Artif Intell Rev
DOI 10.1007/s10462-012-9356-9

Vision based hand gesture recognition for human computer interaction: a survey

Siddharth S. Rautaray · Anupam Agrawal

Abstract As computers become more pervasive in society, facilitating natural human–computer interaction (HCI) will have a positive impact on their use. Hence, there has been growing interest in the development of new approaches and technologies for bridging the human–computer barrier. The ultimate aim is to bring HCI to a regime where interactions with computers will be as natural as an interaction between humans, and to this end, incorporating gestures in HCI is an important research area. Gestures have long been considered as an interaction technique that can potentially deliver more natural, creative and intuitive methods for communicating with our computers. This paper provides an analysis of comparative surveys done in this area. The use of hand gestures as a natural interface serves as a motivating force for research in gesture taxonomies, its representations and recognition techniques, software platforms and frameworks which is discussed briefly in this paper. It focuses on the three main phases of hand gesture recognition i.e. detection, tracking and recognition. Different application which employs hand gestures for efficient interaction has been discussed under core and advanced application domains. This paper also provides an analysis of existing literature related to gesture recognition systems for human computer interaction by categorizing it under different key parameters. It further discusses the advances that are needed to further improvise the present hand gesture recognition systems for future perspective that can be widely used for efficient human computer interaction. The main goal of this survey is to provide researchers in the field of gesture based HCI with a summary of progress achieved to date and to help identify areas where further research is needed.

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

---

S. S. Rautaray (✉) · A. Agrawal
Information Technology, Indian institute of Information Technology, Allahabad, India
e-mail: sr.rgpv@gmail.com

A. Agrawal
e-mail: anupam69@gmail.com

Published online: 06 November 2012

1 Introduction

In the present world, the interaction with the computing devices has advanced to such an extent that as humans it has become necessity and we cannot live without it. The technology has become so embedded into our daily lives that we use it to work, shop, communicate and even entertain our self (Pantic et al. 2008). It has been widely believed that the computing, communication and display technologies progress further, but the existing techniques may become a bottleneck in the effective utilization of the available information flow.

To efficiently use them, most computer applications require more and more interaction. For that reason, human-computer interaction (HCI) has been a lively field of research in the last few years. Firstly based in the past on punched cards, reserved to experts, the interaction has evolved to the graphical interface paradigm. The interaction consists of the direct manipulation of graphic objects such as icons and windows using a pointing device. Even if the invention of keyboard and mouse is a great progress, there are still situations in which these devices are incompatible for HCI. This is particularly the case for the interaction with 3D objects. The 2 degrees of freedom (DOFs) of the mouse cannot properly emulate the 3 dimensions of space. The use of hand gestures provides an attractive and natural alternative to these cumbersome interface devices for human computer interaction. Using hands as a device can help people communicate with computers in a more intuitive way. When we interact with other people, our hand movements play an important role and the information they convey is very rich in many ways. We use our hands for pointing at a person or at an object, conveying information about space, shape and temporal characteristics. We constantly use our hands to interact with objects: move them, modify them, and transform them. In the same unconscious way, we gesticulate while speaking to communicate ideas ('stop', 'come closer', 'no', etc). Hand movements are thus a mean of non-verbal communication, ranging from simple actions (pointing at objects for example) to more complex ones (such as expressing feelings or communicating with others). In this sense, gestures are not only an ornament of spoken language, but are essential components of the language generation process itself. A gesture can be defined as a physical movement of the hands, arms, face and body with the intent to convey information or meaning (Mitra and Acharya 2007).

In particular, recognizing hand gestures for interaction can help in achieving the ease and naturalness desired for human computer interaction. Users generally use hand gestures for expression of their feelings and notifications of their thoughts. Researcher Karam (2006) in his work reported that hand has been widely used in comparison to other body parts for gesturing as it is a natural form of medium for communication between human to human hence can best suited for human computer interaction also as shown in Fig. 1.

The interest in this area has led to a large body of research which has been digested in a number of surveys directly or indirectly related to gesture recognition. Table 1 shows some of the important surveys and articles presented in the area of gesture recognition.

The following comprehensive analysis of the surveys and articles published earlier related to hand gesture recognition could be used for the design, development and implementation of evolved, robust efficient and accurate hand gesture recognition systems for human computer interaction. The key issues addressed in the research articles could in many ways help the researchers in identifying the arid regions of the said area and tapping these arid regions towards advances in more user friendly human computer interaction systems. The remainder of this paper is organized as follows:

Section 2 provides an overview of the enabling technologies available hand gesture recognition.

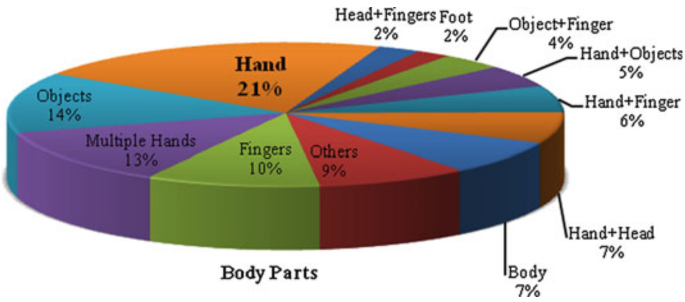

Fig. 1 The *graph* shows the different body parts or objects identified in the literature employed for gesturing (Karam 2006)

Section 3 discusses taxonomies and representations techniques available in literature related to hand gestures.
Section 4 presents the vision based recognition techniques available for recognizing hand gestures.
Section 5 shows the different application domain which uses hand gestures for interaction in different environment.
Section 6 provides comparative analysis between vision based gesture recognition commercial products and software.
Section 7 discusses the software platform/ frameworks used for implementing gesture recognition techniques.
Section 8 presents the state of the art and discussion related to vision based hand gesture recognition techniques.
Section 9 provides a brief analysis and future perspectives for vision based hand gesture recognition HCI.
Section 10 finally summarizes the survey.

2 Enabling technologies for hand gesture recognition

Gesture recognition the term collectively refers to the whole process of tracking human gestures to their representation and conversion to semantically meaningful commands. Research in hand gesture recognition aims to design and development of such systems than can identify explicit human gestures as input and process these gesture representations for device control through mapping of commands as output. Creation and implementation of such efficient and accurate hand gesture recognition systems are aided through two major types of enabling technologies for human computer interaction namely contact based and vision based devices. Figure 2 shows the examples of contact and vision based devices.

Contact based devices employed for hand gesture recognition systems are based on physical interaction of user with the interfacing device i.e. the user needs to be accustomed with the usage of these devices, hence not adaptable to the naïve users. These devices are usually based on technologies like data glove, accelerometers, multi-touch screen etc which uses several detectors. Also there are devices that use only one detector as the accelerometer of the Nitendoc Wii-Remote. This class of contact based devices for hand gesture recognition can be further classified as mechanical, haptics, ultrasonic, inertial and magnetic (Kanniche 2009).

Table 1 Analysis of some comprehensive surveys and articles

Ref.	Year	Scope of analysis	Key findings
Pavlovic et al. (1997)	1997	Covers review of more than 100 papers related to visual interpretation of hand gestures in context to HCI. Method used for modeling, analyzing and recognizing gestures are discussed in detail	Suggests integration of hand gestures with gaze, speech and other naturally related modes of communication in multimodal interface for raising these limitations toward gestural HCI
Wu and Huang (1999a)	1999	A survey on vision based gesture recognition approaches. Focuses on different recognition techniques which comprises of recognition done by modeling the dynamics, modeling the semantics, HMM framework etc	Laid emphasis on the complexity of gesture for which efforts in computer vision, machine learning and psycholinguistics will be needed. Static hand posture recognition techniques try to achieve rotation invariant and view-independent recognition which needs to be more investigated detail
Moeslund and Granum (2001)	2001	Comprehensive review of 130 papers discussing initialization, tracking, pose estimation and recognition of a motion capture system. Performance characteristics related to system functionality and modern advancements in each of these fields are comprehensively evaluated	Problems predominant throughout the domain such as the lack of training data, the large amount of time required for gesture capture, lack of invariance and robustness are explored and possible solutions such as the employment of a approach similar to speech recognition, abstracting the motion layer have to be investigated in detail
Derpanis (2004)	2004	Paper reviews the vision based hand gestures for human computer interaction. Detailed discussion on the feature set, the classification method and the underlying representation of gesture set has been done	Research in the areas of feature extraction, classification methods and gesture representation are needed to be performed in order to acquire the ultimate goal of humans interfacing with human machines on their natural terms
Chaudhary et al. (2011)	2007	Comprehensive survey on gesture recognition techniques particularly focusing on hand and facial movements. Hidden markov models, particle filtering and condensation, finite-state machines, optical flow, skin color and connectionist models discussed in detail	The need for different recognition algorithms depending on the size of the dataset and the gesture performed is identified and various combinations can be drawn out in this regard. From the research it is notable that any developed system should be both flexible and expandable in order to maximize efficiency, accuracy and understandability

Table 1 continued

Ref.	Year	Scope of analysis	Key findings
Wachs et al. (2011)	2011	Discusses soft computing based methods like artificial neural network, fuzzy logic, genetic algorithms etc in designing the hand gesture recognition	Soft computing provides a way to define things which are not certain but with an approximation makes use of learning models and training data. It is effective in getting results where the exact positions of hand or fingers are not possible
Corera and Krishnarajah (2011)	2011	Comprehensive article on vision based hand gesture application. Focuses on different challenges present in vision based gesture recognition systems and their related applications	Aside from technical obstacles like reliability, speed, and low cost implementation hand gesture interaction must also address intuitiveness and gesture spotting. Two handed dynamic hand gesture interaction is promising area for future research
Kanniche (2009)	2012	Survey on tools and techniques used for capturing hand gesture movements. Discusses on logical issues and design consideration for gesture recognition system	It suggests that the way forward is through modularization, scalability and essentially decentralizing the entire approach from gesture capture to recognition

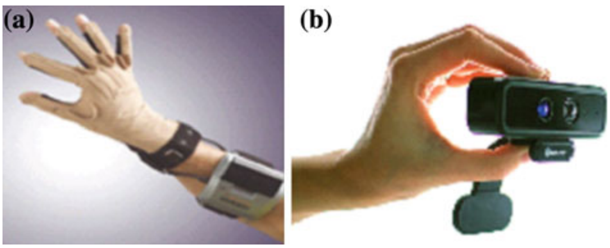

Fig. 2 **a** CyberGlove II (Kevin et al. 2004), **b** SoftKinetic HD camera (SoftKinetic 2012)

Mechanically primed devices are set of equipments to be used by the end user for human computer interaction like "IGS-190" a body suite that captures body gestures, "CyberGlove II" a wireless instrumented glove used for hand gesture recognition (Kevin et al. 2004) as shown in Fig. 2a. Cybergloves and magnetic trackers are also used for trajectory modeling for hand gesture recognition. Haptics primed devices are very commonly used in our daily life based on sense of touch with hardware for human computer interface i.e. multi touch screen devices like Apple iPhone, tablet PC and other devices with multi touch gestural interactions using HMM (Webel et al. 2008).

Ultra-sonic based motion trackers are composed sonic emitters that emit ultrasound, sonic discs that reflect ultrasound and multiple sensors that time the return pulse. The position and orientation of gestures are computed based on propagation, reflection, speed of time and triangulation, respectively (Kanniche 2009). Low resolution and lack of precision are pertained to this set of devices but their applicability to environments having lack of illumination and presence of magnetic obstacles or noise make them usually favored. Inertial primed devices work on the basis of variations of earth's magnetic field for detecting motion. Schlomer et al. (2008) proposed gesture recognition using Wii-controller employing HMM independent of the target system. Bourke et al. (2007) proposed recognition systems to detect the normal gestures which are used in our daily activities using accelerometer. Noury et al. (2003) proposed system for multimodal intuitive media browsing in which the user can learn personalized gestures. Variations of artificial magnetic field for motion detection are measured using magnetic primed devices.

Restrained by the dependence on experienced users the contact based devices do not provide much acceptability, hence vision based devices have been employed for capturing the inputs for hand gesture recognition in human computer interaction. This set of devices relies on captured video sequence by one or several cameras for interpreting and analyzing the motion (Mitra and Acharya 2007). Vision based also uses hand markers for detection of human hand motion and gestures. The hand markers can be further classified as reflective markers which are passive in nature and shines as strobes hit it whereas LED light are active in nature and flashes in sequence. In these systems each camera delivers marker position from its view with a 2D frame which lightens with either strobe lights or normal lights. A preprocessing step is further executed for interpreting the views and positions into a 3D space.

The main challenge of vision-based hand gesture recognition is to cope with the large variety of gestures. Recognizing gestures involve handling a considerable number of degrees of freedom (DoF), huge variability of the 2D appearance depending on the camera view point (even for the same gesture), different silhouette scales (i.e. spatial resolution) and many resolutions for the temporal dimension (i.e. variability of the gesture speed). Moreover, it need also to balance the accuracy-performance-usefulness trade-off according to the type of application, the cost of the solution and several criteria's such as real-time performance, robustness, scalability and user-independence.

In real-time process the system must be able to analyze the image at the frame rate of the input video to give the user instant feedback of the recognized hand gesture. Robustness plays an important role in recognizing different hand gestures successfully under different lighting conditions and cluttered backgrounds. The system should also be robust against in-plane and out-of-plane image rotations. Scalability helps in handling a large gesture vocabulary which can be included with a small number of primitives. This makes the composition of different gesture commands easily controlled by the user. User-independence creates the environment where the system can be handled by different users rather than specific user and should further recognize gestures performed by humans of different sizes and colors.

The above mentioned enabling technology for hand gesture recognition has their advantages and disadvantages. As the contact based can be uncomfortable for user since they require physical contact with the user, still having a verge over the accuracy of recognition and less complexity of implementation goes in favor of these devices. Vision based devices though is user friendly but suffer from configuration complexity and occlusion problems. Some of the major merits and demerits of both enabling technologies has been summarized in Table 2.

The main disadvantage of contact based devices is the health hazards (Schultz et al. 2003), which are caused by its devices like mechanical sensor material which raises symptoms of allergy, magnetic devices which raises risk of cancer etc (Nishikawa et al. 2003). Whereas vision based devices have initial challenge of complex configuration and implementations but are more user friendly and hence more privileged for usage in long run. Reckoning the

Table 2 Comparison between contact-devices and vision-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

above facts about the two streams of enabling technologies this paper focuses on vision based enabling technologies for hand gesture recognition in its further sections.

3 Vision based hand gesture taxonomies and representations

Psychological aspects of gestures based on hand gesture taxonomies and representations are also an important aspect of hand gesture recognition systems. Varying from person to person several hand gesture taxonomies have been proposed in the literature. Gesture acts a medium of communication for non vocal communication in conjunction with or without verbal communication is intended to express meaningful commands. These gestures may be articulated with any of the body parts or with combination of one or many of them. Gestures being major constituent of human communication may serve as an important means for human computer interaction too. Though the significance and meaning associated with different gestures differ very much with cultures having less or invariable or universal meaning for single gesture. For instance different gestures are used for greeting at different geographical separations of the world. For example pointing an extended finger is a common gesture in USA & Europe but it is taken to be as a rude and offensive gesture in Asia. Hence the semantic interpretation of gestures depends strictly on given culture. Theoretically the literature classifies hand gestures into two type's static and dynamic gestures. Static hand gestures are defined as orientation and position of hand in the space during an amount of time without any movement and if a movement is there in the aforementioned time duration it is called dynamic gesture. Dynamic hand gestures include gestures involving body parts like waving of hand while static hand gestures include single formation without movement like jamming the thumb and forefinger to form the "ok" symbol is a static pose which represents static gesture.

Dynamic hand gestures done intentionally for communication are called conscious dynamic gestures, whereas unintentionally (unawareness) done gesture carried out during causal physical activity is known as unconscious dynamic gesture. Figure 3 shows the taxonomy of hand gesture categories. According to research (Hall 1973) 35 % of human communication consists of verbal communication and 65 % is non verbal gesture based communication. Gesture Ottenheimer (2005) can be categorized into five types i.e. emblems, affect displays, regulators, adaptors and illustrators.

Emblematic gestures also referred as emblem or quotable gestures are direct translation of short verbal communication like waving hand for good bye or nodding for assurance. The quotable gestures are specifically culture specific. Gestures conveying emotion or intensions are called affect displays.

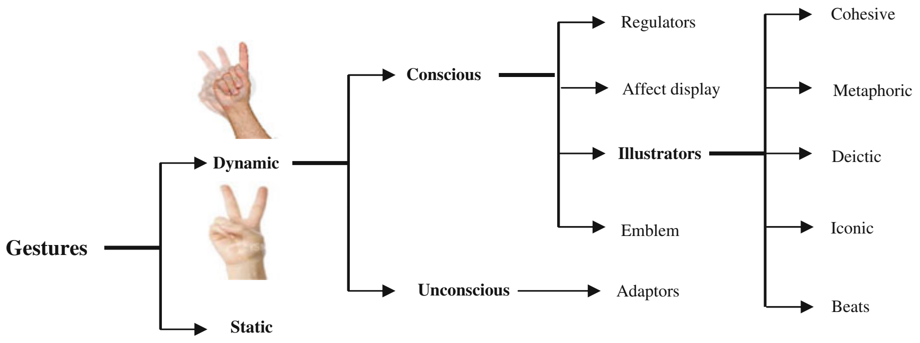

Fig. 3 Vision based hand gesture taxonomies (Kanniche 2009)

The affect displays are generally universal less dependent on culture. Gestures controlling interaction are called regulators. Gestures like headshaking, quickly moving one's leg that enables the release of body tension are called adaptors. Adaptors are generally habit of communicators that are not used intentionally during a communication. Illustrator gestures emphasize the key point in speech to depict the communications pronouncing statements. Being emphasized by the communicators pronouncing statements these gestures are inherently dependent on communicators thought process and speech. These gesticulations could further be categorized into five sub category namely beats, deictic gestures, iconic gestures, metaphoric gestures and cohesive gestures (McNeill 1992).

