

# Recent methods in vision-based hand gesture recognition

Haitham Badi<sup>1</sup>

Received: 29 November 2015 / Accepted: 4 April 2016  
© Springer International Publishing Switzerland 2016

**Abstract** The goal of static hand gesture recognition is to classify the given hand gesture data represented by some features into some predefined finite number of gesture classes. The main objective of this effort is to explore the utility of two feature extraction methods, namely, hand contour and complex moments to solve the hand gesture recognition problem by identifying the primary advantages and disadvantages of each method. Artificial neural network is built for the purpose of classification by using the back-propagation learning algorithm. The proposed system presents a recognition algorithm to recognize a set of six specific static hand gestures, namely: Open, Close, Cut, Paste, Maximize, and Minimize. The hand gesture image is passed through three stages, namely, pre-processing, feature extraction, and classification. In the pre-processing stage, some operations are applied to extract the hand gesture from its background and prepare the hand gesture image for the feature extraction stage. In the first method, the hand contour is used as a feature which treats scaling and translation problems (in some cases). The complex moments algorithm is, however, used to describe the hand gesture and treat the rotation problem in addition to the scaling and translation. The back-propagation learning algorithm is employed in the multi-layer neural network classifier. The results show that hand contour method has a performance of (71.30 %) recognition, while complex moments have a better performance of (86.90 %) recognition rate.

**Keywords** Gesture recognition · Hand gestures · Artificial neural network · Human–computer interaction · Computer vision

## 1 Introduction

Computers have become a key element of our society since their first appearance in the second half of the last century. Surfing the web, typing a letter, playing a video game, or storing and retrieving data are some of the examples of the tasks that involve the use of computers. Computers will increasingly influence our everyday life because of the constant decrease in the price and size of personal computers and the advancement of modern technology. Today, the widespread use of mobile devices such as smart phones and tablets either for work or communication has enabled the people to easily access applications in different domains which include GPS navigation, language learning apps, etc. The efficient use of the most current computer applications requires user interaction. Thus, human–computer interaction (HCI) has become an active field of research in the past few years [1]. On the other hand, input devices have not undergone significant changes since the introduction of the most common computer in the nineteen eighties probably because existing devices are adequate. Computers are tightly integrated with everyday life, and new applications and hardware are constantly introduced as answers to the needs of modern society [2]. The majority of existing HCI devices is based on mechanical devices, such as keyboards, mouse, joysticks, or game pads. However, a growing interest in a class of applications that use hand gestures has emerged, aiming at a natural interaction between the human and various computer-controlled displays [3]. The use of human movements, especially hand gestures, has become an important part of human computer

---

✉ Haitham Badi  
haitham@siswa.um.edu.my

<sup>1</sup> College of Business Informatics, University of Information Technology and Communications, Baghdad, Iraq

intelligent interaction (HCII) in recent years, which serves as a motivating force for research in modeling, analysis, and recognition of hand gestures [4]. The various techniques developed in HCII can be extended to other areas, such as surveillance, robot control, and teleconferencing [4]. The detection and understanding of hand and body gestures is becoming an important and challenging task in computer vision. The significance of the problem can be illustrated easily by the use of natural gestures that are applied with verbal and nonverbal communications [5].

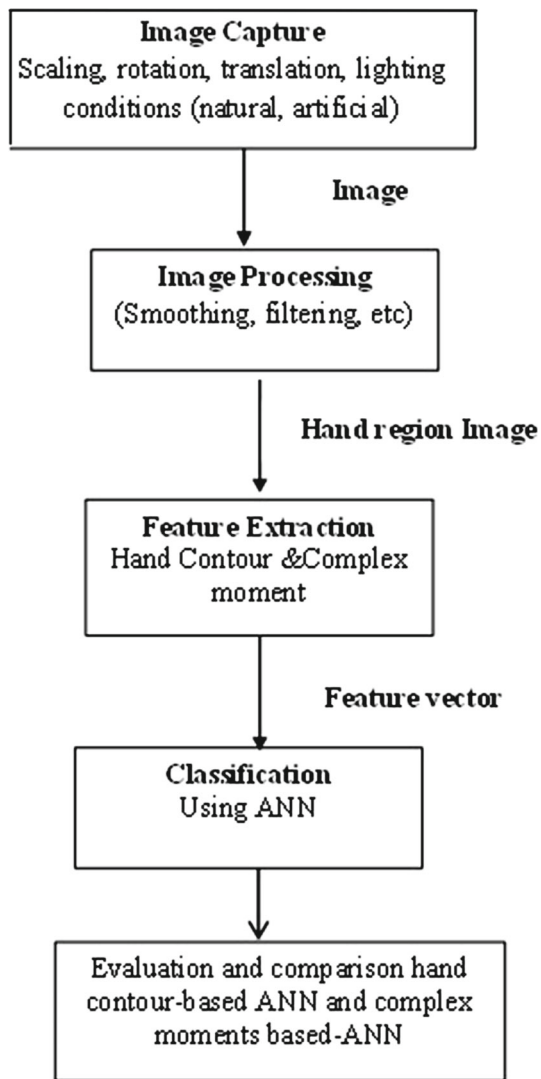
### 1.1 Statement of the problem

Gesture recognition has been adapted for various research applications from facial gestures to complete bodily human action [6]. Several applications have emerged and created a stronger need for this type of recognition system [6]. Static gesture recognition is a pattern recognition problem; as such, an essential part of the pattern recognition pre-processing stage, namely, feature extraction, should be conducted before any standard pattern recognition techniques can be applied. Features correspond to the most discriminative information about the image under certain lighting conditions. A fair amount of research has been performed on different aspects of feature extraction [4, 7–11]. Parvini and Shahabi [9] proposed a method for recognizing static and dynamic hand gestures by analyzing the raw streams generated by the sensors attached to human hands. This method achieved a recognition rate of more than (75 %) on the ASL signs. However, the user needs to use a glove-based interface to extract the features of the hand gestures which limits their usability in real-world applications, as the user needs to use special gloves in order to interact with the system. Another study [10] presented a real-time static isolated gesture recognition application using a hidden Markov model approach. The features of this application were extracted from gesture silhouettes. Nine different hand poses with various degrees of rotation were considered. The drawback of this feature extraction method is the use of skin-based segmentation method which does not work properly in the presence of skin-colored objects in the background. Dong [6] described an approach of vision-based gesture recognition for human–vehicle interaction using the skin-color method for hand segmentation. Similar to the problem in Vieriu [10], the performance of the recognition system is dramatically affected when skin-colored objects are present in the background. Developing a hand gesture recognition system that is capable of working under various conditions is difficult, but it is also more practical because these challenging conditions exist in real-world environment. These conditions include varying illumination and complex background as well as some effects of scaling, translation, and rotation by specific angles [2, 9, 12]. Another criteria that should be considered in the

hand gesture recognition systems that are employed in real-world applications is the computational cost. Some feature extraction methods have the disadvantage of being complicated and therefore consume more time, like Gabor filters with a combination of PCA [8] and the combination of PCA and Fuzzy-C-Mean [13] which are computationally costly which may limit their use in real-world applications. In fact, the trade-off between the accuracy and the computational cost in proposed hand gesture methods should be considered [14]. While most hand gesture systems focus only on the accuracy for hand gesture system assessments [15], it is desirable, in the phase of results evaluation, to consider the two criteria, namely, accuracy and the computational cost in order to identify their strengths and weaknesses and to recommend their potential applications [14].

### 1.2 Scope of the study

This study deals with the problem of developing a vision-based static hand gesture recognition algorithm to recognize the following six static hand gestures: Open, Close, Cut, Paste, Maximize, Minimize. These gestures are chosen because they are commonly used to communicate and can thus be used in various applications, such as, a virtual mouse that can perform six tasks (Open, Close, Cut, Paste, Maximize, Minimize) for a given application. The proposed system consists mainly of three phases: the first phase (i.e., pre-processing), the next phase (i.e., feature extraction) and the final phase (i.e., classification). The first phase includes hand segmentation that aims to isolate hand gesture from the background and removing the noises using special filters. This phase includes also edge detection to find the final shape of the hand. The next phase, which constitutes the main part of this research, is devoted to the feature extraction problem where two feature extraction methods, namely, hand contour and complex moments are employed. These two extraction methods were applied in this study because they used different approaches to extract the features, namely, a boundary-based for hand contour and region-based for complex moments. The feature extraction algorithms deal with problems associated with hand gesture recognition such as scaling, translation and rotation. In the classification phase where neural networks are used to recognize the gesture image based on its extracted feature, we analyze some problems related to the recognition and convergence of the neural network algorithm. As a classification method, ANN has been widely employed especially for real-world applications because of its ability to work in parallel and online training [16]. Thus, an ANN has been a lively field of research [17–20]. In addition, a comparison between the two feature extraction algorithms is carried out in terms of accuracy and processing time (computational cost). This comparison, using these two criteria, is important to identify the strengths



**Fig. 1** Overview of the method used to develop our hand gesture recognition system

and weaknesses of each feature extraction method and assess the potential application of each method. Figure 1 provides an overview of the method used to develop the hand gesture recognition system.

## 2 Background studies

Gesture recognition is an important topic in computer vision because of its wide range of applications, such as HCI, sign language interpretation, and visual surveillance [21]. Krueger [22] was the first who proposed Gesture recognition as a new form of interaction between human and computer in the mid-seventies. The author designed an interactive environment called computer-controlled responsive environment, a space within which everything the user saw or heard

was in response to what he/she did. Rather than sitting down and moving only the users fingers, he/she interacted with his/her body. In one of his applications, the projection screen becomes the wind-shield of a vehicle the participant uses to navigate a graphic world. By standing in front of the screen and holding out the users hands and leaning in the direction in which he/she want to go, the user can fly through a graphic landscape. However, this research cannot be considered strictly as a hand gesture recognition system since the potential user does not only use the hand to interact with the system but also his/her body and fingers, we choose to cite this [22] due to its importance and impact in the field of gesture recognition system for interaction purposes. Gesture recognition has been adapted for various other research applications from facial gestures to complete bodily human action [6]. Thus, several applications have emerged and created a stronger need for this type of recognition system [6]. In their study, Dong [6] described an approach of vision-based gesture recognition for human–vehicle interaction. The models of hand gestures were built by considering gesture differentiation and human tendency, and human skin colors were used for hand segmentation. A hand tracking mechanism was suggested to locate the hand based on rotation and zooming models. The method of hand-forearm separation was able to improve the quality of hand gesture recognition. The gesture recognition was implemented by template matching of multiple features. The main research was focused on the analysis of interaction modes between human and vehicle under various scenarios such as: calling-up vehicle, stopping the vehicle, and directing vehicle, etc. Some preliminary results were shown in order to demonstrate the possibility of making the vehicle detect and understand the humans intention and gestures. The limitation of this study was the use of the skin colors method for hand segmentation which may dramatically affect the performance of the recognition system in the presence of skin-colored objects in the background. Hand gesture recognition studies started as early as 1992 when the first frame grabbers for colored video input became available, which enabled researchers to grab colored images in real time. This study signified the start of the development of gesture recognition because color information improves segmentation and real-time performance is a prerequisite for HCI [23]. Hand gesture analysis can be divided into two main approaches, namely, glove-based analysis, vision-based analysis [24]. The glove-based approach employs sensors (mechanical or optical) attached to a glove that acts as transducer of finger flexion into electrical signals to determine hand posture. The relative position of the hand is determined by an additional sensor. This sensor is normally a magnetic or an acoustic sensor attached to the glove. Look-up table software toolkits are provided with the glove for some data-glove applications for hand posture recognition. This approach was applied by Parvini and Shahabi [9]

to recognize the ASL signs. The recognition rate was (75 %). The limitation of this approach is that the user is required to wear a cumbersome device and generally carry a load of cables that connect the device to a computer [3]. Another hand gesture recognition system was proposed in Swapna [25] to recognize the numbers from 0 to 10 where each number was represented by a specific hand gesture. This system has three main steps, namely, image capture, threshold application, and number recognition. It achieved a recognition rate of 89. The second approach, vision-based analysis, is based on how humans perceive information about their surroundings [24]. In this approach, several feature extraction techniques have been used to extract the features of the gesture images. These techniques include Orientation Histogram [12, 22], Wavelet Transform [26], Fourier Coefficients of Shape [27], Zernic Moment [28], Gabor filter [8, 13, 29], Vector Quantization [30], Edge Codes [31], Hu Moment [32], Geometric feature [33] and Finger-Earth Movers Distance (FEMD) [34]. Most of these feature extraction methods have some limitations. In orientation histogram for example, which was developed by McConnell [35], the algorithm employs the histogram of local orientation. This simple method works well if examples of the same gesture map to similar orientation histograms, and different gestures map to substantially different histograms [12]. Although this method is simple and offers robustness to scene illumination changes, its problem is that the same gestures might have different orientation histograms and different gestures could have similar orientation histograms which affects its effectiveness [36]. This method was used by Freeman and Roth [12] to extract the features of 10 different hand gesture and used nearest neighbor for gesture recognition. The same feature extraction method was applied in another study [2] for the problem of recognizing a subset of American Sign Language (ASL). In the classification phase, the author used a single-layer perceptron to recognize the gesture images. Using the same feature method, namely, orientation histogram, Ionescu et al. [24] proposed a gesture recognition method using both static signatures and an original dynamic signature. The static signature uses the local orientation histograms in order to classify the hand gestures. Despite the limitations of orientation histogram, the system is fast due to the ease of the computing orientation histograms, which works in real time on a workstation and is also relatively robust to illumination changes. However, it suffers from the same fate associated with different gestures having the same histograms and the same gestures having different histograms as discussed earlier. In [13], the authors used Gabor filter with PCA to extract the features and then fuzzy-c-means to perform the recognition of the 26 gestures of the ASL alphabets. Although the system achieved a fairly good recognition accuracy (93.32 %), it was criticized for being computationally costly which may limit its deployment in real-world applications [8]. Another method

extracted the features from color images as in [10] where they presented a real-time static isolated gesture recognition application using a hidden Markov model approach. The features of this application were extracted from gesture silhouettes. Nine different hand poses with various degrees of rotation were considered. This simple and effective system used colored images of the hands. The recognition phase was performed in real-time using a camera video. The recognition system can process 23 frames per second on a Quad Core Intel Processor. This work presents a fast and easy-to-implement solution to the static one hand gesture recognition problem. The proposed system achieved (96.2 %) recognition rate. However, the authors postulated that the presence of skin-colored objects in the background may dramatically affect the performance of the system because the system relied on a skin-based segmentation method. Thus, one of the main weaknesses of gesture recognition from color images is the low reliability of the segmentation process, if the background has color properties similar to the skin [37]. The feature extraction step is usually followed by the classification method, which use the extracted feature vector to classify the gesture image into its respective class. Among the classification methods employed are: Nearest Neighbor [12, 27, 28], Artificial Neural Networks [1, 2, 9], Support Vector Machines (SVMs) [8, 29, 32], Hidden Markov Models (HMMs) [10]. As an example of classification methods, Nearest Neighbor classifier is used as hand recognition method in [27] combined with modified Fourier descriptors (MFD) to extract features of the hand shape. The system involved two phases, namely, training and testing. The user in the training phase showed the system using one or more examples of hand gestures. The system stored the carrier coefficients of the hand shape, and in the running phase, the computer compared the current hand shape with each of the stored shapes through the coefficients. The best matched gesture was selected by the nearest-neighbor method using the MED distance metric. An interactive method was also employed to increase the efficiency of the system by providing feedback from the user during the recognition phase, which allowed the system to adjust its parameters in order to improve accuracy. This strategy successfully increased the recognition rate from (86 %) to (95 %). Nearest neighbor classifier was criticized for being weak in generalization and also for being sensitive to noisy data and the selection of distance measure [38]. To conclude the related works, we can say that hand gesture recognition systems are generally divided into two main approaches, namely, glove-based analysis and vision-based analysis. The first approach, which uses a special gloves in order to interact with the system, and was criticized because the user is required to wear a cumbersome device with cables that connect the device to the computer. In the second approach, namely, the vision-based approach, several methods have been employed to extract the features from



the gesture images. Some of these methods were criticized because of their poor performance in some circumstances. For example, orientation histograms performance is badly affected when different gestures have similar orientation histograms. Other methods such as Gabor filter with PCA suffer from the high computational cost which may limit their use in real-life applications. In addition, the efficiency of some methods that use skin-based segmentation is dramatically affected in the presence of skin-colored objects in the background. Furthermore, hand gesture recognition systems that use feature extraction methods suffer from working under different lighting conditions as well as the scaling, translation, and rotation problems.

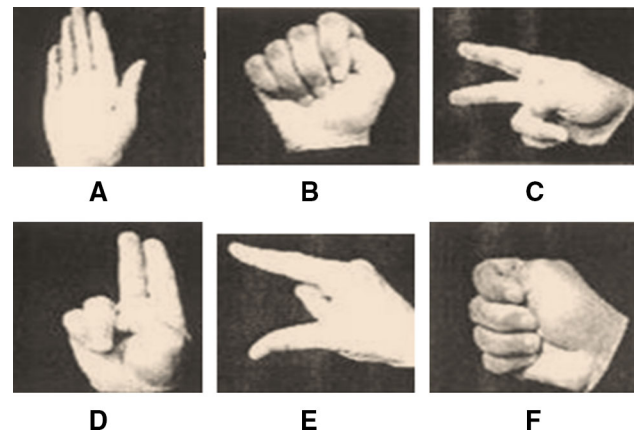
### 3 Methodology

The overview of the hand gesture recognition system (as shown in Fig. 1) consists of the following stages. The first stage is the hand gesture image capture stage where the images are taken using digital camera under different conditions such as scaling, translation and rotation. The second stage is a pre-processor stage in which edge detection, smoothing, and other filtering processes occur. In the next stage, the features of the images of hand gesture are extracted using two methods, namely, hand contour and complex moments. The last stage is the classification using Artificial Neural Network (ANN), where the recognition rate is calculated for both hand contour-based ANN and complex moments-based ANN and comparison is carried out. The following is a description of these stages.

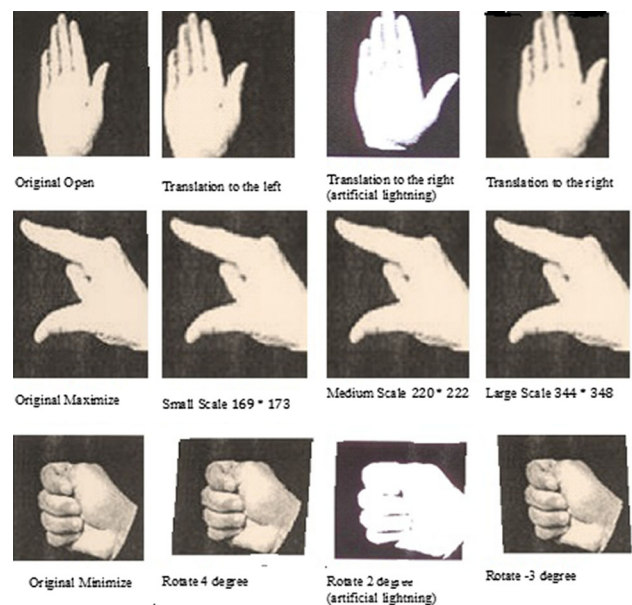
#### 3.1 Hand gesture image capture

The construction of a database for hand gesture (i.e., the selection of specific hand gestures) generally depends on the intended application. A vocabulary of six static hand gestures is made for HCI as shown in Fig. 2.

Each gesture represents a gesture command mode. These commands are commonly used to communicate and can thus be used in various applications such as a virtual mouse that can perform six tasks (Open, Close, Cut, Paste, Maximize, and Minimize) for a given application. The gesture images have different sizes. In image capture stage, we used a digital camera Samsung L100 with 8.2MP and 3x optical zoom to capture the images and each gesture is performed at various scales, translations, rotations and illuminations as follows (see Fig. 3 for some examples): Translation: translation to the right and translation to the left. Scaling: small scale (169173), medium scale (220222) and large scale (344348). Rotation: rotate 4°, rotate 2° and rotate −3°. Original of lightning: original and artificial. The database consists of 30 images for the training set (five samples for each gesture) and



**Fig. 2** Six static hand gestures: Open, Close, Cut, Paste, Maximize and Minimize

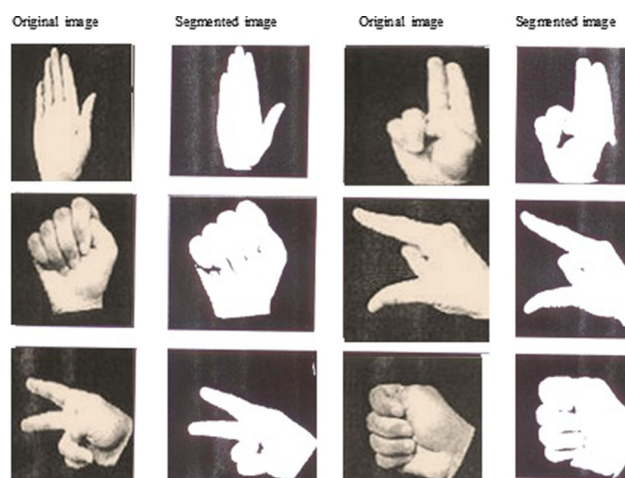


**Fig. 3** Hand gestures images under different conditions

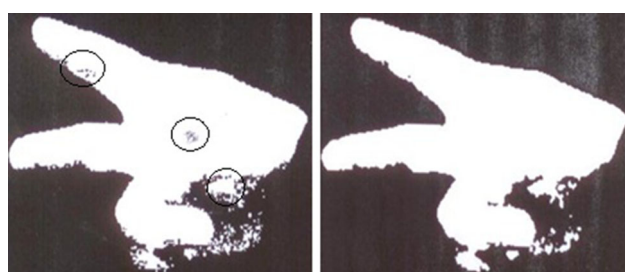
56 images for testing with scaling, translation, and rotation effects. Employing relatively few training images facilitates the measurement of the robustness of the proposed methods, given that the use of algorithms that require relatively modest resources either in terms of training data or computational resources is desirable [39,40]. In addition, Guodong and Dyer [41] considered that using a small data set to represent each class is of practical value especially in problems where it is difficult to get a lot of examples for each class.

#### 3.2 Pre-processing stage

The primary goal of the pre-processing stage is to ensure a uniform input to the classification network. This stage includes hand segmentation to isolate the foreground (hand gesture) from the background and the use of special filters to



**Fig. 4** Hand gesture images before and after segmentation



**Fig. 5** Median filter effect

remove any noise caused by the segmentation process. This stage also includes edge detection to find the final shape of the hand.

### 3.3 Hand segmentation

The hand image is segmented from the background. The segmentation process should be fast, reliable, consistent, and able to achieve optimal image quality suitable for the recognition of the hand gesture. Gesture recognition requires accurate segmentation. A thresholding algorithm is used in this study to segment the gesture image (see Fig. 4). Segmentation is accomplished by scanning the image pixel by pixel and labeling each pixel as an object or a background depending on whether the gray level of that pixel is greater or less than the value of the threshold  $T$ .

### 3.4 Noise reduction

Once the hand gesture image has been segmented, a special filter is applied to remove noise by eliminating all the single white pixels on a black background and all the single black pixels on a white foreground. To accomplish this goal, a median filter is applied to the segmented image as shown in Fig. 5.



**Fig. 6** Sobel edge detection for Open, Close and Cut

### 3.5 Edge detection

To recognize static gestures, the model parameters derived from the description of the shape and the boundary of the hand are extracted for further processing. Sobel was chosen for edge detection. Figure 6 shows some gesture images before and after edge detection operation using Sobel method.

### 3.6 Gesture feature extraction methods

The objective of the feature extraction stage is to capture and distinguish the most relevant characteristics of the hand gesture image for recognition. The selection of good features can greatly affect the classification performance and reduce computational time. The features used must be suitable for the application and the applied classifier. Two different feature extraction methods were used as part of the proposed hand gesture recognition algorithm:

1. Neural networks with hand contour;
2. Neural networks with hand complex moments.

These two extraction methods were applied in this study because they used different approaches to extract the features, namely, a boundary-based

for hand contour and region-based for complex moments. The use of different approaches may help us to identify strengths and weaknesses of each approach. Complex moments are adopted from [42] where the authors proposed a method for image recognition and we have applied this method specifically to the hand gesture recognition problem. The advantage of this method is its ability to extract invariant features that are independent of modifiers such as translation, scaling, rotation, and light conditions. There are other moments methods such as Hu moments [30] but we chose to apply complex moments method on hand gesture recognition to investigate its suitability to solve this problem as, to our best of knowledge, no previous study has applied this method to hand gesture recognition systems. In the Hand Contour method, we combined general and geometric features. The advantage of Boundary-based methods, which are commonly used for feature extraction, is that they are simple to implement and computationally fast [43,44].

#### 4 ANNs with hand contour

In this stage, many processes were performed on the hand gesture image to prepare these images for the subsequent feature extraction stage. These processes were performed using some image processing operations. The effect of these operations is explained below: subsection Hand Gesture Segmentation.

##### 4.1 Training phase

In this phase, the composite feature vectors computed earlier and stored in a feature image database are used as inputs to train the neural networks. The learning process for the five multilayer neural networks is accomplished by using the parameters shown in Table 1.

##### 4.2 Testing phase

After training the five neural networks, the performance is evaluated by using a new set of inputs (test set) and then computing the classification error. The activation function used is the binary-sigmoid function, which always produces out-

**Table 1** Parameters for the five neural networks

Parameters	Values
Input layer	1060 nodes
Hidden layer	100 nodes
Output layer	6
Stop error	0.01
Learning rate	0.9

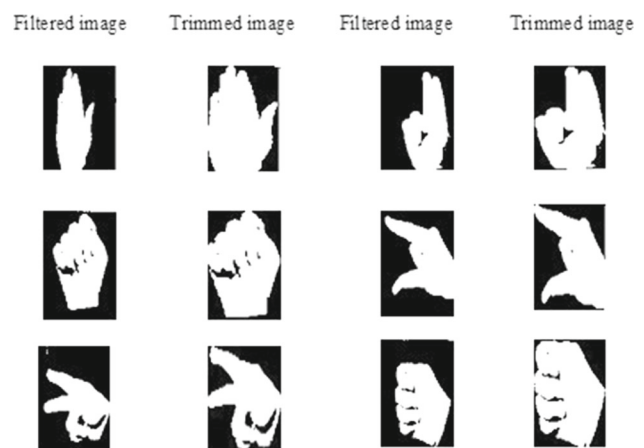
puts between 0 and 1. In our case, the five neural networks are used in a sequential manner, i.e., the test gesture feature image will be entered to the first neural network; if the network successfully recognizes the gesture, the test operation stops. If this network does not recognize the gesture features, the second network will be activated, and so on. If all the five networks fail to identify the feature, a message gesture not recognized appears. Notably, the failure of the neural network to recognize the gesture rather than wrongly recognizing it is directly related to the output of the network, where the recognized image is the one that receives the highest value; in the case where two or more images receive the same highest output value, the network fails to recognize the gesture. In the testing phase, 56 hand gesture images were used to test the system under different light conditions and with the effects of scaling and translation. The system is capable of recognizing and classifying any unknown gesture if such gesture is in the original database.

#### 5 ANNs with complex moment

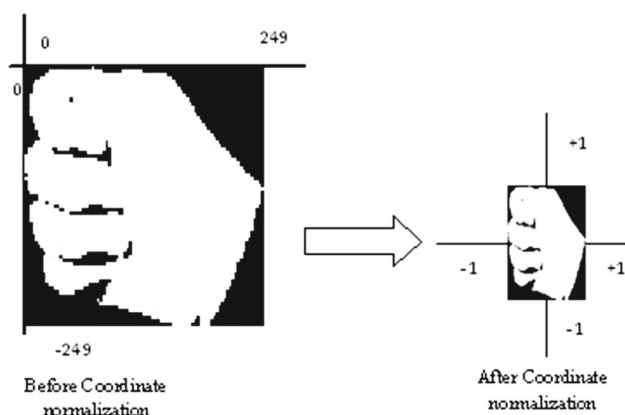
The processing stage in this method includes, in addition to segmentation and noise reduction processes as in the previous method, image trimming for eliminating the empty space and extracting only the region of interest, followed by the normalization process.

##### 5.1 Image trimming effect

The filtered hand gesture image may contain unused space surrounding the hand gesture. Thus, image trimming process is used to extract the hand gesture from its background. The effect of this process is shown in Fig. 7.



**Fig. 7** Image trimming effects



**Fig. 8** Coordinate normalization

**Table 2** Parameters of back-propagation neural networks

Parameters	Values
Input layer	10 nodes
Hidden layer	6 nodes
Output layer	6
Stop error	0.01
Learning rate	0.9

## 5.2 Coordinate normalization

After scaling each image to a fixed size (250, 250), the coordinates for the hand image are normalized between  $[-1, +1]$ . An example of this operation is shown in Fig. 8. For the example below, the coordinate (250, 0) becomes (1, 1) and the coordinate (0, -250) becomes  $(-1, -1)$ .

### 5.2.1 Training phase

After the computation of feature vectors, each one (feature vector) contains 10 translation, scaling, and rotation-invariant elements characterizing the complex moments for the hand gestures. Five similar neural network classifiers are trained with a data set containing 30 feature vectors (training data set). These vectors were computed from the training set that includes five examples for each hand gesture performed by one subject. The learning process for the back-propagation neural networks is accomplished by using the parameters shown in Table 2. The number of nodes in the input layer is equal to the length of the feature vector, while the number of nodes in the output layer is equal to the number of hand gestures. In addition, the number of nodes in the hidden layer is selected based on a trial and error approach, that is, many trials are performed with different number of nodes and the number that gives the best result is selected.

**Table 3** Specificity and sensitivity values for hand gestures (hand contour)

Hand gesture	Specificity	Sensitivity
Open	0.7551	0.5
Close	0.7083	0.78
Cut	0.6808	0.90
Paste	0.7447	0.60
Max	0.7021	0.80
Min	0.7234	0.70

### 5.2.2 Testing phase

After training the five neural networks using the training data consisting of 30 images, the performance is evaluated by applying the testing set on the network inputs and then computing the classification error. The testing process is conducted in the same manner as in the previous method. In this phase, 84 hand gesture images are used to test the system. Each one of the six hand gestures has a number of samples under different light conditions and with effects of scaling, translation and rotation.

## 6 Specificity and sensitivity for hand contour

As shown in Table 3, the sensitivity and accuracy values for gesture classes are the same because they are calculated by the same method. For specificity values, we notice that Open and Cut achieved the highest and lowest value, respectively. For Open, the specificity value is as high as 0.7551, which means that the probability of any image taken from other classes i.e., (Close, Cut, Paste, Max, Min) to be correctly recognized is 0.7551, or simply the average recognition rate of the other classes is (75.51 %), which means that Open has negatively contributed to the overall recognition rate. For sensitivity, which reflects the recognition rate per class, we noticed that Cut/Open has the highest/lowest value 0.9/0.5 which means that Cut/Open has the best/worst recognition rate among the six gestures.

## 7 Scaling and translation in hand contour

From Table 4, we notice that most of the recognition errors are caused by images with Translation (75 %) while a small portion of errors comes from images with Scaling (18.75 %) and less than that from artificial illumination (6.25 %). To evaluate the recognition rate of images with scaling and translation in hand contour which show the results of recognition on images with scaling and translation effects, respectively, that hand contour was able to handle the cases of scaling relatively



**Table 4** Recognition errors of hand gesture with scaling, translation and artificial illumination effects

Hand gesture	Scaling	Translation	Artificial	Total
Open	1 (25 %)	3 (75 %)	0 (0 %)	4
Close	0 (0 %)	2 (100 %)	0 (0 %)	2
Cut	1 (100 %)	0 (0 %)	0 (0 %)	1
Paste	1 (25 %)	2 (50 %)	1 (25 %)	4
Max	0 (0 %)	2 (100 %)	0 (0 %)	2
Min	0 (0 %)	3 (100 %)	0 (0 %)	3
Total	3	12	1	16
Recognition rate (%)	18.75 %	75 %	6.25 %	100

**Table 5** Specificity and sensitivity values for hand gestures (complex moments)

Hand gesture	Specificity	Sensitivity
Open	0.8405	1.00
Close	0.884	0.8
Cut	0.875	0.833
Paste	0.8889	0.75
Max	0.855	0.933
Min	0.8696	0.866

well (25 correct cases out of 28, or 89.29 %) but the problem was related to the translation cases especially for some gestures such as Open, Paste and Min at (0 %), (33.33 %) and (25 %) correctly recognized gestures, hence significantly decreasing the translation recognition rate to 45.45 %. In Table 5 which represents the confusion matrix of the gesture recognition, we notice that all the errors are due to not recognized cases which means that the classifier could not uniquely identify the gesture because the highest likelihood value is shared by two or more gestures.

### 7.1 Specificity and sensitivity for complex moments

As shown in Table 5, the sensitivity and accuracy values for gesture classes are the same because they are calculated by the same method. Table 3 shows that the highest and low-

est for specificity is achieved by Paste (0.8889) and Open, respectively. For Open, the specificity value is 0.8405, the lowest value, while sensitivity is 1.00, the highest value, means that the probability of an image taken from Open gesture class to be correctly recognized is higher (probability is 1) than the average probability of other classes (probability is 0.8405). It also means that Open class positively contributes to the overall recognition rate.

### 7.2 Rotation, scaling and translation for complex moments

Table 6 shows clearly that most of the recognition errors for Complex Moments are attributed to the cases with rotation (81.82 %), while small portion of errors (18.18 %) is caused by cases with scaling effect. Translation was perfectly recognized for all the gestures (error = 0 %).

## 8 Comparison with previous works

In this subsection, our work is compared with three previous related works, namely [1, 2, 9]. These works applied different feature extraction methods but similar classification method. The recognition accuracy of the methods proposed in these works along with our work are not applied on the same data set, and the comparison is just to get a general idea about the performance of other similar works for benchmark-

**Table 6** Error of gesture with (complex moments) translation

Hand gesture	Scaling	Translation	Rotation	Total
Open	0 (0 %)	0 (0 %)	0 (0 %)	0
Close	1 (33.3 %)	0 (0 %)	2 (66.6 %)	3
Cut	0 (0 %)	0 (0 %)	2 (100 %)	2
Paste	0 (0 %)	0 (0 %)	3 (100 %)	3
Max	0 (0 %)	0 (0 %)	1 (100 %)	1
Min	1 (50 %)	0 (0 %)	1 (50 %)	2
Total	2	0	9	11
Recognition rate (%)	18.18 %	0 %	81.82 %	100

ing purposes. Data set used by Parvini and Shahabi [9] and Symeonidis [2] is called American Sign Language (ASL) which has 26 gestures that represent the alphabets but without effects. Just [1] used a data set that has 10 hand gestures images, representing 10 selected alphabets. These images were tested under two different conditions: the first one tests images with uniform background, while the second is with complex background. Our methods and specifically Complex Moments compare favorably with other feature extraction methods and clearly better than some methods such as Orientation Histogram proposed in [2]. The results as obtained by [9], show that more challenging images with complex background achieved a lower recognition rate of (81.25 %) as compared to images with a uniform background (92.79 %). This finding supports the results of with our own research where images with more challenging effects such as rotation achieved as low as (65.38 %) recognition rate. One limitation of the work proposed in [9], is that the user has to wear gloves in order to get the features of the hand gesture. In addition, it seems that the complexity of the data set or the number of gestures in data set makes the recognition task more challenging and this can be seen in the results achieved by the studies that used ASL [2,9], compared with our work [1].

## 9 Conclusion

The primary conclusions can be summarized in the following points:

1. Hand contour-based neural networks training is evidently faster than complex moments-based neural networks training (at least 4 times faster where hand contour-based neural network took roughly between 110 and 160 epochs to converge, whereas complex moment-based neural network required at least between 450 and 900 epochs to convergence). This suggests that the hand contour method is more suitable than the complex moments method in real-world applications that need faster training, such as online training systems.
2. On the other hand, complex moments-based neural networks (86.37 %) proved to be more accurate than the hand contour-based neural networks (71.30 %). In addition, the complex moments-based neural networks are shown to be resistant to scaling (96.15 %) and translation (100 %), and to some extent to rotation (65.38 %) in some gestures (for example: open (100 %), Maximum (80 %). The results indicate that the complex moments method is preferred to the hand contour method because of its superiority in terms of accuracy especially for applications where training speed is not very crucial, such as off-line training applications and desktop applications.

3. Hand contour features are less distinguishable compared to complex moments features. The high number of not recognized cases predicted via the hand contour method makes this evident (11.90 %) of the testing cases for hand contour against (1.20 %) for complex moments). The recognized class cannot be uniquely defined because there are two or more classes (gestures) that have the same high certainty (or probability) value.
4. Neural networks are powerful classifier systems, but they suffer from the problem of overfitting, a problem which was more visible with hand contour method. Less overfitting was observed with the complex methods method, which is considered as an advantage for this method as the learning techniques which avoid the overfitting problem can provide a more realistic evaluation about their future performance based on the training results. In addition, neural networks appear to be more efficient when the number of features or the dimension of the feature vector, which is equal to the number of nodes in the input layer, is moderate (e.g., the complex moments method with 10 nodes is more accurate than the hand contour method with 1060 nodes).
5. The current research aims to provide a generic system that can be customized according to the needs of the user by using each of the six gestures as a specific command. For example, a direct application of the current system is to use it as virtual mouse that has six basic functions, namely, Open, Close, Cut, Paste, Maximize, and Minimize.
6. In addition, the proposed system is flexible; it can be expanded by adding new gestures or reduced by deleting some gestures. For example, you can use four gestures for TV control application, with each gesture being translated into one TV command: Open: to turn on the TV; Close: to turn off the TV; Min: to reduce the sound volume; and Max: to increase the sound volume, and so on.

## References

1. Just, A.: Two-handed gestures for human-computer interaction. *IDIAPRR* **6**(73), 94 (2006)
2. Symeonidis, K.: Hand gesture recognition using neural networks. *Neural Netw.* **13**(5), 1 (1996)
3. Pavlovic, V.I., Sharma, R., Huang, T.S.: Visual interpretation of hand gestures for human-computer interaction: a review. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 677–695 (1997). doi:[10.1109/34.598226](https://doi.org/10.1109/34.598226)
4. Wu, Y., Huang, T.S.: Vision-based gesture recognition: a review. In: Braffort, A., Gherbi, R., Gibet, S., Teil, D., Richardson, J. (eds.) *Gesture-Based Communication in Human-Computer Interaction*, vol. 1739, pp. 103–115. Springer, Berlin (1999)
5. Dadgostar, F., Barczak, A.L., Sarrafzadeh, A.: A color hand gesture database for evaluating and improving algorithms on hand gesture and posture recognition. *Res. Lett. Inf. Math. Sci.* **7**, 127–134 (2005)

6. Dong, G., Yan, Y., Xie, M.: Vision-based hand gesture recognition for human–vehicle interaction. Paper Presented at the Proceedings of the International conference on Control, Automation and Computer Vision (1998)
7. Bretzner, L., Laptev, I., Lindeberg, T.: Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. Paper Presented at the Automatic Face and Gesture Recognition, 2002. Proceedings of Fifth IEEE International Conference on (2002)
8. Gupta, S., Jaafar, J., Ahmad, W.F.W.: Static hand gesture recognition using local gabor filter. *Proc. Eng.* **41**, 827–832 (2012). doi:[10.1016/j.proeng.2012.07.250](https://doi.org/10.1016/j.proeng.2012.07.250)
9. Parvini, F., Shahabi, C.: An algorithmic approach for static and dynamic gesture recognition utilising mechanical and biomechanical characteristics. *Int. J. Bioinform. Res. Appl.* **3**(1), 4–23 (2007)
10. Vieri, R.L., Goras, B., Goras, L.: On HMM static hand gesture recognition. Paper Presented at the 10th International Symposium on Signals, