

Hand Gesture Recognition: A Literature Survey

1. INTRODUCTION

The essential aim of building hand gesture recognition system is to create a natural interaction between human and computer where the recognized gestures can be used for controlling a robot or conveying meaningful information [1]. How to form the resulted hand gestures to be understood and well interpreted by the computer considered as the problem of gesture interaction [2].

Human computer interaction (HCI) also named Man-Machine Interaction (MMI) [3][4] refers to the relation between the human and the computer or more precisely the machine, and since the machine is insignificant without suitable utilize by the human [3]. There are two main characteristics should be deemed when designing a HCI system as mentioned in [3]: functionality and usability. System functionality referred to the set of functions or services that the system equips to the users [3], while system usability referred to the level and scope that the system can operate and perform specific user purposes efficiently [3]. The system that attains a suitable balance between these concepts considered as influential performance and powerful system [3]. Gestures used for communicating between human and machines as well as between people using sign language [5].

International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012
162

Gestures can be static (posture or certain pose) which require less computational complexity [6] or dynamic (sequence of postures) which are more complex but suitable for real time environments [6] [7]. Different methods have been proposed for acquiring information necessary for recognition gestures system [8][9]. Some methods used additional hardware devices such as data glove devices and color markers to easily extract comprehensive description of gesture features [8]. Other methods based on the appearance of the hand using the skin color to segment the hand and extract necessary features [8], these methods considered easy, natural and less cost comparing with methods mentioned before [8].

Some recent reviews explained gesture recognition system applications and its growing importance in our life [10] especially for Human computer Interaction HCI, Robot control, games, and surveillance, using different tools and algorithms [9][11]. This work demonstrates the advancement of the gesture recognition systems, with the discussion of different stages required to build a complete system with less erroneous using different algorithms.

The paper organization is as follows: the following section explains key issues of hand gesture recognition system which are segmentation, features extraction, and recognition. Applications of gesture recognition systems are given in Section 3. Gesture challenges are discussed in section 4. Section 5 provided a literature review of recent hand gesture recognition systems. Drawbacks are given in Section 6, and summary of research results are shown in Section 7, and finally conclusion in section 8.

2. ISSUES TO HAND GESTURE RECOGNITION: EXTRACTION METHODS AND FEATURES EXTRACTION

Most of the researchers classified gesture recognition system into mainly three steps after acquiring the input image from camera(s), videos or even data glove instrumented device. These steps are: Extraction Method, features estimation and extraction, and classification or recognition as illustrated in Figure 1.

Figure 1. Gesture recognition system steps.

2.1. Extraction Method and image pre-processing

Segmentation process is the first process for recognizing hand gestures. It is the process of dividing the input image (in this case hand gesture image) into regions separated by boundaries [12]. The segmentation process depends on the type of gesture, if it is dynamic gesture then the hand gesture need to be located and tracked [12], if it is static gesture (posture) the input image have to be segmented only. The hand should be located firstly, generally a bounding box is used to specify the depending on the skin color [13] and secondly, the hand have to be tracked, for tracking the hand there are two main approaches; either the video is divided into frames and each frame have to be processed alone, in this case the hand frame is treated as a posture and segmented [12], or using some tracking information such as shape, skin color using some tools such as Kalman filter[12].

International Journal of Artificial Intelligence & Applications (IJAlA), Vol.3, No.4, July 2012

163

The common helpful cue used for segmenting the hand is the skin color [12], since it is easy and invariant to scale, translation, and rotation changes [14]. Different tools and methods used skin and non-skin pixels to model the hand. These methods are parametric and non-parametric techniques, Gaussian Model (GM) and Gaussian Mixture Model (GMM) are parametric techniques, and histogram based techniques are non- parametric. However it is affected with illumination condition changes abs different races [6]. Some researches overcome this problem using data glove and colored markers which provide exact information about the orientation and position of palm and fingers [12]. Others used infrared camera [6], and range information generated by special camera Time-of-Flight (ToF) camera [12], although these systems can detect different skin colors under cluttered background but it is affected with changing in temperature degrees besides their expensive cost [6]. The segmentation considered as an open issue problem itself [9]. The color space used in a specific application plays an essential role in the success of segmentation process, however color spaces are sensitive to lighting changes, for this reason, researches tend to use chrominance components only and neglect the luminance components such as r-g, and HS color spaces. However there are some factors that obstacle the segmentation process which is [12]; complex background, illumination changes, low video quality.

[6][15] applied HSV color model which concentrates on the pigments of the pixel, [14] used YCbCr color space. [16] used normalized r-g color space. Some preprocessing operations are applied such as subtraction, edge detection, and normalization to enhance the segmented hand image [8][15]. Figure 2 shows some segmentation method examples.

Figure 2. segmentation method. a) [32], b) [15], c) [14].

2.2. Features Extraction

Good segmentation process leads to perfect features extraction process and the latter play an important role in a successful recognition process [6]. Features vector of the segmented image can be extracted in different ways according to particular application. Various methods have been

International Journal of Artificial Intelligence & Applications (IJAlA), Vol.3, No.4, July 2012

164

applied for representing the features can be extracted. Some methods used the shape of the hand such as hand contour and silhouette [6] while others utilized fingertips position, palm center, etc. [6] created 13 parameters as a feature vector, the first parameters represents the ratio aspect of the bounding box of the hand and the rest 12 parameters are mean values of brightness pixels in the image. [14] used Self-Growing and Self-Organized Neural Gas (SGONG) neural algorithm to capture the shape of the hand, then three features are obtained; Palm region, Palm center, and Hand slope. [16] calculated the Center Of Gravity (COG) of the segmented hand and the distance from the COG to the farthest point in the fingers, and extracted one binary signal (1D) to estimate the number of fingers in the hand region. [15] divided the segmented image into different blocks size and each block represents the brightness measurements in the image. Many experiments were applied to decide the right block size that can achieve good recognition rate [15]. [17][18] used Gaussian pdf to extract geometric central moment as local and global features. Figure 3 shows some applications of feature extraction methods.

Figure 3. features representation. a) The segmented image is partitioned into 11 terraces with 8 regions per

terrace to extract local and global geometric central moment [17][18]. b) Three angles are extracted: RC angle, TC angle, and distance from the palm center [14]. Segmented hand divided into blocks and the brightness factor for each block represents the feature vector (blocks with black area are discarded) [5].

2.3. Gestures Classification

After modeling and analysis of the input hand image, gesture classification method is used to

recognize the gesture. Recognition process affected with the proper selection of features parameters and suitable classification algorithm [7]. For example edge detection or contour operators [9] cannot be used for gesture recognition since many hand postures are generated and could produce misclassification [9]. Euclidean distance metric used to classify the gestures [19][5][17]. Statistical tools used for gesture classification, HMM tool has shown its ability to recognize dynamic gestures [20][13] besides, Finite State Machine (FSM) [21], Learning Vector Quantization [22], and Principal Component Analysis (PCA) [23]. Neural network has been widely applied in the field of extracted the hand shape [14], and for hand gesture recognition [24][25][26]. Other soft computing tools are effective in this field as well, such as Fuzzy C-
International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012
165

Means clustering (FCM) [6], and Genetic Algorithms GAs [27]. Figure 4 explain the architecture of classification system.

Figure 4. Architecture of gesture recognition system [5].

3. APPLICATION AREAS OF HAND GESTURES SYSTEM

Hand gestures recognition system has been applied for different applications on different domains, as mentioned in [7][9] including; sign language translation, virtual environments, smart surveillance, robot control, medical systems etc. overview of some hand gesture application areas are listed below[7][8].

A. Sign Language Recognition:

Since the sign language is used for interpreting and explanations of a certain subject during the conversation, it has received special attention [7]. A lot of systems have been proposed to recognize gestures using different types of sign languages [8]. For example [8] recognized American Sign Language ASL using boundary histogram, MLP neural network and dynamic programming matching. [28] recognized Japanese sign language JSL using Recurrent Neural Network, 42 alphabet and 10 words. [25] recognized Arabic Sign language ArSL using two different types of Neural Network, Partially and Fully Recurrent neural Network.

B. Robot Control:

Controlling the robot using gestures considered as one of the interesting applications in this field [6]. [16] proposed a system that uses the numbering to count the five fingers for controlling a robot using hand pose signs. The orders are given to the robot to perform a particular task [16], where each sign has a specific meaning and represents different function for example, "one" means "move forward", "five" means "stop", and so on.

International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012
166

C. Graphic Editor Control:

Graphic editor control system requires the hand gesture to be tracked and located as a preprocessing operation [7]. [20] used 12 dynamic gestures for drawing and editing graphic system. Shapes for drawing are; triangle, rectangular, circle, arc, horizontal and vertical line for drawing, and commands for editing graphic system are; copy, delete, move, swap, undo, and close [20].

D. Virtual Environments (VEs):

One of the popular applications in gesture recognition system is virtual environments VEs, especially for communication media systems [9]. [29] provided 3D pointing gesture recognition for natural human computer Interaction HCI in a real-time from binocular views. The proposed system is accurate and independent of user characteristics and environmental changes [29].

E. Numbers Recognition:

Another recent application of hand gesture is recognizing numbers. [13] proposed an automatic system that could isolate and recognize a meaningful gesture from hand motion of Arabic numbers from 0 to 9 in a real time system using HMM.

F. Television Control:

Hand postures and gestures are used for controlling the Television device [9]. In [30] a set of hand gesture are used to control the TV activities, such as turning the TV on and off, increasing and decreasing the volume, muting the sound, and changing the channel using open and close hand [30].

G. 3D Modeling

To build 3D modeling, a determination of hand shapes are needed to create, built and view 3D shape of the hand [9]. Some systems built the 2D and 3D objects using hand silhouette [9]. 3D hand modeling can be used for this purpose also which still a promising field of research [9].

4. LITERATURE REVIEW OF GESTURE RECOGNITION SYSTEMS

Hasan [17] applied multivariate Gaussian distribution to recognize hand gestures using non-geometric features. The input hand image is segmented using two different methods [18]; skin color based segmentation by applying HSV color model and clustering based thresholding techniques [18]. Some operations are performed to capture the shape of the hand to extract hand feature; the modified Direction Analysis Algorithm are adopted to find a relationship between statistical parameters (variance and covariance) [17] from the data, and used to compute object (hand) slope and trend [17] by finding the direction of the hand gesture [17], As shown in Figure 5.

International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012
167

Figure 5. computing hand direction [17].

Then Gaussian distinction is applied on the segmented image, and it takes the direction of the hand as shown in figure 6.

Figure 6. Gaussian distribution applied on the segmented image [18].

Form the resultant Gaussian function the image has been divided into circular regions in other words that regions are formed in a terrace shape so that to eliminate the rotation affect [17][18]. The shape is divided into 11 terraces with a 0.1 width for each terrace [17][18]. 9 terraces are resultant from the 0.1 width division which are; (1-0.9, 0.9-0.8, 0.8-0.7, 0.7-0.6, 0.6, 0.5, 0.5-0.4, 0.4-0.3, 0.3-0.2, 0.2-0.1), and one terrace for the terrace that has value smaller than 0.1 and the last one for the external area that extended out of the outer terrace [17][18]. An explanation of this division is demonstrated in Figure 7.

Figure 7. Terraces division with 0.1 likelihood [17][18].

Each terrace is divided into 8 sectors which named as the feature areas, empirically discovered that number 8 is suitable for features divisions [17], To attain best capturing of the Gaussian to fit the segmented hand, re-estimation are performed on the shape to fit capturing the hand object [17], then the Gaussian shape are matched on the segmented hand to prepare the final hand shape for extracting the features, Figure 8 shown this process [17].

International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012
168

Figure 8. Features divisions [18]. a) Terrace area in Gaussian. b) Terrace area in hand image.

After capturing the hand shape, two types of features are extracted to form the feature vector [17][18]; local feature, and global features. Local features using geometric central moments which provide two different moments μ_{00} , μ_{11} as shown by equation (1):

Where μ_x and μ_y is the mean value for the input feature area [17], x and y are the coordinated, and for this, the input image is represented by 88×2 features, as explained in detail in equation (2). While the global features are two features the first and second moments [17][18] that are the computed for the whole hand features area [17]. These feature areas are computed by multiplying feature area intensity plus feature area's map location [17]. In this case, any input image is represented with 178 features [17][18]. The system carried out using 20 different gestures [18], 10 samples for each gesture, 5 samples for training and 5 for testing, with 100% recognition percentage and it decreased when the number of gestures are more than 14 gestures [18]. In [17] 6 gestures are recognized with 10 samples for each gesture. Euclidian distance used for the classification of the feature [17][18].

Kulkarni [31] recognize static posture of American Sign Language using neural networks algorithm. The input image are converted into HSV color model, resized into 80×64 and some image preprocessing operations are applied to segment the hand [31] from a uniform background [31], features are extracted using histogram technique and Hough algorithm. Feed forward Neural Networks with three layers are used for gesture classification. 8 samples are used for each 26 characters in sign language, for each gesture, 5 samples are used for training and 3 samples for testing, the system achieved 92.78% recognition rate using MATLAB language.[31].

Hasan [5] applied scaled normalization for gesture recognition based on brightness factor matching. The input image with is segmented using thresholding technique where the background is black. Any segmented image is normalized (trimmed), and the center mass [5] of the image are

determined, so that the coordinates are shifted to match the centroid of the hand object at the origin of the X and Y axis [5]. Since this method depends on the center mass of the object, the generated images have different sizes [5] see figure 9, for this reason a scaled normalization operation are applied to overcome this problem which maintain image dimensions and the time as well [5], where each block of the four blocks are scaling with a factor that is different from other block's factors. Two methods are used for extraction the features; firstly by using the edge mages, and secondly by using normalized features where only the brightness values of pixels are calculated and other black pixels are neglected to reduce the length of the feature vector [5]. The database consists of 6 different gestures, 10 samples per gesture are used, 5 samples for training International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012

169

and 5 samples for testing. The recognition rate for the normalized feature problem achieved better performance than the normal feature method, 95% recognition rate for the former method and 84% for the latter one [5].

Figure 9. applying trimming process on the input image, followed by scaling normalization process [5]. Wysoski et al. [8] presented rotation invariant postures using boundary histogram. Camera used for acquire the input image, filter for skin color detection has been used followed by clustering process to find the boundary for each group in the clustered image using ordinary contour-tracking algorithm. The image was divided into grids and the boundaries have been normalized. The boundary was represented as chord's size chain which has been used as histograms, by dividing the image into number of regions N in a radial form, according to specific angle. For classification process Neural Networks MLP and Dynamic Programming DP matching were used. Many experiments have implemented on different features format in addition to use different chord's size histogram, chord's size FFT. 26 static postures from American Sign Language used in the experiments. Homogeneous background was applied in the work. Stergiopoulou [14] suggested a new Self-Growing and Self-Organized Neural Gas (SGONG) network for hand gesture recognition. For hand region detection a color segmentation technique based on skin color filter in the YCbCr color space was used, an approximation of hand shape morphology has been detected using (SGONG) network; Three features were extracted using finger identification process which determines the number of the raised fingers and characteristics of hand shape, and Gaussian distribution model used for recognition.

5. DRAWBACKS

In this section, drawbacks of some discussed methods are explained: Orientation histogram method applied in [19] have some problems which are; similar gestures might have different orientation histograms and different gestures could have similar orientation histograms, besides that, the proposed method achieved well for any objects that dominate the image even if it is not the hand gesture [19]. Neural Network classifier has been applied for gestures classification [28][8] but it is time consuming and when the number of training data increase, the time needed for classification are increased too [8]. In [28] the NN required several hours for learning 42 characters and four days to learn ten words [28]. Fuzzy c-means clustering algorithm applied in [6] has some disadvantages; wrong object extraction problem raised if the objects larger than the hand. The performance of recognition algorithm decreases when the distance greater than 1.5 meters between the user and the camera. Besides that, its variation to lighting condition changes and unwanted objects might overlap with the hand gesture. In [16] the system is variation to environment lighting changes which produces erroneous segmentation of the hand region. HMM tools are perfect for recognition dynamic gestures [13] but it is computational consuming.

International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012

170

Other system limitation as listed in [14] where the gestures are made with the right hand only, the arm must be vertical, the palm is facing the camera, and background is plane and uniform. In System limitations restrict the application such as; gestures are made with the right hand only, the arm must be vertical, the palm is facing the camera, background is uniform. In [13] the system could recognize numbers only form 0 to 9. While the system proposed in [16] for controlling a robot, can counts number of active fingers only without regard to which particular fingers are active with a fixed set of commands [16].

6. SUMMARY OF RESEARCH RESULTS

The following tables show summaries of some hand gesture recognition systems. In Table 1 a comparison between recognition methods in hand gesture recognition methods used. Table 2 provides a summary of application areas and invariant vector of some hand gesture recognition systems. Table 3 displays Summary of extraction method, features representation, and recognition of hand gesture recognition systems which are; hand extraction technique, features vector representation, and recognition used in the selected hand gesture recognition systems.