- Beats are short and quick, rhythmics and after repetitive gestures.
- Concrete pointing to real location object or person and abstract pointing to abstract location or period of time are called deictic gestures.
- Hand movements depicting figural representation or actions for example moving hand upward with wiggling gingers 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 gestures is due to interruption of current communicator by any other communicator.

Vision based Hand Gesture Representations: To abstract and model the human body parts motion several hand gesture representations and models have been proposed and implemented by the researchers. The two major categories of hand gesture representation are 3D model based methods and appearance based methods as depicted in Fig. 4.

The 3D model based hand gesture recognition has different techniques for gesture representation namely 3D textured volumetric, 3D geometric model and 3D skeleton model. Appearance based hand gesture representation include color based model, silhouette geometry model, deformable gabarit model and motion based model.

The 3D model based hand gesture representation defines 3D spatial description of human hand for representation with temporal aspect being handled by automation. This automation divides the temporal characteristics of hand gesture into three phases (McNeill 1992) i.e. the preparation or prestroke phase, the nucleus or stroke phase and the retraction or poststroke phase in which every phase corresponds to one or many transitions of spatial states of the 3D human model. One or many cameras focus on the real target and compute parameters spatially matching the real target and then follow its motion during the recognition process in 3D model. Thus the 3D model has an advantage that it updates the model parameters

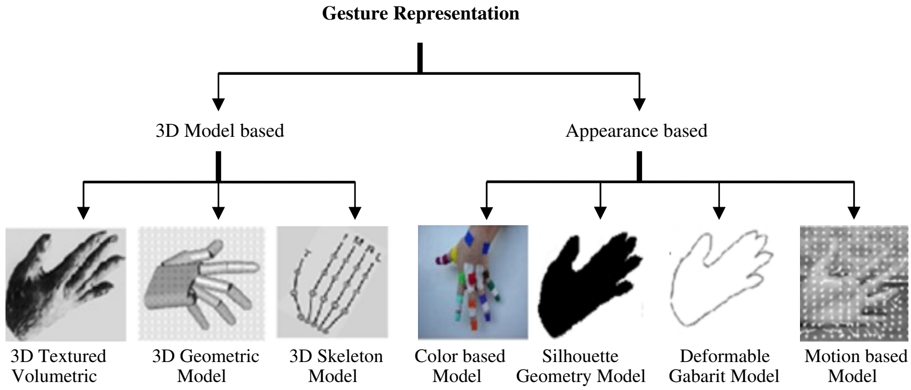

Fig. 4 Vision based hand gesture representations (Bourke et al. 2007)

while checking the matches of transition in temporal model, leading to precise hand gesture recognition and representation, though making it computationally intensive with requirement of dedicated hardware. There are also many methods (Boulay 2007) that combine silhouette extraction with 3D model projection fitting by finding target self oriented. Generally three kinds of model are generally used.

- 3D textured kinematic/volumetric model contains very high details of human body skeleton and skin surface information.
- 3D geometric models are less precise than 3D textures kinematic/volumetric models with respect to skin information but contains essential skeleton information.

Appearance based hand gesture representation methods are though broadly classified into two major subcategories i.e. 2D static model based methods and motion based methods, each sub category is having further variants. The commonly used 2D models include:

- Color based model uses body markers to track the motion of body or body part. As Bretzner et al. (2002) proposed hand gesture recognition employing multi-scale color features, hierarchal 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 were used to recognize hand gestures (Birdal and Hassanpour 2008).
- Deformable gabarit based models: they are generally based on deformable active contours (i.e. snake parameterized with motion and their variants. Ju et al. (1997) used snakes for the analysis of gestures and actions in technical talks for video indexing.
- Motion based models are used for recognition of an object or its motion based on the motion of object in an image sequence. Local motion histogram was introduced by Luo et al. (2008) which uses an Adaboost framework for learning action models.

4 Vision based hand gesture recognition techniques

Most of the complete hand interactive mechanisms that act as a building block for vision based hand gesture recognition system are comprised of three fundamental phases: detection, tracking and recognition. This section of the research survey discusses some of the

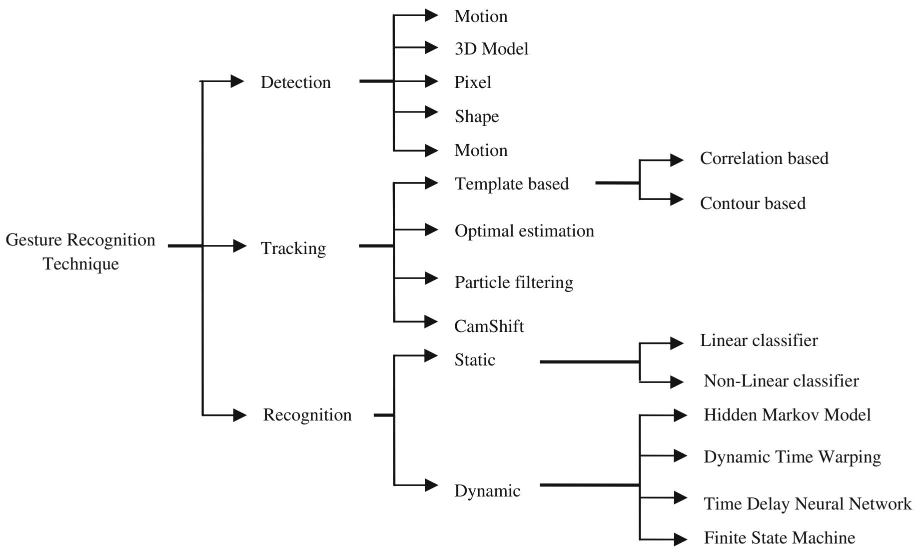

Fig. 5 Vision based hand gesture recognition techniques

prominent vision based hand gesture recognition techniques used by most of the researchers by categorizing under the three verticals representing the three fundamental phases of detection tracking and recognition as shown in Fig. 5.

4.1 Detection

The primary step in hand gesture recognition systems is the detection of hands and the segmentation of the corresponding image regions. This segmentation is crucial because it isolates the task-relevant data from the image background, before passing them to the subsequent tracking and recognition stages. A large number of methods have been proposed in the literature that utilize a several types of visual features and, in many cases, their combination. Such features are skin color, shape, motion and anatomical models of hands. In Cote et al. (2006), Zabulis et al. (2009), a comparative study on the performance of some hand segmentation techniques can be found.

4.1.1 Color

Skin color segmentation has been utilized by several approaches for hand detection. A major decision towards providing a model of skin color is the selection of the color space to be employed. Several color spaces have been proposed including RGB, normalized RGB, HSV, YCrCb, YUV, etc. Color spaces efficiently separating the chromaticity from the luminance components of color are typically considered preferable. This is due to the fact that by employing chromaticity-dependent components of color only, some degree of robustness to illumination changes can be achieved. Terrillon et al. (2000) review different skin chromaticity models and evaluate their performance. To increase invariance against illumination variability some methods (Bradski 1998; Kampmann 1998; Francois and Medioni 1999; Kurata et al. 2001) operate in the HSV (Saxe and Foulds 1996), YCrCb (Chai and Ngan 1998), or YUV (Yang et al. 1998a) colorspaces, in order to approximate the "chromaticity" of skin rather than its apparent color value. They typically eliminate the luminance component, to remove the effect of shadows, illumination changes, as well as modulations of orientation of the skin surface relative to the light source(s). The remaining 2D color vector is nearly constant for skin regions and a 2D histogram of the pixels from a region containing skin shows a strong peak at the skin color.

The perceived color of human skin varies greatly across human races or even between individuals of the same race. Additional variability may be introduced due to changing illumination conditions and/or camera characteristics. Therefore, color-based approaches to hand detection need to employ some means for compensating for this variability. In Yang and Ahuja (1998), Sigal et al. (2004), an invariant representation of skin color against changes in illumination is pursued, but still with not conclusive results. In Yang et al. (1998b), an adaptation technique estimates the new parameters for the mean and covariance of the multivariate Gaussian skin color distribution, based on a linear combination of previous parameters. However, most of these methods are still sensitive to quickly changing or mixed lighting conditions.

In general, color segmentation can be confused by background objects that have a color distribution similar to human skin. A way to cope with this problem is based on background subtraction (Rehg and Kanade 1994). However, background subtraction is typically based on the assumption that the camera system does not move with respect to a static background. To solve this problem, some research (Utsumi and Ohya 1998; Blake et al. 1999), has looked into the dynamic correction of background models and/or background compensation methods.

4.1.2 Shape

The characteristic shape of hands has been utilized to detect them in images in multiple ways. Much information can be obtained by just extracting the contours of objects in the image. If correctly detected, the contour represents the shape of the hand and is therefore not directly dependent on viewpoint, skin color and illumination. On the other hand, the expressive power of 2D shape can be hindered by occlusions or degenerate viewpoints. In the general case, contour extraction that is based on edge detection results in a large number of edges that belong to the hands but also to irrelevant background objects. Therefore, sophisticated post-processing approaches are required to increase the reliability of such an approach. In this spirit, edges are often combined with skin color and background subtraction motion cues.

Local topological descriptors have been used to match a model with the edges in the image. In Belongie et al. (2002), the shape context descriptor is proposed, which characterizes a particular point location on the shape. This descriptor is the histogram of the relative polar coordinates of all other points. Detection is based on the assumption that corresponding points on two different shapes will ideally have a similar shape context. In Song and Takatsuka (2005), the fingertip of the user was detected in both images of a calibrated stereo pair. In these images, the two points at which this tip appears establish a stereo correspondence, which is utilized to estimate the fingertip's position is 3D space. In turn, this position is utilized by the system to estimate the distance of the finger from the desk and, therefore, determine if the user is touching it. In Argyros and Lourakis (2006), stereoscopic information is used to provide 3D positions of hand centroids and fingertips but also to reconstruct the 3D contour of detected and tracked hands in real time. In Yin and Xie (2003) stereo correspondences of multiple fingertips have been utilized to calibrate a stereo pair. The main disadvantage in the use of fingertips as features is that they can be occluded by the rest of the hand. A solution to this occlusion problem involves the use of multiple cameras (Rehg and Kanade 1994; Lee and Kunii 1995). Other solutions are based on the estimation of the occluded fingertip positions, based on the knowledge of the 3D model of the gesture in question (Shimada et al. 1998; Wu et al. 2001; Wu and Huang 1999b; Rehg and Kanade 1995). Approach in Segen and Kumar (1999), utilizes as input hand images against a homogeneous and planar background. The illumination is such that the hand's shadow is cast on the background plane. By corresponding high-curvature features of the hand's silhouette and the shadows, depth cues such as vanishing points are extracted and the hand's pose is estimated.

4.1.3 Pixel values

Significant work has been carried out on finding hands in grey level images based on their appearance and texture. In Wu and Huang (2000), the suitability of a number of classification methods for the purpose of view-independent hand posture recognition was investigated. Several methods (Cui et al. 1995; Cui and Weng 1996; Triesch and Malsburg 1996; Triesch and Von der Malsburg 1998) attempt to detect hands based on hand appearances, by training classifiers on a set of image samples. The basic assumption is that hand appearance differs more among hand gestures than it differs among different people performing the same gesture. Still, automatic feature selection constitutes a major difficulty. More recently, methods based on a machine learning approach called boosting have demonstrated very robust results in face and hand detection. Due to these results, they are reviewed in more detail below. Boosting is a general method that can be used for improving the accuracy of a given learning algorithm (Schapire 2002). It is based on the principle that a highly accurate or "strong" classifier can be derived through the linear combination of many relatively inaccurate or\weak" classifiers. In general, an individual weak classifier is required to perform only slightly better than random. As proposed in Viola and Jones (2001) for the problem of hand detection, a weak classifier might be a simple detector based on basic image block differences efficiently calculated using an integral image.

The AdaBoost algorithm (Freund and Schapire 1997) provides a learning method for finding suit-able collections of weak classifiers. For training, it employs an exponential loss function that models the upper bound of the training error. The method utilizes a training set of images that consists of positive and negative examples (hands and non-hands, in this case), which are associated with corresponding labels. Weak classifiers are added sequentially into an existing set of already selected weak classifiers in order to decrease the upper bound of the training error. It is known that this is possible if weak classifiers are of a particular form (Friedman et al. 2000). However, this method may result in an excessive number of weak classifiers. The problem is that AdaBoost does not consider the removal of selected weak classifiers that no longer contribute to the detection process. The FloatBoost algorithm proposed in Li and Zhang (2004) extends the original AdaBoost algorithm, in that it removes an existing weak classifier from a strong classifier if it no longer contributes to the decrease of the training error. In the same context, the final detector can be divided into a cascade of strong classifier layers (Viola and Jones 2001).

4.1.4 3D model

A category of approaches utilize 3D hand models for the detection of hands in images. One of the advantages of these methods is that they can achieve view- independent detection.

The employed 3D models should have enough degrees of freedom to adapt to the dimensions of the hand(s) present in an image. Different models require different image features to construct feature-model correspondences. Point and line features are employed in kinematic hand models to recover angles formed at the joints of the hand (Rehg and Kanade 1995; Shimada et al. 1998; Wu and Huang 1999b; Wu et al. 2001). Hand postures are then estimated provided that the correspondences between the 3D model and the observed image features are well established. Various 3D hand models have been proposed in the literature. Some approaches (Lee and Kunii 1995; Heap and Hogg 1996) utilize a deformable model framework to fit a 3D model of the hand to image data. The fitting is guided by forces that attract the model to the image edges, balanced by other forces that tend to preserve continuity and evenness among surface points. In Lee and Kunii (1995), the process is enhanced with anatomical data of the human hand that are incorporated into the model. Also, to fit the hand model to an image of a real hand, characteristic points on the hand are identified in the images, and virtual springs are implied which pull these characteristic points to goal positions on the hand model.

4.1.5 Motion

Motion is a cue utilized by a few approaches to hand detection. The reason is that motion-based hand detection demands for a very controlled setup, since it assumes that the only motion in the image is due to hand movement. Indeed, early works (e.g. Cui and Weng 1996; Freeman and Weissman 1995) assumed that hand motion is the only motion occurring in the imaged environment. In more recent approaches, motion information is combined with additional visual cues. In the case of static cameras, the problem of motion estimation reduces to that of background maintenance and subsequent subtraction. For example in Cutler and Turk (1998), Martin et al. (1998) such information is utilized to distinguish hands from other skin-colored objects and cope with lighting conditions imposed by colored lights. The difference in luminance of pixels from two successive images is close to zero for pixels of the background. By choosing and maintaining an appropriate threshold, moving objects are detected within a static scene. In Yuan et al. (1995), a novel feature, based on motion residue, is proposed. Hands typically undergo non-rigid motion, because they are articulated objects. Consequently, hand detection capitalizes on the observation that for hands, inter-frame appearance changes are more frequent than for other objects such as clothes, face, and background.

4.2 Tracking

If the detection method is fast enough to operate at image acquisition frame rate, it can be used for tracking as well. However, tracking hands is notoriously difficult since they can move very fast and their appearance can change vastly within a few frames. Tracking can be defined as the frame-to-frame correspondence of the segmented hand regions or features towards understanding the observed hand movements. The importance of robust tracking is twofold. First, it provides the inter-frame linking of hand/finger appearances, giving rise to trajectories of features in time. These trajectories convey essential information regarding the gesture and might be used either in a raw form (e.g. in certain control applications like virtual drawing the tracked hand trajectory directly guides the drawing operation) or after further analysis (e.g. recognition of a certain type of hand gesture). Second, in model-based methods, tracking also provides a way to maintain estimates of model parameters variables and features that are not directly observable at a certain moment in time.

4.2.1 Template based

This class of methods exhibits great similarity to methods for hand detection. Members of this class invoke the hand detector at the spatial vicinity that the hand was detected in the previous frame, so as to drastically restrict the image search space. The implicit assumption for this method to succeed is that images are acquired frequently enough.

Correlation-based feature tracking: In Crowley et al. (1995) correlation-based template matching is utilized to track hand features across frames. Once the hand(s) have been detected in a frame, the image regions in which they appear is utilized as the prototype to detect the hand in the next frame. Again, the assumption is that hands will appear in the same spatial neighborhood. This technique is employed for a static camera in Darrell et al. (1996), to obtain characteristic patterns or signatures of gestures, as seen from a particular view. The work in Darrell et al. (1996) deals also with variable illumination. A target is viewed under various lighting conditions. Then, a set of basis images that can be used to approximate the appearance of the object viewed under various illumination conditions is constructed. Tracking simultaneously solves for the affine motion of the object and the illumination.

Real-time performance is achieved by pre-computing "motion templates" which are the product of the spatial derivatives of the reference image to be tracked and a set of motion fields. Some approaches detect hands as image blobs in each frame and temporally correspond blobs that occur in proximate locations across frames. Approaches that utilize this type of blob tracking are mainly the ones that detect hands based on skin color, the blob being the correspondingly segmented image region (e.g. Birk et al. 1997; Argyros and Lourakis 2004a). Blob-based approaches are able to retain tracking of hands even when there are great variations from frame to frame.

Contour based tracking: Deformable contours or "snakes" have been utilized to track hand regions in successive image frames (Cootes and Taylor 1992). Typically, the boundary of this region is determined by intensity or color gradient. Nevertheless, other types of image features (e.g. texture) can be considered. The technique is initialized by placing a contour near the region of interest. The contour is then iteratively deformed towards nearby edges to better fit the actual hand region. This deformation is performed through the optimization of energy functional that sums up the gradient at the locations of the snake while, at the same time, favoring the smoothness of the contour. When snakes are used for tracking, an active shape model is applied to each frame and the convergence of the snake in that frame is used as a starting point for the next frame. Snakes allow for real-time tracking and can handle multiple targets as well as complex hand postures. They exhibit better performance when there is sufficient contrast between the background and the object (Cootes et al. 1995). On the contrary, their performance is compromised in cluttered backgrounds. The reason is that the snake algorithm is sensitive to local optima of the energy function, often due to ill foreground/background separation or large object displacements and/or shape deformations between successive images.

