid ellipsoidal head model for head tracking, where the optical flow was interpreted in terms of rigid motions. Cascia et al. [62] presented a 3D cylinder head model and formulated head tracking as an image registration problem in the cylinder's texture map image. To avoid the troubles of 3D model maintenance and camera calibration, view-based 2D face models have also been proposed, such as Active Appearance Model [63] and bunch graph model of Gabor jets [64]. Tu et al. [65] investigated head pose tracking in low-resolution video by modeling facial appearance variations online with incremental weighted PCA.

We introduced in [66] a probabilistic framework for simultaneous head tracking and pose estimation. By embedding the pose variable into the motion state, head

Fig. 6 Simultaneous head tracking and pose estimation using particle filtering. (**a**) An input frame. (**b**) Particles are resampled and propagated in the location space. (**c**) Weighted resampling is performed with respect to the skin-color-based importance function. (**d**) Particles are weighted by the shape likelihood function, and the particles with high likelihoods are resampled (we show here 10 particles for illustration) for propagation and evaluation in the pose space. (**e**) The particles are evaluated in the pose space, and the final result is obtain by the MAP estimation

Fig. 7 Head tracking and pose estimation results in one sequence

pose tracking and recognition were formulated as a sequential maximum a posteriori (MAP) estimation problem solved by particle filtering. Faces were represented by ellipses bounding them. We adopted the partitioned sampling [67] to divide the state space into partitions, allowing efficient tracking with a small number of particles. Some intermediate results in one example frame are shown in Fig. 6. Figure 7 shows some examples of head tracking and pose estimation in one sequence. Based on our approach, a real-time head control interface for a robotic wheelchair was implemented.

Adaptation to changing appearance and scene conditions is a critical property a hand or head tracker should satisfy. Ross et al. [68] represented the target in a low-dimensional subspace which is adaptively updated using the tracking results. In [69], Grabner et al. introduced the online boosting for adaptive tracking, which allows online updating of discriminative features of the target object. Compared to the approaches using a fixed target model such as [70], these adaptive trackers are more robust to appearance changes in video sequences. One main drawback of these adaptive approaches is their susceptibility to drift, i.e., gradually adapting to nontargets, because the target model is updated according to the tracked results, which could be with errors. To address this problem, a mechanism for detecting or correcting drift should be introduced. In [71], global constraints on the overall appearance of the face were added. Grabner et al. [72] introduced an online semi-supervised boosting to alleviate the problem. They formulated the update process in a semi-supervised fashion which uses the labeled data as a prior and the data collected during tracking as unlabeled samples.

3.3 Gesture Recognition

Human gestures include static configurations and postures (e.g., hand posture, head pose, facial expression, and body posture) and dynamic gestures (e.g, hand gesture, head gestures like shaking and nodding, facial action like raising the eyebrows, and body gestures). Therefore, gesture recognition can be categorized as static posture recognition and dynamic gesture recognition. A static posture is represented by a single image, while a dynamic gesture is represented by a sequence of images.

In [17, 18, 73], Freeman et al. adopted steerable filters to derive local orientations of the input image and then used the orientation histogram to represent hand posture. The local orientation measurements are less sensitive to lighting changes. Figure 8 illustrates the orientation histograms of several hand postures. To make it work in complex background, in [74], we first derived the hand contour based on skin color and then computed the orientation histograms of hand contour for posture recognition. The process is shown in Fig. 9. In [75], Gabor Jets were adopted as local image description for hand posture recognition in complex backgrounds. Fourier descriptors were exploited in [43] to represent the segmented hand shape.

Starner et al. [10] developed a wearable hand control device for home appliances. By placing a camera on the user body, occlusion problems can be minimized. To make the system work in a variety of lighting conditions, even in the dark, the camera is ringed by near Infrared LEDs and has an infrared-pass filter mounted in the front (see Fig. 10). The prototype can recognize four hand poses (Fig. 10) and six dynamic gestures. Region-growing was used to segment hand region, and a set of eight statistics were extracted from the blob for posture description. In [76, 77], Kösch et al. presented a mobile gesture interface that allows control of wearable computer with hand postures. They used a texture-based approach to classify tracked hand regions into seven classes (six postures and "no known hand posture"). A gesture control interface for CE devices was presented in [78], in which seven hand postures were defined. In the prototype, the hand is segmented using a skin color model in the YCbCr color space, and moment invariants are extracted for posture recognition using a neural network classifier.

