CHAPTER 2

Background and Related Work

In this chapter, the theoretical background related to the hand gesture recognition in the fields of human-computer interaction and other related areas are presented. The gesture recognition methods are broadly divided into two approaches such as Data-Glove based approaches and Vision based approaches [1]. The Data-Glove based approaches make use of the mechanical or optical sensors which are connected to a glove. This glove is worn by the user during gesticulation that converts the finger flexions into electrical signals, thus recognizing the hand posture. The main disadvantage of this method is that it obstructs the ease and naturalness of the user because it has load of cables which are connected from the glove to the computer. This prevents the users to move freely. On the contrary, the vision based approach need only a camera for interaction between humans and computers which makes the system simpler and easy to use.

2.1 Data-Glove Based Approaches

Some of the works related to data glove based approaches have been discussed in this section. Few of the data gloves used in literatures [3, 6, 7, 8, 9, 11] for developing the systems are shown in Fig. 2.1. Kumar *et al.*, [2] used DG5 VHand 2.0 data gloves for hand gesture recognition. This data glove consist of five flex sensors for sensibility. A bi-flex bend sensor was also used in this work which was a type of sensor that changes resistance when it bends.

The experiment was carried out using single hand which may be considered as the drawback of the system. Kim *et al.*, [3] used KHU-1 data glove along with Kinematic chain theory to extract the features like joints from hand. The glove consisted of 3 accelerometer sensor, a controller and a Bluetooth. Finally, rule based algorithm was used for gesture identification. A data glove named Cyberglove was used by Weissmann and Salomon [4]. This data glove generated angles made by 18 joints of the hand. Features extracted using this glove were the angle between the neighboring fingers, wrist pitch, thumb rotation which were then trained using ANN.

Cavallo [5] used a data glove where 18 markers were attached to it. Among the 18 markers, 15 were used for fingers and 3 were used as reference. The image captured was then classified based on singular value decomposition (SVD). Saengsri *et al.*, [6] in his paper on Thai Sign Language Recognition used '5DT Data Glove 14 Ultra' data glove that was attached with 14 sensors. Out of these 14 sensors, 10 sensors were put on fingers and rest 4 sensors between the fingers which measures flexures and abductions respectively. The achieved accuracy rate of the system was 94%. A data glove namely Pohelmus 3 SPACE-position trackers were used for recognition of Chinese Sign Language [7]. Out of the three trackers, two trackers were used at wrist of each hand and the third tracker was placed at the back of the signer which was considered to be the reference point. A sensory glove, Cyberglove and a 'Flock of Birds®' 3-D motion tracker were used to recognize American Sign Language [8]. Sensory glove was used to find the hand shape by calculating the joint angle. 18 sensors were present in the Cyberglove that was used for measuring the bending angles at various positions. Finally the motion tracker was used for measuring the trajectory of hand's motion. The data glove '5DT data gloves 14 Ultra' and an accelerometer were used by Ibarguren *et al.*, [9] for recognition of Spanish sign language. Sensors in the data glove provided information about the finger flexures and abductions between fingers. The accelerometer provides the information about the hand movements such as 3-D acceleration, angular orientation and velocity. A CyberGlove, a Polhemus, six DOF position sensor and a geolocation system were used for controlling a robot through hand gestures [10]. The data glove used in this system measured 18 joint angles of the fingers and wrist at the frequency of 30Hz. The wrist position and orientation were determined using the positioning system that consists of a transmitter and receiver pair. The position and orientation of the hand was tracked when the receiver detected the magnetic fields which were emitted by the transmitter. A Z-glove™ and a dataglove™ has been developed by Zimmerman *et al.*, [11] to allow the direct manipulation of the computer generated objects. These gloves measure the finger bending, position and orientation of hand. A real time human hand simulation system was designed using DG5-VHand 2.0 data gloves by Teleb and Chang [12]. This data glove comprises of five bend sensors and three axes accelerometer which measure the movements and orientations of the hand. An AcceleGlove and a two-link arm skeleton was used by Rebollar *et al.*, for detection of hand from the input image [13]. AcceleGlove consisted of sensors and wires that were mounted on the leather glove for improving the robustness. It was used for detecting hand shapes. The two-link arm skeleton consists of one dual-axis accelerometer which detected arm elevation and arm rotation, two resistive angular sensors: one present on shoulder measures the forearm rotation and another placed at elbow measures the forearm flexion. A data glove was developed which consists of Flex sensors attached to the fingers of the glove [14]. It is observed that with the change in the finger position, the resistance changes. The kit is provided with a switch to select either teach mode or learn mode. In teach mode, the database for different letters of American Sign Language (ASL) are saved and in learn mode, the signer can compare the sign with the existing signs in database. However, the equipment is relatively expensive.

2.2 Vision Based Approaches

Early research on vision-based hand tracking and gesture recognition usually used markers or colored gloves [15, 16]. In current state-of-art systems, the researches in vision-based hand tracking and gesture recognition techniques are more focused on using bare hand and identifying the hand gestures without the help of any markers and gloves. However, obtaining highly accurate results is a challenging task for any vision based approaches [17]. Such systems suffers from some of difficulties such as: the user and camera are to be independent and invariant against background changes, transformations, and lighting conditions to attain real-time performance. Since the human hand is a deformable and articulated object, therefore it may increase the difficulty in the segmentation process and recognition stage.

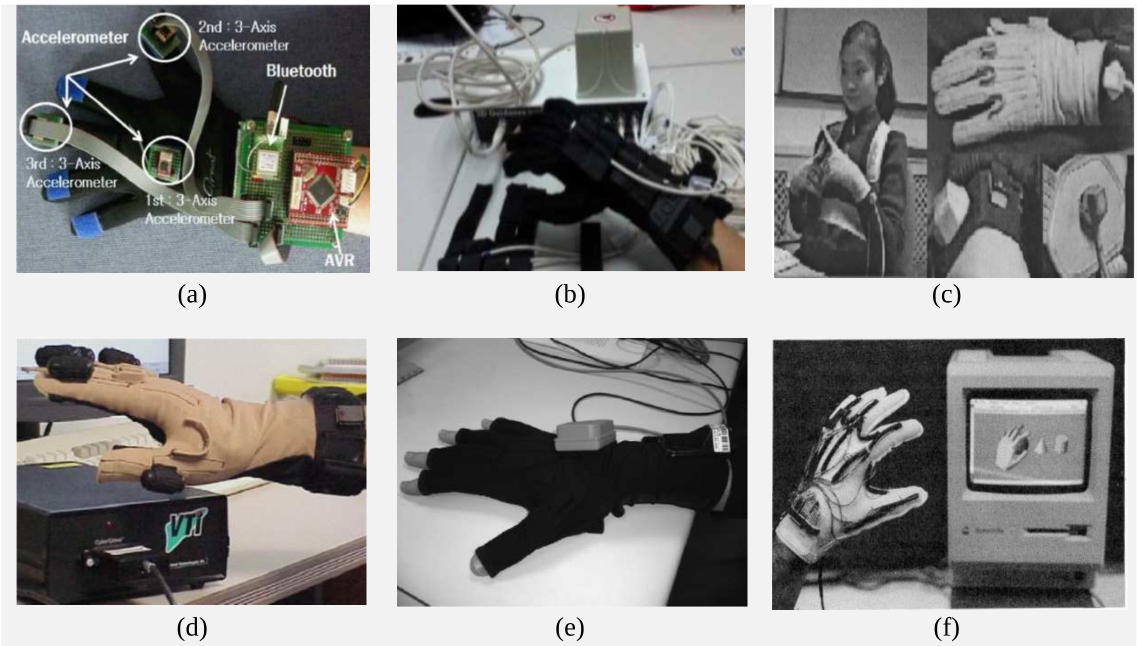


FIGURE 2.1 Few of the data gloves used in various existing recognition systems (a) KHU-1 data glove [3] (b) 5DT Glove 14 Ultra [6] (c) Pohelmus 3 SPACE-position trackers [7] (d) Cyberglove and Flock of Birds® [8] (e) 5DT Glove 14 Ultra [9] (f) Z-glove [11]

Thus, vision-based approaches using bare hand should carefully deal with the above difficulties to make the system robust and more natural for human-computer interface. Vision based hand gesture recognition system are comprised of four fundamental phases: detection, tracking, feature extraction and recognition. The summary of the different techniques available in the existing literatures for every phase of a vision based system are shown in Fig. 2.2. The literatures related to each phases are discussed in next subsections.