4.2.2 Optimal estimation

Feature tracking has been extensively studied in computer vision. In this context, the optimal estimation framework provided by the Kalman filter (Kalman 1960) has been widely employed in turning observations (feature detection) into estimations (extracted trajectory). The reasons for its popularity are real-time performance, treatment of uncertainty, and the provision of predictions for the successive frames. In Argyros and Lourakis (2004b), the target is retained against cases where hands occlude each other, or appear as a single blob in the image, based on a hypothesis formulation and validation/rejection scheme. The problem of multiple blob tracking was investigated in Argyros and Lourakis (2004a), where blob tracking is performed in both images of a stereo pair and blobs are corresponded, not only across frames, but also across cameras. The obtained stereo information not only provides the 3D locations of the hands, but also facilitates the potential motion of the observing stereo pair which could be thus mounted on a robot that follows the user. In Utsumi and Ohya (1999), hands are tracked from multiple cameras, with a Kalman filter in each image, to estimate the 3D hand postures. Snakes integrated with the Kalman filtering framework have been used for tracking hands (Terzopoulos and Szeliski 1992). Robustness against background clutter is achieved in Peterfreund (1999), where the conventional image gradient is combined with optical flow to separate the foreground from the background.

4.2.3 Particle filtering

Particle filters have been utilized to track the position of hands and the configuration of fingers in dense visual clutter. In this approach, the belief of the system regarding the location of a hand is modeled with a set of particles. The approach exhibits advantages over Kalman filtering, because it is not limited by the unimodal nature of Gaussian densities that cannot represent simultaneous alternative hypotheses. A disadvantage of particle filters is that for complex models such as the human hand many particles are required, a fact which makes the problem intractable especially for high-dimensional models. Therefore, other assumptions are often utilized to reduce the number of particles. For example in Isard and Blake (1998), dimensionality is reduced by modeling commonly known constraints due to the anatomy of the hand.

The CONDENSATION algorithm (Isard and Blake 1998) which has been used to learn to track curves against cluttered backgrounds, exhibits better performance than Kalman filters, and operates in real-time. It uses "factored sampling", previously applied to the interpretation of static images, in which the probability distribution of possible interpretations is represented by a randomly generated set. Condensation uses learned dynamical models, together with visual observations, to propagate this random set over time. The result is highly robust tracking of agile motion. In Laptev and Lindeberg (2001), the state space is limited to 2D translation, planar rotation, scaling and the number of outstretched fingers. Extending the CONDENSATION algorithm the work in Mammen et al. (2001), detects occlusions with some uncertainty. In Perez et al. (2002), the same algorithm is integrated with color information; the approach is based on the principle of color histogram distance, but within a probabilistic framework, the work in introduces a new Monte Carlo tracking technique. In general, contour tracking techniques, typically, allow only a small subset of possible movements to maintain continuous deformation of contours.

4.2.4 CamShift

This algorithm is based on the principle of mean shift algorithm. Mean-shift is a kernel-based tracking method which uses density-based appearance models to represent targets. The method tracks targets by finding the most similar distribution pattern in a frame sequences with its sample pattern by iterative searching. It has been widely used because of its relative simplicity and low computational cost, but mean-shift would fail in changing the track window‘s scale, as targets move toward or away from the camera (Yilmaz et al. 2006; Bradski and Kaehler 2008). Based on mean-shift, continuous adaptive mean-shift (CamShift) was proposed to overcome the problem. CamShift adaptively adjusts the track window's size and the distribution pattern of targets during tracking. CamShift algorithm can be used to track the distribution of any kind of feature that represents the target in a lightweight, robust and efficient way (Bradski and Kaehler 2008). Most researchers though, use color data to represent targets for CamShift (Yin et al. 2009), this would give a low complexity and practical performance to the method. While CamShift performs well with objects that have a simple and constant appearance, it is not robust in more complex scenes. For example, when background has similar color distribution with a target, or when a target moves in front of different color background objects, the tracker is very likely to fail. In another case, when the initial search window contains some parts of the background, due to poor object detection, it would normally cause the CamShift's search window to drift and diverge. The problem of search window drift is inherent to many probability-based trackers, since these techniques only track the peak of a probability distribution and not the composition of probabilities. Several approaches combine other simple tracking methods with CamShift to improve the tracking performance, the approaches proposed in Xiangyu and Xiujuan (2010), Huang and Hong (2011) for example, combine CamShift algorithm with Kalman filter. In Xiangyu and Xiujuan (2010), the possible positions of a target are predicted by Kalman filter, and then CamShift is used to search and match the target in the predicted areas.

4.3 Recognition

The overall goal of hand gesture recognition is the interpretation of the semantics that the hand(s) location, posture, or gesture conveys. Vision based hand gesture recognition techniques can be further classified under static and dynamic gestures. To detect static gestures (i.e. postures), a general classifier or a template-matcher can be used. However, dynamic hand gestures have a temporal aspect and require techniques that handle this dimension like Hidden Markov Models (HMM) unless the temporal dimension is modeled through the hand gesture representation (e.g. motion based model). The static hand gestures are further classified into linear learner and non-linear learner. The former is suited for linearly separable data and the latter for the other cases. Another way to categorize learning algorithms is to consider their outcome. Thus, it distinguishes supervised learning (i.e. matching samples to labels), unsupervised learning (i.e. only sample clusters without labels), semi-supervised learning (i.e. mix of labelled and un-labelled data), reinforcement learning (i.e. learns policies given observations, transduction (i.e. supervised learning with prediction, (Li and Wechsler 2005) and learning to learn (i.e. learns his own inductive bias based on previous experience (Baxter 2000). The choice of the learning algorithm depends mainly on the chosen hand gesture representation. For example, Ren and Zhang (2009a) propose to recognize static hand gestures by learning the contour line's Fourier descriptor of a segmentation image obtained by mean shift algorithm.

With learning algorithms, automata-based methods are the most common approaches detailed in the literature. For instance, FSMs, HMMs, PNF (i.e. Past-Now-Future network) are sort of automaton with a set of states and a set of transitions. The states represent static hand gestures (i.e. postures) and transitions represent allowed changes with temporal and/or probabilistic constraints. A dynamic hand gesture is then considered as a path between an initial state and a final state. Lee and Kim (1999) propose an approach for gesture recognition using HMM based threshold. Lu and Little (2006) present a method for recognizing human gestures using PCA-HOG global descriptor. The main limitation of the approaches based on automata is that the gesture model must be modified when a new gesture needs to be recognized. Moreover, the computational complexity of such approaches is generally huge since it is proportional to the number of gestures to be recognized which is not the case for methods based on other techniques. Some of the common techniques used for static and dynamic hand gesture recognition are as follows:

4.3.1 K-means

The k-means problem is to determine k points called centers so as to minimize the clustering error, defined as the sum of the distances of all data points to their respective cluster centers. This classification finds statistically similar groups in multi-spectral space (Francois and Medioni 1999; Lu and Little 2006). The algorithm starts by randomly locating k clusters in spectral space. Each pixel in the input image group is then assigned to the nearest cluster centre and the cluster centre locations are moved to the average of their class values. This process is then repeated until a stopping condition is met. The stopping condition may either be a maximum number of iterations (specified by the user) or a tolerance threshold which designates the smallest possible distance to move cluster centres before stopping the iterative process. The most commonly used algorithm for solving this problem is the Lloyd's k-means algorithm (Lloyd 1982; Kanungo et al. 2002) which iteratively assigns the patterns to clusters and computes the cluster centers.

MacQueens k-means algorithm (MacQueen 1967) is a two-pass variant of the Lloyd's k-means algorithm: choose the first k patterns as the initial k centers and the assign each of the remaining N–k patterns to the cluster whose center is closest. Calculate the new centers of the clusters obtained. Assign each of the N patterns to one of the k clusters obtained in step 1 based on its distance from the cluster centers and recompute the centers.

4.3.2 K-nearest neighbor

It is a method for classifying objects based on closest training examples in the feature space. k-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computations are deferred until classification (Thirumuruganathan 2010). An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors; k is a positive integer, typically small. If k = 1, then the object is simply assigned to the class of its nearest neighbor. In binary (two class) classification problems, it is helpful to choose k to be an odd number as this avoids tied votes. The same method can be used for regression, by simply assigning the property value for the object to be the average of the values of its k nearest neighbors. It can be useful to weight the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. The neighbors are taken from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. In order to identify neighbors, the objects are represented by position vectors in a multidimensional feature space. It is usual to use the Euclidean distance, though other distance measures, such as the Manhattan distance could in principle be used instead. The k-nearest neighbor algorithm is sensitive to the local structure of the data.

4.3.3 Mean shift clustering

It is a nonparametric clustering technique (Derpanis 2005) which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters.

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. For each data point in the feature space, one performs a gradient ascent procedure on the local estimated density until convergence. The stationary points of this procedure represent the modes of the distribution. Furthermore, the data points associated (at least approximately) with the same stationary point are considered members of the same cluster.

4.3.4 Support vector machine

SVM is a non-linear classifier (Burges 1998) which is often reported as producing superior classification results compared to other methods. The idea behind the method is to non-linearly map the input data to some high dimensional space, where the data can be linearly separated, thus providing great classification (or regression) performance. One of the bottlenecks of the SVM is the large number of support vectors used from the training set to perform classification (regression) tasks. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. In addition to performing linear classification, SVMs can efficiently perform non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

4.3.5 Hidden Markov model

HMM were introduced in the mid 1990s, and quickly became the recognition method of choice, due to its implicit solution to the segmentation problem. In describing hidden Markov models (Ramage 2007) it is convenient first to consider Markov chains. Markov chains are simply finite-state automata in which each state transition arc has an associated probability value; the probability values of the arcs leaving a single state sum to one. Markov chains impose the restriction on the finite-state automaton that a state can have only one transition arc with a given output; a restriction that makes Markov chains deterministic. A hidden markov model (HMM) can be considered a generalization of a Markov chain without this Markov chain restriction (Charniak 1993). Since HMMs can have more than one arc with the same output symbol, they are nondeterministic, and it is impossible to directly determine the state sequence for a set of inputs simply by looking at the output (hence the “hidden” in “hidden Markov model”). More formally, a HMM is defined as a set of states of which one state is the initial state, a set of output symbols, and a set of state transitions. Each state transition is represented by the state from which the transition starts, the state to which transition moves, the output symbol generated, and the probability that the transition is taken (Charniak 1993). In the context of hand gesture recognition, each state could represent a set of possible hand positions. The state transitions represent the probability that a certain hand position transitions into another; the corresponding output symbol represents a specific posture and sequences of output symbols represent a hand gesture. One then uses a group of HMMs, one for each gesture, and runs a sequence of input data through each HMM. The input data, derived from pixels in a vision-based solution can be represented in many different ways, the most common by feature vectors (Staner and Pentland 1995a). The HMM with the highest forward probability determines the users' most likely gesture. An HMM can also be used for hand posture recognition; see Liang and Ouhyoung (1996) for details.

4.3.6 Dynamic time warping

It has long been used to find the optimal alignment of two signals. The DTW algorithm (Senin 2008) calculates the distance between each possible pair of points out of two signals in terms of their associated feature values. It uses these distances to calculate a cumulative distance matrix and finds the least expensive path through this matrix. This path represents the ideal warp—the synchronization of the two signals which causes the feature distance between their synchronized points to be minimized. Usually, the signals are normalized and smoothed before the distances between points are calculated. DTW has been used in various fields, such as speech recognition, data mining, and movement recognition (Andrea 2001; Gavrila and Davis 1995). Previous work in the field of DTW mainly focused on speeding up the algorithm, the complexity of which is quadratic in the length of the series. Examples are applying constraints to DTW (Eamonn and Pazzani 2001), approximation of the algorithm (Stan and Philip 2004) and lower bounding techniques (Keogh and Ratanamahatana 2005). Eamonn and Pazzani (2001) proposed a form of DTW called Derivative DTW (DDTW). Here, the distances calculated are not between the feature values of the points, but between their associated first order derivatives. In this way, synchronization is based on shape characteristics (slopes, peaks) rather than simple values. Most work, however, only considered one-dimensional series.

4.3.7 Time delay neural networks

These are special artificial neural networks which focus on working with continuous data making the architecture adaptable to online networks hence advantageous to real time applications. Theoretically, time delay neural networks are also considered as an extension of multi-layer perceptron. TDNN Sigal et al. (2004), Wohler and Anlauf (1999) is based on time delays which gives individual neurons the ability to store the history of their input signals. Therefore the network can adapt to sequence of patterns. Due to the concept of time delay, each neuron has access not only to present input at time t but also to the inputs at time $t_1, t_2, \ldots, t_n$. Therefore each neuron can detect relationship between the current and former input values which might be a typical pattern in the input signal. Also, the network is able to approximate functions that are derived from time sampled history of input signal. Learning of typical TDNN can be accomplished by standard back propagation as well as its variants.

4.3.8 Finite state machine

A finite state machine (Holzmann) is one that has a limited or finite number of possible states (an infinite state machine can be conceived but is not practical). A finite state machine can be used both as a development tool for approaching and solving problems and as a formal way of describing the solution for later developers and system maintainers. There are a number of ways to show state machines, from simple tables through graphically animated illustrations. Usually, the training of the model is done off-line, using many possible examples of each gesture as training data, and the parameters (criteria or characteristics) of each state in the FSM are derived. The recognition of hand gestures can be performed online using the trained FSM. When input data (feature vectors such as trajectories) are supplied to the gesture recognizer, the latter decides whether to stay at the current state of the FSM or jump to the next state based on the parameters of the input data. If it reaches a final state, we say that a gesture has been recognized. Table 3 provides a comparison between vision based hand gesture recognition techniques.

5 Application domains

Vision based hand gestures recognition systems since its early days of exploration and research have found vital applications to a wide range of real life and real time scenarios. The evolution of human computer interaction has been paced up with the advances in pervasive computing and real time application scenarios of computing devices. The application domain are subcategorized under core and advanced applications.

Though the recent or advanced applications of interaction systems are evolved formats of classical applications of core area to refined ones with latest avenues tapping the arid environments of applications. This section provides a brief overview of some of the core and advanced application domains of vision based hand gesture recognition systems. When we say core applications we are counting the core areas of computing applications that include information processing and visualization as the premier followed by robotics and sign language, virtual reality etc as shown in Fig. 6. While the advanced applications include other related application domains like tablet PC, games, medicine environment, augmented reality etc as shown in Fig. 7. Some of the application domains are discussed below.

In desktop applications, hand gestures can offer a substitute interaction medium for mouse and keyboard (Iannizzotto et al. 2001; Stotts et al. 2004a) as shown in Fig. 8a. Many hand gestures for desktop computing tasks involve manipulating graphic objects (Bolt and Herranz 1992) or annotating and editing documents using pen-based gestures (Cohen et al. 1997). Smith and Schraefel (2004) also use pen gestures, where circular motion creates a radial-scrolling effect for navigating through documents, while some authors (Lenman et al. 2002) make marking menu selections using stroke gestures. Mouse gestures are also used for various applications including web browsing tasks (Moyle and Cockburn 2002). But most of the gestures that are seen in desktop applications employ direct input devices such as a pen or mouse (Cohen et al. 1997). Similar gestures are commonly used for tablet computers in specialized applications for air traffic control rooms (Chatty and Lecoanet 1996), adaptive technology (Ward et al. 2000) and musical score editing (Forsberg et al. 1998). Gestures using non-direct input devices for desktop computing also include nodding to respond to dialogue boxes (Davis and Vaks 2001); however most of the applications for desktop domains use the standard direct-input devices.