Dynamic gestures are characterized by the spatio-temporal motion structures in image sequences. A static posture can be regarded as a state of a dynamic gesture. Handwriting with a pen or mouse in 2D input devices is dynamic gestures that had been well studied [2]; many commercial systems of pen-based gesture recognition have been available since the 1970s. However, compared with pen-based gestural system, the visual interpretation of dynamic gestures is much more complex and difficult. Two main difficulties are: (1) temporal segmentation ambiguity, i.e., how to decide the starting and ending points of continuous gestures. The existing systems usually require a starting position in time and/or space or use static pose to segment gestures. (2) spatial-temporal variability. This is because gestures vary among individuals, which even vary from instance to instance for a given individual.

Many methods used in speech recognition can be borrowed for dynamic gesture recognition because of the similarity of the domains, for example, Dynamic Time Warping (DTW) and Hidden Markov Model [43, 79]. Other approaches, including

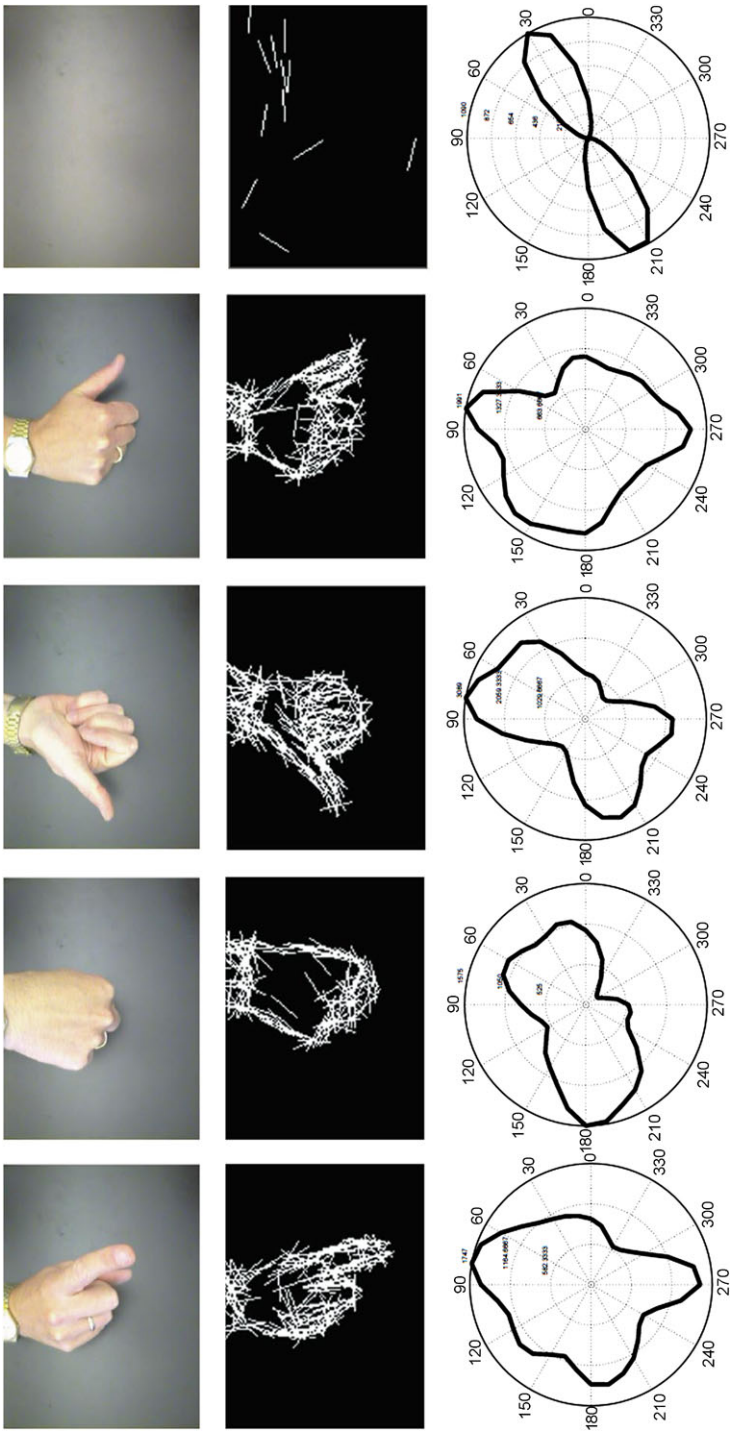

Fig. 8 Hand posture recognition based on orientation histogram matching [18] (*top*), the orientation maps (*middle*), and the orientation histograms plotted in polar coordinates