2.2.1 Detection of hand

The primary step in any hand gesture recognition systems is the detection of hand from the background. This step is required before the subsequent stages of process such as tracking and recognition stages. A large number of methods have been proposed in the literature that uses several types of visual features and, in many cases, their combination. The features used in the literatures are skin color, 3D model of hand and motion etc.

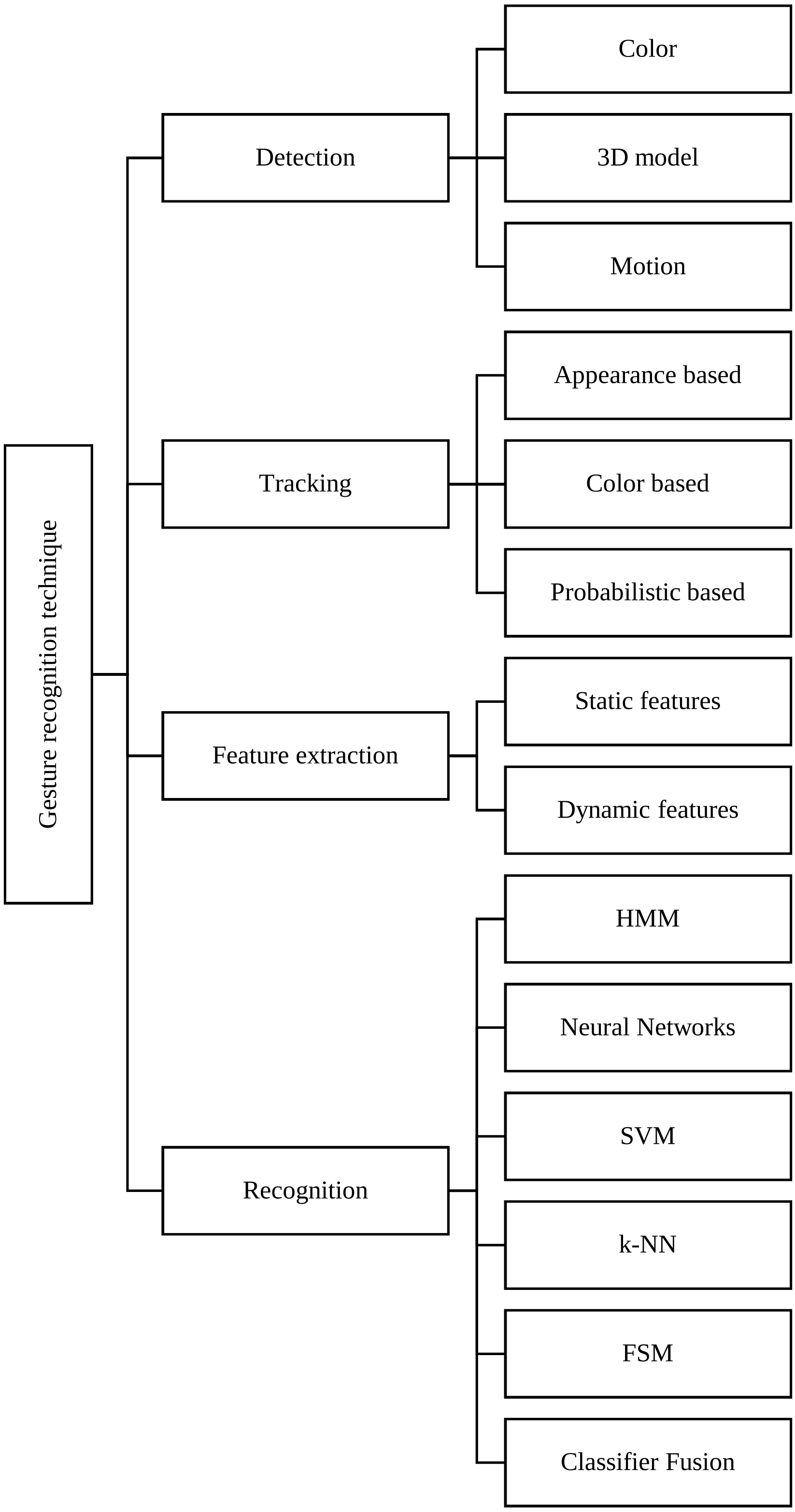


FIGURE 2.2 Hand gesture recognition steps with different techniques available in the literature

Color based detection: Many researchers extracted the skin colored region from the input image to obtain the desired hand region. 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 such as RGB, normalized RGB, HSV [18], YCrCb [19], YUV [20], etc., that have been used for developing the skin color model. YCbCr color model was used to distinguish skin colored pixels from the background in some of the researches [21-23]. The required portion of the hand was extracted using this color model that was filtered by median filter and finally processed by a smoothing filter. Malima *et al.*, [24] used the Red/Green ratio to determine the skin colored regions for robotic application. Initially, the centre of gravity of the hand was searched and then the farthest distance from the centre was calculated. In this way, they have identified the fingertips of the hand. Manigandan and Jackin [25] used the same steps as used in the existing literature [24], except that the RGB input was converted to HSV color space before doing the further processing. Fang *et al.*, [26] applied Adaptive Boost algorithm for detecting hand from the input image. This algorithm was able to detect the overlapped hand. The Haar-like features were extracted and other processing like face subtraction, skin detection and contour comparison algorithm were used for detecting the hand region [27]. The experiments were carried out in cluttered background.

However, using only color as information to segment the hand can create confusion between the background objects that have a color distribution similar to human skin. A way to minimize this problem is to use the background subtraction technique [28]. The first frame of the video was considered as thebackground for the entire processing of the system [29]. The first frame did not contain the target hand. Then this background was subtracted with the successive frames of the input video to detect the moving objects. Finally, skin colored pixels was extracted out using the CamShift algorithm to detect the hand from the objects. Tewari and Srivastava [30] first converted the input RGB image to grayscale image. The pixels of the hand regions could be extracted out easily as it was carried out in a controlled environment where signer used black dress, black bandage with background as black. Salleh and Ramli [31] used background subtraction to detect the hand. It was observed that there are three binary linked regions i.e., the face and two hands. To select the two hands, the maximum values of the region connected was identified using binary linked object (BLOB) analysis. However, background subtraction is typically based on the assumption that both the camera and background are static. To solve such problem, some researchers [32, 33] have proposed to make dynamic correction of background models.

The color of the human skin varies greatly across the human races or even between individuals of the same race. Moreover, variations may also result from the changing illumination conditions or camera characteristics. Therefore, the color based approaches should also take care of such problems for compensating such variability. Some researchers [34, 35] proposed a model for detecting skin color which was independent of the changes in illumination.

3D model based detection: The advantage of the 3D hand models 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 [36]. Kinematic hand models employ point and line features in order to recover angles formed at the joints of the hand [37-40]. Various 3D hand models have been proposed in the literature having their own advantages and drawbacks.

In some of the researches, deformable model framework is utilized to fit a 3D model of the hand to image data [41, 42]. 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. The process is enhanced with anatomical data of the human hand which is incorporated into the model [41]. Moreover, 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.

Motion based detection: Motion is one of the cue used by few literatures to detect hand. In more recent approaches, motion information is combined with additional visual cues like color cues to obtain the desired result. Researchers used motion information to distinguish hands from other skin-colored objects [43, 44]. The system is also independent of the illumination problem. But, the drawback of the system is that the camera and background should remain static so that the hand may be detected easily. The difference in the luminance of pixels from two successive frames of the input video is close to zero for pixels of the background. Thus, by choosing the appropriate threshold, moving objects (hand)are detected well in a static background. A novel feature based on motion residue is proposed by Yuan *et al.* [45]. It is observed that hands normally undergo non-rigid motion, because they are articulated objects. Therefore, the hand is detected by exploiting the information that for hands, the inter frame appearance changes more frequently as compared to other objects such as face, clothes and background. Spatiotemporal based characteristics were generated from the video sequence [46]. Frame differencing was performed between successive frames of the video sequence and then skin filtering was performed to extract the skin colored regions. Chen *et al.*, introduced a model which included both skin filtering and motion information [47]. Skin colored objects like face and hand of the signer were detected. Then the face region was extracted from the image using a suitable face detection algorithm. The frame differencing was used to locate the moving objects in the surrounding. Finally, the results were combined to obtain the hand region.