Hand gestures for virtual reality applications have experienced one of the greatest levels of uptake in computing. Virtual reality interactions use gestures to enable realistic manipulations of virtual objects using ones hands, for 3D display interactions (Sharma et al. 1996) as shown in Fig. 8d or 2D displays that simulate 3D interactions (Gandy et al. 2000). We identified three sub-categories of virtual reality applications where gestures are primarily employed: non-immersed interactions, where the users are not represented within the virtual world, semi-immersive interactions, where a user is represented in the virtual world as an avatar, and fully-immersive interactions where the user interacts from the perspective of being inside the virtual world. Osawa et al. (2000) used hand gestures to arrange virtual objects and to navigate around a 3D information space such as a graph, using a stereoscopic display. These interactions do not require any representations of the user and user performs

Table 3 Comparison between different vision based hand gesture recognition techniques

Technique	Principle	Parameters	Advantages	Limitations
K-means	To determine k points called . centers as the sum of the distances of all data points to their respective cluster centers	Cluster centre locations	Computationally faster, produce tighter clusters	Prediction of K is difficult for fixed number of clusters. Different initial partitions result in different final clusters
K-nearest neighbor	Closest training of the feature space using instance-based learning	Class of nearest neighbor	Easy to implement. Lowest complexity. Carefully chosen features give good results	Sensitive to arbitrary attributes
Mean shift clustering	Nonparametric clustering technique based on prior knowledge of clusters	Empirical probability density function	Spherical clusters not assumed works on a single parameter (window size). Robust to outliers	Ambiguity in optimal parameter selection. Computationally expensive, does not scale well with dimension of feature space
Support vector machine	Non-linearly map the input data dimensional space, having linearly separated data for classification	High-dimensional feature spaces	Higher prediction accuracy. Robust for errors in training examples. Fast evaluation of the learned target function	Long training time, complex learned function (weights), domain knowledge incorporation not easy
Hidden Markov model	Generalization of a Markov chain without Markov chain restriction. Set of states, transitions represent the set of possible hand positions	Pixels in a vision-based input	Easily extended to deal with strong TC tasks. Embedded re-estimation is possible easy to understand	Large assumptions about the data. Huge number of parameters needs to be set. Training data required is large
Dynamic time warping	Optimal alignment of features is found and ideal warp is obtained based on cumulative distance matrix	Shape characteristics.	Reliable time alignment robust to noise and easy to implement	Complexity is quadratic, distance metric needs to be defined. Dynamic time warping to strong TC tasks not achieved

Table 3 continued

Technique	Principle	Parameters	Advantages	Limitations
Time delay neural network	Special artificial neural networks based on time delays giving individual neurons ability to store history making the system adapt to sequence of patterns	Time sampled history of input signal	Faster learning, invariance under time or space translation. Faster execution	Lacking Robustness. Based on typical patterns of input that is inconsistent
Finite state machine	Finite state machine having limited or finite number of possible states	Feature vectors such as trajectories	Easy to implement. Efficient predictability, low processor overhead	Not Robust, ridged conditions of implementation

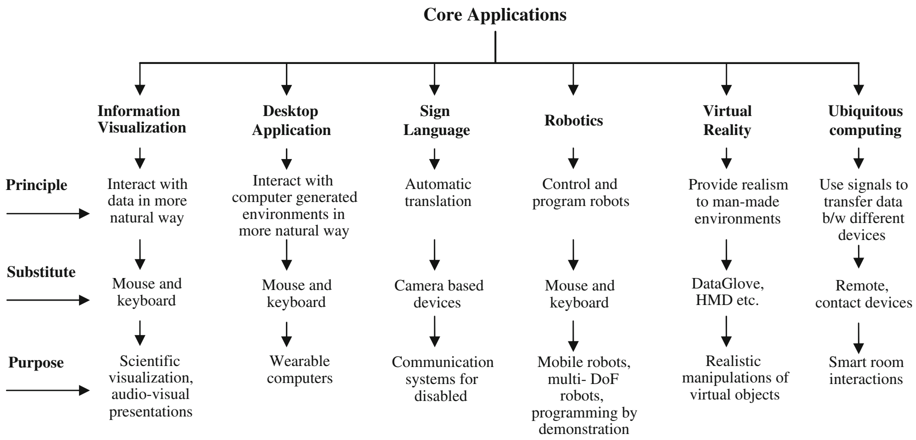

Fig. 6 Core applications

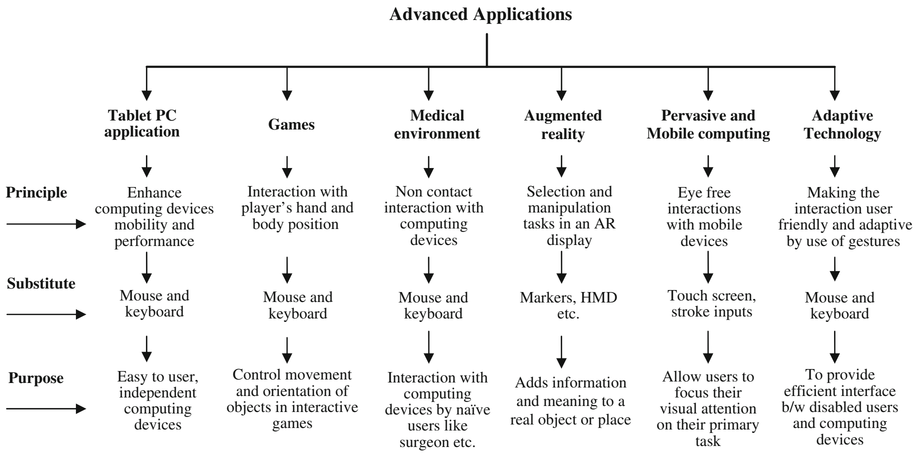

Fig. 7 Advanced applications

gestures primarily for navigation and manipulation tasks. The hand gestures used to interact with control robots are similar to fully-immersed virtual reality interactions, however the worlds are often real, presenting the operator with video feed from cameras located on the robot (Goza et al. 2004). Here, gestures can control a robots hand and arm movements to reach for and manipulate actual objects, as well its movement through the world.

Early work on gestures demonstrated how distance interactions for displays or devices could be enabled within smart room environments. Examples of smart room interactions use gestures to signal the transfer of data between different devices (Swindells et al. 2002). As smart room technologies became more sophisticated, so did the notion of using perceptual style input to enable gestures to control smart room applications (Crowley and Jolle Coutaz 2000), which included controlling lights, entertainment units or appliances (Wilson and Shafer 2003; Nickel and Stiefelhagen 2003) and interactions with large screen displays (Hardenberg and Berard 2001).

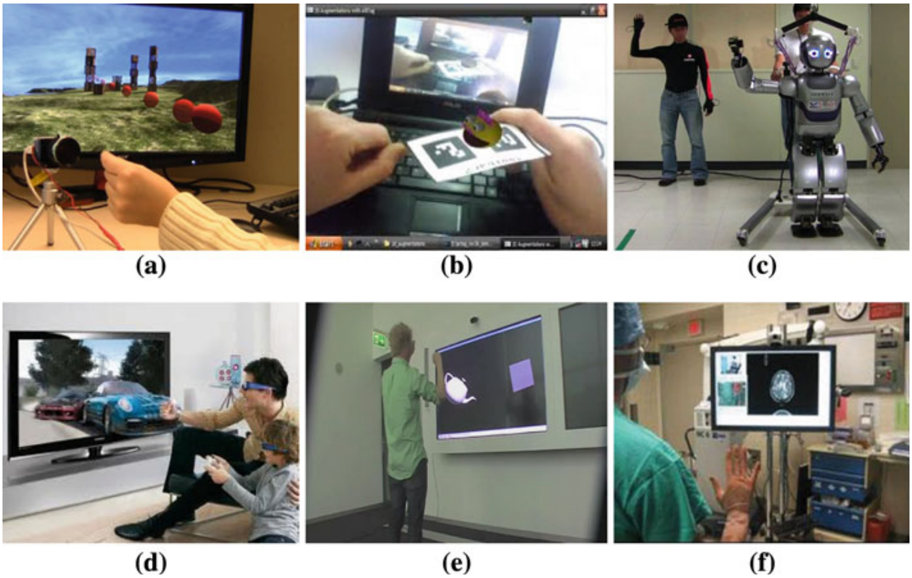

Fig. 8 Some key application domain for vision based hand gesture recognition. **a** Desktop computing interaction using hand gestures (Lenman et al. 2002). **b** Augmented reality based interaction using hand gestures (Radkowski and Stritzke 2012). **c** A robotic learning using hand gestures (Goza et al. 2004). **d** Virtual reality based gaming using hand gestures (Sharma et al. 1996). **e** Hand gesture for interaction with large screen displays (Cao and Balakrishnan 2003). **f** Surgeons interacting with computing devices using hand gestures in medical environment (Schultz et al. 2003)

Finally, we look at hand gestures for computer games. Freeman et al. (1996) tracked a player's hand or body position to control movement and orientation of interactive game objects such as cars, while Segen and Kumar (1998a) tracked motion to navigate a game environment. Konrad et al. (2003) used gestures to control the movement of avatars in a virtual world, Paiva et al. (2002) employed pointing gestures as a virtual-reality game input, and PlayStation 2 has introduced the Eye Toy, a camera that tracks hand movements for interactive games (Eyetoy 2003).

Augmented realty applications often user markers, consisting of patterns printed on physical objects, which can more easily be tracked using computer vision, and that are used for displaying virtual objects in augmented reality displays as shown in Fig. 8b. Buchmann et al. (2004) used such markers in combination with bare hand movements to recognize gestures for selection and manipulation tasks in an AR display.

Hand gestures are also used in computer supported collaborative work (CSCW) applications to enable multiple users to interact with a shared display, using a variety of computing devices such as desktop or tabletop displays (Wu and Balakrishnan 2003) or large screen displays (Cao and Balakrishnan 2003) as shown in Fig. 8e. Notes and annotations can be shared within groups using strokes either locally or for remote interactions (Stotts et al. 2004b). Annotations can be transmitted using live video streams, to enable remote collaborations between students and instructors (Ou et al. 2003).

Hand gestures can enable eyes-free interactions with mobile devices that allow users to focus their visual attention on their primary task (Schmandt et al. 2002; Lumsden and Brewster 2003). PDA's augmented with touch sensitive screens can also interpret finger gestures or strokes as input, and provide audio output to the user to support eyes-free interactions with mobile devices. Computer technology is now ubiquitous within automotive design, but gestures have not yet received a great deal of attention in the research literature. Alpern and Minardo (2003) explored the use of gesturing with telematics to enable secondary task interactions to reduce the distraction caused to the primary task of driving. A number of hand gesture recognition techniques for human vehicle interface have been proposed time to time. Additional literature explores gesture for telematics applications (Pickering 2005) to minimize distraction while driving. The primary motivation of research into the use of hand gestures for in-vehicle secondary controls is broadly based on the premise that taking the eyes off the road to operate conventional secondary controls can be reduced by using hand gestures.

Computer information technology is increasingly penetrating into the medical domain. It is important that such technology be used in a safe manner to avoid serious mistakes leading to possible fatal incidents. Keyboards and mice are today's principle method of human computer interaction. Unfortunately, it has been found that a common method of spreading infection from one person to another involves computer keyboards and mice in intensive care units (ICUs) used by doctors and nurses (Schultz et al. 2003) as shown in Fig. 8f. In a setting like an operating room (OR), touch screen displays must be sealed to prevent the buildup of contaminants, and require smooth surfaces for easy cleaning with common cleaning solutions. A gesture plays an important role in such situations for interaction with computing devices. In face mouse (Nishikawa et al. 2003), a surgeon can control the motion of the laparoscope by simply making the appropriate face gesture, without hand or foot switches or voice input. Graetzel et al. (2004) developed a computer vision system that enables surgeons to perform standard mouse functions like pointer movement and button presses with hand gestures.

Other systems (Wachs et al. 2002) suggest a tele-operated robotic arm using hand gestures for multipurpose tasks. Gestures are not the most common technique for adaptive interfaces since they require movement, and this may not be conducive to some physical disabilities, although some technology, such as the DataGlove, has been used to measure and track hand impairment in disabled users. Segen and Kumar (1998b) extends previous work to develop a system for wheelchair navigation using gestures, while the gesture–pendant (Gandy et al. 2000) was also extended for adaptive interfaces for home emergency services, enabling control of devices for home patients with vision or physical impairment. Since gestures typically require movements, they may not be the primary choice for adaptive interactions; however some gestures can require only minimal motion, and can require reduced mobility than a mouse or keyboard for text entry or desktop computer interactions (Ward et al. 2000). Pausch and Williams (1990) demonstrate the sign language which may also be considered adaptive technology for users with hearing impairment, however current systems are not yet capable of interpreting large vocabulary sets fluidly, and can experience tracking limitations in the computer-vision implementations (Fang et al. 2003; Bowden et al. 2003).

Hand gestures can be used for analyzing and annotating video sequences of technical talks. Such a system is presented in Ju et al. (1997), Charniak (1993). Speaker's gestures such as pointing or writing are automatically tracked and recognized to provide a rich annotation of the sequence that can be used to access a condensed version of the talk. Given the constrained domain a simple "vocabulary" of actions is defined, which can easily be recognized based on the active contour shape and motion. The recognized actions provide a rich annotation of the sequence that can be used to access a condensed version of the talk from a web page. Also, hand gestures are an attractive method for communication with the deaf and dumb. One of the most structured sets of gestures is those belonging to any of the several sign languages. In sign language, each gesture already has assigned meaning, and strong rules of context and grammar may be applied to make recognition tractable. Starner and Pentland (1995b) described an extensible system which used a single color camera to track hands in real time and interpret American Sign Language (ASL).

For achieving natural human computer interaction for virtual environments, Berry (1998) integrated controlling gestures into the virtual environment BattleField. In this system hand gestures are used not only for navigating the VE, but also as an interactive device to select and move the virtual objects in the BattleField. Another system (Zeller et al. 1997) where hand gestures serve as the input and controlling device of the virtual environment is presented.

The significant factor which makes vision based hand gesture recognition more practical for widespread use in different application domain as discussed above is due to its decrease in use of hardware and processing cost.

6 Commercial products and software's

The hand gestures considered as a promising research area for researchers in designing a natural and intuitive method for HCI. This section discusses and provides a comparison between different commercial products and software's based on vision based gesture recognition that has paved their way from HCI research labs to market as fully fledged commercial products. Table 4 compares some of the latest vision based hand gesture recognition commercial products and software's available for interacting with the computing world. IISU SoftKinetic (2012) is a product SoftKinectic based on real-time 3D gesture recognition technology that are used to be build gesture recognition application compatible with all 3D cameras and platform. Another product that uses the power of intuitive hand gesture to control the wide array of consumer electronics and digital devices is available in the market by eyeSight's (eyeSight's 2012).

These devices utilize advanced image processing and machine vision algorithms to track the users hand gesture and then convert them into commands. PointGrab's PointGrab's (2012) enables the integration of hand gestures with application such as games and customized user interfaces that are based on unique sophisticated hand shape and motion detection algorithms working together with a standard 2D camera. HandGET HandGKET (2011) has the toolkit that facilities integration of hand gesture control with games and VE applications. The middleware used in this product recognizes user's hand gestures and generates keyboard and mouse events to control applications based on computer vision techniques. Mgestyk developed in 2009 (Li and Wechsler 2005; Mgestyk 2009) employs software for hand gesture processing and 3D camera to interact with computer for operating games and applications. Based on the technology of 3D camera for computer vision, camera in mobile device and pointing frame GestureTek (GestureTek 2008) product is used for applications like controlling pc, mobile or console applications using camera or phone. Wii Nintendo (2006) has wireless and motion sensitive remote with game console and may be used for game with any pc etc. Another product HandVu (2003) launched in 2003 is also used for interaction with computer to operate games and application based on real-time gesture recognition using computer vision techniques. OMRON Corporation (OMRON 2012) developed a new hand gesture recognition technology capable of simultaneously recognizing the position, shape, and motion of a person's hand or finger by referencing a camera-recorded image. By combining this technology with OMRON's core facial image sensing technology, gesture recognition can be started automatically based on the analysis of interrelation between the position or direction of the face and the hand shape or position.

Table 4 Comparison between different vision based gesture recognition products and software

Year	SoftKinetic (2012)	PointGrab's (2012)	HandGKET (2011)	Mgestyk (2009)	GestureTek (2008)	Wii Nintendo (2006)	Microsoft Kinect (2012)	OMRON (2012)	eyeSight's (2012)	HandVu (2003)	Eyetoy (2003)
	2012	2012	2011	2009	2008	2006	2012	2012	2012	2003	2003
Platforms											
Windows	√	√	√	√	√	√	√	√	√	√	√
Linux	√	√						√			
Android		√						√			
Languages											
Java					√	√					
Python								√	√		
C++	√	√	√	√		√	√	√	√	√	√
C#	√		√	√			√		√		
Other		√			√		√			√	√
Input device											
Camera	√	√	√	√	√	√	√	√	√	√	√
Multitouch					√			√			√
Wii Remote						√					
Windows 7 touch	√	√				√	√		√		
TVIO	√	√		√							
Other			√					√	√	√	
Features											
Provides standard gestures	√	√	√	√	√	√	√	√	√	√	
Allows custom gestures	√										

Table 4 continued

Year	SoftKinetic (2012)	PointGrab's (2012)	HandGKET (2011)	Mgestyk (2009)	GestureTek (2008)	Wii Nintendo (2006)	Microsoft Kinect (2012)	OMRON (2012)	eyeSight's (2012)	HandVu (2003)	Eyetoy (2003)
	2012	2012	2011	2009	2008	2006	2012	2012	2012	2003	2003
Allows custom input devices					√				√	√	√
Provides gesture aware UI widget set		√	√			√	√	√			√
Support for non touch events	√				√	√	√				
Dynamically loaded gestures							√				

These commercial products developed and used are though in initial phases of acceptance which may still be made robust with user requirements and feedback. Though efficiency and user friendliness are the basic criteria's considered for design and development of such products the technological constraints make them liable to various lack of abilities, that has to be supported by the research and development in technological areas associated. The need for evolution of these products are manifold including cost effectiveness, robust under different real life and real time application environments, effective and end user acceptability.

7 Software platforms/frameworks

When implementing a technique/algorithm for developing an application which detects, tracks and recognize hand gestures, the main thing to consider is the methodology which to recognize the gestures. This section discusses platforms which supports gesture recognition in various methods and further in development from small to medium scale software applications.

7.1 OpenCV

It is an open source computer vision programming functions library aimed at developing applications based on real time computer vision technologies. This framework has BSD license which enables usage of the framework for both commercial and research purposes. OpenCV Bradski and Kaehler (2008) was originally developed in C but now consists of full libraries for C++, Python and supports for Android platform. It also supports hardware for Linux, MacOS X and Windows platforms providing extensive cross platform compatibility. OpenCV Eclipse IDE, C++ Builder, DevCpp IDE support provides developer's easy access to build applications with any of the IDEs listed above. Some example applications in OpenCV library are object identification, segmentation and recognition, face recognition, gesture recognition, motion tracking, mobile robotics etc. With this extensive features this framework also requires a strong knowledge in both development and integration methods to create a software application.

7.2 MATLAB

Matrix laboratory (MATLAB) is a numerical computing environment and fourth generation programming language. It was developed by MathWorks. MATLAB MATLAB allows implementation of algorithms, matrix manipulations, plotting of functions and data, creation of user interfaces and interfacing with programs written in other languages including C, C++, Java and Fortran. It also supports hardware for Linux, MacOS X and Windows platforms providing extensive cross platform compatibility. MATLAB provides Image Processing Toolbox™ which provides a comprehensive set of reference-standard algorithms and graphical tools for image processing, analysis, visualization, and algorithm development. Different task like image enhancement, image deblurring, feature detection, noise reduction, image segmentation, geometric transformations, image registration, object tracking, recognition etc can be performed using the toolbox. Many toolbox functions are multithreaded to take advantage of multicore and multiprocessor computers.