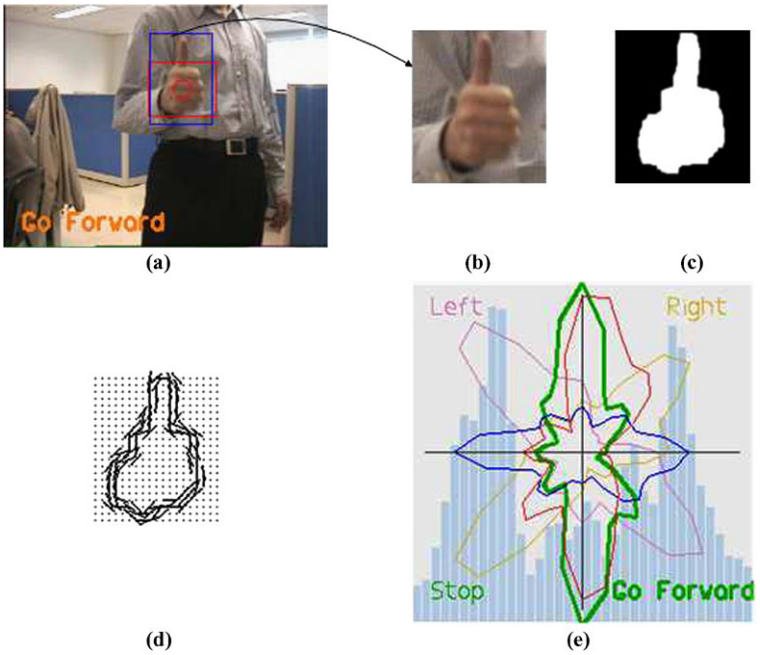

Fig. 9 Hand posture recognition using the orientation histogram of hand contour. (**a**) hand localization by tracking; (**b**) rectangle bounding hand region; (**c**) hand contour segmented based on skin color; (**d**) local orientations of hand contour; (**e**) posture recognition by matching the orientation histogram of hand contour (plotted in polar coordinates)

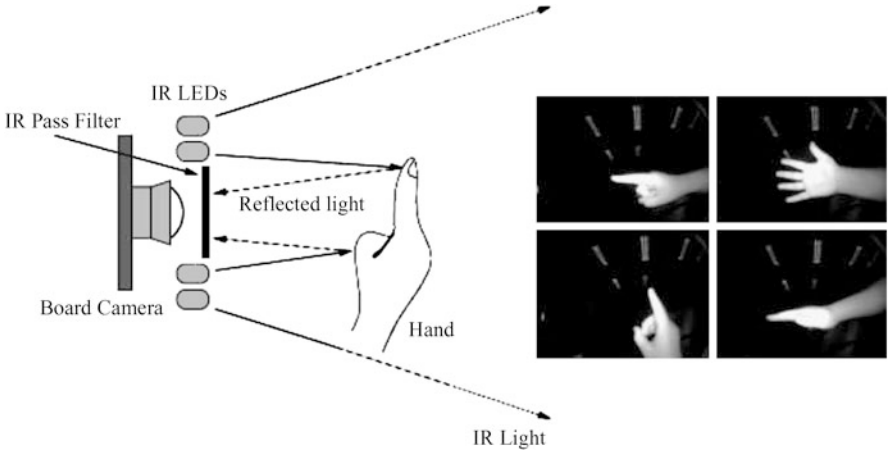

Fig. 10 The Gesture Pendant developed in [10]. (*Left*) sideview of the infrared setting; (*Right*) the four hand poses

Finite State Machines (FSM) [80], Dynamic Bayesian Networks (DBN) [81], and Time Delayed Neural Network TDNN) [82], have also been introduced to model the temporal transitions in gesture state spaces. Gesture recognition can also been addressed by trajectory matching [83–85]. The trajectory templates are first learned from training samples; in the testing phase, the input trajectory is matched with learned templates. Black et al. [83, 84] adopted a particle-filtering-based probabilistic framework for dynamic gesture recognition.

4 Gesture Control: Products and Applications

In recent years, many commercial gesture control systems or prototypes have been developed. Gesture control has been implemented in many CE devices. In this section, we present a review on current gesture control products and applications. We first introduce some gesture control products and solutions.

GestureTek [86] is one of the leading companies working on gesture recognition technologies. By converting hand or finger movements into mouse control, their product GestPoint provides a touch-free "point-to-click" control interface. Two cameras are used to capture and track hand or finger movements inside a control frame. For reliable tracking in varied lighting conditions and even with poor illumination, IR lighting is utilized. GestureTek's Illuminate series provides surface computing with a touch-free gesture control, which enables users navigate dynamic content by pointing fingers or waving hands (shown in Fig. 11). Their GestFX series allows the users to control the visual content projected on the floor, wall, or table space with their body motion; an example is shown in Fig. 11.