2.2.2 Tracking approaches

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 one of the extremely difficult task in a hand gesture recognition system since the gesticulation speed of the users are not same. The most problem occurs when the hand moves very fast and their appearance change largely within few frames. A robust tracking is necessary in order to provide good performance of the system in later stages. Tracking provides the inter-frame linking of hand appearances, thus generating the trajectory of a gesture. It is from this trajectory that the features are extracted in the later stages. These trajectories carries information regarding the gesture and might be used either in raw form or after further analysis. Moreover, tracking also provides a way to maintain estimates of model parameters variables and the features that are not directly observable at a certain moment in time [36].

Color based approach: Guo *et al.*, [48] proposed a hand tracking system using skin filtering, pixel based hierarchical feature AdaBoosting, and codebook background cancelation. But, the background has to be known a priori. Koh *et al.*, developed a color model using skin so as to track the hand gestures [49]. A hand appearance model was constructed which considered both shape and color information using Active Appearance Model (AAM). Mahalanobis distance was used during initialization of the system to verify the user's hand and the appearance model. Then, the skin color model is constructed using Gaussian distribution. Color histogram was extracted and used as the information to track an object [50]. Mixtures of Gaussians were used for developing the model for color distribution of the object [51-53]. But the drawback of this color based technique is that it fails if there is presence of objects in the background with similar color as that of the hand.

Probabilistic approach: In last decade, many researchers have adopted the probabilistic approaches to track hand [54-60]. The blobs are computed in some literatures [54, 55] which is used for tracking hand. The next location of the hand is predicted using Kaman filter. The measurement noise used in Kalman filter is assumed to be Gaussian for the system developed in these papers. Moreover, the gesticulation should be performed with constant velocity which restricts the natural speed of the user. Multiple cameras were used to track the hand using Kalman filter running in each frame of the video to estimate the hand postures [61]. Peterfreund [62] developed a robust technique to handle the cluttered background. The conventional image gradient is combined with optical flow to separate the foreground from the background. Asaari *et al.*, integrated Adaptive Kalman Filter and Eigenhand to track hand under different challenging environment [63]. But the algorithms fails in presence of large scale variations and pose changes.

Particle filters have been utilized to track the position of hands. Here, the location of a hand is modeled with a set of particles. The Condensation algorithm provides better performance than Kalman filters [56, 64]. This algorithm performs well against cluttered and dynamic backgrounds. It uses “factored sampling”, where the probability distribution of possible interpretations is represented by a randomly generated set. This algorithm uses visual observations and learned dynamical models to propagate this random set over time. Mammen *et al.*, [65] extended the Condensation algorithm to detect target object under occlusions. The same algorithm is integrated with color information within a probabilistic framework by Perez *et al.* [66]. This technique introduces a new Monte Carlo tracking algorithm.

Appearance based approach: Comaniciu et al. proposed a model to track hand using color histogram [50]. They used the color histogram of the detected hand as the mean shift input to locate and track hand in the video sequences. Based on mean-shift algorithm, the continuous adaptive mean-shift (CamShift) has been developed. CamShift- an improved version of mean shift algorithm was widely used for tracking objects [67, 68]. This algorithm has found to be track hand efficiently in a simple background scene but it cannot give the same result when the target is occluded with other skin colored objects. CamShift adaptively adjusts the track window’s size and the distribution pattern of targets during tracking. CamShift algorithm can track the distribution of any kind of feature representing the target in an efficient way [69]. There are many techniques where the CamShift was combined with other simple tracking methods to improve the tracking performance. For example, in literatures [70, 71], CamShift algorithm was combined with Kalman filter. The possible positions of a target are predicted by Kalman filter, and then the CamShift is used to search and match the target in the predicted areas [70].

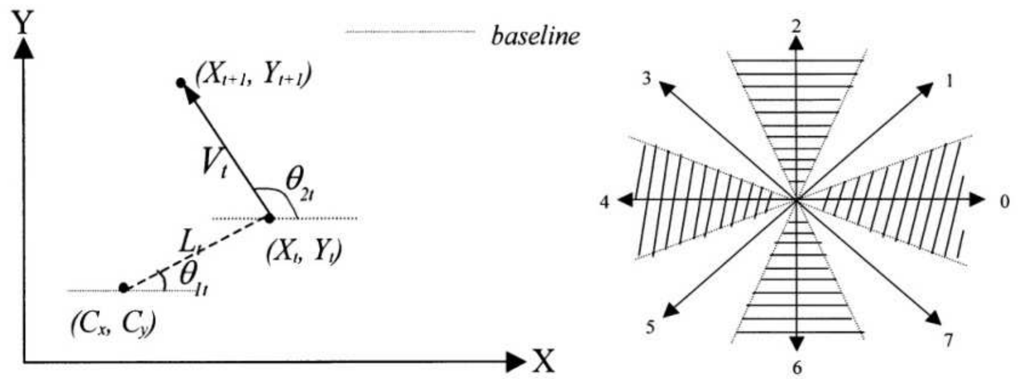
Shi and Tomasi [72] selected the corner points with high intensities as the features to track target object. Though good tracking results have been observed but the feature points goes on decreasing with succeeding video frames. This happens due to change in illumination or change in appearance of the hand. Kolsch and Turk [73] introduced a KLT tracker based hand tracking algorithm. This tracker fails when there is change in the shape of the hand during gesticulation. Porikli *et al.*, [74] proposed a tracker to track using the covariance matrix representation [75] and Riemannian manifold was used for modeling the updated mechanism. In their system, the target object was represented with the set of features as covariance matrix. The correlation of spatial and statistical properties were captured between the same representations. For every consecutive frame of the input video, a candidate region was searched which had covariance matrix similar to the target object. The model then receives the information and updates the system with the changes in the appearance of the hand. But this tracker will fail to respond if target and background has very less variations. An eigenspace approach based tracking system named as Eigen tracking was developed by Black and Jepson [76]. This tracker uses a subspace constancy assumptions for estimating the hand motion. This technique requires pre-training of the eigenbasis which increases the tracking time of the system. Moreover, the eigenbasis are not updated, thus the system cannot work in environment suffering from illumination changes.

2.2.3 Feature extraction

Extraction of suitable robust features is required for improving the performance of the system. In some papers, single feature were used for developing the gesture recognition system. Elmezain *et al.*, proposed a system to recognize both isolated and continuous gestures [77]. They used orientation feature in the feature extraction stage which provided the direction of motion between consecutive trajectory points. The angle of orientations was quantized using codeword ranging from 1 to 18. Kao and Fahn [78] also used similar orientation feature in the feature extraction stage to develop a real-time hand gesture recognition system. The orientation features obtained were quantized using codeword from 1 to 12 and finally the gestures were classified using HMM.

Many researchers [79-81] have used location, orientation and velocity as features in their feature extraction stage. Xu *et al.*, designed a hand gesture recognition system for robotic application [79]. They used three features such as orientation between consecutive trajectory points represented by 1-8 codeword, location and velocity. Chamfer distance matching technique and skin color model were used for hand segmentation. Yoon *et al.*, [80] also used the combination of features such as location, orientation and velocity as shown in Fig. 2.3. Orientation feature includes orientation from center to each point of the gesture trajectory and orientation between each consecutive trajectory points. Elmezain *et al.*, [81] used two location features such as distance from center to all the trajectory points of the gesture; and distance from start point to all the trajectory points of the gesture path. Three orientation features used by them were: orientation from center to each point of the trajectory; orientation between successive points and orientation between first point and each points in the trajectory. Finally, the velocity feature was combined with the orientation and location features.

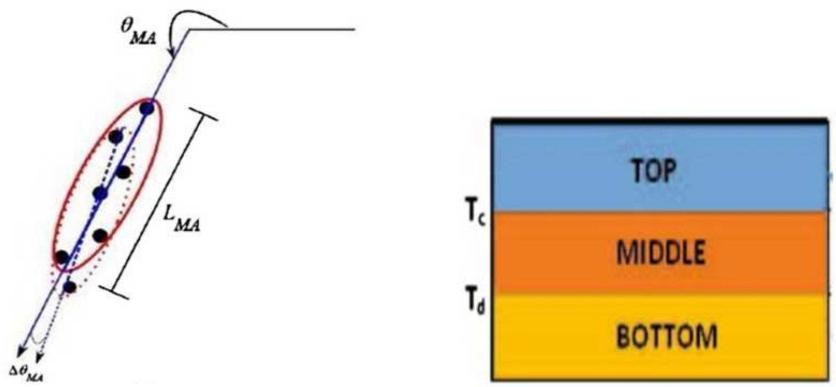
Recently, many researchers combined multiple features to improve the performance of the system. Bhuyan *et al.*, used two types of features: four static and two dynamic features to recognize the gestures [82]. Static features included key trajectory point selection, trajectory length, location and orientation features. The orientation feature includes: the number of significant curves and orientation of start and end of the gesture trajectory. The dynamic features include the velocity and acceleration features. Average velocity, maximum trajectory velocity, minimum trajectory velocity, number of maxima and number of minima are features calculated from the velocity features.