7.3 A Forge.NET

The A Forge .NET Framework (A Forge.NET) is an open source framework based on C#.NET designed for developers to provide a platform to develop applications in Computer Vision and Artificial Intelligence. This framework supports image processing, neural networks, genetic algorithms, machine learning and robotics. The framework is published under LGPL v3 License. This framework has gained popularity because its ease of use, effectiveness and short learning period. This framework is compatible with .NET Framework 2.0 and above. This framework can be easily integrated with Microsoft Visual Studio IDE for development. The framework consists of set of libraries and sample applications that can be used as a basis to develop a gesture recognition application.

7.4 iGesture

It is a well reputed and older gesture recognition framework is iGesture (iGesture). This framework is a Java based and focused on extendibility and cross-platform reusability. The distinctive feature of the iGesture framework is that it supports both developers and designers to develop new hand gesture recognition algorithms. iGesture integrated framework includes the gesture recognition framework and 'I' gesture tool component to create custom gestures sets. This makes it better compared to other frameworks as the other frameworks have predefined gestures and the developers are limited to those gestures. Also iGesture tools provide the ability to evaluate the usability, performance, effectiveness of new and existing hand gesture recognition algorithms. The main disadvantage of this framework is its long learning period because of the extensive usages the framework offers the developers must have a good understanding of the principals and methods of using this framework in software applications.

8 State of the art and discussion

The research in hand gesture recognition has evolved a lot since its advent with increased usage of computing devices in our day to day life. This section of the survey is an effort to accumulate some of the previous work done in the field of human computer interaction. The effort is to classify the state of the art on the basis of the discussion done in section III, IV, V and VII *(i.e. gesture representations: 3D model based or appearance based, techniques: static, dynamic and further highlighting the key methods discussed, software platform, background (BG.) conditions and application/ task performed)* between man and machine interaction over the period of years from 2005 to 2012 till the study has been done. The exhaustive list of the work done in the aforementioned era of development has been given in the following Table 5. Though the list may not be exclusive enough the effort has been made to make it as exhaustive as possible. The detailed analysis of the table presents a lot of interesting facts about the ongoing research in field of human computer interaction in general and vision based hand gesture recognition in particular.

The literature presents various interesting facts that compare and contrast the two object representation techniques i.e. 3D model and appearance based. The former representation technique is generally based on computer aided design through wired model of the object. While the later segments the potential region having object of interest from the given input sequence. Though the 3D model allows for real time object representation along with low computing efforts the major difficulty with the approach is that the system is limited by the

Table 5 Analysis of some vital literature related to vision based hand gesture recognition systems

Ref.	Year	Represent		Techniques			Software platform	BG. condition	Application/ task
		3D	App.	Static	Dynamic	Key methods			
Cheng et al. (2012)	2012	√			√	PCA, SVM	M	S	Digit gestures
Sangineto and Cupelli (2012)	2012		√		√	Hybrid	OC	C	Drawing lines
Wang et al. (2012)	2012	√			√	Hybrid	OC	S	Dutch sign language
Radkowski and Stritzke (2012)	2012	√			√	AR Tool Kit	OC	S	3D modeling of technical systems
Tran and Trivedi (2012)	2012	√			√	KC- GMM	M	S	Hand activity recognition
Rautaray and Agrawal (2012)	2012		√	√		Hybrid	OC	S	Interaction in virtual environment
Reale et al. (2011)	2011		√	√		Hybrid	OC	S	Visualization, navigation
Várkonyi-Kóczy and Tusor (2011)	2011	√		√	√	Fuzzy-neural network	M	S	Control smart environment
Gorce et al. (2011)	2011	√		√		Hybrid	M	O/Iv	Hand pose recognition
Sajjawiso and Kanongchaiyos (2011)	2011	√		√		Hybrid	M	NM	Hand pose recognition
Henia and Bouakaz (2011)	2011	√			√	PCA	M	S	Animate 3D hand model
Ionescu et al. (2011a)	2011		√		√	Hybrid	OC	S	Physical game controller

Table 5 continued

Ref.	Year	Represent		Techniques			Software platform	BG. condition	Application/ task
		3D	App.	Static	Dynamic	Key methods			
Yang et al. (2011)	2011		√	√		KNN, Fourier descriptor	OC	S	Controlling Microsoft PPT
Huang et al. (2011a)	2011		√		√	Hybrid	NM	S	VR based MIVS
Ionescu et al. (2011b)	2011		√	√	√	Hybrid	NM	S/Iv	Controlling TV and set top box
Bergh and Gool (2011)	2011		√	√		ANMM	OC	O/Iv/C	Manipulate 3D gesture models
Bao et al. (2011)	2011		√		√	Hybrid	OC	C	26 alphabetical hand gesture
Bellarbi et al. (2011)	2011		√		√	Hybrid	OC	S	Virtual keyboard
Hackenberg et al. (2011)	2011		√	√	√	Hybrid	OC	S	Manipulating objects
Du et al. (2011)	2011	√		√		Hybrid	NM	S	Virtual hand interaction system
Rautaray and Agrawal (2011)	2011		√	√		Hybrid	OC	S	Controlling VLC media player
Visser and Hopf (2011)	2011		√	√		Hybrid	NM	S	3D application
He et al. (2011)	2011		√	√		Hybrid	NM	S	Gaming application
De Tan and Geo (2011)	2011		√	√		Hybrid	OC	NM	Manipulating VR objects

Table 5 continued

Ref.	Year	Represent		Techniques			Software platform	BG. condition	Application/ task
		3D	App.	Static	Dynamic	Key methods			
Ho et al. (2011)	2011	√		√		Hybrid	NM	O	3D gesture modeling
Huang et al. (2011b)	2011		√	√	√	Gabor filter, SVM	NM	Iv	Hand pose recognition
Hsieh et al. (2010)	2010		√		√	Hybrid	OC	O/Iv/C	Control mouse for desktop app.
Pang et al. (2010)	2010		√	√		Hybrid	OC	Iv	Control mouse for desktop app.
Vo et al. (2010)	2010		√		√	Hybrid	OC	NM	Control mouse for desktop app
Lee and Hong (2010)	2010		√		√	Hybrid	NM	S	Chinese chess game
Murthy and Jadon (2010)	2010		√	√		Neural network	M	Iv	Hand gesture modeling
Hu et al. (2010)	2010		√	√		AAM	OC	S	3D hand gesture Modeling
Zhu and Pun (2010)	2010		√		√	Hybrid	OC	S	Digit recognition
Appenrodt et al. (2010)	2010	√		√		ICP, Genetic Algorithm	OC	O/C	Hand modeling
Lin et al. (2010)	2010	√			√	Hybrid	OC	S	Controlling media player
Elmezain et al. (2010)	2010		√		√	HMM, CRF	OC	S	Digit recognition
Rodriguez et al. (2010)	2010		√		√	Hybrid	NM	Iv	3D manipulation of object

Table 5 continued

Ref.	Year	Represent		Techniques			Software platform	BG. condition	Application/ task
		3D	App.	Static	Dynamic	Key methods			
Deng et al. (2010)	2010		√		√	Shape context based	OC	S	Hand gesture modeling
Yuan et al. (2010)	2010		√		√	HMM	NM	S	Desktop application
Suka et al. (2010)	2010		√		√	DBN, NN	NM	C	Sign language
Ibarguren et al. (2010)	2010	√		√		Hybrid	NM	S	Sign language
Elmezain et al. (2009)	2009		√		√	HMM	NM	C	Arabic number recognition
Alon et al. (2009)	2009		√		√	Hybrid	OC	O/Iv/C	American sign language
Lee and Park (2009)	2009		√	√		Hybrid	OC	S	Controlling game
Rashid et al. (2009)	2009		√		√	HMM, SVM	OC	S	ASL Alphabets
Zhao et al. (2009)	2009		√	√		Fourier des., BPNN	NM	S	Virtual reality application
Chung et al. (2009)	2009		√	√		Haar Wavelet	OC	S	Alphabetic hand gesture
Yun and Peng (2009)	2009		√	√		Viola Jones, SVM	OC	S	Hand gesture model recognition
Bernardes et al. (2009)	2009		√	√		Hybrid	OC	S	Game control
Bergh et al. (2009)	2009		√	√		ANMM, Haarlet based	OC	S	Manipulating 3D objects

Table 5 continued

Ref.	Year	Represent		Techniques			Software platform	BG. condition	Application/ task
		3D	App.	Static	Dynamic	Key methods			
Ren and Zhang (2009b)	2009		√	√		MEB- SVM	OC	C/Iv	Classifying hand gesture
Choras (2009)	2009		√	√		RT, Hybrid	NM	S	Sign language
Verma and Dev (2009)	2009		√		√	FSM, Fuzzy logic	NM	S	Hand gesture modeling
Liu and Zhang (2009)	2009		√	√		SVM, Adaboost	OC	S	Controlling cursor for desktop app.
Just and Marcel (2009)	2009		√		√	HMM, IOHMM	OC	C/Iv	Hand gesture modeling
Bandera et al. (2009)	2009	√			√	Dynamic time warping	NM	S	Hand gesture modeling
Varona et al. (2009)	2009	√		√		KNN, Hybrid	OC	S	Video game control
Yi et al. (2009)	2009	√		√		Hybrid	NM	O/C	Articulated hand signs
Sawah et al. (2008)	2008	√		√		Hybrid	M	C/Iv	Hand gesture modeling
Salinas et al. (2008)	2008	√			√	PCA, HMM, SVM	M	C	Sign language
Chen et al. (2008)	2008		√	√		Adaboost, SCFG	NM	S	Hand gesture modeling

Table 5 continued

Ref.	Year	Represent		Techniques			Software platform	BG. condition	Application/ task
		3D	App.	Static	Dynamic	Key methods			
Ge et al. (2008)	2008	√			√	DLLE, PNN	NM	NM	Hand gesture modeling
Malassiotis and Strintzis (2008)	2008	√		√		Eigenspace, PCA	M	Iv	Hand gesture modeling
Vafadar and Behrad (2008)	2008		√	√		KNN, BPNN	M	Iv	Hand gesture modeling
Alijanpour et al. (2008)	2008		√	√		Hybrid	M	NM	Alphabet recognition
Liu et al. (2008)	2008		√	√		SVM	NM	C/Iv	Hand gesture modeling
Howe et al. (2008)	2008		√		√	Hybrid	NM	S	Desktop application
Ongkittikul et al. (2008)	2008		√	√		K means, Particle filter	OC	C	Hand gesture modeling
Modler and Myatt (2008)	2008		√		√	TDNN	NM	C/Iv	Recognizing hand movements
Suk et al. (2008)	2008		√		√	DBM	OC	S	Sign language
Paulraj et al. (2008)	2008		√		√	Hybrid	OC	C	Sign language
Chen and Tseng (2007)	2007		√	√		SVM	OC	Iv	Finger guessing game

Table 5 continued

Ref.	Year	Represent		Techniques			Software platform	BG. condition	Application/ task
		3D	App.	Static	Dynamic	Key methods			
Chen et al. (2007)	2007		√	√		Adaboost, Hybrid	OC	C	Hand gesture modeling
Fang et al. (2007)	2007		√	√		Hybrid	OC	S	Hand navigation interface
Vámossy et al. (2007)	2007		√	√		Hybrid	NM	NM	Sign language
Sawah et al. (2007)	2007	√			√	DBN	OC	S	Hand gesture modeling
Kapralos et al. (2007)	2007		√		√	HMM	OC	C/Iv	Remote learning application
Kuzmanić and Zanchi (2007)	2007		√	√		DTW- LCSS	NM	C/Iv	Alphabet recognition
Conci et al. (2007)	2007		√	√		Hybrid	NM	S	Interactive virtual blackboard
Chalechale and Naghdy (2007)	2007		√	√		K-means, Bayesian rule	NM	C	Hand gesture modeling
Berci and Szolgay (2007)	2007	√		√		Hybrid	NM	S	Desktop application
Shimizu et al. (2007)	2007	√		√		Hybrid	OC	C	Hand gesture modeling
Patwardhan and Roy (2007)	2007		√	√		Predictive Eigen tracker	NM	C	Hand gesture modeling

Table 5 continued

Ref.	Year	Represent		Techniques			Software platform	BG. condition	Application/ task
		3D	App.	Static	Dynamic	Key methods			
Habib and Mufti (2006)	2006		√		√	Hybrid	NM	C	Sign language
Bhuyan et al. (2006)	2006		√	√		DTW	OC	Iv/C	Sign Language
Binh and Ejima (2006)	2006		√		√	2DHMM	M	S	American Sign Language
Prieto et al. (2006)	2006		√	√		Neural network	M	S	Desktop application
Ge et al. (2006)	2006	√			√	PNN, Hybrid	M	S	Hand gesture modeling
Bimbo et al. (2006)	2006		√	√		Hybrid	OC	S	Manipulating multimedia app.
Tseng et al. (2006)	2006		√		√	Hybrid	OC	S	Gaming application
Sriboonruang et al. (2006)	2006		√	√		Hybrid	OC	NM	Gaming control
Vilaplana and Coronado (2006)	2006	√		√		Neural Network	M	NM	Hand gesture modeling
Chang (2006)	2006		√	√		KNN	M	C	Static hand gesture modeling
Ong et al. (2006)	2006		√		√	Bayesian Network	M	C	Sign language
Juang and Ku (2005)	2005		√		√	Hybrid	NM	C/Iv	Hand gesture modeling

Table 5 continued

Ref.	Year	Represent		Techniques			Software platform	BG. condition	Application/ task
		3D	App.	Static	Dynamic	Key methods			
Wu et al. (2005)	2005	√		√		Hybrid	NM	NM	Hand gesture modeling
Wachs et al. (2005)	2005		√		√	Hybrid	OC	C	Sign language
Yi et al. (2005)	2005	√		√		Fuzzy C Means	NM	S	Visualization application
Gastaldi et al. (2005)	2005	√			√	HMM	NM	C	Alphabet recognition
Joslin et al. (2005)	2005		√		√	Hybrid	OC	C	Sign language
Liu and Lovell (2005)	2005		√	√		PCA	OC	C	Hand gesture modeling
Chalechale et al. (2005)	2005		√	√		Hybrid	OC	S	Desktop application
Hossain and Jenkin (2005)	2005		√		√	Fuzzy logic, HMM	NM	S	Sign language
Alon et al. (2005)	2005		√		√	Hybrid	NM	C	Hand gesture modeling
Juang et al. (2005)	2005		√		√	DSTW	NM	C	Digit recognition
Sclaroff et al. (2005)	2005		√		√	HMM, fuzzy logic	NM	C	Sign language
Teng et al. (2005)	2005		√	√		DTW	OC	C	Digit recognition
Ferscha et al. (2005)	2005		√		√	Hybrid	NM	Iv	Chinese sign language
Licsar and Sziranyi (2005)	2005		√		√	Hybrid	OC	S	Desktop application

M Matlab, *OC* OpenCv, *S* simple, *O* occlusion, *C* cluttered, *Iv* illumination variation, *NM* not mentioned shapes it can deal with. The appearance based methods use templates to correlate the gestures to a predefined set of template gestures; this leads to the simplicity of the parameter computations. 3D model based approaches articulate real time representations but a prototype for estimating the position and orientation of the gestures need to be developed for accurate and robust representation. Although the 3D model based approach provides promising results, posture and gesture estimation is still not that much advanced to provide flexible as well as reliable performance. Most of the researchers have preferred the appearance based representation of the gestures for low level task and applications related desktop and gamming applications as shown in Table 5. Though the 3D model based representations are accurate enough but the difficulties faced by the researchers in the implementation of these representations finds them suitable for high level gesture modeling task and applications.

As far as the techniques for recognizing hand gestures are concerned, this survey indicates that static as well as dynamic gestures have been preferred by different researchers depending upon the requirement of the application. This makes both static and dynamic gestures equally important. Though both the hand gesture recognitions are used in varied applications in the literature they have their own pros and cons. The computations and complexity increases with the implementations of dynamic hand gestures while the accuracy of gesture recognition strengthens with the implementations of static gestures. Dynamic gestures are more useful and thus preferred for applications having frequent change in the gesture adding to the dynamic nature of gesture being recognized. The Table 5 also highlighted the method used by different researchers for recognizing hand gestures. As the hand gesture recognition system consist of different phases i.e. detection, tracking and recognition as discussed in section IV, this makes it difficult to highlight single technique used for hand gesture recognition system. Based on the survey done on previous work it shows that most of the system designed and implemented consists of combination of two or more standard method which can be termed as hybrid methods.

The platform used for implementing different techniques depends on the user familiarity and requirement of the application as needed for online or offline processing. It was noted that the most of the research papers doesn't provide details about the platform used which makes it difficult to analyze the performance of platforms for specific task/application. The survey suggests that OpenCv is generally used for real time applications as it requires less computational time taken for processing. Further being an open source computer vision programming functions library it available easily. MATAB is generally used for analyzing and recognizing high level gesture modeling. With the availability of image processing toolbox in MATLAB it provides different build in functions for detection, tracking and recognition. Further comparison and analysis between different techniques used for gesture modeling can be performed in robust manner through MATLAB.

One of the major challenges in vision based hand gesture recognition is to recognize the hand gestures effectively in different background conditions. Background may vary from place to place depending on the environment conditions. Background conditions changes due to varying illumination conditions, occlusion, dynamic or moving objects in background, cluttered or distorted objects in background scene etc. In designing the real time recognition system these conditions should be taken into consideration as these noises will be present in real time scenario which affects the robustness of the system in recognizing the hand gestures.