Toshiba has been actively working on vision-based gesture control. Their Qosmio laptops support hand gesture control. For example, forming a fist allows the user to

Fig. 11 The Illuminate series (*left*) and the GestFX system (*right*) from GestureTek [86]

Fig. 12 Gesture interaction systems from Fraunhofer [90]: (*Left*) iPoint Explorer and (*Right*) iPoint Presenter

move the cursor around the screen, and pressing the thumb down on top of the fist makes a selection. In IFA 2008, Toshiba showed gesture control for TVs. In their systems, a single webcam is used to sense the user's hand movement at the distance of 1–3 meters.

Mgestyk [87] have developed 3D gesture control solutions, using 3D cameras provided by 3DV Systems [88]. Since gesture recognition is performed directly on 3D depth data, their 3D system can capture small hand movements accurately, even depth-based gestures. Any data beyond a certain depth (such as people walking in the background) can be ignored. The system is reliable to lighting variations and even works in total darkness without using lighting sources. Softkinetic [89] has also been working on 3D gesture recognition solutions, based on a depth-sensing camera.

Fraunhofer Institute for Telecommunications HHI [90] has developed gesture interaction systems using a pair of stereo cameras. Their hand tracker can measure the 3D position of the user's fingertips at a rate of 50 Hz. They combine camera sensor with other sensors for reliable performance. For example, in the iPoint Explorer system (Fig. 12), ultrasonic sensors and two cameras are utilized for reliable sensing. In the iPoint Presenter system (Fig. 12), IR lights and cameras are adopted for detection and tracking of multiple fingers. LM3LABS [91] is also working on gesture interaction using stereo camera sensors.

Gesture control can be applied for most of CE devices including TV/displays, game consoles, mobile phones, and so on. In the following, we discuss current gesture control applications for CE devices.

TVs or Displays Many companies have recently introduced gesture control for TVs or displays. As mentioned above, Toshiba developed gesture-control interface for TV. In CES 2008, JVC showed a TV that reacts on hand claps and hand gestures.

Fig. 13 The EyeMobile engine from GestureTek [86] tracks three movements: shake, rock, and roll

The user can move his/her hand as pointer; the icon in the screen is clicked by bending and extending fingers. Samsung also introduced a gesture-control TV in CES 2008, based on the WAVEscape platform using a stereo near-IR vision system. In CES 2009, Hitachi showed a gesture-controlled TV, which integrates the 3D sensors provided by Canesta [92] and gesture recognition software from GestureTek.

Gaming Sony's PlayStation comes with Eye Toy, a set-top camera, which enables players to interact with games using full-body motion. With built-in LED lights, Eye Toy works when the room is with poor illumination. Microsoft also supports game-control games in their Xbox 360. Both Sony and Microsoft licensed GestureTek's patents on gesture control. Microsoft Xbox 360 will support more gesture interaction by using 3D cameras from 3DV Systems. Freeverse [93] developed gesture-based ToySight for Apple's iSight camera.

Mobile Phones Many mobile phones, including Sony Ericsson, Nokia, HTC, and Apple iPhone, started to support gesture control. For example, for Sony Ericsson Z555, the user can let it go mute or snooze the alarm by waving the hand to the build-in camera. GestureTek has developed middleware for gesture control on mobile phones. Their EyeMobile engine measures movement when a user shakes, rocks, or rolls the device (shown in Fig. 13). EyeMobile can also track a person's movements in front of the device. Samsung filed patents on gesture control for mobile phones and devices, where the predefined finger motions captured by the camera are translated into on-screen control. EyeSight [94] developed vision algorithms on the mobile phone that can detect and recognize 3D hand motions in front of the camera.

Automobiles Gesture control can be used in automotive environments for controlling applications such as CD-player and telephone, which reduces visual and mental distraction by allowing the driver to keep the eyes on the road. Many car manufacturers have developed gesture-control interfaces [95]. A prototype implemented in a BMW limousine [11] can recognize 17 hand gestures and 6 head gestures using IR lighting and camera. Head gestures are recognized to detect shaking and nodding for communicating approval or rejection, while hand gestures provide a way to skip CD-tracks or radio channels and to select shortcut functions (Fig. 14). Another system, called iGest [96], can recognize 16 dynamic and 6 static gestures. General Motors [97] has developed iWave, a gesture-based car navigation and entertainment system. In the Gesture Panel [98], as shown in Fig. 15, a camera is aimed at a grid of IR LEDs to capture gestures that are made between the camera and the grid.