(a) (b)

FIGURE 2.3 Feature extraction [80] (a) three features: location (L_t), orientation (θ_{1t}, θ_{2t}) and velocity (V_t) (b) chain code

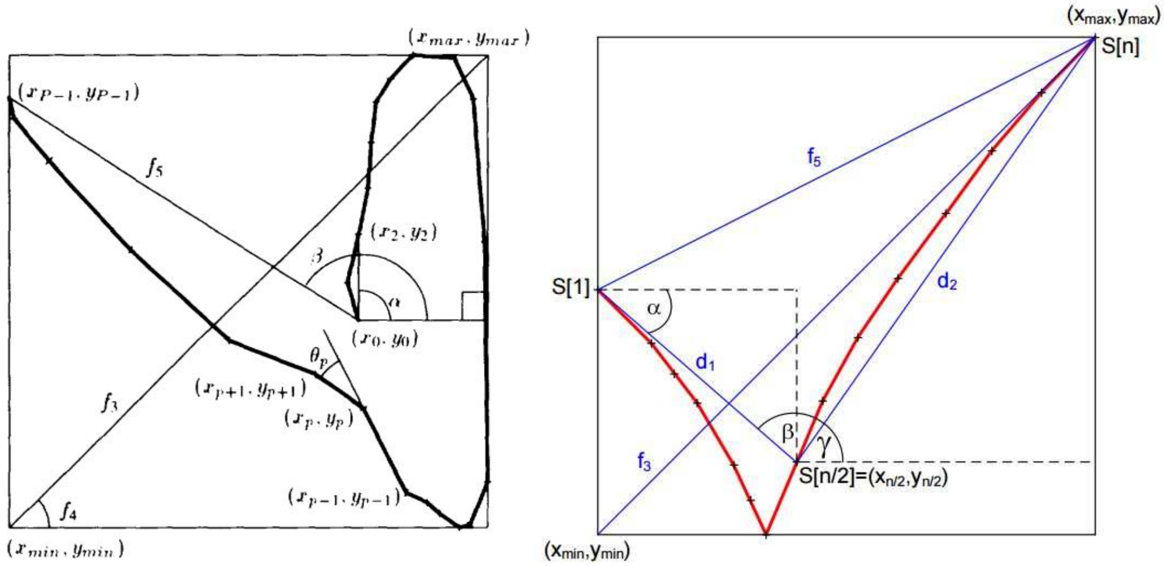
Bhuyan *et al.*, [83] extended the feature set used in their paper [82] by adding the standard deviation of the speed feature to recognize the gestures from the same set of database. A gesture recognition system was developed using Conditional Random Fields (CRF) based classifier model to recognize continuous sequence of gestures [84]. They used ellipse fitting technique in which the ellipse was fitted over every six consecutive trajectory points. From the ellipse, features such as orientation of the major axis of each ellipse, length of the major axis of each ellipse were extracted. Another feature namely position feature was extracted where the position of the start and end of the gesture trajectory was found from the gesture region which was divided into three horizontal sections: Top/Middle/Bottom. An accuracy of 96% was achieved for isolated gestures. Combination of two types of features: hand shape and hand direction was used by Li *et al.*, [85]. Hand shape includes the distance between the fingers of the hand and hand direction feature includes acceleration, velocity and orientation features.



(a) (b)

FIGURE 2.4 Feature extraction [84] (a) orientation of major axis of the ellipse (b) region to indicate position of the hand

Rubine developed 13 features [86] as shown in Fig. 2.5(a). the features are: cosine and sine of the initial angle with respect to the x-axis (cosα, sinα), the length of bounding box diagonal (f_3), the angle of the bounding box (f_4), the distance between first and last point (f_5), the cosine and sine of the angle between the first and last point (cosβ, sinβ), the total gesture length, the total traversed angle, maximum speed squared and stroke duration. Signer *et al.*, [87] extended the features used by Rubine [86] and included 11 features in addition to Rubine's feature. The additional features used are: number of stop points, distance from start to centre point in relation to the diagonal (d_1/f_3), direction of the first half of the stroke (sinα), direction of the second half of the stroke (cosγ), angle between the first and second half of the stroke (cosβ), distance from the start to end point in relation to the diagonal (f_5/f_3), total number of strokes, straightness, total distance between strokes, angle between strokes and stroke distance in relation to each other.



(a) (b)

FIGURE 2.5 (a) Rubine features [86] (b) E-Rubine features [87]

2.2.4 Modelling and recognition of gestures

Some of the mostly used classification techniques for dynamic hand gesture recognition are as follows a) Hidden Markov Model (HMM) b) Neural Networks c) SVM d) k-NN e) Finite State Machine (FSM) f) Classifier fusion.

Hidden Markov Model: HMM were introduced in the mid-1990s. It quickly became one of the most widely used recognition method due to its inherent solution to the segmentation problem. In a HMM, Markov chains are simply finite-state automata in which each state transition arc has an associated probability value [88]. 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” [36]. A Markov chain without the above Markov chain restriction is a HMM [89]. HMMs are non-deterministic as it can have more than one arc with the same output symbol. Also, it is not possible to directly determine the state sequence for a set of inputs by simply looking at the output, so the term ‘hidden’ is used in the HMM. 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.

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. Then a group of HMMs may be used, one for each gesture to run a sequence of input data through each HMM. Chen *et al.*, [90] used HMM based recognizers for identifying the best likelihood gesture model for a given pattern. The variations in gesture from a reference pattern reduce the likelihood of the gesture with the model. An HMM based threshold model was used to filter out the patterns which were having less likelihood values [91]. The direction of the hand movement is used to represent the spatio-temporal sequences of gestures. The method reliably detects an end point of a gesture and with the use of backtracking the start point was estimated properly.

Marcel *et al.*, [92] extended the HMM algorithm and developed Input/Output Hidden Markov Model (IOHMM), for hand gesture recognition. IOHMM is based on a non-homogeneous Markov chain where the emission and transition probabilities depend on the input. The IOHMM learns to map the observations, input sequences, output sequences, and the gesture classes for all the observations using a supervised discriminant learning. IOHMM is a discriminative approach as it directly models posterior probabilities as compared to the HMMs. But they performed the experiments for only binary class. Just and Marcel [93] performed experiments for large database in which there were 7 to 16 classes. The study was made for the recognition of single and double handed gestures. A comparative analysis of HMM and IOHMM concluded that HMM performed better than IOHMM for large number of classes. Conditional Random Fields (CRF) is a widely used tool nowadays. It is advantageous in comparison to HMM because CRF does not consider strong independent assumptions about the observations and can be trained with a fewer samples than HMM [84, 94]. Beh *et al.*, [95] used HMM as the classifier. The hand motion trajectories which composed of unique series of straight and curved segments are modelled as a connected series of states. A state splitting algorithm was proposed by Siddiqi *et al.*, [96] which was based on expectation-maximization algorithm. A method was developed by Ulas and Yildiz [97] to find the optimal structure of HMM. For this purpose, they incremented the number of states, subsequently measured the likelihood values and tried to find the optimal structure.

Neural Networks: Neural Networks have become a growing classifier with great influence over pattern recognition history [98-100]. Nowadays, the time delay neural networks (TDNN) are widely used. “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” [36]. A TDNN network was used to learn the2D motion trajectories [101, 102]. The classification in TDNN is dynamic as the network sees only a small window of the input motion pattern. The window slides over the input data while the network makes a series of local decisions. These local decisions are temporally integrated into a global decision at the output layer.