The applications/task in research papers helps new researchers in the field of HCI to think uses of vision based gesture recognition beyond laboratory (restricted) conditions for implementation in real time environment. Different application domains have been discussed in the literature as shown in section V. In the literature as shown in Table 5 this survey generally finds two types of research areas (1) *application oriented* such as desktop, virtual reality applications etc. and (2) *task oriented* which leads to high level modeling and analyzing of gestures. Though the survey does not present any orientation towards some specific application for the hand gesture recognition systems but still the most preferred application domain which uses hand gestures for interaction is said to be for desktop applications. The suggested reason for the same could be distributed nature of the desktop applications as they present a varied scope of implementations. The next set of preferred application of hand gesture recognition system as depicted by the current survey can be said to be for gaming and sign language. Gaming though supported by a large consumer base of different age groups ranging from infants to teenagers where as sign language based applications have high humanitarian grounds for acceptability. Hand gestures are also used as an interface for applications related to entertainment by many authors but still with a very low cost of acceptability by the end users of the gesture recognition systems.

9 Vision based hand gesture recognition analysis for future perspectives

This section provides an analysis of the previous sections discussed in the paper along with an insight for the future perspectives of vision based hand gesture recognition for human computer interaction. Hand gesture recognition techniques rely heavily on the core image processing techniques for detection tracking and recognition.

Recognition techniques limitations

There are many limiting factors influencing the uses of their core technologies for real time systems as discussed in previous sections III and IV having key issues being highlighted here. Though the color based technology are used for segmentation in the detection phase but generally the color based segmentation could be confused by presence of objects of similar color distribution as hat of hand. Within the shape based core technologies of hand gesture recognition information is obtained by extracting contours of objects but within the limitations of weak classifier implementation generating faulty results. 3D hand models used for detection of hand achieve view independent detection but the fitting is guided by the forces that pull the detecting characteristic points away from goal positions on the hand models. Motion based hand detection is not so much favored because of the undertaken assumption that movement in the image is only because of the hand movements. The core technologies of hand detection if improved to operate on high frame rate acquisition then it becomes effective for tracking phase also. Co-relation based feature tracking uses template matching that is not very effective under variations illumination conditions. Contour based tracking also suffers from limitation of smoothness of contours. Optimal tracking operates an observation to transform them into estimations but is not robust enough to operated in the cluttered background. Similarly the advance technologies for hand gesture recognition are also having different sets of limitations and overheads with their evolution. The overall performance of any gesture recognition system is very much dependent on these set of techniques used for its implementation. Hence, it is required to find the optimal set of techniques at different phases that is very much application dependent for which the system has been developed. Though there are many limitations of the varied application domains of hand gesture recognition system that is discussed as follows:

Application domain constraints

It is often assumed that the application domains for which the hand gesture recognition systems are implemented would be restricted to single application only. As the research in HCI till date concentrates on the design and development of application specific gesture recognition systems in general and hand gesture recognition systems in particular. The core concept of limitation because of which most of the hand gesture recognition systems developed are application specific is the cognitive mapping of the gesture to command operating within the application. These cognitive mappings of gesture to command are easier in the case of system being developed for single application. The complexity associated with conversion of this cognitive mapping of gestures to command for different applications had hindered the evolution of application independent hand gesture recognition systems. Hence it is one of the prime concerns for future of gesture recognition system to be application independent.

Real-time challenges

The present survey has found across the literature the tendency of the developed hand gesture recognition systems trying to attain specific performance accuracy against various real time challenges faced during the design and implementation of these systems. These set of real time challenges varied from variations in illumination conditions to occlusion problems to real time compatibility of performance along with forward and backward compatibility among the technologies implemented. Nevertheless some of these real time challenges are worked upon to a certain extent by some of the authors but still no robust framework for solution to all of these real time challenges has been proposed. Efforts are need to be organized for the design and development of a framework that generates a hand gesture recognition system satisfying all the real time challenges posed by these systems. Without any detailed level of performance defined within the framework it would be really difficult to develop an optimal solution based system for various real time challenges. The static and dynamic sets of background from which the hand gestures need to be segmented are also one of the prime real time challenges that need to be addressed for the wide applicability of these systems.

Robustness

Evolution for robustness of a hand gesture recognition system is a complicated task. As there is no standard baseline algorithms that could accurately define the quantitative or qualitative robustness of any gesture recognition system specifically. Neither there is present any formal performance comparison framework for the recognition systems. Still based on the typical problems faced the robustness within the hand gesture recognition system could be defined under three majors vertical of user, gesture and application with specifications of the conditional assumptions taken for development of the systems. Being user adaptive is one of the prime requirements of any hand gesture recognition system for its wide acceptability. This includes the systems to be independent of the type of user, experience of user with such systems and the compatibility of user with the system. Secondly the gesture used by the system needs to be user friendly with high intuitiveness and lower stress and fatigue. The system also needs to be gesture independent in terms of its cognitive mapping with the set of commands. This means the system must be compatible of switching the cognitive mapping of same gesture to different set of commands and vice versa.

10 Conclusion

Over the past few years the use of natural human hand gestures for interaction with computing devices has continued to be a thriving area of research. This survey has identified more than two hundred fifty recent related publications in major conferences and journals. Increased activity in this research area has been a driven by both scientific challenge of recognizing hand gestures and the demands of potential applications related to desktop and tablet PC applications, virtual reality etc. This survey is an endeavor to provide the upcoming researchers in the field of human computer interaction a brief overview of the core technologies related to and worked upon in the recent years of research. There are well known limitations of the core technologies that need to be addressed and provide the scope for future research and development. Analysis of the comprehensive surveys and articles indicates that the techniques implemented for hand gesture recognition are often sensitive to poor resolution, frame rate, drastic illumination conditions, changing weather conditions and occlusions among other prevalent problems in the hand gesture recognition systems. The survey enlists some the common enabling technologies of hand gesture recognition and advantages and disadvantage related to them. The paper list out some of the vision based commercial products and software's for hand gesture recognition available in market.

Over the last decade numerous methods for hand gesture taxonomies and representations have been evaluated for the core technologies proposed in the hand gesture recognition systems. However the evaluations are not dependent on the standard methods in some organized format but have been done on the basis of more usage in the hand gesture recognition systems. Hence the analysis of the detailed survey presented in the paper states the fact that the appearance based hand gesture representations are more preferred than the 3D based gesture representations in the hand gesture recognition systems. Though there are vast amount of information and research publications available in both the techniques but due to complexity of implementation the 3D model based representations are less preferred. The state of art for applications of the hand gesture recognition systems present desktop applications to be the most implemented application for hand gesture recognition systems. Future research in the field of hand gesture recognition systems provide an opportunity for the researchers to come up with efficient systems overcoming the disadvantages associated with the core technologies in the current state of art for enabling technologies gesture representations and gesture recognition systems as a whole. The industrial applications also require specific advances in the man to machine and machine to machine interactions. The potential related to the application of hand gesture recognition systems in day to day life always keeps inspiring the advances required to realize the reliable efficient accurate and robust gesture recognition systems.

References

A Forge.NET (2012) http://www.aforgenet.com/framework/

Alijanpour N, Ebrahimnezhad H, Ebrahimi A (2008) Inner distance based hand gesture recognition for devices control. In: International conference on innovations in information technology, pp 742–746

Alon J, Athitsos V, Yuan Q, Sclaroff S (2005) Simultaneous localization and recognition of dynamic hand gestures. In: IEEE workshop on motion and video computing (WACV/MOTION'05), pp 254–260

Alon J, Athitsos V, Yuan Q, Sclaroff S (2009) A unified framework for gesture recognition and spatiotemporal gesture segmentation. IEEE Trans Pattern Analy Mach Intell 31(9):1685–1699

Alpern M, Minardo K (2003) Developing a car gesture interface for use as a secondary task. In: CHI '03 extended abstracts on human factors in computing systems. ACM Press, pp 932–933

Andrea C (2001) Dynamic time warping for offline recognition of a small gesture vocabulary. In: Proceedings of the IEEE ICCV workshop on recognition, analysis, and tracking of faces and gestures in real-time systems, July–August, p 83

Appenrodt J, Handrich S, Al-Hamadi A, Michaelis B (2010) Multi stereo camera data fusion for fingertip detection in gesture recognition systems. In: International conference of soft computing and pattern recognition (SoCPaR), 2010, pp 35–40

Argyros A, Lourakis MIA (2004a) Real-time tracking of multiple skin-colored objects with a possibly moving camera. In: Proceedings of the European conference on computer vision, Prague, pp 368–379

Argyros A, Lourakis MIA (2004b) 3D tracking of skin-colored regions by a moving stereoscopic observer. Appl Opt 43(2):366–378

Argyros A, Lourakis MIA (2006) Binocular hand tracking and reconstruction based on 2D shape matching. In: Proceedings of the international conference on pattern recognition (ICPR), Hong-Kong

Bandera JP, Marfil R, Bandera A, Rodríguez JA, Molina-Tanco L, Sandoval F (2009) Fast gesture recognition based on a two-level representation. Pattern Recogn Lett 30:1181–1189

Bao J, Song A, Guo Y, Tang H (2011) Dynamic hand gesture recognition based on SURF tracking. In: International conference on electric information and control engineering (ICEICE), pp 338–341

Baxter J (2000) A model of inductive bias learning. J Artif Intell Res 12:149–198

Bellarbi A, Benbelkacem S, Zenati-Henda N, Belhocine M (2011) Hand gesture interaction using color-based method for Tabletop interfaces. In: IEEE 7th international symposium on intelligent signal processing (WISP), pp 1–6

Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intell 24(4):509–522

Berci N, Szolgay P (2007) Vision based human–machine interface via hand gestures. In: 18th European conference on circuit theory and design (ECCTD 2007), pp 496–499

Bergh M, Gool L (2011) Combining RGB and ToF cameras for real-time 3D hand gesture interaction. In: Workshop on applications of computer vision (WACV), IEEE, pp 66–72

Bergh MV, Meier EK, Bosch'e F, Gool LV (2009) Haarlet-based hand gesture recognition for 3D interaction, workshop on applications of computer vision (WACV), pp 1–8

Bernardes J, Nakamura R, Tori R (2009) Design and implementation of a flexible hand gesture command interface for games based on computer vision. In: 8th Brazilian symposium on digital games and entertainment, pp 64–73

Berry G (1998) Small-wall, a multimodal human computer intelligent interaction test bed with applications, Dept. of ECE, University of Illinois at Urbana-Champaign, MS thesis

Bhuyan MK, Ghoah D, Bora PK (2006) A framework for hand gesture recognition with applications to sign language. In: Annual IEEE India conference, pp 1–6

Bimbo AD, Landucci L, Valli A (2006) Multi-user natural interaction system based on real-time hand tracking and gesture recognition. In: 18th International conference on pattern recognition (ICPR'06), pp 55–58

Binh ND, Ejima T (2006) A new approach dedicated to hand gesture recognition. In: 5th IEEE international conference on cognitive informatics (ICCI'06), pp 62–67

Birdal A, Hassanpour R (2008) Region based hand gesture recognition. In: 16th International conference in central Europe on computer graphics, visualization and computer vision, pp 1–7

Birk H, Moeslund TB, Madsen CB (1997) Real-time recognition of hand alphabet gestures using principal component analysis. In: Proceedings of the Scandinavian conference on image analysis, Lappeenranta

Blake A, North B, Isard M (1999) Learning multi-class dynamics. In: Proceedings advances in neural information processing systems (NIPS), vol 11, pp 389–395

Bolt RA, Herranz E (1992) Two-handed gesture in multi-modal natural dialog. In: Proceedings of the 5th annual ACM symposium on user interface software and technology, ACM Press, pp 7–14

Boulay B (2007) Human posture recognition for behavior understanding. PhD thesis, Universit'e de Nice-Sophia Antipolis

Bourke A, O'Brien J, Lyons G (2007) Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. Gait & Posture 26(2):194–199. http://www.sciencedirect.com/science/article/B6T6Y-4MBCJHV-1/2/f87e4f1c82f3f93a3a5692357e3fe00c

Bowden R, Zisserman A, Kadir T, Brady M (2003) Vision based interpretation of natural sign languages. In: Exhibition at ICVS03: the 3rd international conference on computer vision systems. ACM Press, pp 1–2

Bradski G (1998) Real time face and object tracking as a component of a perceptual user interface. In: IEEE workshop on applications of computer vision. Los Alamitos, California, pp 214–219

Bradski G, Kaehler A (2008) Learning OpenCV, O'Reilly, pp 337–341

Bretzner L, Laptev I, Lindeberg T (2002) 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

Buchmann V, Violich S, Billinghurst M, Cockburn A (2004) Fingartips: gesture based direct manipulation in augmented reality. In: 2nd international conference on computer graphics and interactive techniques, ACM Press, pp 212–221

Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Kluwer, Boston 1–43

Cao X, Balakrishnan R (2003) Visionwand: interaction techniques for large displays using a passive wand tracked in 3d. In: 'UIST '03: proceedings of the 16th annual ACM symposium on User Interface software and technology. ACM Press, New York, pp 173–182

Chai D, Ngan K (1998) Locating the facial region of a head and-shoulders color image. In: IEEE international conference on automatic face and gesture recognition, pp 124–129, Piscataway

Chalechale A, Naghdy G (2007) Visual-based human–machine interface using hand gestures. In: 9th International symposium on signal processing and its applications (ISSPA 2007), pp 1–4

Chalechale A, Safaei F, Naghdy G, Premaratn P (2005) Hand gesture selection and recognition for visual-based human–machine interface. In: IEEE international conference on electro information technology, pp 1–6

Chang CC (2006) Adaptive multiple sets of CSS features for hand posture recognition. Neuro Comput 69:2017–2025

Charniak E (1993) Statistical language learning. MIT Press, Cambridge

Chatty S, Lecoanet P (1996) Pen computing for air traffic control. In: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press, pp 87–94

Chaudhary A, Raheja JL, Das K, Raheja S (2011) Intelligent approaches to interact with machines using hand gesture recognition in natural way: a survey. Int J Comput Sci Eng Survey (IJCSES) 2(1):122–133

Chen YT, Tseng KT (2007) Developing a multiple-angle hand gesture recognition system for human machine interactions. In: 33rd annual conference of the IEEE industrial electronics society (IECON), pp 489–492

Chen Q, Georganas ND, Petriu EM (2007) Real-time vision-based hand gesture recognition using Haar-like features. In: Conference on instrumentation and measurement technology (IMTC 2007), pp 1–6

Chen Q, Georganas ND, Petriu M (2008) Hand gesture recognition using Haar-like features and a stochastic context-free grammar. IEEE Trans Instrum Meas 57(8):1562–1571

Cheng J, Xie X, Bian W, Tao D (2012) Feature fusion for 3D hand gesture recognition by learning a shared hidden space. Pattern Recogn Lett 33:476–484

Choras RS (2009) Hand shape and hand gesture recognition. In: IEEE symposium on industrial electronics and applications (ISIEA 2009), pp 145–149

Chung WK, Wu X, Xu Y (2009) A real time hand gesture recognition based on Haar wavelet representation. In: International conference on robotics and biomimetics, Bangkok, pp 336–341

Cohen PR, Johnston M, McGee D, Oviatt S, Pittman J, Smith I, Chen L, Clow J (1997) Quickset: multimodal interaction for distributed applications. In: Proceedings of the fifth ACM international conference on Multimedia, ACM Press, pp 31–40

Conci N, Ceresato P, De Natale FGB (2007) Natural human–machine interface using an interactive virtual blackboard. In: IEEE international conference on image processing, pp 181–184

Cootes TF, Taylor CJ (1992) Active shape models smart snakes. In: British machine vision conference, pp 266–275

Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) Active shape models—their training and applications. Comput Vis Image Underst 61(1):38–59

Corera S, Krishnarajah N (2011) 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.pdf

Cote M, Payeur P, Comeau G (2006) Comparative study of adaptive segmentation techniques for gesture analysis in unconstrained environments. In: IEEE international workshop on imagining systems and techniques, pp 28–33

Crowley JL, Jolle Coutaz FB (2000) Perceptual user interfaces: things that see. Commun ACM 43(3):54–64

Crowley J, Berard F, Coutaz J (1995) Finger tracking as an input device for augmented reality. In: International workshop on gesture and face recognition, Zurich

Cui Y, Weng J (1996) Hand sign recognition from intensity image sequences with complex background. In: Proceedings of the IEEE computer vision and pattern recognition (CVPR), pp 88–93

Cui Y, Swets D, Weng J (1995) Learning-based hand sign recognition using shoslf-m. In: International workshop on automatic face and gesture recognition, Zurich, pp 201–206

Cutler R, Turk M (1998) View-based interpretation of real-time optical flow for gesture recognition. In: Proceedings of the international conference on face and gesture recognition. IEEE Computer Society, Washington, pp 416–421

Darrell T, Essa I, Pentland A (1996) Task-specific gesture analysis in real-time using interpolated views. IEEE Trans Pattern Anal Mach Intell 18(12):1236–1242

Davis JW, Vaks S (2001) A perceptual user interface for recognizing head gesture acknowledgements. In: Proceedings of the 2001 workshop on perceptive user interfaces. ACM Press, pp 1–7
De Tan T, Geo ZM (2011) Research of hand positioning and gesture recognition based on binocular vision. In: EEE international symposium on virtual reality innovation 2011, pp 311–315
Deng LY, Lee DL, Keh HC, Liu YJ (2010) Shape context based matching for hand gesture recognition. In: IET international conference on frontier computing. Theory, technologies and applications, pp 436–444
Derpanis KG (2004) A review of vision-based hand gestures. http://cvr.yorku.ca/members/gradstudents/kosta/publications/file_Gesture_review.pdf
Derpanis KG (2005) Mean shift clustering, Lecture Notes. http://www.cse.yorku.ca/~kosta/CompVis_Notes/mean_shift.pdf
Du H, Xiong W, Wang Z (2011) Modeling and interaction of virtual hand based on virtools. In: International conference on multimedia technology (ICMT), pp 416–419
Eamonn K, Pazzani MJ (2001) Derivative dynamic time warping. In: First international SIAM international conference on data mining, Chicago
Elmezain M, Al-Hamadi A, Michaelis B (2009) Hand trajectory-based gesture spotting and recognition using HMM. In: 16th IEEE international conference on image processing (ICIP 2009), pp 3577–3580
Elmezain M, Al-Hamadi A, Sadek S, Michaelis M (2010) Robust methods for hand gesture spotting and recognition using hidden Markov models and conditional random fields. In: IEEE international symposium on signal processing and information technology (ISSPIT), pp 133–136
EyeSight's (2012) http://www.eyesight-tech.com/
Eyetoy (2003) http://asia.gamespot.com/eyetoy-play/
Fang G, Gao W, Zhao D (2003) Large vocabulary sign language recognition based on hierarchical decision trees. In: Proceedings of the 5th international conference on multimodal interfaces. ACM Press, pp 125–131
Fang Y, Wang K, Cheng J, Lu H (2007) A real-time hand gesture recognition method. In: IEEE international conference on multimedia and expo, pp 995–998
Ferscha A, Resmerita S, Holzmann C, Reichor M (2005) Orientation sensing for gesture-based interaction with smart artifacts. Comput Commun 28:1552–1563
Forsberg A, Dieterich M, Zeleznik R (1998) The music notepad. In: Proceedings of the 11th annual ACM symposium on user interface software and technology, ACM Press, pp 203–210
Francois R, Medioni G (1999) Adaptive color background modeling for real-time segmentation of video streams. In: International conference on imaging science, systems, and technology, Las Vegas, pp 227–232
Freeman W, Weissman C (1995) Television control by hand gestures. In: International workshop on automatic face and gesture recognition, Zurich, pp 179–183
Freeman W, Tanaka K, Ohta J, Kyuma K (1996) Computer vision for computer games. In: Proceedings of the second international conference on automatic face and gesture recognition, pp 100–105
Freund Y, Schapire R (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55(1):119–139
Friedman J, Hastie T, Tibshiranim R (2000) Additive logistic regression: a statistical view of boosting. Ann Stat 28(2):337–374
Gandy M, Starner T, Auxier J, Ashbrook D (2000) The gesture pendant: a self illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In: '4th IEEE international symposium on wearable computers, IEEE Computer Society, pp 87–94
Gastaldi G, Pareschi A, Sabatini SP, Solari F, Bisio GM (2005) A man-machine communication system based on the visual analysis of dynamic gestures. In: IEEE international conference on image processing (ICIP 2005), pp 397–400
Gavrila DM, Davis LS (1995) Towards 3-d model-based tracking and recognition of human movement: multi-view approach. In: IEEE international workshop on automatic face- and gesture recognition. IEEE Computer Society, Zurich, pp 272–277
Ge SS, Yang Y, Lee TH (2006) Hand gesture recognition and tracking based on distributed locally linear embedding. In: IEEE conference on robotics, automation and mechatronics, pp 1–6
Ge SS, Yang Y, Lee TH (2008) Hand gesture recognition and tracking based on distributed locally linear embedding. Image Vis Comput 26:1607–1620
GestureTek (2008) http://www.gesturetek.com/
Gorce MDL, Fleet DJ, Paragios N (2011) Model-based 3D hand pose estimation from monocular video. IEEE Trans Pattern Anal Mach Intell 33(9):1793–1805
Goza SM, Ambrose RO, Diftler MA, Spain IM (2004) Telepresence control of the nasa/darpa robonaut on a mobility platform. In: Conference on human factors in computing systems. ACM Press, pp 623–629

Graetzel C, Fong TW, Grange S, Baur C (2004) A non-contact mouse for surgeon-computer interaction. Technol Health Care 12(3):245–257

Habib HA, Mufti M (2006) Real time mono vision gesture based virtual keyboard system. IEEE Trans Consumer Electron 52(4):1261–1266

Hackenberg G, McCall R, Broll W (2011) Lightweight palm and finger tracking for real-time 3D gesture control. In: IEEE virtual reality conference (VR), pp 19–26

Hall ET (1973) The silent language. Anchor Books. ISBN-13: 978-0385055499

HandGKET (2011) https://sites.google.com/site/kinectapps/kinect

HandVu (2003) http://www.movesinstitute.org/~kolsch/HandVu/HandVu.html

Hardenberg CV, Berard F (2001) Bare-hand human–computer interaction. Proceedings of the ACM workshop on perceptive user interfaces. ACM Press, pp 113–120

He GF, Kang SK, Song WC, Jung ST (2011) Real-time gesture recognition using 3D depth camera. In: 2nd International conference on software engineering and service science (ICSESS), pp 187–190

Heap T, Hogg D (1996) Towards 3D hand tracking using a deformable model. In: IEEE international conference automatic face and gesture recognition, Killington, pp 140–145

Henia OB, Bouakaz S (2011) 3D Hand model animation with a new data-driven method. In: Workshop on digital media and digital content management, IEEE, pp 72–76

Ho MF, Tseng CY, Lien CC, Huang CL (2011) A multi-view vision- based hand motion capturing system. Pattern Recogn 44:443–453

Holzmann GJ (1925) Finite state machine: Ebook. http://www.spinroot.com/spin/Doc/Book91_PDF/F1.pdf

Hossain M, Jenkin M (2005) Recognizing hand-raising gestures using HMM. In: 2nd Canadian conference on computer and robot vision (CRV'05), pp 405–412

Howe LW, Wong F, Chekima A (2008) Comparison of hand segmentation methodologies for hand gesture recognition. In: International symposium on information technology (ITSim 2008), pp 1–7

Hsieh CC, Liou DH, Lee D (2010) A real time hand gesture recognition system using motion history image. In: 2nd International conference on signal processing systems (ICSPS), pp 394–398

Hu K, Canavan S, Yin L (2010) Hand pointing estimation for human computer interaction based on two orthogonal-views. In: International conference on pattern recognition 2010, pp 3760–3763

Huang S, Hong J (2011) Moving object tracking system based on camshift and Kalman filter. In: International conference on consumer electronics, communications and networks (CECNet), pp 1423–1426

Huang D, Tang W, Ding Y, Wan T, Wu X, Chen Y (2011a) Motion capture of hand movements using stereo vision for minimally invasive vascular interventions. In: Sixth international conference on image and graphics, pp 737–742

Huang DY, Hu WC, Chang SH (2011b) Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination. Expert Syst Appl 38(5):6031–6042

Iannizzotto G, Villari M, Vita L (2001) Hand tracking for human-computer interaction with gray level visual glove: turning back to the simple way. In: Workshop on perceptive user interfaces, ACM digital library, ISBN 1-58113-448-7

Ibarguren A, Maurtua I, Sierra B (2010) Layered architecture for real time sign recognition: hand gesture and movement. Eng Appl Artif Intell 23:1216–1228

iGesture (2012) http://www.igesture.org/

Ionescu D, Ionescu B, Gadea C, Islam S (2011a) 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

Ionescu D, Ionescu B, Gadea C, Islam S (2011b) 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

Isard M, Blake A (1998) Condensation—conditional density propagation for visual tracking. Int J Comput Vis 29(1):5–28

Joslin C, Sawah AE, Chen Q, Georganas N (2005) Dynamic gesture recognition. In: Conference on instrumentation and measurement technology, pp 1706–1711

Ju SX, Black MJ, Minneman S, Kimber D (1997) Analysis of gesture and action in technical talks for video indexing, Technical report, American Association for Artificial Intelligence. AAAI Technical Report SS-97-03

Juang CF, Ku KC (2005) A recurrent fuzzy network for fuzzy temporal sequence processing and gesture recognition. IEEE Trans Syst Man Cybern Part B Cybern 35(4):646–658

Juang CF, Ku KC, Chen SK (2005) Temporal hand gesture recognition by fuzzified TSK-type recurrent fuzzy network. In: International joint conference on neural networks, pp 1848–1853

Just A, Marcel S (2009) A comparative study of two state-of-the-art sequence processing techniques for hand gesture recognition. Comput Vis Image Underst 113:532–543

Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82:35–42
Kampmann M (1998) Segmentation of a head into face, ears, neck and hair for knowledge-based analysis-synthesis coding of video-phone sequences. In: Proceedings of the international conference on image processing (ICIP), vol 2, Chicago, pp 876–880
Kanniche MB (2009) Gesture recognition from video sequences. PhD Thesis, University of Nice
Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY (2002) An efficient k-means clustering algorithm: analysis and implementation. IEEE Trans Pattern Anal Mach Intell 24(7):881–892
Kapralos B, Hogue A, Sabri H (2007) Recognition of hand raising gestures for a remote learning application. In: Eight international workshop on image analysis for multimedia interactive services (WIAMIS'07), pp 1–4
Karam M (2006) A framework for research and design of gesture-based human computer interactions. PhD Thesis, University of Southampton
Keogh E, Ratanamahatana CA (2005) Exact indexing of dynamic time warping. Knowl Inf Syst 7(3):358–386
Kevin NYY, Ranganath S, Ghosh D (2004) Trajectory modeling in gesture recognition using cybergloves and magnetic trackers. In: TENCON 2004. IEEE region 10 conference, pp 571–574
Konrad T, Demirdjian D, Darrell T (2003) Gesture + play: full-body interaction for virtual environments. In: 'CHI '03 extended abstracts on human factors in computing systems. ACM Press, pp 620–621
Kurata T, Okuma T, Kourogi M, Sakaue K (2001) The hand mouse: GMM hand-color classification and mean shift track-ing. In: International workshop on recognition, analysis and tracking of faces and gestures in real-time systems, Vancouver, pp 119–124
Kuzmanić A, Zanchi V (2007) Hand shape classification using DTW and LCSS as similarity measures for vision-based gesture recognition system. In: International conference on "Computer as a Tool (EUROCON 2007)", pp 264–269
Laptev I, Lindeberg T (2001) Tracking of multi-state hand models using particle filtering and a hierarchy of multi-scale image features. In: Proceedings of the sScale-space'01, volume 2106 of Lecture Notes in Computer Science, p 63
Lee DH, Hong KS (2010) Game interface using hand gesture recognition. In: 5th international conference on computer sciences and convergence information technology (ICCIT), pp 1092–1097
Lee H-K, Kim JH (1999) An hmm-based threshold model approach for gesture recognition. IEEE Trans Pattern Anal Mach Intell 21(10):961–973
Lee J, Kunii TL (1995) Model-based analysis of hand posture. IEEE Comput Graphics Appl 15(5):77–86
Lee D, Park Y (2009) Vision-based remote control system by motion detection and open finger counting. IEEE Trans Consumer Electron 55(4):2308–2313
Lenman S, Bretzner L, Thuresson B (2002) Using marking menus to develop command sets for computer vision based hand gesture interfaces. In: Proceedings of the second Nordic conference on human–computer interaction, ACM Press, pp 239–242
Li F, Wechsler H (2005) Open set face recognition using transduction. IEEE Trans Pattern Anal Mach Intell 27(11):1686–1697
Li S, Zhang H (2004) Multi-view face detection with ○oat-boost. IEEE Trans Pattern Anal Mach Intell 26(9):1112–1123
Liang R-H, Ouhyoung M (1996) A sign language recognition system using hidden Markov model and context sensitive search. In: Proceedings of the ACM symposium on virtual reality software and technology'96, ACM Press, pp 59–66
Licsar A, Sziranyi T (2005) User-adaptive hand gesture recognition system with interactive training. Image Vis Comput 23:1102–1114
Lin SY, Lai YC, Chan LW, Hung YP (2010) Real-time 3D model-based gesture tracking for multimedia control. In: International conference on pattern recognition, pp 3822–3825
Liu N, Lovell BC (2005) Hand gesture extraction by active shape models. In: Proceedings of the digital imaging computing: techniques and applications (DICTA 2005), pp 1–6
Liu Y, Zhang P (2009) Vision-based human–computer system using hand gestures. In: International conference on computational intelligence and security, pp 529–532
Liu Y, Gan Z, Sun Y (2008) Static hand gesture recognition and its application based on support vector machines. In: Ninth ACIS international conference on software engineering, artificial intelligence, networking, and parallel/distributed computing, pp 517–521
Lloyd S (1982) Least squares quantization in PCM. IEEE Trans Inf Theory 28(2):129–137
Lu W-L, Little JJ (2006) Simultaneous tracking and action recognition using the pca-hog descriptor. In: The 3rd Canadian conference on computer and robot vision, 2006. Quebec, pp 6–13

Lumsden J, Brewster S (2003) A paradigm shift: alternative interaction techniques for use with mobile & wearable devices. In: Proceedings of the 2003 conference of the centre for advanced studies conference on collaborative research. IBM Press, pp 197–210

Luo Q, Kong X, Zeng G, Fan J (2008) Human action detection via boosted local motion histograms. Mach Vis Appl. doi:10.1007/s00138-008-0168-5

MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: The proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol 1, pp 281–297

Malassiotis S, Strintzis MG (2008) Real-time hand posture recognition using range data. Image Vis Comput 26:1027–1037

Mammen JP, Chaudhuri S, Agrawal T (2001) Simultaneous tracking of both hands by estimation of erroneous observations. In: Proceedings of the British machine vision conference (BMVC), Manchester

Martin J, Devin V, Crowley J (1998) Active hand tracking. In: IEEE conference on automatic face and gesture recognition, Nara, Japan, pp 573–578

MATLAB (2012) http://www.mathworks.in/products/matlab/

McNeill D (1992) Hand and mind: what gestures reveal about thought. University Of Chicago Press. ISBN: 9780226561325

Mgestyk (2009) http://www.mgestyk.com/

Microsoft Kinect (2012) http://www.microsoft.com/en-us/kinectforwindows/

Mitra S, Acharya T (2007) Gesture recognition: a survey. IEEE Trans Syst Man Cybern (SMC)Part C Appl Rev 37(3):311–324

Modler P, Myatt T (2008) Recognition of separate hand gestures by time delay neural networks based on multi-state spectral image patterns from cyclic hand movements. In: IEEE international conference on systems, man and cybernetics (SMC 2008), pp 1539–1544

Moeslund T, Granum E (2001) A survey of computer vision based human motion capture. Comput Vis Image Underst 81:231–268

Moyle M, Cockburn A (2002) Gesture navigation: an alternative 'back' for the future. In: Human factors in computing systems, ACM Press, New York, pp 822–823

Murthy GRS, Jadon RS (2010) Hand gesture recognition using neural networks. In: 2nd International advance computing conference (IACC), IEEE, pp 134–138

Nickel K, Stiefelhagen R (2003) Pointing gesture recognition based on 3d-tracking of face, hands and head orientation. In: ICMI '03: proceedings of the 5th international conference on multimodal interfaces. ACM Press, New York, pp 140–146

Nishikawa A, Hosoi T, Koara K, Negoro D, Hikita A, Asano S, Kakutani H, Miyazaki F, Sekimoto M, Yasui M, Miyake Y, Takiguchi S, Monden M (2003) FAce MOUSe: a novel human-machine interface for controlling the position of a laparoscope. IEEE Trans Robotics Autom 19(5):825–841

Noury N, Barralon P, Virone G, Boissy P, Hamel M, Rumeau P (2003) A smart sensor based on rules and its evaluation in daily routines. In: Engineering in medicine and biology society, 2003. Proceedings of the 25th annual international conference of the IEEE, vol 4, pp 3286–3289

OMRON (2012) http://www.omron.com/

Ong SCW, Ranganath S, Venkatesh YV (2006) Understanding gestures with systematic variations in movement dynamics. Pattern Recogn 39:1633–1648

Ongkittikul S, Worrall S, Kondoz A (2008) Two hand tracking using colour statistical model with the K-means embedded particle filter for hand gesture recognition. In: 7th Computer information systems and industrial management applications, pp 201–206

Osawa N, Asai K, Sugimoto YY (2000) Immersive graph navigation using direct manipulation and gestures. In: ACM symposium on virtual reality software and technology. ACM Press, pp 147–152

Ottenheimer HJ (2005) The anthropology of language: an introduction to linguistic anthropology. Wadsworth Publishing. ISBN-13: 978-0534594367

Ou J, Fussell SR, Chen X, Setlock LD, Yang J (2003) Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks. In: Proceedings of the 5th international conference on Multimodal interfaces. ACM Press, pp 242–249

Paiva A, Andersson G, Hk K, Mourao D, Costa M, Martinho C (2002) SenToy in fantasyA: designing an affective sympathetic interface to a computer game. Pers Ubiquitous Comput 6(5–6):378–389

Pang YY, Ismail NA, Gilbert PLS (2010) A real time vision-based hand gesture interaction. In: Fourth Asia international conference on mathematical/analytical modeling and computer simulation, IEEE, pp 237–242

Pantic M, Nijholt A, Pentland A, Huanag TS (2008) Human-centred intelligent human–computer Interaction (HCI^2): how far are we from attaining it?. Int J Auton Adapt Commun Syst 1:168–187

Patwardhan KS, Roy SD (2007) Hand gesture modelling and recognition involving changing shapes and trajectories, using a predictive EigenTracker. Pattern Recogn Lett 28:329–334

Paulraj MP, Yaacob S, Desa H, Hema CR (2008) Extraction of head and hand gesture features for recognition of sign language. In: International conference on electronic design, pp 1–6

Pausch R, Williams RD (1990) Tailor: creating custom user interfaces based on gesture. In: Proceedings of the 3rd annual ACM SIGGRAPH symposium on user interface software and technology. ACM Press, pp 123–134

Pavlovic VI, Sharma R, Huang TS (1997) Visual interpretation of hand gestures for human–computer interaction: a review. Trans Pattern Anal Mach Intell 19(7):677–695

Perez P, Hue C, Vermaak J, Gangnet M (2002) Color-based probabilistic tracking. In: Procedings of the European conference on computer vision, Copenhagen, pp 661–675

Peterfreund N (1999) Robust tracking of position and velocity with Kalman snakes. IEEE Trans Pattern Anal Mach Intell 10(6):564–569

Pickering CA (2005) Gesture recognition driver controls. IEE J Comput Control Eng 16(1):27–40

