DOI: 10.6977/IJoSI.202309_7(7).0003
S. Saboo, J. Singha/ Int. J. Systematic Innovation, 7(7),47-70(2023)

Dynamic hand gesture tracking and recognition: Survey of different phases

Shweta Saboo[1*], Joyeeta Singha[2]
[1] Department of Electronics and Communication Engineering, JECRC, Jaipur, Rajasthan, INDIA
[2] Department of Electronics and Communication Engineering, The LNMIIT, Jaipur, Rajasthan, INDIA
* Corresponding author E-mail: shweta.saboo.y18pg@lnmiit.ac.in

(Received 30 November 2022; Final version received 24 August 2023; Accepted 1 September 2023.)

Abstract

Hand gesture plays an important role in controlling various appliances and gadgets nowadays. Recognition of proper gestures with the help of multiple techniques is vital for the hardware interfaced with it. Work has been done on various steps of the process of hand gesture recognition. Starting with video acquisition and pre-processing, hand detection and tracking, and feature extraction finally lead to classification and recognition. This paper provides a detailed review of state-of-art techniques used in recent hand gesture recognition techniques. We have also discussed the advantages and disadvantages of various techniques and the reason behind moving to another method. It is hoped that this study might provide researchers with a comprehensive description of the hand gesture recognition techniques that may help in pattern recognition, computer vision, and artificial intelligence.

Keywords: Computer vision, Hand Gesture, Hand Gesture Dataset, Hand Tracking, Machine learning, Recognition.

1. Introduction

In the growing world of artificial intelligence and image processing, the demand for controlling objects remotely and security has been raised a lot. Image processing needs to be done for visualization, sharpening of images & regeneration, retrieval of images, pattern measurement, and recognition of images. Many algorithms quickly detect and track various gestures in static conditions but do not support dynamic image recognition due to multiple challenges like invariant features, movement between gestures, automatic filtering, feature segmentation, matching techniques, mixed gestures, and complex dynamic backgrounds.

Early research on vision-based hand tracking and gesture recognition usually used markers or colored gloves (Mistry & Chang, 2009, Wang & Poppovic, 2009). Current research in vision-based hand tracking and gesture recognition techniques is more focused on using bare hands and identifying hand gestures without the help of any markers and gloves. However, obtaining highly accurate results is challenging for any vision-based approach (Rautaray, 2012). Systems using bare hands suffer from some difficulties, such as the user and camera needing to be independent and invariant against the dynamic background, transformations, and variable lighting conditions for real-time performance. Human hands can change their shape; therefore, there are chances of difficulty arising in detection and recognition. Work done addressing issues in the research articles related to vision-based recognition techniques helped the researchers identify the key issues and problems and work further on them to reduce and make them more user-friendly. This paper constitutes the research done in the existing literature about various steps to be followed in developing a robust hand gesture recognition system. This study will help to gain knowledge for the upcoming researchers about the process that needs to be followed to develop a system that dynamically recognizes hand gestures and can help in a wide range of applications.

The papers cited as references have been collected by first finding the literature associated with the topic. Various hand gesture recognition system steps have been categorized, and the related literature has been searched and analyzed. After the analysis, those papers were systematically evaluated based on the results and optimized techniques used in them. If the paper is found relevant based on the requirements, the paper is selected and cited. When collecting papers, several factors are considered, including the research question, the scope of the study, and the type of publication required. The search terms used depend on the research question and the scope of the study. A comprehensive list of relevant search terms is compiled, which may include keywords, phrases, and subject headings related to the research topic. The inclusion and exclusion criteria are developed to ensure that only relevant studies are selected for the analysis. The criteria may include factors such as the type of publication, the date of publication, the study design, the sample size, and the quality of the study. After the initial search, the titles and abstracts of the articles are screened to determine their relevance to the research question. The full texts of the relevant articles are then retrieved and assessed for eligibility based on the inclusion and exclusion criteria. Finally, the selected articles are read in detail, and the relevant data are extracted for analysis.

The paper is organized as follows:

Section 2 gives an overall view of the hand gesture recognition system. Section 3 discusses various types of approaches used in Hand Detection. Section 4 presents different types of hand tracking methods. Section 5 provides a list of various features available in previous papers for the feature extraction stage. Section 6 presents classifiers used in machine learning for proper recognition of gestures. Section 7 provides a final summary of the survey.

2. Overview of the Hand Gesture Recognition System

Fig. . 1 shows various steps that play an important role in the Hand gesture recognition process. Gesture recognition involves tracking human gestures to represent them and convert them into meaningful commands Among the various phases of a hand gesture recognition system, video acquisition, and pre-processing steps depend on the applications for which the system is developed. For contactless handling of devices, proper tracking and recognition are required in which a pre-requisite condition of accurate image pre-processing and detection should be fulfilled. Extracted features should be meaningful and also need to justify the proper method of the recognition process.

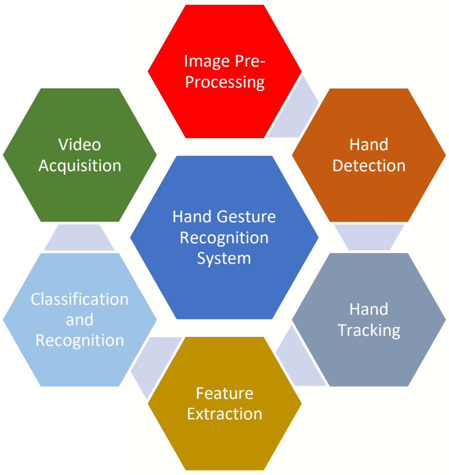


Fig. 1. Overall phases in a hand gesture recognition system

3. Detection

The hand gesture recognition system is initialized by the process of hand detection and extraction from the background. Successful execution of this step will lead to proper tracking and recognition of gestures. Skin color detection, 3D model-based, and motion-based detection are among the methods available in the literature. Fig. . ure 2 shows the various color detection techniques evolving with time.

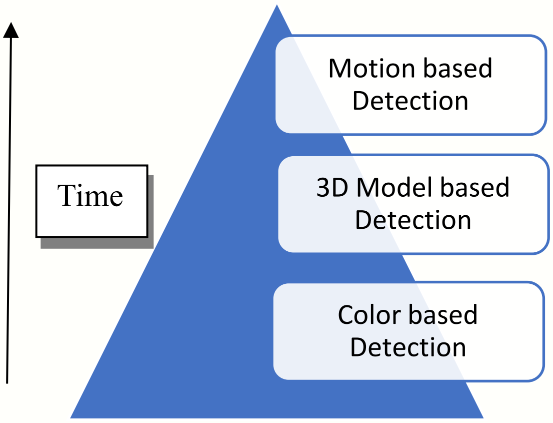


Fig. 2. Different Approaches for Detection

3.1Color based detection：The skin-colored region is extracted successfully from the input image to obtain many researchers' desired hands. Successful detection includes a proper selection of color models, and a variety of them are available, like Red-Green-Blue (RGB), normalized RGB, Hue Saturation Value (HSV) (Saxe & Foulds, 1996), YCrCb (Chai & Ngan 1998, YUV Yang et al., 1997), etc. The YCbCr color model distinguished skin-colored pixels from the background (Yu et al., 2010, Rekha & Majumder et al., 2011, Panwar & Mehra, 2011). The required hand portion was extracted using this color model, filtered by a median filter, and finally processed by a smoothing filter. (Malima et al., 2006) used the Red/Green ratio to determine the skin-colored regions for robotic application. Initially, the center of gravity of the hand was searched, and then the farthest distance from the center was calculated. In this way, the fingertips of the hand have been identified. Manigandan and Jackin (Manigandan & Jackin, 2010) used the same steps as (Malima et al., 2006), except the RGB input was converted to HSV color space before further processing. Fang et al. (Fang & Lu, 2007) applied the Adaptive Boost algorithm to detect hands from the input image captured during image acquisition. This algorithm was able to detect the overlapped hand. The Haar-like features were extracted, and other processing, like face subtraction, skin detection, and a contour comparison algorithm, was used to detect the hand region (Dardas & Georganas, 2011). The experiments were carried out in a 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 (Rehg & Kanade, 1994). The first frame of the video was considered the background for the entire processing of the system (Hsieh et al., 2012). 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 were extracted using the CamShift algorithm to detect the hand from the objects. Tewari and Srivastava (Tewari & Srivastava, 2012) first converted the input RGB image to a grayscale image. The pixels of the hand regions could be extracted easily as it was carried out in a controlled environment where the signer used a black dress and black bandage with a black background. Salleh and Ramli (Ramli, 2012) used background subtraction to detect the hand. It was observed that there are three binary-linked regions, i.e., the face and two hands. The maximum values of the region connected were identified using binary linked object (BLOB) analysis to select the two hands. To solve the background subtraction problem when the camera and background are in a static position, some researchers (Utsumi & Ohya, 1998, Blake et al., 1998) have proposed to make dynamic corrections of background models.

Human skin color varies naturally, but skin color variations in images may also result from changing illumination conditions or camera characteristics. Therefore, the color-based approaches should also consider such problems for compensating such variability. Some researchers (Yang & Ahuja, 1998, Sigal et al., 2004) proposed a model for detecting skin color independent of the changes in illumination. Yoon et al. (Yoon et al.,2001) proposed various Hidden Markov Model (HMM) models for recognition systems having multiple hand gestures. Detection is done based on skin color and motion. Features used comprise combined and weighted location, angle, and velocity. A set of 2400 trained and 2400 untrained gestures have been used for training and testing.

Rahman, Purnama, and Purnomo (Rahman et al.,2014) proposed a new system combining two skin color models for each pixel, forming a vector containing Hue, Saturation, Cb, and Cr color elements. From the proposed detection models, 93.89% true Positif Rate and 10.75% false Positif Rate are calculated. This paper describes three categories of skin color detection. The Explicit Range method determines the pixel class for the predetermined color range. A nonparametric method calculates the variance of a single-color model but cannot handle all kinds of skin color distribution, and the third one is the Parametric method. The database consists of 50 images with varied sizes and ten images. The detection phase converts every pixel of an image from HSV to the YCbCr color model and forms a vector consisting of H, S, Cb, and Cr, which gives results in the form of mean and covariance. Mahalanobis distance (D) is calculated using threshold T=0.5Ʈ. Yushan et al. (Yu et al., 2016) used a method in which the image is firstly pre-processed using Haar-like features. Pixel sum and difference of particular region are considered such that the image transforms to a gray image, and then the histogram equalization process is performed.

Kaur, Anuranjan, and Nair (Kaur & Nair, 2018) present a genuine time hand gesture recognition system that detects hand gestures in midair and controls the appliance corresponding to input gestures. Real-time hand detection is done with the help of the Histogram of Oriented Gradients (HOG) feature in MATLAB. Mayyadah et al. (Mahmood et al., 2018) considered approximately 200 images and captured data images with the help of HP pavilion dv6. In the detection phase, subtraction between the Region of Interest (ROI) background and ROI hand gesture is made, and afterward, the input image is transformed into a grayscale image. Maleika et al. ([Heenaye-Mamode et al.,2019) developed an application to categorize and recognize the classical “Bharatanatyam” dance hand gestures. A customized database of these dance movements was prepared with 900 images, among which 450 were for the training set. An equal number is taken for the testing set, consisting of 15 instances for each hand gesture. For the detection process, the use of Chain Codes along with a Histogram of Oriented Gradients (HOG) was proposed.

3.2 3D model-based detection：The advantage of the 3D hand models is that they support detection, which is view independent. Dimensions of the hand in the image should be adapted using 3D models with sufficient degrees of freedom. Image features are the basic requirement for constructing feature models (Rautaray & Agrawal, 2015). Kinematic hand models employ point and line features to recover angles formed at the joints of the hand (Rehg,1995, Shimada,1998, Wu & Huang, 1999, Wuelt, 2001). There are various 3D models present in literature by researchers, along with their advantages and disadvantages. In some of the research, a framework having a deformable model is utilized to fit the hand 3D hand model to the available image (Lee & Kunii, 1995, Heap & Hogg, 1996). The image edge model gets attracted by the forces guiding the filling, and other balanced forces are available by which continuity and evenness are preserved among surface points.

3.3 Motion based detection: A few pieces of literature used methods using motion to detect hand. Motion information provides better results when combined with additional color cues and successfully distinguishes hands from other skin-colored objects (Cutler 1998, Martin et al.,1998). The system also works successfully under changes in illumination, but the camera and background require being static for easy detection of hand. 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, moving objects (hands) are detected well in a static background by choosing the appropriate threshold. A novel feature based on motion residue is proposed by Yuan et al. (Yuan, 2005). It is observed that hands are articulated objects, so they have non-rigid motion. Therefore, the hand is detected by exploiting the information that for hands, appearance among the frames changes more frequently as compared to other objects such as the face, clothes, and background. Spatiotemporal-based characteristics were generated from the video sequence (Yang, 2010). 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. (Kim, 2001) proposed combining skin filtering and motion information models. The detection phase included all the skin-colored objects like face and hand, and a suitable face detection algorithm resulted in the extraction of the face. The frame differencing was used to locate the moving objects in the surroundings. Finally, the results were combined to obtain the hand region. Aditya Ramamoorthy et al. (Chaudhurya, 2000) developed a recognition engine that recognizes dynamic gestures despite individual variations. For shape change detection, the centroid of the contour points is calculated, and the variance of the contour points from the centroid is obtained.

Table 1. Techniques Used in Detection

DETECTION		
Color based Detection	Background subtraction	Rehg et al., 1994
	HSV Color space	Saxe et al., 1996
	YCrCb color space	Chai et al., 1998
	YUV color space	Yang et al., 1998
	Dynamic correction of background models.	Utsumi et al., 1998
	Color system conversion from RGB to YIQ	Yoon et al., 2001
	Red/Green ratio to determine the skin-colored regions	Malima et al., 2006
	Adaptive Boost algorithm to detect overlapped hand	Fang et al., 2007
	YCbCr color model detects and extracts skin-colored pixels from the background	Yu et al., 2010
	RGB input was converted to HSV color space before processing	Manigandan et al., 2010
	Input RGB image to grayscale image	Tewari et al., 2012
	Maximum values of the region connected were identified using binary linked object (BLOB) analysis	Ramli et al., 2012
	A combination of two skin color models forms a vector containing color elements of H, S, Cb, and Cr for each pixel	Rahman et al., 2014
	Haar-like features like pixel sum and subtraction	Yushan et al., 2016
	Histogram of Oriented Gradients (HOG)	Kaur et al., 2018
	Difference between the ROI background and ROI hand gesture	Mahmood et al., 2018
	Combination of Chain Codes and Histogram of Oriented Gradients (HOG)	Yang et al., 2019
	RGB frames combined with hand segmentation masks	Garcia et al., 2021
3D model based Detection	Use of point and line features to recover angles formed at joints	Rehg et al., 1995
	Use of anatomical data of human hand	Lee et al., 1995
	Deformable model framework	Heap et al., 1996
	Construction of feature model correspondences	Rautaray et al., 2015
Motion based Detection	Employed changes in interframe appearance	Yuan et al., 1995
	Two successive frames pixel luminance difference	Cutler et al., 1998
	Centroid of the contour points and variance calculation	Chaudhurya et al., 2000
	Frame differencing	Kim et al., 2001
	Spatial information of hand	Bhuyan et al., 2006
	Use of spatiotemporal based characteristics	Quan et al., 2010
	Use of both color and motion features	Chen et al., 2018
	Angular-velocity method	Shantakumar et al., 202

Ratios of different variances of different shapes are calculated. It is observed that at the time of shape change, the ratio of the variance of the predicted and observed contours will be above the expected value. Based on a small area of 10*10 pixels, the mean and variance of hand color are estimated. Ghoan et al. (Bhuyan, 2006) describe a gesture recognition system that recognizes broad classes of hand gestures in a vision setup. Firstly, gestures having only one global motion are recognized with the help of spatial information of hand in the form of motion trajectory along with some static and dynamic features. The concept of object-based video abstraction is used for segmenting video frames. (Chen et al., 2018) proposed a hand-tracking method that uses the strategy of proposal selection based on temporal information, hand detection, and human pose estimation. Some of the important techniques have been categorized and shown in Table 1.

4. Tracking

The detection method can be used for tracking if it is fast enough to operate at an image acquisition frame rate. One of the most challenging tasks in a hand gesture recognition system is tracking due to the variable gesticulation speed of the users. The tracking system should be robust enough to perform well even when the hand moves quickly. Tracking provides the interconnection between the hand appearances of consecutive frames, thus generating the trajectory of a gesture. The features are extracted from this trajectory in the later stages. Tracking also maintains model parameter estimates and features that can be observed afterward (Rautaray, 2018).

4.1 Color based approach： (Guo et al., 2011) suggested a new hand-tracking system that uses skin filtering pixel-based hierarchical feature AdaBoosting and is used with background cancelation. (Koh et al., 2009) proposed a color model to track hand gestures with the help of skin. The active appearance model helps construct a hand appearance model that considers color and shape information. During the initialization of the system, Mahalanobis distance was used, which helps verify the user's hand and appearance model. The skin color model is constructed using Gaussian distribution.

Color histogram was extracted and used as the information to track an object (Comaniciu, 2003). Mixtures of Gaussians were used to develop the model for the color distribution of the object (Jepson, 2003, Zhou, 2004, McKenna, 1999). However, the drawback of this color-based technique is that it fails if there is the presence of things in the background with a similar color as that of the hand.

4.2 Probabilistic approach: In the last decade, many researchers have adopted probabilistic approaches to track hands (Binh, 2005, Imagawa, 1998, Isard, 1998, Shan, 2007, Weng, 2006, Zhang, 2009, Zheng, 2009). The blobs are computed in some literature (Binh, 2005, Imagawa, 1998), which is used for

tracking hands. The following location of the hand is predicted using the Kalman filter. The measurement noise used in the 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 a Kalman filter running in each video frame to estimate the hand postures (Utsumi 1999). Peterfreund (Peterfreund 1999) developed a robust technique to handle the cluttered background. The foreground can be separated from the background by combining the conventional image gradient with optical flow. (Asaari et al.) combined Adaptive Kalman Filter and Eigen hand features to track hands under various perplexing conditions. However, the algorithm is not successful in the presence of large-scale variations and changes in poses.

One of the methods to track the hand position is particle filters, in which the hand location is modeled with a particle set. The Condensation algorithm performs better than Kalman filters (Isard & Blake 1998). This algorithm performs well against cluttered and dynamic backgrounds. It uses "factored sampling," where a randomly generated set represents the probability distribution of possible interpretations. This algorithm uses visual observations and learned dynamical models to propagate this random set over time. (Mammen et al. 2001) extended the Condensation algorithm to detect target objects under occlusions. The same algorithm is combined with color information within a probabilistic framework by (Perez et al. 2002). This technique proposes a new Monte Carlo tracking algorithm.

(Bhuyan et al. 2006) proposed a new model-based method for tracking hand motion in complex scenes is being designed in this paper. The motion vector estimation process takes place in the type of tracking algorithm used. An object tracker forms the core of this algorithm which matches a two-dimensional binary model of the

DOI: 10.6977/IJoSI.202309_7(7).0003

gesture with subsequent frames using the Hausdorff distance measure. Frames are segmented in the gesture sequence to form object video planes where the hand is considered a video object. The hand pixels are assigned a binary value of ‘1’, and the background pixels are assigned a value of ‘0’. Then, the trajectory is estimated using the centroid of the hand being detected by the above process.

4.3 Appearance based approach：(Comaniciu et al. 2003) used a color histogram to develop a hand tracker model. Hand detection is done by calculating color histogram, which is used as mean shift and locates hands in video frames and tracking. A new type of algorithm named CamShift (Continuous adaptive mean-shift) has been proposed as an enhanced form of mean shift algorithm used for object tracking (Nadgeri 2010, Bradski 1998).

This algorithm tracks the hand efficiently in normal backgrounds, but it does not provide an accurate result when the hand occlusion occurs with other skin-colored objects. The track window’s size is adjusted by the CamShift algorithm. CamShift algorithm can track any feature distribution representing the target successfully (Bradski 2008). There are many techniques where the CamShift was combined with various other tracking methods, which led to improved tracking efficiency. For example, in literature (Wang 2010, Huang 2011), the CamShift algorithm was combined with the Kalman filter. The Kalman filter predicts the possible positions of a target, and then the CamShift is used to search and match the target in the predicted areas (Wang 2010).

(Shi and Tomasi 1994) chose high-intensity corner points as features for tracking the target object. This led to successful tracking and results, but as soon as the number of frames increased, feature points decreased. This can be due to illumination changes or changes in hand appearance. (Kolsch & Turk 2004) introduced an algorithm based on a tracker based on Kanade Lucas Tomasi (KLT). However, the KLT tracker does not yield good results at the time of hand shape change during gesticulation. (Porikli et al. 2006) proposed a tracker working on the concept of a covariance matrix (Tuzel et al.,2006), and the Riemannian manifold was used for modeling the updated mechanism. In their system, the target object was represented with a set of features as a covariance matrix. For every consecutive frame of the input video, a candidate region searched that had a 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 only respond if the target and background have a few variations. An Eigen space approach-based tracking system named Eigen tracking was developed by (Black and Jepson 1998). This tracker uses subspace constancy assumptions for estimating hand motion. This technique requires pre-training of the Eigen basis, which increases the tracking time of the system. Moreover, the Eigen basis needs to be updated. Thus the system cannot work in an environment suffering from illumination changes.

(Yushan et al. 2016) proposed a method that combines the CamShift algorithm and Haar-like feature detection. This method successfully gives output for tracking and classifying hand gestures in images acquired in a dynamic environment. During the initial stage, a Haar-like classifier is employed, which acquires the color of the hand. To track the acquired hand, Camshift, along with a 2-D Kalman filter, is used. This algorithm solves the problem of lost tracking due to hand occlusion and skin color disturbances to a great extent. A recognition rate of 99.5% is obtained by using the proposed system.

(Xiu, Su, and Pan 2018) Try to reduce problems in which accuracy decreases because of similar target color and background color or if the target is covered. The tracking algorithm is improved and is based on CamShift, which has the advantage of the Mean shift algorithm, in which the window size can be changed as the size of the target changes. The proposed algorithm starts with the Kalman filter, which tracks the target and stores its motion information by prediction method. The Bhattacharya coefficient is calculated using the image histogram features. If any target occlusion exists, the Kalman filtering algorithm will repeat itself and predict the target motion in the next frame. After this process, CamShift will try to find the target near the expected location by the probability distribution map of the image. The target's position can be accurately located by the size of the tracking window, which should exceed the threshold value to calculate the perfect target position. Fig. . 3 shows examples of some of the video gestures formed using tracking techniques.

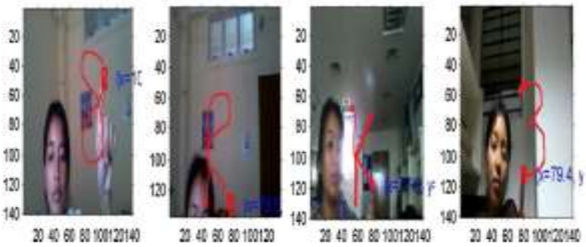


Fig. 3. Tracking of some of the gestures

The problems in the previous two steps are resolved by improved CamShift, which means shift operation is used as a tracking principle. The last frame is used for the window position search of the next frame. If the background is, the same tracking fails as the background starts acting as a target. With the changes made in CamShift Algorithm, the background is changed due to the available value of thresholds for the camshaft window, which decides the background interference problem. If background interference is detected, the target contour is extracted by three steps:

- The image should be binary.
- Noise interference should be reduced.
- Canny edge detection is used for target detection.

(Chen and Zhu 2018) designed a hand tracker with the self-correcting capability to re-initialize the tracker position by integrating the human pose and hand detection information. Hand positions and an approximate center of the human body in the starting frame are used for initialization to give a perfect starting point. Hand tracker uses contour points and Harris corners, a pixel-based skin detection method to recover the tracked hand in subsequent frames based on information from the hand detector and wrist position estimator. Table 2 demonstrates the evolution of tracking methodologies used in Hand Gesture recognition systems.

Table 2. Evolution of tracking methodologies used in Hand Gesture Recognition

TRACKING		
Color based approach	Gaussian distribution and Color Histogram	Comaniciu et al., 2003
	Use of color distribution of the object	Jepson et al., 2003
	Active Appearance Model and Mahalanobis Distance	Koh et al., 2009
	Skin filtering pixel based hierarchical feature Ada-Boosting	Guo et al., 2012
Probabilistic approach	Kalman filter running in each frame of video	Utsumi et al., 1999
	Optical flow along with image gradient	Peterfreund et al., 1999
	Condensation algorithms	Isard et al., 1998
	Extended condensation algorithm	Mammen et al., 2001
	Condensation algorithm integrated with color information	Perez et al., 2002
	Computation of blobs and Kalman Filter	Binh et al., 2005
	Motion vector estimation	Bhuyan et al., 2006
	Adaptive Kalman Filter along with Eigen feature tracking	Asaari et al., 2015
Appearance based approach	High intensity with corner points is being selected as features for tracking the target object	Shi et al., 1994
	Camshift Algorithm	Bradski et al., 1998
	An Eigen space approach based tracking system	Black et al., 1998
	KLT tracker	Kolsch et al., 2004
	Covariance matrix representation	Porikli et al., 2006
	Riemannian manifold	Tuzel et al., 2006
	Adaptive mean shift	Nadgeri et al., 2010
	CamShift combined with kalman Filter	Xiangyu et al., 2010
	CamShift and Haar-like feature detection	Yushan et al., 2016
	Hand tracker having self-correcting capability that can re-initialize the tracker position	Chen et al., 2018

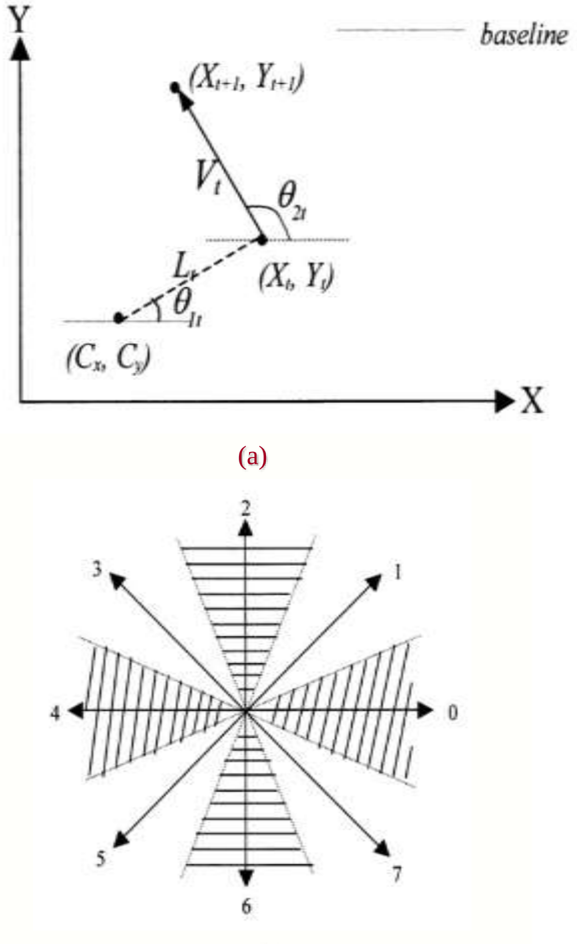
5. Feature Extraction

A proper feature matrix consisting of robust features is required for improved recognition. A few papers used a single feature to develop the gesture recognition system. (Elemicin *et al.2008*) tried to recognize both isolated and continuous gestures with the help of the orientation feature. With the help of this feature, gesture motion direction can be calculated using trajectory points. The quantization process used code words from 1 to 18 on the orientation angle. (Kao and Fahn 2011) also used the orientation feature to design a real-time hand gesture recognition system. Gestures were classified using HMM after the quantization process.

Location, orientation, and velocity features were among the most used features by researchers (Bhuyan 2008, Bhuyan 2014, Li 2016). (Xu *et al. 2015*) proposed a novel hand gesture recognition system for robotic applications using features like orientation, chain code by 1-8 code words, location, and velocity. (Yoon *et al.* 2001) used a combination of location, orientation, and velocity, as shown in Fig. .   4. The orientation feature calculates the direction of gesture motion from the center to all gesture points in the trajectory and between the trajectory points. (Elmezain *et al.* 2009) proposed the use of two location features and three orientation features which were combined with velocity features.

Many researchers have tried to improve the system's performance by combining multiple features. (Bhuyan *et al. 2006*) used static and dynamic features to construct the feature matrix, which helps recognize the gestures. A few static features include trajectory point selection, location, trajectory length, orientation, and location features. The number of significant curves start, and end of the gesture trajectory forms the orientation feature. The dynamic features include the velocity and acceleration features. Fig. . ures 4 and 5 show samples of some of the features extracted and used in the feature extraction process.

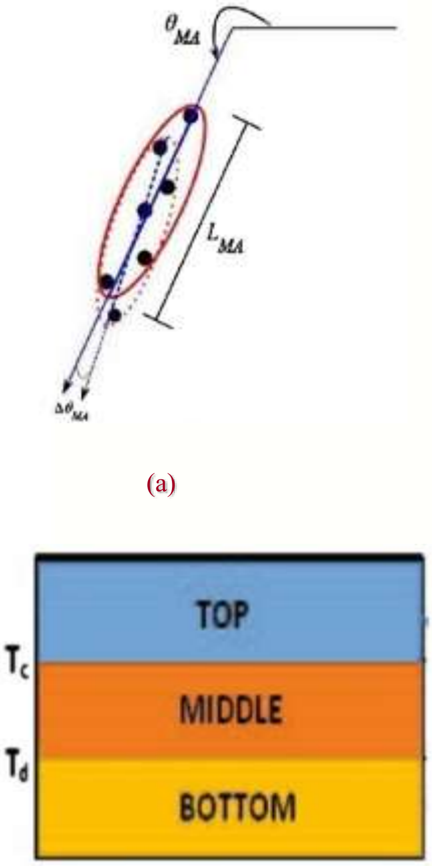
(Bhuyan *et al.* 2008) increased features to be included in the feature matrix they used (Bhuyan 2006), like the standard deviation of the speed feature for gesture recognition. A conditional Random Fields (CRF) based classifier model was proposed, which helped recognize continuous hand gestures (Bhuyan 2014). In Fig. .   5, a technique was used in which the ellipse was adjusted over every six trajectory points named as ellipse fitting technique. From the ellipse, features such as orientation and major axis length of all ellipses were extracted. The positioning feature was extracted where the start and end position of the gesture trajectory was found and divided into top, middle, and bottom horizontal sections. For isolated gestures, 96% recognition accuracy was achieved. A combination of two types of features: hand shape and hand direction, was used by (Li *et al.* 2016). Hand shape includes the distance between the fingers of the hand, and hand direction feature includes acceleration, velocity, and orientation features



(a)

(b)

Fig. 4. Feature extraction (a) three features: location (Lt), orientation (θ1t, θ2t), and velocity (Vt) (b) chain code

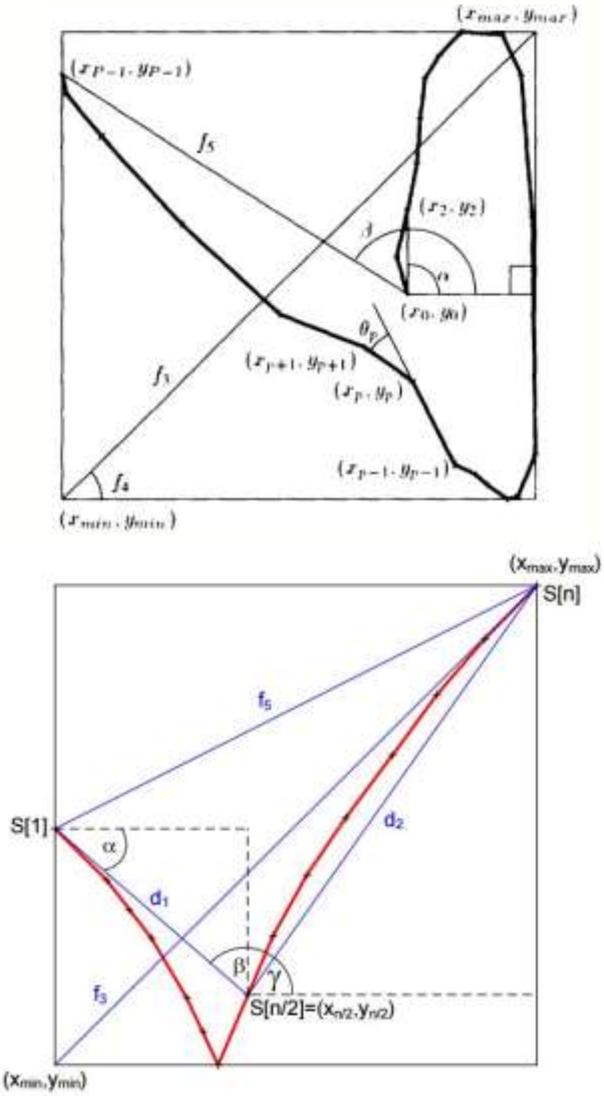


(a)

(b)

Fig. 5. Feature extraction (a) orientation of the major axis of the ellipse (b) region to indicate the position of the hand

Rubine developed 13 features (Rubine 1991), as shown in Fig. . 6(a). The features are: initial angle sine and cosine along the x-axis (cosα, sinα), bounding box length and angle (f3 and f4), first and last point distance (f5), total traversed angle and gesture length, sine-cosine angle of the angle between first and last point (cosβ, sinβ) and duration of the stroke. Features of Rubine 1991 were extended by (Signer *et al.* 2011), and 11 new features were added to the feature list provided by Rubine. Some of these are the direction of the first and second half of the stroke (sinα and cosγ), the number of breakpoints available, distance from the start to center point to the diagonal (d1/f3), distance from the beginning to the end-point to diagonal (f5/f3), the total number and total length of gestures and stroke distance from each other along with straightness as shown in Fig. . 6(b).



(a) (b)

Fig. 6. (a) Rubine features [(b) E-Rubine features

Another important extracted feature is stated as the start of the dynamic gesture or end of the same, which is determined by multiple centroids overlapping each other, stating the stationary position of the hand. Another feature is the coordinate vector of the centroid, which is associated with each gesture, especially alphabetical letters that indicate a timing sequence starting with 1. According to the thumb rule, 5-second gestures have been used where a complete gesture is recorded in 5-12 s including all hand positions at 30 frames per second.

Gesture recognition faces problems like occlusion, and to solve this problem, a new feature Comp-LOP, has been developed (Cen 2017). The new feature extracted is a complex form-local orientation plane (Comp-LOP) based on complex form. This feature provides a more elaborated version of the orientation using the complex form with an angle of 45°. It is better than a histogram of oriented gradients and local binary patterns. It reduces the occlusion problem as the representation of orientation pixels gives a precise direct relationship between them. (Tang *et al*. 2018) used the corner point positions of arbitrary trajectories as structured information and assigned different weights to feature space. Position and orientation features are combined with a Dynamic time warping approach to make trajectories more discriminative.

(Singla *et al.2019*) have taken the normalized sequence of captured 3D space coordinates as input, and the sequence of features is computed along the trajectory. Gesture direction, curvature, aspect, curliness, slope, and lines are some of the features that have been calculated and used to develop feature space and recognition. (Misra *et al.2019*) presented novel spatiotemporal trajectory features that provide output as the gestures' structural values. These features include Area of two-halves (ATH), Local geometrical area ratios (LGAR), and curve-area features (CAF). The gesture is divided into two halves equally, and the Area of each half gives the output as an ATH feature. For calculating the LGAR feature, the ratio between enclosed areas in each case of the stroke length is measured. The Area between them is the starting point, and the first point representing a sharp transition is measured as CAF. Some of the features used in existing literature have been tabulated in Table 3.

Table 3. Various features used in the Feature Extraction process

FEATURES	
Initial angle sine and cosine along the x-axis (cosα, sinα), bounding box length and angle (f3 and f4), first and last point distance (f5), total traversed angle and gesture length, sine-cosine angle of angle between first and last point (cosβ, sinβ) and duration of the stroke	Rubine et al., 1991
Location, orientation, and velocity feature combination	Yoon et al., 2001
The direction of the first and second half of the stroke (sinαand cosγ), number of breakpoints available, distance from the start to centre point concerning the diagonal (d1/f3), distance from start to the end point with respect to diagonal (f5/f3), the total number and total distance of gestures and stroke distance.	Signer et al., 2007
Orientation Feature	Elmezain et al., 2008
Length of trajectory, selection of trajectory points, location, orientation features; significant curves and orientation of start and end of the gesture trajectory, Standard deviation of the speed feature	Bhuyan et al., 2008
Distance between centre to trajectory points and start point to trajectory points	Al-Hamadi et al., 2009
Quantized orientation feature	Kao et al., 2011
Orientation between consecutive trajectory points (1-8 code words), location, and velocity	Xu et al., 2014
Orientation of the major axis of each ellipse, major axis length of each ellipse	Bhuyan et al., 2014
Distance between the fingers of the hand; acceleration, velocity, and orientation features	Li et al., 2016
Centroid coordinate vector	Prashan et al., 2017
Complex form-local orientation plane	Cen et al., 2017
Corner point positions of arbitrary trajectories	Tang et al., 2018
Gesture direction, curvature, aspect, curliness, slope, and linearity.	Singla et al., 2019
Area of two-halves, Local geometrical area ratios, and curve-area features	Misra et al. 2019
Euclidean Distance, Instantaneous velocity, and polarity	Yadav et al., 2021
Shape matching, Velocity change, Displacement of centroid between successive frames	Choudhary et al., 2021

6. Modelling And Recognition of Gestures

Researchers have worked and addressed recognition techniques which are as follows a) Hidden Markov Model (HMM) b) Neural Networks c) SVM d) k-NN e) Finite State Machine (FSM) f) Classifier fusion g) Naïve Bayes and Extreme Machine Learning (ELM) Neuro-Fuzzy (NF) and Voronoi based Classifier.

6.1 Hidden Markov Model：After the introduction of HMM in the early 1990s quickly became one of the most widely used recognition methods due to its inherent solution to the segmentation problem. In the HMM model, Markov chains are used simply as finite-state automata with a probability value associated with each arc (Rabiner 1986). Circular segment likelihood estimations leave a single state whole to one state. A Markov chain without the above Markov chain restriction is an HMM (Charniak 1993). HMMs are non-deterministic as the same output symbol represents more than one arc. An HMM can be defined as a combination of states which comprises the initial state, output symbols, and state transition.

With regards to hand motion acknowledgment, each state represents a lot of conceivable hand positions. (Chen *et al.2003*) distinguished the best probability signal model using HMM-based recognizers. Patterns having fewer likelihood values were filtered out using an HMM-based threshold model, and the hand movement direction was used for representing the sequences of gestures.

(Marcel *et al.2000*) extended the HMM algorithm, which resulted in the development of the Input/output Hidden Markov Model (IOHMM) used for hand gesture recognition. The IOHMM uses supervised discrimination learning with input/output sequences, observations, and gesture classes. IOHMM directly models posterior probabilities as compared to the HMMs. However, they performed the experiments for only binary classes. (Just and Marcel 2009) performed experiments for an extensive database with 7 to 16 classes. The study resulted in the recognition of all types of hand gestures. A comparative analysis of HMM and IOHMM established that for a large number of classes, HMM performed better than IOHMM.

Conditional Random Fields (CRF) is a widely used tool nowadays. It is advantageous compared to HMM because CRF does not consider solid independent assumptions about the observations and can be trained with fewer samples than HMM (Yang 2006). (Beh *et al.2014*) used HMM as the classifier. The hand motion trajectories are a combination of straight and curved segments. A state-splitting algorithm was proposed by (Siddiqi *et al.2007*) based on an expectation-maximization algorithm. A method was developed by (Ulas and Yildiz 2009) to find the optimum structure of HMM. For this purpose, they increased the number of states, subsequently measured the possible values, and tried to find the optimal structure.

6.2 Neural Networks： Neural Networks are upcoming classifier models that provide good pattern recognition results (Bishop 1995 [98], Haykin 2009, Bamwenda 2019 [100]). Gradient features per alphabet from the hand gesture images can be trained, tested, and validated using Artificial Neural Networks (ANN) (Bamwenda 2019 [100]). Artificial Neural Networks (ANN) refer to the simulations performed on the computer to complete several machine learning tasks such as pattern recognition, clustering, and classification. The time delay neural networks (TDNN) are one of the latest available models, which are counterfeit neural systems working with constant information making the engineering versatile to online systems and henceforth favorable to continuous applications. For 2D motion trajectories, TDNN networks have been used (Suykens 1999, Gopalan 2009). TDNN supports dynamic order as there is a little window of the information movement design. A process based on a double-channel convolutional neural network (DC-CNN) is suggested to improve the recognition rate (Mohammed 2011). The steps include the preprocessing, removal of noise and edge location of images to obtain hand-edge pictures. CNN separates the hand motion and edge pictures and denotes them as two different information channels. Each channel has a different weight but the same number of convolutional layers. Lastly, all the features are fused, and the Softmax classifier classifies the output.

6.3 Support Vector Machine：SVM is a supervised learning model in which optimization of class separation hyperplane such that there is the maximum distance between the hyperplane and the available pattern. The class separation hyperplane is optimized to maximize the space between the pattern and the hyperplane separating the classes (Dominio 2014, Thirumuruganathan 2010). (Gopalan and Dariush 2009)

cropped the regions corresponding to the extracted skin-colored pixels. Some of the extracted features include distance and angle, which utilize contour points by connecting each other through an inner shape distance context algorithm, leading to the recognition of gesture by SVM. In the testing stage, the detected hand gesture is classified by a multiclass SVM classifier. A recognition rate of 96.23% under diverse, challenging environments like variable illumination, dynamic background, etc., was obtained.

A Library of SVM (LIBSVM) was used to recognize hand gestures (Dominio 2014). The parameters used for SVM were Radial Basis Function (RBF) as the kernel function. A grid search approach and cross-validation on the training set tuned the other classifier parameters. (Parama Sridevi *et al.* 2018) presented a new sign language interpreter which verbalizes American Sign Language. Features of real-time video sequences of hand gestures were compared with the stored features of database images for better accuracy. MATLAB is used for generating output and depends on predicting the values representing the highest resemblance. This model also helps fill the communication gap between speaking and hearing-impaired people and those without them. For classification purposes, Quadratic SVM is used, which provides about 85% accuracy.

6.4 K-Nearest Neighbour：This method classifies objects based on feature space training examples. K-Nearest Neighbor (k-NN) is a kind of instance-based learning in which the function is approximated locally, and all computations are delayed until classification (Ge 2008, Oka 2002). This classifier solves classification and regression problems using supervised machine-learning methods. This algorithm assumes that similar things are available in close proximity and thus captures the idea of similarity. Training and testing have been done for different values of K. The k-NN algorithm must be run several times to select the correct and appropriate value of K. The value of K, which reduces the number of errors while making predictions accurately, is selected. The value of K is chosen to be odd if the numbers of classes are odd to avoid the situation of the draw of votes.

The maximum vote of its neighbors characterizes an object, and the item is allotted to the class that belongs to its k closest neighbors where k is a positive number. If k=1, an object is assigned to its nearest neighbor class. Relapse also uses a similar technique and allows the property estimation for the item to the normal estimations of k closest neighbors. The neighbors are taken from items of the correct order. Euclidean separation can be used to distinguish neighbors and leads to the nearby structure of the information.

6.5 Finite State Machine：FSM is a technique that consists of a finite number of possible states. It can help develop a tool for solving problems and describing solutions for developers and maintainers. The gestures were decomposed into four fixed-order distinct phases resulting in the development of an FSM model for classification (Davis 1994). A sequential signature of hand motion is extracted, after which the hand gestures are classified using an FSM (Yeasin 2000). The dominant motion was estimated from an image sequence using motion energy. The FSM model was developed using the positions of the user's hand and head centers (Hong 2000). For the recognition of a continuous hand gesture recognition system, features like hand motion chain codes, the relation between the two hands according to their position, and the relation between face and hands were given as input to the dynamic Bayesian network model (Suk 2010).

6.6 Classifier Fusion：Recognition accuracy obtained by traditional individual classifiers can be improved using classifier combining techniques (Thai 2012, Kang 2009). (Dinh et al. 2006) developed a hand gesture recognition system that used a cascade of classifiers trained by AdaBoost and Harr wavelet coefficient features. (Burger et al. 2008) recognized hand shapes by proposing a belief-based method for SVM fusion. This method outperforms the classical methods by reducing the mistakes by 1/5.

A combination of HMM and Recurrent Neural Networks (RNN) was used by (Ng *et al.2002*), which provided improved performance compared to the performance of the individual classifier, such as HMM or RNN. Features used are based on Fourier descriptors and act as input to the RBF network. HMM, and RNN take motion information and pose likelihood vector from the RBF network as input. The final result is obtained from the combination of the classifiers' outputs. A combination of AdaBoost and rotation forest was used by (Wang *et al.2012*) for the recognition of hand gestures. Improved performance of the fusion technique is being observed. A combination of HMM and ANN models is being proposed and provides better results by improving the accuracy by 2-3% (Corradini 2002).

6.7 Naïve Bayes and ELM： (Singha et al.2016 [122]) classified dynamic gestures using the Naïve Bayes classifier based on the Bayes theorem, which operates independently between the features. Assuming the feature vector represented by x = [xa… xn] T and the class one of c classes w1…wc. For the model proposed, n=40 and c=40. Minimum classification error is assured if the class with the largest posterior probability P is decided upon. Kernel and multivariate multinomial distribution (MVMN) are the two distributions that can be used in the training phase to obtain the highest accuracy model. A cross-validation process using 5-fold was also performed for testing. Extreme Learning Machine (ELM) based classification uses the feed-forward neural networks and nonlinear mappings that use a gradient descent approach for weights and bias optimization (Mohammed 2011, Liu 2016, Chen 2015).

6.8 Neuro-fuzzy classifier (NF) and Voronoi diagram-based classifier (VDBC)： A new type of classification model named Voronoi diagram-based classifier (VDBC) and neuro-fuzzy (NF) classifier was proposed, and the accuracy was improved (Misra 2019). Multiple layers like fuzzy membership, fuzzification, defuzzification, normalization, and output are present in the neuro-fuzzy classifiers. NF model is developed by the linguistic hedges (LHs) formed by the fuzzy sets. VDBC uses the ad-hoc approach of classification in which all the classes are handled at the same time. VDBC is designed using the Voronoi diagram model in which the training space is divided into multiple regions with a seed set which is S = s1, s2… sn, also known as discriminative functions.

(Yang and Liu 2019) have tried to improve the recognition accuracy by introducing an online classifier that adjusts each feature value in the tracking target model in accordance with the object change and situation. The learning algorithm produces a sequence of classifiers, which are F= (f_1……f_T), where the video image frame becomes accessible one frame by one frame. Using the first frame information, f_1 is trained, and f_i (for $i > 1$) is the i-th classifier learning after seeing the i-th frame image. Table 4 shows different classifiers used in the recognition methods of hand gesture recognition systems.

Table 4. Different Classifiers used in the recognition method

RECOGNITION		
HMM	Use of Markov Chain having probabilistic value	Rabiner et al., 1986
	HMM without Markov chain restriction	Charniak, 1993
	HMM based threshold model	Lee et al., 1999
	Input/output Hidden Markov Model (IOHMM)	Marcel et al., 2000
	HMM based recognizers for identifying the best likelihood model	Chen et al., 2003
	Conditional Random Fields (CRF)	Yang et al., 2006
	Expectation-maximization algorithm	Siddiqui et al., 2007
	Comparison of HMM and IOHMM model	Just et al., 2009
	Optimum HMM by incrementing the number of states	Ulas et al., 2009
	Use of HMM as a classifier	Beh et al., 2014
Neural Networks	Dynamic TDNN	Yang et al., 1998
	Time delay neural networks	Ahuja et al., 2002
	Artificial Neural Networks	Bamwenda et al., 2019
	Double Channel CNN	Wu et al., 2019
Support Vector Machine	Multiclass SVM classifier	Suykens et al., 1999
	Class separation hyper plane	Hsu et al., 2002
	Library of SVM (LIBSVM)	Dominio et al., 2014
	Quadratic SVM	Sridevi et al., 2018[121]
k-NN	k-Nearest Neighbor (k-NN) (Lazy or instance-based learning)	Thirumuruganathan, 2010
FSM	FSM model by decomposing the gestures	Davis et al., 1994
	FSM model by temporal signature of hand motion	Yeasin et al., 2000
	FSM model with the help of the position of centers of user's hand	Hong et al., 2000
	Dynamic Bayesian network model	Suk et al., 2010
CLASSIFIER FUSION	HMM and ANN based models	Corradini, 2001
	Fusion of HMM and Recurrent Neural Networks (RNN)	Ngand et al., 2002
	Boosted cascade of classifiers trained by AdaBoost and informative Haar wavelet coefficients	Dinh et al., 2006
	SVM fusion (Belief based)	Burger et al., 2008
	Traditional Single classifier	Hai et al., 2012

	AdaBoost and rotation forest fusion	Wang et al., 2012
	Naïve Bayes and ELM	Singha et al., 2016
	NF and VDBC classifiers	Misra et al., 2019

7. Conclusions

Hand gesture recognition has great potential to extend the application in contactless Human-computer interaction (HCI), which is currently done with the help of keyboards, mice, or joysticks. To increase flexibility in the application for disabled people, these systems provide a good and developing platform. HCI systems will be easy to use; user-friendly and can provide easy access to a vast range of applications.

The initial step in any hand gesture recognition is to detect and extract the hand from the background. There are various difficulties during this phase, such as a cluttered background, illumination problems, occlusion, etc. The second problem is the tracking of the hand. To achieve the correct gesture trajectory, the hand must be tracked correctly in every video frame. This phase is affected by different scenarios like varying gesticulation speed and pattern.

The third problem is detecting and removing unwanted hand movements, which may be intentional (self-co-articulation) or unintentional (hand trembling). Detecting these unwanted strokes will make the system easier to recognize. The fourth problem is to develop a robust feature set for the system. The last issue is to develop a system that should be able to recognize a continuous sequence of data that are connected by self-co-articulation, movement epenthesis, and other unwanted hand movements.

Multiple researchers worked to solve these issues differently. This paper briefly surveys most of the significant work carried out in 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. This survey has tried identifying more than one hundred and thirty research publications. A lot of potential is present in the hand gesture recognition system inspiring the researchers to design efficient and accurate gesture recognition systems.

References

Bamwenda, J., & Özerdem, M. S. (2019). Static hand gesture recognition system using artificial neural networks and support vector machine. Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 10(2), 561-568.

Beh, J., Han, D., & Ko, H. (2014). Rule-based trajectory segmentation for modeling hand motion trajectory. Pattern Recognition, 47(4), 1586-1601.

Benitez-Garcia, G., Prudente-Tixteco, L., Castro-Madrid, L. C., Toscano-Medina, R., Olivares-Mercado, J., Sanchez-Perez, G., & Villalba, L. J. G. (2021). Improving real-time hand gesture recognition with semantic segmentation. Sensors, 21(2), 356.

Bhuyan, M. K., Ajay Kumar, D., MacDorman, K. F., & Iwahori, Y. (2014). A novel set of features for continuous hand gesture recognition. Journal on Multimodal User Interfaces, 8, 333-343.

Bhuyan, M. K., Bora, P. K., & Ghosh, D. (2008). Trajectory guided recognition of hand gestures having only global motions. World Academy of science, engineering, and technology, 21, 753-764.

Bhuyan, M. K., Ghoah, D., & Bora, P. K. (2006, September). A framework for hand gesture recognition with applications to sign language. In 2006 Annual IEEE India Conference (pp. 1-6). IEEE.

Bhuyan, M. K., Ghosh, D., & Bora, P. K. (2006, June). Feature extraction from 2D gesture trajectory in dynamic hand gesture recognition. In 2006 IEEE Conference on Cybernetics and Intelligent Systems (pp. 1-6). IEEE.

Binh, N. D., Shuichi, E., & Ejima, T. (2005). Real-time hand tracking and gesture recognition system. Proc. GVIP, 19-21.

Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford University Press.

Black, M. J., & Jepson, A. D. (1998). Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. International Journal of Computer Vision, 26, 63-84.

Blake, A., North, B., & Isard, M. (1998). Learning multi-class dynamics. Advances in neural information processing systems, 11.

Bradski, G. R. (1998, October). Real time face and object tracking as a component of a perceptual user interface. In Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No. 98EX201) (pp. 214-219). IEEE.

Bradski, G., & Kaehler, A. (2008). Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc.".

Burger, T., Aran, O., Urankar, A., Caplier, A., & Akarun, L. (2008). A Dempster-Shafer theory based combination of classifiers for hand gesture recognition. In Computer Vision and Computer Graphics. Theory and Applications: International Conference VISIGRAPP 2007, Barcelona, Spain, March 8-11, 2007. Revised Selected Papers (pp. 137-150). Springer Berlin Heidelberg.

Cen, M., & Jung, C. (2017). Complex form of local orientation plane for visual object tracking. IEEE Access, 5, 21597-21604.

Chai, D., & Ngan, K. N. (1998, April). Locating facial region of a head-and-shoulders color image. In Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition (pp. 124-129). IEEE.

Charniak, E. (1993). Statistical language learning MIT Press. Google Scholar Google Scholar Digital Library Digital Library.

Chaudhurya, S., Banerjeeb, S., Ramamoorthya, A., & Vaswania, N. (2000). Recognition of dynamic hand gestures. The Journal of The Pattern Recognition Society. Department of Electrical Engineering and Department of Computer Science Engineering, IIT Delhi.

Chen, F. S., Fu, C. M., & Huang, C. L. (2003). Hand gesture recognition using a real-time tracking method and hidden Markov models. Image and vision computing, 21(8), 745-758.

Chen, Q., & Zhu, F. (2018, July). Long term hand tracking with proposal selection. In 2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (pp. 1-6). IEEE.

Chen, X., & Koskela, M. (2015). Skeleton-based action recognition with extreme learning machines. Neurocomputing, 149, 387-396.

Choudhury, A., Talukdar, A. K., Sarma, K. K., & Bhuyan, M. K. (2021). An adaptive thresholding-based movement epenthesis detection technique using hybrid feature set for continuous fingerspelling recognition. SN Computer Science, 2, 1-21.

Comaniciu, D., Ramesh, V., & Meer, P. (2003). Kernel-based object tracking. IEEE Transactions on pattern analysis and machine intelligence, 25(5), 564-577.

Corradini, A. (2002, May). Real-time gesture recognition by means of hybrid recognizers. In Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop, GW 2001 London, UK, April 18–20, 2001 Revised Papers (pp. 34-47). Berlin, Heidelberg: Springer Berlin Heidelberg.

Cutler, R., & Turk, M. (1998, April). View-based interpretation of real-time optical flow for gesture recognition. In Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition (pp. 416-421). IEEE.

Dardas, N. H., & Georganas, N. D. (2011). Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. IEEE Transactions on Instrumentation and Measurement, 60(11), 3592-3607.

Davis, J., & Shah, M. (1994). Recognizing hand gestures. In Computer Vision—ECCV'94: Third European Conference on Computer Vision Stockholm, Sweden, May 2–6, 1994 Proceedings, Volume I 3 (pp. 331-340). Springer Berlin Heidelberg.

Dinh, T. B., Dang, V. B., Duong, D. A., Nguyen, T. T., & Le, D. D. (2006, February). Hand gesture classification using boosted cascade of classifiers. In 2006 International Conference onResearch, Innovation and Vision for the Future (pp. 139-144). IEEE.

Dominio, F., Donadeo, M., & Zanuttigh, P. (2014). Combining multiple depth-based descriptors for hand gesture recognition. Pattern Recognition Letters, 50, 101-111.

Elmezain, M., Al-Hamadi, A., & Michaelis, B. (2009). Hand gesture recognition based on combined features extraction. International Journal of Electrical and Computer Engineering, 3(12), 2389-2394.

Elmezain, M., Al-Hamadi, A., Appenrodt, J., & Michaelis, B. (2008, December). A hidden Markov model-based continuous gesture recognition system for hand motion trajectory. In 2008 19th International Conference on Pattern Recognition (pp. 1-4). IEEE.

Fang, Y., Wang, K., Cheng, J., & Lu, H. (2007, July). A real-time hand gesture recognition method. In 2007

IEEE International Conference on Multimedia and Expo (pp. 995-998). IEEE.

Ge, S. S., Yang, Y., & Lee, T. H. (2008). Hand gesture recognition and tracking based on distributed locally linear embedding. Image and Vision Computing, 26(12), 1607-1620.

Gopalan, R., & Dariush, B. (2009, October). Toward a vision based hand gesture interface for robotic grasping. In 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 1452-1459). IEEE.

Guo, J. M., Liu, Y. F., Chang, C. H., & Nguyen, H. S. (2011). Improved hand tracking system. IEEE Transactions on Circuits and Systems for Video Technology, 22(5), 693-701.

Haykin, S. (2009). Neural networks and learning machines, 3/E. Pearson Education India.

Heap, T., & Hogg, D. (1996, October). Towards 3D hand tracking using a deformable model. In Proceedings of the Second International Conference on Automatic Face and Gesture Recognition (pp. 140-145). IEEE.

Heenaye-Mamode Khan, M., Ittoo, N., & Coder, B. K. (2019). Hand Gestures Categorisation and Recognition. In Information Systems Design and Intelligent Applications: Proceedings of Fifth International Conference INDIA 2018 Volume 2 (pp. 295-304). Springer Singapore.

Hong, P., Turk, M., & Huang, T. S. (2000, March). Gesture modeling and recognition using finite state machines. In Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580) (pp. 410-415). IEEE.

Hsieh, C. T., Yeh, C. H., Hung, K. M., Chen, L. M., & Ke, C. Y. (2012, June). A real time hand gesture recognition system based on DFT and SVM. In 2012 8th International Conference on Information Science and Digital Content Technology (ICIDT2012) (Vol. 3, pp. 490-494). IEEE.

Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. IEEE Transactions on Neural Networks, 13(2), 415-425.

Huang, S., & Hong, J. (2011, April). Moving object tracking system based on camshift and Kalman filter. In 2011 International Conference on Consumer Electronics, Communications and Networks (CECNet) (pp. 1423-1426). IEEE.

Imagawa, K., Lu, S., & Igi, S. (1998, April). Color-based hands tracking system for sign language recognition. In Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition (pp. 462-467). IEEE.

Isard, M., & Blake, A. (1998). CONDENSATION--conditional density propagation for visual tracking. International journal of computer vision, 29(1), 5.

Isard, M., & Blake, A. (1998, January). A mixed-state condensation tracker with automatic model-switching. In Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271) (pp. 107-112). IEEE.

Jepson, A. D., Fleet, D. J., & El-Maraghi, T. F. (2003). Robust online appearance models for visual tracking. IEEE transactions on pattern analysis and machine intelligence, 25(10), 1296-1311.

Just, A., & Marcel, S. (2009). A comparative study of two state-of-the-art sequence processing techniques for hand gesture recognition. Computer Vision and Image Understanding, 113(4), 532-543.

Kang, S., & Park, S. (2009). A fusion neural network classifier for image classification. Pattern Recognition Letters, 30(9), 789-793.

Kao, C. Y., & Fahn, C. S. (2011). A human-machine interaction technique: hand gesture recognition based on hidden Markov models with trajectory of hand motion. Procedia Engineering, 15, 3739-3743.

Kaur, S., & Nair, N. (2018, January). Electronic Device Control Using Hand Gesture Recognition System for Differently Abled. In 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 371-375). IEEE.

Kim, I. C., & Chien, S. I. (2001). Analysis of 3D hand trajectory gestures using stroke-based composite hidden Markov models. Applied Intelligence, 15, 131-143.

Koh, E., Won, J., & Bae, C. (2009, May). On-premise skin color modeling method for vision-based hand tracking. In 2009 IEEE 13th International Symposium on consumer electronics (pp. 908-909). IEEE.

Kolsch, M., & Turk, M. (2004, June). Fast 2D hand tracking with flocks of features and multi-cue integration. In 2004 Conference on Computer Vision and Pattern Recognition Workshop (pp. 158-158). IEEE.

Lee, H. K., & Kim, J. H. (1999). An HMM-based threshold model approach for gesture recognition. IEEE Transactions on pattern analysis and machine intelligence, 21(10), 961-973.

Lee, J., & Kunii, T. L. (1995). Model-based analysis of hand posture. IEEE Computer Graphics and Applications, 15(5), 77-86.

Li, K., Zhou, Z., & Lee, C. H. (2016). Sign transition modeling and a scalable solution to continuous sign language recognition for real-world applications. ACM Transactions on Accessible Computing (TACCESS), 8(2), 1-23.

Liu, H., Yu, L., Wang, W., & Sun, F. (2016). Extreme learning machine for time sequence classification. Neurocomputing, 174, 322-330.

Mahmood, M. R., Abdulazeez, A. M., & Orman, Z. (2018, October). Dynamic hand gesture recognition system for Kurdish sign language using two lines of features. In 2018 International Conference on Advanced Science and Engineering (ICOASE) (pp. 42-47). IEEE.

Malima, A. K., Özgür, E., & Çetin, M. (2006). A fast algorithm for vision-based hand gesture recognition for robot control.

Mammen, J. P., Chaudhuri, S., & Agrawal, T. (2001, September). Simultaneous Tracking of Both Hands by Estimation of Erroneous Observations. In BMVC (pp. 1-10).

Manigandan, M., & Jackin, I. M. (2010, June). Wireless vision based mobile robot control using hand gesture recognition through perceptual color space. In 2010 International Conference on Advances in Computer Engineering (pp. 95-99). IEEE.

Marcel, S., Bernier, O., Viallet, J. E., & Collobert, D. (2000, March). Hand gesture recognition using input-output hidden Markov models. In proceedings fourth IEEE International Conference on automatic face and gesture recognition (Cat. No. PR00580) (pp. 456-461). IEEE.

Martin, J., Devin, V., & Crowley, J. L. (1998, April). Active hand tracking. In Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition (pp. 573-578). IEEE.

McKenna, S. J., Raja, Y., & Gong, S. (1999). Tracking colour objects using adaptive mixture models. Image and vision computing, 17(3-4), 225-231.

Misra, S., & Laskar, R. H. (2019). Development of a hierarchical dynamic keyboard character recognition system using trajectory features and scale-invariant holistic modeling of characters. Journal of Ambient Intelligence and Humanized Computing, 10(12), 4901-4923.

Mistry, P., Maes, P., & Chang, L. (2009). WUW-wear Ur world: a wearable gestural interface. In CHI'09 extended abstracts on Human factors in computing systems (pp. 4111-4116).

Mohammed, A. A., Minhas, R., Wu, Q. J., & Sid-Ahmed, M. A. (2011). Human face recognition based on multidimensional PCA and extreme learning machines. Pattern recognition, 44(10-11), 2588-2597.

Mohd Asaari, M. S., Rosdi, B. A., & Suandi, S. A. (2015). Adaptive Kalman Filter Incorporated Eigenhand (AKFIE) for real-time hand tracking system. Multimedia Tools and Applications, 74, 9231-9257.

Nadgeri, S. M., Sawarkar, S. D., & Gawande, A. D. (2010, November). Hand gesture recognition using CAMSHIFT algorithm. In 2010 3rd International Conference on Emerging Trends in Engineering and Technology (pp. 37-41). IEEE.

Ng, C. W., & Ranganath, S. (2002). Real-time gesture recognition system and application. Image and Vision Computing, 20(13-14), 993-1007.

Oka, K., Sato, Y., & Koike, H. (2002). Real-time fingertip tracking and gesture recognition. IEEE Computer Graphics and Applications, 22(6), 64-71.

Panwar, M., & Mehra, P. S. (2011, November). Hand gesture recognition for human computer interaction. In 2011 International Conference on Image Information Processing (pp. 1-7). IEEE.

Pérez, P., Hue, C., Vermaak, J., & Gangnet, M. (2002). Color-based probabilistic tracking. In Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part I 7 (pp. 661-675). Springer Berlin Heidelberg.

Peterfreund, N. (1999). Robust tracking of position and velocity with Kalman snakes. IEEE transactions on pattern analysis and machine intelligence, 21(6), 564-569.

Porikli, F., Tuzel, O., & Meer, P. (2006, June). Covariance tracking using model update based on means on Riemannian manifolds. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (Vol. 1, pp. 728-735).

Premaratne, P., Yang, S., Vial, P., & Ifthikar, Z. (2017). Centroid tracking based dynamic hand gesture recognition using discrete Hidden Markov Models. Neurocomputing, 228, 79-83.

Rabiner, L., & Juang, B. (1986). An introduction to hidden Markov models. IEEE assp magazine, 3(1), 4-16.

Rahman, M. A., Purnama, I. K. E., & Purnomo, M. H. (2014, August). Simple method of human skin detection using HSV and YCbCr color spaces. In 2014 International Conference on Intelligent Autonomous Agents, Networks and Systems (pp. 58-61). IEEE.

Ramli, S. (2012, July). GMT feature extraction for representation of BIM sign language. In 2012 IEEE Control and System Graduate Research Colloquium (pp. 43-48). IEEE.

Rautaray, S. S. (2012). Real time hand gesture recognition system for dynamic applications. International Journal of ubicomp (IJU), 3(1).

Rautaray, S. S., & Agrawal, A. (2015). Vision based hand gesture recognition for human computer interaction: a survey. Artificial intelligence review, 43, 1-54.

Rehg, J. M., & Kanade, T. (1994, November). Digiteyes: Vision-based hand tracking for human-computer interaction. In Proceedings of 1994 IEEE workshop on motion of non-rigid and articulated objects (pp. 16-22). IEEE.

Rehg, J. M., & Kanade, T. (1995, June). Model-based tracking of self-occluding articulated objects. In Proceedings of IEEE International Conference on Computer Vision (pp. 612-617). IEEE.

Rekha, J., Bhattacharya, J., & Majumder, S. (2011, December). Shape, texture, and local movement hand gesture features for Indian sign language recognition. In 3rd international conference on trends in information sciences & computing (TISC2011) (pp. 30-35). IEEE.

Rubine, D. (1991). Specifying gestures by example. ACM SIGGRAPH computer graphics, 25(4), 329-337.

Saxe, D., & Foulds, R. (1996, September). Toward robust skin identification in video images. In Proceedings of the Second International Conference on Automatic Face and Gesture Recognition (pp. 379-384). IEEE.

Shan, C., Tan, T., & Wei, Y. (2007). Real-time hand tracking using a mean shift embedded particle filter. Pattern recognition, 40(7), 1958-1970.

Shanthakumar, V. A., Peng, C., Hansberger, J., Cao, L., Meacham, S., & Blakely, V. (2020). Design and evaluation of a hand gesture recognition approach for real-time interactions. Multimedia Tools and Applications, 79, 17707-17730.

Shi, J. (1994, June). Good features to track. In 1994 Proceedings of IEEE conference on computer vision and pattern recognition (pp. 593-600). IEEE.

Shimada, N., Shirai, Y., Kuno, Y., & Miura, J. (1998, April). Hand gesture estimation and model refinement using monocular camera-ambiguity limitation by inequality constraints. In Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition (pp. 268-273). IEEE.

Siddiqi, S. M., Gordon, G. J., & Moore, A. W. (2007, March). Fast state discovery for HMM model selection and learning. In Artificial Intelligence and Statistics (pp. 492-499). PMLR.

Sigal, L., Sclaroff, S., & Athitsos, V. (2004). Skin color-based video segmentation under time-varying illumination. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(7), 862-877.

Signer, B., Norrie, M. C., & Kurmann, U. (2011). iGesture: A Java framework for the development and deployment of stoke-based online Gesture recognition algorithms. Technical Report/ETH Zurich, Department of Computer Science, 561.

Singha, J., & Laskar, R. H. (2016). Recognition of global hand gestures using self co-articulation information and classifier fusion. Journal on Multimodal User Interfaces, 10(1), 77-93.

Singha, J., Misra, S., & Laskar, R. H. (2016). Effect of variation in gesticulation pattern in dynamic hand gesture recognition system. Neurocomputing, 208, 269-280.

Singla, A., Roy, P. P., & Dogra, D. P. (2019). Visual rendering of shapes on 2D display devices guided by hand gestures. Displays, 57, 18-33.

Sridevi, P., Islam, T., Debnath, U., Nazia, N. A., Chakraborty, R., & Shahnaz, C. (2018, December). Sign Language recognition for Speech and Hearing Impaired by Image processing in Matlab. In 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 1-4). IEEE.

Suk, H. I., Sin, B. K., & Lee, S. W. (2010). Hand gesture recognition based on dynamic Bayesian network framework. Pattern recognition, 43(9), 3059-3072.

Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural processing letters, 9, 293-300.

Tang, J., Cheng, H., Zhao, Y., & Guo, H. (2018). Structured dynamic time warping for continuous hand trajectory gesture recognition. Pattern Recognition, 80, 21-31.

Tewari, D., & Srivastava, S. K. (2012). 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 (IJEAT), 2(2), 165-170.

Thai, L. H., Hai, T. S., & Thuy, N. T. (2012). Image classification using support vector machine and artificial neural network. International Journal of Information Technology and Computer Science, 4(5), 32-38.

Thirumuruganathan, S. (2010). A detailed introduction to K-nearest neighbor (k-NN) algorithm. Retrieved March, 20, 2012.

Tuzel, O., Porikli, F., & Meer, P. (2006). Region covariance: A fast descriptor for detection and classification. In Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part II 9 (pp. 589-600). Springer Berlin Heidelberg.

Ulas, A., & Yildiz, O. T. (2009, December). An incremental model selection algorithm based on cross-validation for finding the architecture of a hidden Markov model on hand gesture data sets. In 2009 International Conference on Machine Learning and Applications (pp. 170-177). IEEE.

Utsumi, A., & Ohya, J. (1998, June). Image segmentation for human tracking using sequential-image-based hierarchical adaptation. In Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231) (pp. 911-916). IEEE.

Utsumi, A., & Ohya, J. (1999, June). Multiple-hand-gesture tracking using multiple cameras. In Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149) (Vol. 1, pp. 473-478). IEEE.

Wang, G. W., Zhang, C., & Zhuang, J. (2012). An application of classifier combination methods in hand gesture recognition. Mathematical Problems in Engineering, 2012.

Wang, R. Y., & Popović, J. (2009). Real-time hand-tracking with a color glove. ACM transactions on graphics (TOG), 28(3), 1-8.

Wang, X., & Li, X. (2010, December). The study of MovingTarget tracking based on Kalman-CamShift in the video. In The 2nd International Conference on Information Science and Engineering (pp. 1-4). IEEE.

Weng, S. K., Kuo, C. M., & Tu, S. K. (2006). Video object tracking using adaptive Kalman filter. Journal of Visual Communication and Image Representation, 17(6), 1190-1208.

Wu, X. Y. (2020). A hand gesture recognition algorithm based on DC-CNN. Multimedia Tools and Applications, 79(13-14), 9193-9205..

Wu, Y., & Huang, T. S. (1999, September). Capturing articulated human hand motion: A divide-and-conquer approach. In Proceedings of the seventh IEEE International Conference on computer vision (Vol. 1, pp. 606-611). IEEE.

Wu, Y., Lin, J. Y., & Huang, T. S. (2001, July). Capturing natural hand articulation. In Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001 (Vol. 2, pp. 426-432). IEEE.

Xiu, C., Su, X., & Pan, X. (2018, June). Improved target tracking algorithm based on Camshift. In 2018 Chinese Control and Decision Conference (CCDC) (pp. 4449-4454). IEEE.

Xu, D., Wu, X., Chen, Y. L., & Xu, Y. (2015). Online dynamic gesture recognition for human robot interaction. Journal of Intelligent & Robotic Systems, 77(3-4), 583-596.

Yadav, K. S., Misra, S., Khan, T., Bhuyan, M. K., & Laskar, R. H. (2020). Segregation of meaningful strokes, a pre-requisite for self co-articulation removal in isolated dynamic gestures. IET Image Processing, 15(5), 1166-1178.

Yang, J., Lu, W., & Waibel, A. (1997). Skin-color modeling and adaptation. In Computer Vision—ACCV'98: Third Asian Conference on Computer Vision Hong Kong, China, January 8–10, 1998 Proceedings, Volume II 3 (pp. 687-694). Springer Berlin Heidelberg.

Yang, M. H., & Ahuja, N. (1998, June). Extraction and classification of visual motion patterns for hand gesture recognition. In Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231) (pp. 892-897). IEEE.

Yang, M. H., & Ahuja, N. (1998, October). Detecting human faces in color images. In Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No. 98CB36269) (Vol. 1, pp. 127-130). IEEE.

Yang, M. H., Ahuja, N., & Tabb, M. (2002). Extraction of 2D motion trajectories and its application to hand gesture recognition. IEEE Transactions on pattern analysis and machine intelligence, 24(8), 1061-1074.

Yang, Q. (2010, June). Chinese sign language recognition based on video sequence appearance modeling. In 2010 5th IEEE Conference on Industrial Electronics and Applications (pp. 1537-1542). IEEE.

Yang, R., & Sarkar, S. (2006, August). Detecting coarticulation in sign language using conditional random fields. In 18th International conference on pattern recognition (ICPR'06) (Vol. 2, pp. 108-112). IEEE.

Yang, W., Liu, Y., Zhang, Q., & Zheng, Y. (2019). Comparative object similarity learning-based robust visual tracking. IEEE Access, 7, 50466-50475.

Yeasin, M., & Chaudhuri, S. (2000). Visual understanding of dynamic hand gestures. Pattern Recognition, 33(11), 1805-1817.

Yoon, H. S., Soh, J., Bae, Y. J., & Yang, H. S. (2001). Hand gesture recognition using combined features of location, angle, and velocity. Pattern recognition, 34(7), 1491-1501.

Yu, C., Wang, X., Huang, H., Shen, J., & Wu, K. (2010, October). Vision-based hand gesture recognition using combinational features. In 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (pp. 543-546). IEEE.

Yu, Y., Bi, S., Mo, Y., & Qiu, W. (2016, June). Real-time gesture recognition system based on Camshift algorithm and Haar-like feature. In 2016 IEEE International Conference on Cyber Technology in automation, Control, and intelligent systems (CYBER) (pp. 337-342). IEEE.

Yuan, Q., Sclaroff, S., & Athitsos, V. (2005, January). Automatic 2D hand tracking in video sequences. In 2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)-Volume 1 (Vol. 1, pp. 250-256). IEEE.

Zhang, Q. Y., Zhang, M. Y., & Hu, J. Q. (2009). Hand Gesture Contour Tracking Based on Skin Color Probability and State Estimation Model. Journal of Multimedia, 4(6).

Zheng, W., & Bhandarkar, S. M. (2009). Face detection and tracking using a boosted adaptive particle filter. Journal of Visual Communication and Image Representation, 20(1), 9-27.

Zhou, S. K., Chellappa, R., & Moghaddam, B. (2004). Visual tracking and recognition using appearance-adaptive models in particle filters. IEEE Transactions on Image Processing, 13(11), 1491-1506.