Support Vector Machine: SVM is a well-known supervised learning method. The class separation hyperplane is optimized in such a way that the distance between pattern and the hyperplane separating the classes is maximized [27, 103, 104]. Gopalan and Dariush [105] found the skin colored pixels and regions corresponding to such pixels were cropped out. Features extracted were distance, angle by which each points on the contour was related to one another by Inner distance shape context algorithm and finally gesture was recognized by SVM. Multiclass SVM classifier was used in the testing stage to classify the detected hand posture. This postures were captured from a webcam after constructing visual words vector for keypoints of the small image that contains the detected hand gesture only [27]. Experimental results show that the system has achieved satisfactory performance with high classification accuracy of 96.23% under different challenging environment like variable scale, orientation and illumination conditions, and cluttered background. A Library of SVM (LIBSVM) was used to recognize the hand gestures [106]. The parameters used for SVM were Radial Basis Function (RBF) as the kernel function and the other classifier parameters were tuned by a grid search approach and cross-validation on the training set.

k-Nearest Neighbor: It is a method for classifying objects based on closest training examples in the feature space. k-Nearest Neighbor (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 [107-110]. 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.

Finite State Machine: FSM is a technique that has a limited or finite number of possible states. A finite state machine can be used both as a development tool for solving problems and as a formal way of describing the solution for later developers and system maintainers [36]. The gestures were decomposed into four distinct phases which occurred in a fixed order, and hence developed a FSM model for recognition [111]. A temporal signature of hand motion is extracted and then the hand gesture are modelled using an FSM [112]. The dominant motion was estimated from an image sequence using the concept of motion energy. The positions of the centers of users head and hands were used to develop the FSM model [113]. A dynamic Bayesian network model is proposed for the recognition of continuous hand gestures by using the features such as direction codes for hand motion, positional relation between the two hands, and the positional relation between face and hands [114].

Classifier Fusion: In recent literatures, classifier combining techniques are becoming popular and has shown to outperform the traditional single classifiers [115, 116]. However, there are very few research papers available in the area of gesture recognition to avail the benefits of classifier combining methods. Dinh *et al.*, [117] proposed a hand gesture recognition system that uses boosted cascade of classifiers trained by AdaBoost and informative Harr wavelet coefficients as features. Burger *et al.*, developed a belief-based method for SVM fusion to recognize hand shapes [118]. Moreover, the method was integrated into a wider classification framework that allows taking into account other sources of information by expressing the same in belief theories' formulation. The experimental results have shown that this method was better than the classical methods in avoiding more than 1/5 of the mistakes. A combination of HMM and Recurrent Neural Networks (RNN) was used by Chan *et al.*, [119] which provided better performance compared to the performance of the individual classifiers such as HMM or RNN. The shape features used are based on Fourier descriptors, which are the inputs to RBF network for an initial pose classification. The pose likelihood vector from the RBF network and the motion information is fed as an input to the HMM and RNN. Outputs from the classifiers are combined to obtain the final result.Wang *et al.*, used a combination of AdaBoost and rotation forest to recognize hand gestures [120]. They have also observed that the performance of the combined technique was more efficient than the individual technique. HMM and ANN based models were also combined for hand gesture recognition. It was found that the combination of the classifier models outperform the individual classifiers [121]. Thus, it may be pointed out that the combination of suitable set of features along with the better classifier combining techniques may provide improved performance for hand gesture recognition system.

2.3 Continuous Hand Gesture Recognition

In a continuous hand gesture recognition, gesture spotting is the most crucial stage. Gesture spotting is used to derive the start and end point of a gesture so as to obtain the gesture boundary. Yoon *et al.*, used time varying based spotting rule where the users were trained to stop for few seconds before the gesture starts and after it ends [80]. The pause region indicates the start and end region and the gesture between the two pause region was referred to as a pure gesture. Bhuyan *et al.*, developed a model where ellipse was fitted to the gesture trajectory points to obtain the features from it [84]. The start and end points of a gesture was found by calculating the length of the major axis of the ellipse. The length was found to be very small for the ellipse fitted at the first six and last six points. But, the gestures used were single stroked gestures, therefore the above gesture spotting method could not be used in self co-articulated gestures. Fig. 2.6 shows some example of the continuous gesture sequence. It can be observed that by only removing the movement epenthesis, gesture spotting can easily be obtained. Velocity information was used to separate the movement epenthesis from the continuous gesture sequence [122]. They observed that the movement epenthesis has velocity greater the normal gesture strokes as shown in Fig. 2.7. Thus, the gesture spotting may be performed using velocity information. Elmezain *et al.*, [77] performed gesture spotting by keeping a constraint to the user during gesticulation. The constraint was a zero codeword should be followed after every gesture in a sequence as shown in Fig. 2.8.

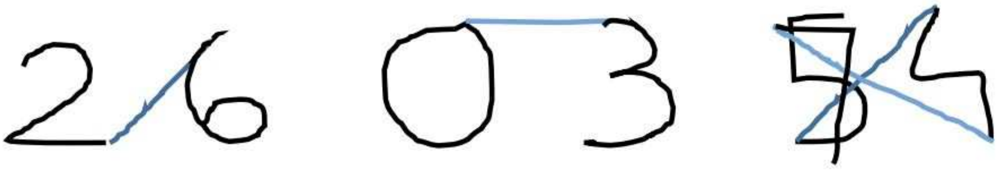
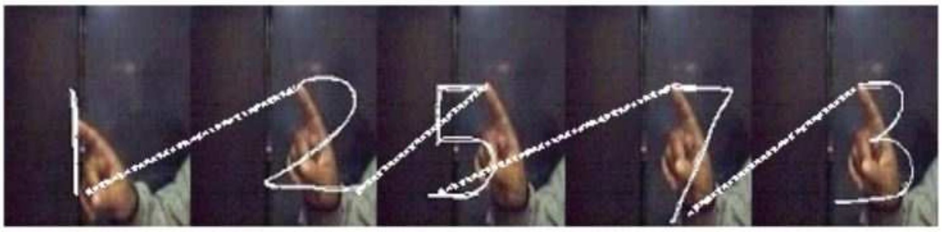
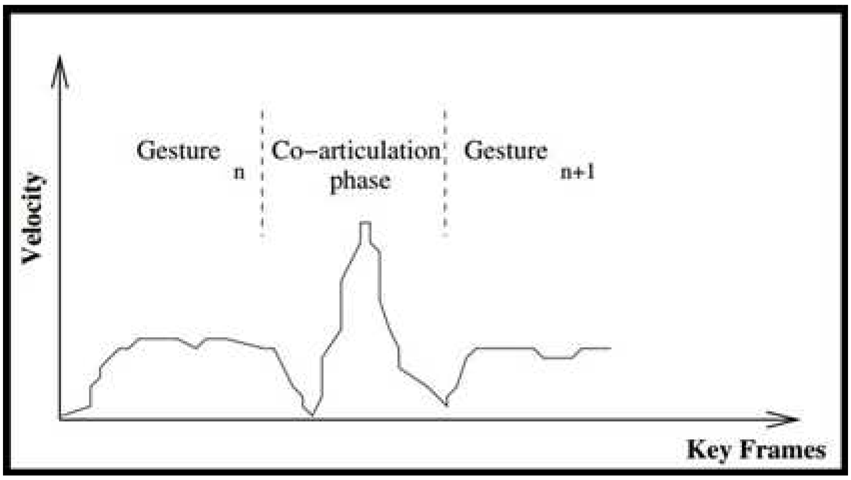


FIGURE 2.6 Continuous gesture sequence of paper [84]



(a)



(b)

FIGURE 2.7 (a) Continuous gesture sequence (b) velocity profile of continuous gesture having sequence of two gestures [122]

2.4 Summary

Hand gesture recognition has gained large attention due to its potential applications in contactless human computer interaction (HCI). Interactions between human and computer are currently performed using keyboards, mouse or joysticks. These tools are different from our natural way of interaction with computers. Moreover, they do not provide enough flexibility for a number of practical applications. Thus, recognizing the hand gestures may help in achieving the ease and naturalness desired for HCI.

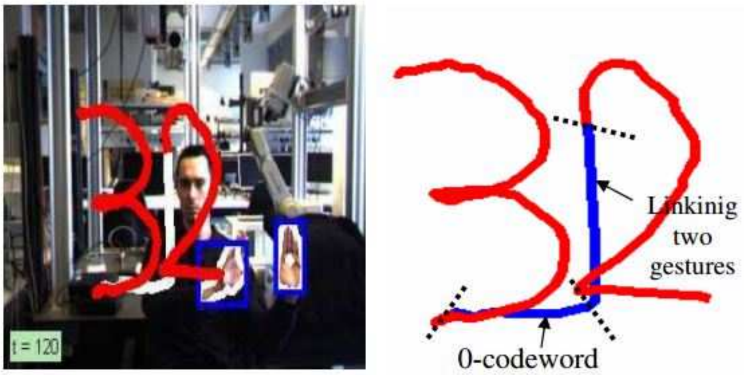


FIGURE 2.8 Continuous gesture sequence of paper [77]

The first problem in any hand gesture recognition is to detect the hand from the background. There are various difficulties during this phase such as cluttered background, illumination problem, occlusion, etc. The second problem is the tracking of hand. The hand needs to be tracked correctly in every frame of the video in order to achieve the correct gesture trajectory. This phase is affected by different scenarios like varying gesticulation speed and pattern. The third problem is the detection and removal of the unwanted hand movements which may be intentionally done (self co-articulation) or unintentionally done (hand trembling). Detecting these unwanted strokes will make the system easier for recognition. The fourth problem is to develop strong feature set for the system. The last problem is to develop a system which should be able to recognize continuous sequence of data that are connected by self co-articulation, movement epenthesis and other unwanted hand movements.

To address all these issues, different research work has been carried out. This chapter presents a brief review of most of the relevant work carried out in the area of hand gesture recognition. The various models proposed by the researchers to design the hand gesture recognition system and the techniques to improve the performance have been presented. In the next chapters, various systems are developed for different practical scenarios such as presence and absence of variation in gesticulation speed and gesticulation pattern.

References

[1] N. H. Dardas, “Real-time hand gesture detection and recognition for human computer interaction,” Doctoral dissertation, University of Ottawa, 2012.

[2] P. Kumar, S. S. Rautaray, and A. Agrawal, “Hand Data Glove: A new generation real-timemouse for human-computer interaction,” in *Proc. 1st International Conference on Recent Advances in Information Technology*, Dhanbad, pp. 750-755, 15-17 March 2012.

[3] J. H. Kim, N. D. Thang, and T. S.Kim, “3-D Hand Motion Tracking and Gesture Recognition Using a Data Glove,” in *Proc. IEEE International Symposium on Industrial Electronics*, pp. 1013-1018, 5-8 July 2009.

[4] J. Weissmann and R. Salomon, “Gesture recognition for virtual reality applications using data gloves and neural networks,” in *Proc. IEEE International Joint Conference on Neural Networks*, pp. 2043-2046, 1999.

[5] A. Cavallo, “Primitive Actions Extraction for a Human Hand by using SVD,” in *Proc. IEEE 9th International Symposium on Intelligent Systems and Informatics*, Subotica, pp. 413-419, 8-10 September 2011.

[6] S. Saengsri, V. Niennattrakul, and C. A. Ratanamahatana, “TFRS: Thai Finger-Spelling Sign Language Recognition System,” in *Proc. 2nd International Conference on Digital Information and Communication Technology and its Applications*, Bangkok, pp. 457-462, 16-18 May 2012.

[7] W. Gao, G. Fang, D. Zhao, and Y.Chen, “A Chinese Sign Language Recognition System based on SOFM/SRN/HMM,” *Pattern Recognition*, vol. 37, no. 12, pp. 2389-2402, December 2004.

[8] C. Oz and M. C. Leu, “Linguistic Properties based on American Sign Language Isolated word Recognition with Artificial Neural Networks using a Sensory Glove and Motion Tracker,” *Neurocomputing*, vol. 70, no. 16-18, pp. 2891-2901, 2007.

[9] A. Ibarguren, I. Maurtua, and B. Sierra, “Layered Architecture for Real Time Sign Recognition: Hand Gesture and Movement,” *Engineering Applications of Artificial Intelligence*, vol. 23, no. 7, pp. 1216-1228, October 2010.

[10] S. Iba, J. M. V. Weghe, C. J. J. Paredis, and P. K. Khosla, “An architecture for gesture-based control of mobile robots,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 851-857, 17-21 October 1999.

[11] T. G. Zimmerman, J. Lanier, C. Blanchard, S. Bryson, and Y. Harvill, “A hand gesture interface device,” in *ACM SIGCHI Bulletin*, vol. 18, no. 4, pp. 189-192. ACM, 1987.

[12] H. Teleb and G. Chang, “Data glove integration with 3D virtual environments,” in *Proc. International Conference on Systems and Informatics*, pp. 107-112, 12-20 May 2012.

[13] J. L. H. Rebollar, N. Kyriakopoulos, and R. W. Lindeman, “A new instrumented approach for translating American sign language into sound and text,” in *Proc. 6th IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul, pp. 547-552, 17-19 May 2004.

[14] K. Kadam, R. Ganu, A. Bhosekar, and S. D. Joshi, “American sign language interpreter,” in *Proc. IEEE 4th International Conference on technology for Education*, pp. 157-159, 18-20 July 2012.

[15] P. Mistry, P. Maes, L. Chang, “WUW-wear Ur world: a wearable gestural interface,” in *Proc. 27th International Conference Extended Abstracts on Human Factors in Computing Systems*, Boston, USA, pp. 4111-4116, 2009.

[16] R. Y. Wang and J. Popović, “Real-time hand-tracking with a color glove,” *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 63, 2009.

[17] S. S. Rautaray and A. Agrawal, “Real time hand gesture recognition system for dynamic applications,” *International Journal of UbiComp*, vol. 3, no. 1, pp. 21–31, 2012.

[18] D. Saxe and R. Foulds, “Toward robust skin identification in video images,” in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, Killington, pp. 379–384, 1996.

[19] D. Chai and K. Ngan, "Locating the facial region of a head and-shoulders color image," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, Piscataway, pp. 124–129, 1998.

[20] J. Yang, W. Lu, and A. Waibel, "Skin-color modeling and adaptation," Springer Berlin Heidelberg, pp. 687–694, 1998.

[21] C. Yu, X. Wang, H. Huang, J. Shen, and K. Wu, "Vision-based hand gesture recognition using combinational features," in *Proc. IEEE 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Germany, pp. 543-546, 15-17 October 2010.

[22] J. Rekha, J. Bhattacharya, and S. Majumder, "Shape, texture and local movement hand gesture features for Indian sign language recognition," in *Proc.3rd International Conference on Trendz in Information Sciences and Computing*, Chennai, pp. 30-35, 8-9 December 2011.

[23] M. Panwar and P. S. Mehra, "Hand gesture recognition for human computer interaction," in *Proc. International Conference on Image Information Processing*, Shimla, pp. 1-7, 3-5 November 2011.

[24] A. Malima, E. Ozgur, and M. Cetin, "A fast algorithm for vision-based hand gesture recognition for robot control," in *Proc. IEEE 14th Signal Processing and Communications Applications*, pp. 1-4, 17-19 April 2006.

[25] M. Manigandan and I. M.Jackin, "Wireless vision based mobile robot control using hand gesture recognitionthrough perceptual color space," in *Proc. International Conference on Advances in Computer Engineering*, pp. 95-99, 20-21 June 2010.

[26] Y. Fang, K. Wang, J. Cheng, and H. Lu, "A real-time hand gesture recognition method," in *Proc. IEEE International Conference on Multimedia and Expo*, pp. 995-998, 2-5 July 2007.

[27] N. H. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 11, pp. 3592-3607, 2011.

[28] J. Rehg and T. Kanade, "Digiteyes: vision-based hand tracking for human-computer interaction," in *Proc. Workshop on Motion of Non-rigid and Articulated Bodies*, Austin Texas, pp. 16–24, 1994.

[29] C. T.Hsieh, C. H. Yeh, K. M.Hung, L. M. Chen, and C. Y. Ke, "A real time hand gesture recognition system based on DFT and SVM," in *Proc.8th International Conference on Information Science and Digital Content Technology*, pp. 490-494, 26-28 June 2012.

[30] D. Tewari and S. K. Srivastava, "A visual recognition of static hand gestures in Indian sign language based on kohonen self-organizing map algorithm," *International Journal of Engineering and Advanced Technology*, vol. 2, no. 2, pp. 165-170, 2012.

[31] S. Ramli, "GMT feature extraction for representation of BIM sign language," in *Proc. IEEE Control and System Graduate Research Colloquium*, pp. 43-48, 16-17 July 2012.

[32] A. Utsumi and J. Ohya, "Image segmentation for human tracking using sequential-image-based hierarchicaladaptation," in *Proc. IEEE Computer Vision and Pattern Recognition(CVPR)*, pp. 911–916, 1998.