PointGrab's (2012) http://www.pointgrab.com/

Prieto A, Bellas F, Duro RJ, López-Peña F (2006) An adaptive visual gesture based interface for human machine interaction in intelligent workspaces. In: IEEE international conference on virtual environments, human–computer interfaces, and measurement systems, pp 43–48

Radkowski R, Stritzke C (2012) Interactive hand gesture-based assembly for augmented reality applications. In: ACHI 2012: the fifth international conference on advances in computer–human interactions, IARIA, pp 303–308

Ramage D (2007) Hidden Markov models fundamentals, Lecture Notes. http://cs229.stanford.edu/section/cs229-hmm.pdf

Rashid O, Al-Hamadi A, Michaelis B (2009) A framework for the integration of gesture and oosture recognition using HMM and SVM. In: IEEE international conference on intelligent computing and intelligent systems (ICIS 2009), pp 572–577

Rautaray SS, Agrawal A (2011) A novel human computer interface based on hand gesture recognition using computer vision techniques. In: International conference on intelligent interactive technologies and multimedia (IITM-2011), pp 292–296

Rautaray SS, Agrawal A (2012) Real time hand gesture recognition system for dynamic applications. Int J UbiComp 3(1):21–31

Reale MJ, Canavan S, Yin L, Hu K, Hung T (2011) 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

Rehg J, Kanade T (1994) Digiteyes: vision-based hand tracking for human–computer interaction. In: Workshop on motion of non-rigid and articulated bodies, Austin Texas, pp 16–24

Rehg J, Kanade T (1995) Model-based tracking of self-occluding articulated objects. In: Proceedings of the international conference on computer vision (ICCV), pp 612–617

Ren Y, Zhang F (2009a) Hand gesture recognition based on meb-svm. In: Second international conference on embedded software and systems, IEEE Computer Society, Los Alamitos, pp 344–349

Ren Y, Zhang F (2009b) Hand gesture recognition based on MEB-SVM. In: International conferences on embedded software and systems, pp 344–349

Rodriguez S, Picon A, Villodas A (2010) Robust vision-based hand tracking using single camera for ubiquitous 3D gesture interaction. In: IEEE symposium on 3D user interfaces (3DUI), pp 135–136

Sajjawiso T, Kanongchaiyos P (2011) 3D hand pose modeling from uncalibrate monocular images. In: Eighth international joint conference on computer science and software engineering (JCSSE), pp 177–181

Salinas RM, Carnicer RM, Cuevas FJ, Poyato AC (2008) Depth silhouettes for gesture recognition. Pattern Recogn Lett 29:319–329

Sangineto E, Cupelli M (2012) Real-time viewpoint-invariant hand localization with cluttered backgrounds. Image Vis Comput 30:26–37

Sawah AE, Joslin C, Georganas ND, Petriu EM (2007) A framework for 3D hand tracking and gesture recognition using elements of genetic programming. In: Fourth Canadian conference on computer and robot vision (CRV'07), pp 495–502

Sawah AE, Georganas ND, Petriu EM (2008) A prototype for 3-D hand tracking and posture estimation. IEEE Trans Instrum Meas 57(8):1627–1636

Saxe D, Foulds R (1996) Toward robust skin identification in video images. In: IEEE international conference on automatic face and gesture recognition, pp 379–384

Schapire R (2002) The boosting approach to machine learning: an overview. In: MSRI workshop on nonlinear estimation and classification

Schlomer T, Poppinga B, Henze N, Boll S (2008) Gesture recognition with a wii controller. In: TEI '08: proceedings of the 2nd international conference on Tangible and embedded interaction. ACM, New York, pp 11–14

Schmandt C, Kim J, Lee K, Vallejo G, Ackerman M (2002) Mediated voice communication via mobile ip. In: Proceedings of the 15th annual ACM symposium on User interface software and technology. ACM Press, pp 141–150

Schultz M, Gill J, Zubairi S, Huber R, Gordin F (2003) Bacterial contamination of computer keyboards in a teaching hospital. Infect Control Hosp Epidemiol 4(24):302–303

Sclaroff S, Betke M, Kollios G, Alon J, Athitsos V, Li R, Magee J, Tian TP (2005) Tracking, analysis, and recognition of human gestures in video. In: 8th International conference on document analysis and recognition, pp 806–810

Segen J, Kumar S (1998a) Gesture VR: vision-based 3d Hand interface for spatial interaction. In: Proceedings of the sixth ACM international conference on multimedia. ACM Press, pp 455–464

Segen J, Kumar S (1998b) Video acquired gesture interfaces for the handicapped. In: Proceedings of the sixth ACM international conference on multimedia. ACM Press, pp 45–48

Segen J, Kumar SS (1999) Shadow gestures: 3D hand pose estimation using a single camera. In: Proceedings of the IEEE computer vision and pattern recognition (CVPR), pp 479–485

Senin P (2008) Dynamic time warping algorithm review, technical report. http://csdl.ics.hawaii.edu/techreports/08-04/08-04.pdf

Sharma R, Huang TS, Pavovic VI, Zhao Y, Lo Z, Chu S, Schulten K, Dalke A, Phillips J, Zeller M, Humphrey W (1996) Speech/gesture interface to a visual computing environment for molecular biologists. In: International conference on pattern recognition (ICPR '96) volume 7276. IEEE Computer Society, pp 964–968

Shimada N, Shirai Y, Kuno Y, Miura J (1998) Hand gesture estimation and model refinement using monocular camera ambiguity limitation by inequality constraints. In: IEEE international conference on face and gesture recognition, Nara, pp 268–273

Shimizu M, Yoshizuka T, Miyamoto H (2007) A gesture recognition system using stereo vision and arm model fitting. In: International congress series 1301, Elsevier, pp 89–92

Sigal L, Sclaroff S, Athitsos V (2004) Skin color-based video segmentation under time-varying illumination. IEEE Trans Pattern Anal Mach Intell 26(7):862–877

Smith GM, Schraefel MC (2004) The radial scroll tool: scrolling support for stylus- or touch-based document navigation. In: Proceedings of the 17th annual ACM symposium on User interface software and technology, ACM Press, pp 53–56

SoftKinetic, IISU SDK (2012) http://www.softkinetic.com/Solutions/iisuSDK.aspx

Song L, Takatsuka M (2005) Real-time 3D ¯nger pointing for an augmented desk. In: Australasian conference on user interface, vol 40. Newcastle, pp 99–108

Sriboonruang Y, Kumhom P, Chamnongthai K (2006) Visual hand gesture interface for computer board game control. In: IEEE tenth international symposium on consumer electronics, pp 1–5

Stan S, Philip C (2004) Fastdtw: toward accurate dynamic time warping in linear time and space. In: KDD workshop on mining temporal and sequential data

Staner AT, Pentland A (1995a) Visual recognition of American sign language using hidden Markov models. Technical Report TR-306, Media Lab, MIT

Starner T, Pentland A (1995b) Real time American sign language recognition from video using hidden Markov models, Technical Report 375, MIT Media Lab

Stotts D, Smith JM, Gyllstrom K (2004a) Facespace: endo- and exo-spatial hypermedia in the transparent video face top. In: 15th ACM conference on hypertext & hypermedia. ACM Press, pp 48–57

Stotts D, Smith JM, Gyllstrom K (2004b) Facespace: endo- and exo-spatial hypermedia in the transparent video facetop. In: Proceedings of the fifteenth ACM conference on hypertext & hypermedia. ACM Press, pp 48–57

Suk H, Sin BK, Lee SW (2008) Robust modeling and recognition of hand gestures with dynamic Bayesian network. In: 19th international conference on pattern recognition, pp 1–4

Suka H, Sin B, Lee S (2010) Hand gesture recognition based on dynamic Bayesian network framework. Pattern Recogn 43:3059–3072

Swindells C, Inkpen KM, Dill JC, Tory M (2002) That one there! Pointing to establish device identity. In: Proceedings of the 15th annual ACM symposium on user interface software and technology. ACM Press, pp 151–160

Teng X, Wu B, Yu W, Liu C (2005) A hand gesture recognition system based on local linear embedding. J Vis Lang Comput 16:442–454

Terrillon J, Shirazi M, Fukamachi H, Akamatsu S (2000) Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In: Proceedings of the international conference on automatic face and gesture recognition (FG), pp 54–61

Terzopoulos D, Szeliski R (1992) Tracking with Kalman Snakes. MIT Press, Cambridge 3–20

Thirumuruganathan S (2010) A detailed introduction to K-nearest neighbor (KNN) algorithm. http://saravananthirumuruganathan.wordpress.com/2010/05/17/a-detailed-introduction-to-k-nearest-neighbor-knn-algorithm/
Tran C, Trivedi MM (2012) 3-D posture and gesture recognition for interactivity in smart spaces. IEEE Trans Ind Inform 8(1):178–187
Triesch J, Malsburg C (1996) Robust classification of hand postures against complex background. In: IEEE automatic face and gesture recognition, Killington, pp 170–175
Triesch J, Von der Malsburg C (1998) A gesture interface for human-robot-interaction. In: Proceedings of the international conference on automatic face and gesture recognition (FG). IEEE, Nara, Japan, pp 546–551
Tseng KT, Huang WF, Wu CH (2006) Vision-based finger guessing game in human machine interaction. In: IEEE international conference on robotics and biomimetics, pp 619–624
Utsumi A, Ohya J (1998) Image segmentation for human tracking using sequential-image-based hierarchical adaptation. In: Proceedings IEEE computer vision and pattern recognition (CVPR), pp 911–916
Utsumi A, Ohya J (1999) Multiple-hand-gesture tracking using multiple cameras. In: Proceedings of the IEEE computer vision and pattern recognition (CVPR), Colorado, pp 473–478
Vafadar M, Behrad A (2008) Human hand gesture recognition using spatio-temporal volumes for human–computer interaction. In: International symposium in telecommunications, pp 713–718
Vámossy Z, Tóth A, Benedek B (2007) Virtual hand—hand gesture recognition system. In: 5th International symposium on intelligent systems and informatics, pp 97–102
Várkonyi-Kóczy AR, Tusor B (2011) Human–computer interaction for smart environment applications using fuzzy hand posture and gesture models. IEEE Trans Instrum Meas 60(5):1505–1514
Varona J, Jaume-i-Capó A, Gonzàlez J, Perales FJ (2009) Toward natural interaction through visual recognition of body gestures in real-time. Interact Comput 21:3–10
Verma R, Dev A (2009) Vision based hand gesture recognition using finite state machines and fuzzy logic. In: International conference on ultra modern telecommunications & workshops (ICUMT '09), pp 1–6
Vilaplana JM, Coronado JL (2006) A neural network model for coordination of hand gesture during reach to grasp. Neural Netw 19:12–30
Viola P, Jones M (2001) Robust real-time object detection. In: IEEE workshop on statistical and computational theories of vision, Vancouver
Visser M, Hopf V (2011) 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
Vo N, Tran Q, Dinh TB, Dinh TB, Nguyen QM (2010) An efficient human–computer interaction framework using skin color tracking and gesture recognition. In: IEEE RIVF international conference on computing and communication technologies, research, innovation and vision for the future (RIVF), pp 1–6
Wachs J, Stern H, Edan Y, Kartoun U (2002) Real-time hand gestures using the fuzzy-C-means Algorithm. In: Proceeding of WAC 2002, Florida
Wachs JP, Stern H, Edan Y (2005) Cluster labeling and parameter estimation for the automated setup of a hand-gesture recognition system. IEEE Trans Syst Man Cybern PART A Syst Humans 35(6):932–944
Wachs JP, Kolsch M, Stern H, Edan Y (2011) Vision-based hand-gesture applications. Commun ACM 54:60–71
Wang GW, Zhang C, Zhuang J (2012) An application of classifier combination methods in hand gesture recognition. Mathematical Problems in Engineering Volume 2012, Hindawi Publishing Corporation, pp 1–17. doi:10.1155/2012/346951
Ward DJ, Blackwell AF, MacKay DJC (2000) Dasher—a data entry interface using continuous gestures and language models. In: Proceedings of the 13th annual ACM symposium on user interface software and technology, ACM Press, pp 129–137
Webel S, Keil J, Zoellner M (2008) Multi-touch gestural interaction in x3d using hidden markov models. In: VRST '08: proceedings of the 2008 ACM symposium on vVirtual reality software and technology. ACM, New York, pp 263–264
Wii Nintendo (2006) http://www.nintendo.com/wii
Wilson A, Shafer S (2003) Xwand: UI for intelligent spaces. In: Proceedings of the conference on Human factors in computing systems. ACM Press, pp 545–552
Wohler C, Anlauf JK (1999) An adaptable time-delay neural-network algorithm for image sequence analysis. IEEE Trans Neural Netw 10(6):1531–1536
Wu M, Balakrishnan R (2003) Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In: Proceedings of the 16th annual ACM symposium on user interface software and technology. ACM Press, pp 193–202
Wu Y, Huang T (1999a) Vision-based gesture recognition: a review. In: Gesture-based communications in HCI, Lecture Notes in Computer Science, vol 1739. Springer, Berlin

Wu Y, Huang TT (1999b) Capturing human hand motion: a divide-and-conquer approach. In: Proceedings of the international conference on computer vision (ICCV), Greece, pp 606–611
Wu Y, Huang TS (2000) View-independent recognition of hand postures. In: Proceedings of the IEEE computer vision and pattern recognition (CVPR), vol 2. Hilton Head Island, SC, pp 84–94
Wu Y, Lin J, Huang T (2001) Capturing natural hand articulation. In: Proceedings of the international conference on computer vision (ICCV), Vancouver, pp 426–432
Wu Y, Lin J, Huang TS (2005) Analyzing and capturing articulated hand motion in image sequences. IEEE Trans Pattern Anal Mach Intell 27(12):1910–1922
Xiangyu W, Xiujuan L (2010) The study of moving target tracking based on Kalman CamShift in the video. In: 2nd International conference on information science and engineering (ICISE), pp 1–4
Yang M, Ahuja N (1998) Detecting human faces in color images. In: Proceedings of the international conference on image processing (ICIP), Piscataway, pp 127–130
Yang J, Lu W, Waibel A (1998a) Skin-color modeling and adaptation. In: ACCV, pp 687–694
Yang J, Lu W, Waibel A (1998b) Skin-color modeling and adaptation. In: ACCV, pp 687–694
Yang J, Xu J, Li M, Zhang D, Wang C (2011) A real-time command system based on hand gesture recognition. In: Seventh international conference on natural computation, pp 1588–1592
Yi B, Harris FC Jr, Wang L, Yan Y (2005) Real-time natural hand gestures. Comput Sci Eng IEEE 7(3):92–97
Yi X, Qin S, Kang J (2009) Generating 3D architectural models based on hand motion and gesture. Comput Ind 60:677–685
Yilmaz JA, Javed O, Shah M (2006) Object tracking: a survey. ACM Comput Surv 38:13
Yin M, Xie X (2003) Estimation of the fundamental matrix from uncalibrated stereo hand images for 3D hand gesture recognition. Pattern Recogn 36(3):567–584
Yin J, Han Y, Li J, Cao A (2009) Research on real-time object tracking by improved CamShift. In: International symposium on computer network and multimedia technology, pp 1–4
Yuan Q, Sclaroff S, Athitsos V (1995) Automatic 2D hand tracking in video sequences. In: IEEE workshop on applications of computer vision, pp 250–256
Yuan R, Cheng J, Li P, Chen G, Xie C, Xie Q (2010) View invariant hand gesture recognition using 3D trajectory. In: Proceedings of the 8th world congress on intelligent control and automation, Jinan, pp 6315–6320
Yun L, Peng Z (2009) An automatic hand gesture recognition system based on Viola–Jones method and SVMs. In: Second international workshop on computer science and engineering, pp 72–76
Zabulis X, Baltzakis H, Argyros A (2009) Vision-based Hand gesture recognition for human–computer interaction. In: The Universal Access Handbook. LEA
Zeller M et al (1997) A visual computing environment for very large scale biomolecular modeling. In: Proceedings of the IEEE international conference on application specific systems, architectures and processors (ASAP), Zurich, pp 3–12
Zhao S, Tan W, Wu C, Liu C, Wen S (2009) A Novel interactive method of virtual reality system based on hand gesture recognition. In: Chinese control and decision conference (CCDC ’09), pp 5879–5822
Zhu HM, Pun CM (2010) Movement tracking in real-time hand gesture recognition. In: 9th IEEE/ACIS international conference on computer and information science, pp 241–245

Author Biographies

Siddharth S. Rautaray is currently a Ph.D. scholar in the Department of Information Technology at Indian Institute of Information Technology, Allahabad, India. His current research interests include Human-Computer Interactions, User Interface Design and Computer Vision Image Processing. He has published around 15 research papers in different international and national journals and conferences. He is a member of the IEEE Computer Society, ACM SIGCHI and CSI.

Anupam Agrawal is presently working as Professor of Computer Science and Information Technology at Indian Institute of Information Technology Allahabad (IIIT-A). Before joining IIIT-A in the year 2000, he was working as scientist ‘D’ at DEAL, DRDO, Govt. of India, Dehradun. He received his M.Tech. degree in Computer Science and Engineering from Indian Institute of Technology Madras, Chennai and Ph.D. degree in Information Technology from Indian Institute of Information Technology Allahabad (in association with Indian Institute of Technology, Roorkee). He was a postdoctoral researcher at the Department of Computer Science & Technology, University of Bedfordshire (UK) during which he had contributed significantly in two major European projects. His research interests include Computer Vision, Image Processing, Visual Computing, Soft-Computing and Human-Computer Interaction. He has more than 75 publications related to these areas in various international journals and conference proceedings, and has co-authored one book. He is on the review

board for various international journals including IEEE, Springer, MDPI, Taylor & Francis and Elsevier. He is currently serving as a Principal Investigator of an international (Indo-UK) Project. He is a member of ACM (USA), senior member of IEEE (USA) and a fellow of IETE (India). He is also serving as Chairman of the ACM Chapter at IIIT-A.