Fig. 14 BMW's gesture control prototype [11]. (*Left*) skipping audio tracks by hand gestures; (*Right*) reference coordinate system for hand gestures with interaction area

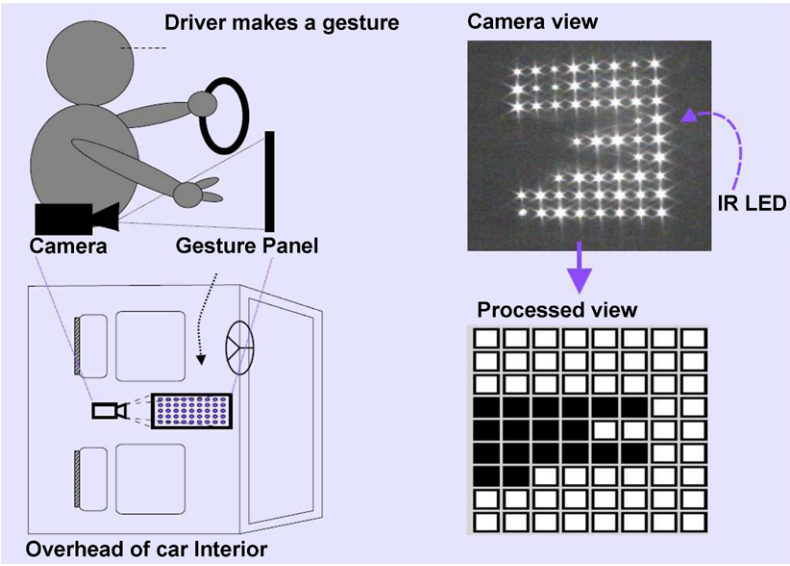

Fig. 15 Gesture Panel [98]. (*Left*) over-head and side view of the placement; (*Right*) camera view of a gesture and the corresponding binary representation

5 Conclusions

Gesture control provides a promising direction for natural user interfaces. Recent years have witnessed much interest on gesture control in CE industry. In this chapter, we present a overview on gesture control technologies. Different sensing technologies are discussed, among which vision-based gesture sensing and interpretation is more powerful, more general, and less unobtrusive. We review computer vision research on each component of vision-based gesture recognition. Latest developments on gesture control products and applications are also presented.

One trend is to use stereo or 3D depth-sensing sensors for gesture recognition. Many difficulties with normal cameras are avoided with 3D sensors, for example,

background noises and lighting variations. Although currently the 3D sensors are more expensive than normal cameras, more and more low-cost 3D sensing technologies are becoming commercially available.

With advance in sensing hardware and computer vision algorithms, vision-based gesture recognition technologies will become eventually mature for industrial applications. We believe that gesture control will be in widespread use in numerous applications in near future. We also believe that future user interfaces may ultimately combine vision, voice, and other modalities as we humans do, leading to multimodal interaction.

References

1. Pavlovic, V.I., Sharma, R., Huang, T.S.: Visual interpretation of hand gestures for human–computer interaction: a review. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 677–695 (1997)
2. Turk, M.: Gesture recognition. In: Stanney, K. (ed.) Handbook of Virtual Environment Technology. Lawrence Erlbaum Associates, Hillsdale (2001)
3. Mitra, S., Acharya, T.: Gesture recognition: a survey. IEEE Trans. Syst. Man Cybern., Part C, Appl. Rev. **37**(3), 311–324 (2007)
4. Geer, D.: Will gesture-recognition technology point the way? Computer **37**(10), 20–23 (2004)
5. McLeod, R.G.: Gesture control: new wave in ce products. PC World **8** (2008)
6. Webel, S., Keil, J., Zoellner, M.: Multi-touch gestural interaction in X3D using hidden Markov models. In: ACM Symposium on Virtual Reality Software and Technology, pp. 263–264 (2008)
7. imatt. http://www.imatte.com/
8. Cybernet. http://www.cybernet.com/
9. Ogris, G., Stiefmeier, T., Junker, H., Lukowicz, P., Troster, G.: Using ultrasonic hand tracking to augment motion analysis based recognition of manipulative gestures. In: IEEE International Symposium on Wearable Computers, pp. 152–159 (2005)
10. Starner, T., Auxier, J., Ashbrook, D., Candy, M.: The gesture pendant: a self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In: International Symposium on Wearable Computers, pp. 87–94 (2000)
11. Althoff, F., Lindl, R.: Walchshäusl. Robust multimodal hand- and head gesture recognition for controlling automotive infotainment systems. In: VDI-Tagung: Der Fahrer im 21. Jahrhundert (2005)
12. Navisense. http://www.navisense.com/
13. Elliptic labs. http://www.ellipticlabs.com/
14. Stiefmeier, T., Ogris, G., Junker, H., Lukowicz, P., Troster, G.: Combining motion sensors and ultrasonic hands tracking for continuous activity recognition in a maintenance scenario. In: IEEE International Symposium on Wearable Computers, pp. 97–104 (2006)
15. Butler, A., Izadi, S., Hodges, S.: Sidesight: Multi-"touch" interaction around small devices. In: ACM Symposium on User Interface Software and Technology, pp. 201–204 (2008)
16. Celluon. http://www.celluon.com/
17. Freeman, W.T., Weissman, C.: Television control by hand gestures. In: Proceedings of the IEEE International Workshop on Automated Face and Gesture Recognition (FG'95), Zurich, Switzerland, June 1995, pp. 179–183 (1995)
18. Freeman, W.T., Anderson, D., Beardsley, P., Dodge, C., Kage, H., Kyuma, K., Miyake, K., Roth, M., Tanaka, K., Weissman, C., Yerazunis, W.: Computer vision for interactive computer graphics. IEEE Comput. Graph. Appl. **18**(3), 42–53 (1998)
19. Yang, M.-H., Kriegman, D., Ahuja, N.: Detecting faces in images: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **24**(1), 34–58 (2002)

20. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 511–518 (2001)
21. Viola, P., Jones, M.: Robust real-time face detection. Int. J. Comput. Vis. **57**(2), 137–154 (2004)
22. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. **55**(1), 119–139 (1997)
23. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. Mach. Learn. **37**(3), 297–336 (1999)
24. Lienhart, R., Kuranov, D., Pisarevsky, V.: Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In: DAGM 25th Pattern Recognition Symposium, Madgeburg, Germany, September 2003, pp. 297–304 (2003)
25. Li, S.Z., Zhang, Z.: Floatboost learning and statistical face detection. IEEE Trans. Pattern Anal. Mach. Intell. **26**(9), 1–12 (2004)
26. Liu, C.: A Bayesian discriminating features method for face detection. IEEE Trans. Pattern Anal. Mach. Intell. **25**(6), 725–740 (2003)
27. Shih, P., Liu, C.: Face detection using discriminating feature analysis and support vector machine. Pattern Recognit. **39**(2), 260–276 (2006)
28. Garcia, C., Delakis, M.: Convolutional face finder: a neural architecture for fast and robust face detection. IEEE Trans. Pattern Anal. Mach. Intell. **26**(11), 1408–1423 (2004)
29. Chen, Y.-N., Han, C.-C., Wang, C.-T., Jeng, B.-S., Fan, K.-C.: A CNN-based face detector with a simple feature map and a coarse-to-fine classifier. IEEE Trans. Pattern Anal. Mach. Intell. (2010)
30. Wu, Y., Huang, T.S.: View-independent recognition of hand postures. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 88–94 (2000)
31. Zhu, X., Yang, J., Waibel, A.: Segmenting hands of arbitrary color. In: IEEE International Conference on Automatic Face & Gesture Recognition (FG), pp. 446–453 (2000)
32. Bretzner, L., Laptev, I., Lindeberg, T.: Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. In: Proceedings of the IEEE International Conference on Automated Face and Gesture Recognition (FG'02), pp. 405–410 (2002)
33. Kölsch, M., Turk, M.: Robust hand detection. In: IEEE International Conference on Automatic Face & Gesture Recognition (FG), pp. 614–619 (2004)
34. Ong, E.J., Bowden, R.: A boosted classifier tree for hand shape detection. In: IEEE International Conference on Automatic Face & Gesture Recognition (FG), pp. 889–894 (2004)
35. Zondag, J.A., Gritti, T., Jeanne, V.: Practical study on real-time hand detection. In: IEEE International Workshop on Social Signal Processing (2009)
36. Stenger, B., Woodley, T., Kim, T.-K., Hernandez, C., Cipolla, R.: AIDIA—adaptive interface for display interaction. In: British Machine Vision Conference (BMVC), vol. 2, pp. 785–794 (2008)
37. Stenger, B., Woodley, T., Kim, T.-K., Cipolla, R.: A vision-based system for display interaction. In: British Computer Society Conference on Human–Computer Interaction, pp. 163–168 (2009)
38. Yilmaz, A., Javed, O., Shah, M.: Object tracking: a survey. ACM Comput. Surv. **38**(4), 13 (2006)
39. McAllister, G., McKenna, S.J., Ricketts, I.W.: Hand tracking for behaviour understanding. Image Vis. Comput. **20**(12), 827–840 (2002)
40. Isard, M., Blake, A.: Condensation—conditional density propagation for visual tracking. Int. J. Comput. Vis. **29**(1), 5–28 (1998)
41. Oka, K., Sato, Y., Koike, H.: Real-time tracking of multiple fingertips and gesture recognition for augmented desk interface systems. In: Proceedings of the IEEE International Conference on Automated Face and Gesture Recognition (FG'02), pp. 411–416 (2002)
42. Laptev, I., Lindeberg, T.: Tracking of multi-state hand models using particle filtering and a hierarchy of multi-scale image features. In: Proceedings of the IEEE Workshop on Scale-Space and Morphology (2001)
43. Ng, C.W., Ranganath, S.: Real-time gesture recognition system and application. Image Vis. Comput. **20**(13–14), 993–1007 (2002)

44. Huang, F., Reid, I.: Probabilistic tracking and recognition of non-rigid hand motion. In: Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03), pp. 60–67 (2003)
45. Isard, M., Blake, A.: ICONDENSATION: unifying low-level tracking in a stochastic framework. In: Proceedings of the European Conference on Computer Vision (ECCV'98), Freiburg, Germany, January 1998. vol. 1, pp. 893–908 (1998)
46. Yang, M.H., Ahuja, N., Tabb, M.: Extraction of 2d motion trajectories and its application to hand gesture recognition. IEEE Trans. Pattern Anal. Mach. Intell. **24**(8), 1061–1074 (2002)
47. Kölsch, M., Turk, M.: Fast 2d hand tracking with flocks of features and multi-cue integration. In: IEEE Conference on Computer Vision and Pattern Recognition Workshop (2004)
48. Yuan, Q., Sclaroff, S., Athitsos, V.: Automatic 2d hand tracking in video sequences. In: Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV'05) (2005)
49. Arulampalam, M., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Process. **50**(2), 174–189 (2002)
50. Shan, C., Tan, T., Wei, Y.: Real-time hand tracking using a mean shift embedded particle filter. Pattern Recognit. **40**(7), 1958–1970 (2007)
51. Bradski, G.R.: Computer vision face tracking for use in a perceptual user interface. Intel Technol. J. Q. **2** (1998)
52. Comaniciu, D., Ramesh, V., Meer, P.: Real-time tracking of non-rigid objects using mean shift. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 142–149 (2000)
53. Lu, S., Metaxas, D., Samaras, D., Oliensis, J.: Using multiple cues for hand tracking and model refinement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03), pp. II: 443–450 (2003)
54. Bray, M., Koller-Meier, E., Van Gool, L.: Smart particle filtering for 3d hand tracking. In: Proceedings of the IEEE International Conference on Automated Face and Gesture Recognition (FG'04), pp. 675–680 (2004)
55. Sudderth, E.B., Mandel, M.I., Freeman, W.T., Willsky, A.S.: Visual hand tracking using nonparametric belief propagation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) (2004)
56. Chang, W.Y., Chen, C.S., Hung, Y.P.: Appearance-guided particle filtering for articulated hand tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05) (2005)
57. Stenger, B., Thayananthan, A., Torr, P.H.S., Cipolla, R.: Model-based hand tracking using a hierarchical Bayesian filter. IEEE Trans. Pattern Anal. Mach. Intell. **28**(9), 1372–1384 (2006)
58. Zhou, H., Huang, T.S.: Tracking articulated hand motion with eigen dynamics analysis. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV'03), pp. 1102–1109 (2003)
59. Tsap, L.V.: Gesture-tracking in real time with dynamic regional range computation. Real-Time Imaging **8**(2), 115–126 (2002)
60. Inaguma, T., Saji, H., Nakatani, H.: Hand motion tracking based on a constraint of three-dimensional continuity. J. Electron. Imaging **14**(1), 013021 (2005)
61. Basu, S., Essa, I., Pentland, A.: Motion regularization for model-based head tracking. In: Proceedings of the IEEE Conference on Pattern Recognition (ICPR'96), Vienna, Austria, pp. 611–616 (1996)
62. Cascia, M.L., Sclaroff, S., Athitsos, V.: Fast, reliable head tracking under varying illumination: an approach based on registration of texture-mapped 3d models. IEEE Trans. Pattern Anal. Mach. Intell. **22**(4), 322–336 (2000)
63. Cootes, T.F., Walker, K., Taylor, C.J.: View-based active appearance models. In: Proceedings of the IEEE International Conference on Automated Face and Gesture Recognition (FG'00), Grenoble, France, March 2000, pp. 227–233 (2000)
64. Kruger, N., Potzsch, M., von der Malsburg, C.: Determination of faces position and pose with a learned representation based on labeled graphs. Image Vis. Comput. **15**(8), 665–673 (1997)

65. Tu, J., Huang, T.S., Tao, H.: Accurate head pose tracking in low resolution video. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG'06), Southampton, UK, May 2006
66. Wei, Y., Shan, C., Tan, T.: An efficient probabilistic framework for simultaneous head pose tracking and recognition in video sequences. Technical report, National laboratory of Pattern Recognition, Chinese Academy of Sciences (2004)
67. MacCormick, J., Blake, A.: A probabilistic exclusion principle for tracking multiple objects. Int. J. Comput. Vis. **39**(1), 57–71 (2000)
68. Ross, D., Lim, J., Lin, R.-S., Yang, M.-H.: Incremental learning for robust visual tracking. Int. J. Comput. Vis. **77**(1–3), 125–141 (2008)
69. Grabner, H., Grabner, M., Bischof, H.: Real-time tracking via on-line boosting. In: British Machine Vision Conference (BMVC), pp. 47–56 (2006)
70. Avidan, S.: Support vector tracking. IEEE Trans. Pattern Anal. Mach. Intell. **26**(8), 1064–1072 (2004)
71. Kim, M., Kumar, S., Pavlovic, V., Rowley, H.: Face tracking and recognition with visual constraints in real-world videos. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2008)
72. Grabner, H., Leistner, C., Bischof, H.: Semi-supervised on-line boosting for robust tracking. In: European Conference on Computer Vision (ECCV), pp. 234–247 (2008)
73. Freeman, W.T., Roth, M.: Orientation histograms for hand gesture recognition. In: Proceedings of the IEEE International Workshop on Automated Face and Gesture Recognition (FG'95), Zurich, Switzerland, June 1995, pp. 296–301 (1995)
74. Shan, C.: Vision-based hand gesture recognition for human–computer interaction. Master's thesis, National laboratory of Pattern Recognition, Chinese Academy of Sciences (2004)
75. Triesch, J., von der Malsburg, C.: A system for person-independent hand posture recognition against complex backgrounds. IEEE Trans. Pattern Anal. Mach. Intell. **23**(12), 1449–1453 (2001)
76. Kölsch, M., Turk, M., Höllerer, T., Chainey, J.: Vision-based interfaces for mobility. Technical Report TR 2004-04, University of California at Santa Barbara (2004)
77. Kölsch, M.: Vision Based Hand Gesture Interfaces for Wearable Computing and Virtual Environments. PhD thesis, University of California, Santa Barbara (2004)
78. Premaratne, P., Nguyen, Q.: Consumer electronics control system based on hand gesture moment invariants. IET Comput. Vis. **1**(1), 35–41 (2007)
79. Ren, H., Xu, G.: Human action recognition with primitive-based coupled-HMM. In: Proceedings of the IEEE Conference on Pattern Recognition (ICPR'02), vol. II, pp. 494–498 (2002)
80. Hong, P., Turk, M., Huang, T.S.: Gesture modeling and recognition using finite state machines. In: Proceedings of the IEEE International Conference on Automated Face and Gesture Recognition (FG'00), pp. 410–415 (2000)
81. Pavlovic, V.: Dynamic Bayesian Networks for Information Fusion with Applications to Human–Computer Interfaces. PhD thesis, University of Illinois at Urbana-Champaign (1999)
82. Yang, M.H., Ahuja, N.: Recognizing hand gesture using motion trajectories. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'99), January 1999, vol. I, pp. 466–472 (1999)
83. Black, M.J., Jepson, A.D.: Recognition temporal trajectories using the CONDENSATION algorithm. In: Proceedings of the IEEE International Conference on Automated Face and Gesture Recognition (FG'98), Japan, pp. 16–21 (1998)
84. Black, M.J., Jepson, A.D.: A probabilistic framework for matching temporal trajectories: CONDENSATION-based recognition of gestures and expressions. In: Proceedings of the European Conference on Computer Vision (ECCV'98) (1998)
85. Psarrou, A., Gong, S., Walter, M.: Recognition of human gestures and behaviour based on motion trajectories. Image Vis. Comput. **20**(5–6), 349–358 (2002)
86. Gesturetek. http://www.gesturetek.com/
87. Mgestyk. http://www.mgestyk.com/
88. 3dv systems. http://www.3dvsystems.com/

89. Softkinetic. http://www.softkinetic.net/
90. fraunhofer hhi. http://www.hhi.fraunhofer.de
91. Lm3labs. http://www.lm3labs.com/
92. Canesta. http://www.canesta.com/
93. Freeverse. http://www.freeverse.com/
94. eyesight. http://www.eyesight-tech.com/
95. Pickering, C.: Gesture recognition could improve automotive safety. Asia-Pacific Engineer—Automotive-Design (2006)
96. Canzler, S., Akyol, U.: GeKomm – Gestenbasierte Mensch–Maschine Kommunikation im Fahrzeug. PhD thesis, Rheinisch-Westfälische Technische Hochschule Aachen (2000)
97. Gm-cmu. http://gm.web.cmu.edu/demos/
98. Westeyn, T., Brashear, H., Atrash, A., Starner, T.: Georgia tech gesture toolkit: supporting experiments in gesture recognition. In: International Conference on Multimodal Interfaces (ICMI), pp. 85–92 (2003)