[33] A. Blake, B. North, and M. Isard, "Learning multi-class dynamics," *Advances in Neural Information Processing Systems*, vol 11, pp. 389–395, 1999.

[34] M. Yang and N. Ahuja, "Detecting human faces in color images," in *Proc. International Conference on Image Processing* , Piscataway, pp. 127–130, 1998.

[35] L. Sigal, S. Sclaroff, and V. Athitsos, "Skin color-based video segmentation under time-varying illumination," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, pp. 862–877, 2004.

[36] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1-54, 2015.

[37] J. Rehg and T. Kanade, "Model-based tracking of self-occluding articulated objects," in *Proc. International Conference on Computer Vision*, pp. 612–617, 1995.

[38] N. Shimada, Y. Shirai, Y. Kuno, and J. Miura, "Hand gesture estimation and model refinement using monocular camera ambiguity limitation by inequality constraints," in *Proc. IEEE International Conference on Face and Gesture Recognition,* Japan, pp. 268–273, 1998.

[39] Y. Wu and T. T. Huang, "Capturing human hand motion: a divide-and-conquer approach," in *Proc. International Conference on Computer Vision*, Greece, pp. 606–611, 1999.

[40] Y. Wu, J. Lin, and T. Huang, "Capturing natural hand articulation," in *Proc. International Conferenceon Computer Vision (ICCV)*, Vancouver, pp. 426–432, 2001.

[41] J. Lee and T. L. Kunii, "Model-based analysis of hand posture," *IEEE Computer Graphics and Applications*, vol. 15, no. 5, pp. 77–86, 1995.

[42] T. Heap and D. Hogg, "Towards 3D hand tracking using a deformable model," in *Proc. IEEE International Conference Automatic on Face and Gesture Recognition*, Killington, pp. 140–145, 1996.

[43] R. Cutler and M. Turk, "View-based interpretation of real-time optical flow for gesture recognition," in *Proc. IEEE International Conference on Face and Gesture recognition*, pp. 416–421, 1998.

[44] J. Martin, V. Devin and J. Crowley, "Active hand tracking," in *Proc. IEEE Conference on Automatic Face and Gesture Recognition*, Japan, pp. 573–578, 1998.

[45] Q. Yuan, S. Sclaroff, and V. Athitsos, "Automatic 2D hand tracking in video sequences," in *Proc. IEEE Workshop on Applications of Computer Vision*, pp. 250–256, 1995.

[46] Y. Quan, "Chinese sign language recognition based on video sequence appearance modeling," in *Proc.5th IEEE Conference on Industrial Electronics and Applications*, pp. 1537-1542, 15-17 June 2010.

[47] C. Kim and S. I. Chien, "Analysis of 3D hand trajectory gestures using stroke based composite hidden markov models," *Applied Intelligence*, vol. 15, no. 2, pp. 131-143, 2001.

[48] J. M. Guo, Y. F. Liu, and C. H. Chang, "Improved hand tracking system," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 5, pp. , May 2012.

[49] E. Koh, J. Won, and C. Bae, "On-premise skin color modeing method for vision-based hand tracking," *International Symposium on Consumer Electronics*, pp. 908–909, 2009.

[50] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Mach. Intelligence*, vol. 25, no. 5,pp. 234–240, 2003.

[51] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi, "Robust online appearance models for visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.25, no. 10, pp. 1296–1311, 2003.

[52] S. Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particles filters," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1491–1506, 2004.

[53] S. McKenna, J. Sabri, and S. Gong, "Tracking color objects using adaptive mixture models," *Image and Vision Computing*, vol. 17, no. 3,pp. 225–231, 1999.

[54] N. D. Binh, E. Shuichi, and T. Ejima, "Real time hand tracking and gesture recognition system," in *Proc. International Conference on Graphic, Vision and Image Processing*, pp. 111–115, 2005.

[55] K. Imagawa, S. Lu, and S. Igi, "Color based hands tracking system for sign language recognition," in *Proc. International Conference on Face and Gesture Recognition*, pp. 462–467, 1998.

[56] M. Isard and A. Blake, "Condensation-conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[57] C. Shan, T. Tan, and Y. Wei, "Real-time hand tracking using a mean shift embedded particle filter," *Pattern Recognition*, vol. 40, no. 7, pp. 1958–1970, 2007.

[58] S. K. Weng, C. M. Kuo, and S. K. Tu, "Video object tracking using adaptive Kalman filter," *Journal of Visual Communication and Image Representation*, vol. 17, no. 6, pp. 1190–1208, 2006.

[59] Q. Y. Zhang, M. Y. Zhang, and J. Q. Hu, "Hand gesture contour tracking based on skin color probability and state estimation model," *Journal of Multimedia*, vol. 4, no. 6, pp. 349–355, 2009.

[60] W. Zheng and S. M. Bhandarkar, "Face detection and tracking using a boosted adaptive particle filter," *Journal of Visual Communication and Image Representation*, vol.20, no. 1, pp. 9–27, 2009.

[61] A. Utsumi and J. Ohya, "Multiple-hand-gesture tracking using multiple cameras," in *Proc. IEEE Computer Vision and Pattern Recognition*, Colorado, pp. 473–478, 1999.

[62] N. Peterfreund, “Robust tracking of position and velocity with Kalman snakes,” *IEEE Transactions on Pattern Analysis andMachine Intelligence*, vol. 10, no. 6, pp. 564–569, 1999.

[63] M. S. M. Asaari, B. A. Rosdi, and S. A. Suandi, “Adaptive Kalman filter incorporated eigenhand (AKFIE) for real-time hand tracking system,” *Multimedia Tools and Applications*, vol. 74, no. 21, pp. 9231-9257, 2015.

[64] M. Isard and A. Blake, “A mixed-state condensation tracker with automatic model-switching,” in *Proc. 6th International Conference on Computer Vision*, Bombay, India, pp. 107-112, January 1998.

[65] J. P. Mammen, S. Chaudhuri, and T. Agrawal, “Simultaneous tracking of both hands by estimation of erroneous observations,” in *Proc. British Machine Vision Conference (BMVC)*, Manchester, pp. 1-10, 2001.

[66] P. Perez, C. Hue, J. Vermaak, and M. Gangnet, “Color-based probabilistic tracking,” in *Proc. European conference on computer vision*, Copenhagen, pp. 661–675, 2002.

[67] S. M. Nadgeri, S. D. Sawarkar, and A. D. Gawande, “Hand gesture recognition using CAMSHIFT algorithm,” in *Proc. 3rd International Conference on Emerging trends in Engineering and Technology*, pp. 37-41, November 2010.

[68] G. R. Bradski, “Computer vision face tracking as a component of a perceptual user interface,” in *The Workshop on Applications of Computer Vision*, Princeton, NJ, pp. 214–219, 1998.

[69] G. Bradski, “KAEHLER: Learning Open CV,” O‘Reilly, pp. 337–341, 2008.

[70] W. Xiangyu and L. Xiujuan, “The study of moving target tracking based on Kalman camshift in the video,” in *Proc. 2nd International conference on information science and engineering (ICISE)*, pp. 1–4, 2010.

[71] S. Huang and J. Hong, “Moving object tracking system based on camshift and Kalman filter,” in *Proc. International conference on consumer electronics, communications and networks*, pp. 1423–1426, 2011.

[72] J. Shi and C. Tomasi, “Good features to track,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593-600, 1994.

[73] M. Kolsch and M. Turk, “Fast 2D hand tracking with flocks of features and multi-cue integration,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pp. 158, June 2004.

[74] F. Porikli, O. Tuzel, and P. Meer, “Covariance tracking using model update based on means on riemannian manifolds,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 728–735, 2006.

[75] O. Tuzel, F. Porikli, and P. Meer, “Region covariance: a fast descriptor for detection and classification,” in *Proc. European Conference on Computer Vision*, pp. 589–600, 2006.

[76] M. J. Black and A. D. Jepson, “Eigen tracking: robust matching and tracking of articulated objects using a view-based representation,” *International Journal of Computer Vision*, vol. 26, no. 1, pp. 63–84, 1998.

[77] M. Elmezain, A. Al-Hamadi, J. Appenrodt, and B. Michaelis, “A hidden markov model-based continuous gesture recognition system for hand motion trajectory,” in *Proc. 19th International Conference on Pattern Recognition*, pp.1-4, 2008.

[78] C. Y. Kao and C. S. Fahn, “A human-machine interaction technique: hand gesture recognition based on hidden Markov models with trajectory of hand motion,” *Procedia Engineering*, vol. 15, pp. 3739-3743, 2011.

[79] D. Xu, X. Wu, Y. L. Chen, and Y. Xu, “Online dynamic gesture recognition for human robot interaction,” *Journal of Intelligent and Robotic Systems*, vol. 77, no. 3-4, pp. 583-596, 2014.

[80] H. S. Yoon, J. Soh, Y. J. Bae, and H. S. Yang, “Hand gesture recognition using combined features of location, angle and velocity,” *Pattern Recognition*, vol. 34, no. 7, pp. 1491-1501, 2001.

[81] M. Elmezain, A. Al-Hamadi, and B. Michaelis, “Hand gesture recognition based on combined features extraction,” *World Academy of Science, Engineering and Technology*, vol. 60, pp. 395, 2009.

[82] M. K. Bhuyan, D. Ghosh, and P. K. Bora, "Feature extraction from 2D gesture trajectory in dynamic hand gesture recognition," in Proc. *IEEE Conference on Cybernetics and Intelligent Systems*, pp. 1-6, 2006.

[83] M. K. Bhuyan, P. K. Bora, and D. Ghosh, "Trajectory guided recognition of hand gestures having only global motions," *International Journal of Computer Science*, vol. 2, no. 9, pp. 753-764, 2008.

[84] M. K. Bhuyan, D. A. Kumar, K. F. MacDorman, and Y. Iwahori, "A novel set of features for continuous hand gesture recognition," *Journal on Multimodal User Interfaces*, vol. 8, no. 4, pp. 333-343, 2014.

[85] K. Li, Z. Zhou, and C. H. Lee, "Sign transition modeling and a scalable solution to continuous sign language recognition for real-world applications," *ACM Transactions on Accessible Computing*, vol. 8, no. 2, pp. 7:1-7:23, 2016.

[86] B. Rubine, "Specifying gestures by example," in *Proc. ACM 18th International Conference on Computer Graphics and Interactive Techniques*, USA, pp. 329-337, July 1991.

[87] B. Signer, M. C. Norrie, and U. Kurmann, "iGesture: A Java framework for the development and deployment of stroke-based online gesture recognition algorithms," Technical Report TR561, ETH Zurich, 2007.

[88] R. Rabiner and B. H. Juang, "An introduction to hidden Markov models," *IEEE Signal Processing Magazine*,vol. 3, no. 1, pp. 4-16, 1986.

[89] E. Charniak, "Statistical language learning," MIT Press, Cambridge, 1993.

[90] S. Chen, C. M. Fu, and C. L. Huang, "Hand gesture recognition using a real-time tracking method and hidden markov models," *Image and Vision Computing,* vol. 21, pp. 745-758, 2003.

[91] H. Lee and J. H. Kim, "An HMM based threshold model approach for gesture recognition,"*IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 961-973, 1999.

[92] S. Marcel, O. Bernier, J. E. Viallet, and D. Collobert, "Hand gesture recognition using input/output hidden markov models," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 456-461, 2000.

[93] A. Just and S. Marcel, "A comparative study of two state-of-the-art sequence processing techniques for hand gesture recognition," *Computer Vision and Image Understanding*, vol. 113, no. 4, pp. 532-543, 2009.

[94] R. Yang and S. Sarkar, "Detecting co-articulation in sign language using Conditional Random Fields," in *Proc. IEEE International Conference on Pattern Recognition (ICPR)*, pp. 108-112, HongKong, 2006.

[95] J. Beh, D. Han, and H. Ko, "Rule-based trajectory segmentation for modeling hand motion trajectory," *Pattern Recognition*, vol. 47, no. 4, pp. 1586-1601, 2014.

[96] S. M. Siddiqi, G. J. Gordon, and A. W. Moore, "Fast state discovery for HMM model selection and learning," in *Proc. International Conference on Artificial Intelligence and Statistics*, pp. 492-499, 2007.

[97] A. Ulas and O. T. Yildiz, "An incremental model selection algorithm based on cross validation for finding the architecture of a hidden Markov model on hand gesture data sets," in *Proc. International Conference on Machine Learning and* Applications, pp. 170-177, 2009.

[98] M. Bishop, "Neural networks for pattern recognition," Clarendon Press, Oxford, 1995.

[99] S. Haykin, "Neural networks and learning machines," vol. 3, Upper Saddle River: Pearson Education, 2009.

[100] B. D. Ripley, "Pattern recognition and neural networks" University Press, Cambridge, 1996.

[101] H. Yang, N. Ahuja, and M. Tabb, "Extraction of 2D motion trajectories and its application to hand gesture recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1061-1074, 2002.

[102] H. Yang and N. Ahuja, "Extraction and classification of visual motion patterns for hand gesture recognition," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, USA, pp. 892-897, 1998.

[103] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415-425, 2002.

[104] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293-300, 1999.

[105] R. Gopalan and B. Dariush, "Towards a vision based hand gesture interface for robotic grasping", in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1452-1459, October 10-15, 2009.

[106] F. Dominio, M. Danadeo, and P. Zanuttigh, "Combining multiple depth-based descriptors for hand gesture recognition," *Pattern Recognition Letters*, vol. 50, pp. 101-111, 2014.

[107] S. Thirumuruganathan, "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/, 2010.

[108] S.S., Yang, Y. and T.H. Lee, "Hand gesture recognition and tracking based on distributed locally linear embedding, " *Image and Vision Computing*, vol. 26, no. 12, pp. 1607-1620, 2008.

[109] R.Y. Wang and J. Popović, "Real-time hand-tracking with a color glove." *ACM transactions on graphics (TOG)*, vol. 28, no. 3, pp. 63. ACM, 2009.

[110] K. Oka, Y. Sato, and H. Koike, "Real-time fingertip tracking and gesture recognition." *IEEE Computer graphics and Applications*, vol. 22, no. 6, pp. 64-71.

[111] J. Davis and M. Shah, "Recognizing hand gestures," in *Proc. European Conference on Computer Vision*, pp. 331-340, 1994.

[112] M. Yeasin and S. Chaudhuri, "Visual understanding of dynamic hand gestures," *Pattern Recognition*, vol. 33, no. 11, pp. 1805-1817, 2000.

[113] P. Hong, M. Turk, and T. S. Huang, "Gesture modeling and recognition using finite state machines," in *Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 410-415, 2000.

[114] I. Suk, B. K. Sin, and S. W. Lee, "Hand gesture recognition based on dynamic Bayesian network framework," *Pattern Recognition*, vol. 43, no. 9, pp. 3059-3072, 2010.

[115] T. S. Hai and N. T. Thuy, "Image classification using support vector machine and artificial neural network," *International Journal of Information Technology and Computer Science*, vol. 4, no. 5, pp. 32–38, 2012.

[116] S. Kang and S. Park, "A fusion neural network classifier for image classification," *Pattern Recognition Letters*, vol.30, no. 9, pp. 789–793, 2009.

[117] T. B. Dinh, V. B. Dang, D. A. Duong, T. T. Nguyen, and D. D. Le, "Hand gesture classification using boosted cascade of classifiers," in *Proc. International Conference on Research, Innovation and Vision for the Future*, Vietnam, pp. 139-144, 2006.

[118] T. Burger, O. Aran, A. Urankar, A. Caplier, and L. Akarun, "A Dempster-Shafer theory based combination of classifiers for hand gesture recognition," *Communications in Computer and Information Science*, vol. 21, pp. 137-150, 2008.

[119] C. W. Ngand S. Ranganath, "Real-time gesture recognition system and application," *Image and Vision Computing*, vol.20, pp. 993-1007, 2002.

[120] G. W. Wang, C. Zhang, and J. Zhuang, "An application of classifier combination methods in hand gesture recognition," *Mathematical Problems in Engineering*, pp. 1-17, 2012.

[121] A. Corradini, "Real-time gesture recognition by means of hybrid recognizers," *International Gesture Workshop*, GW 2001 London, UK, pp. 34-47, 2001.

[122] M. K. Bhuyan, D. Ghosh, and P. K. Bora, "Continuous hand gesture segmentation and co-articulation detection," In Proc. *Computer Vision, Graphics and Image Processing*, Springer Berlin Heidelberg, pp. 564-575, 2006.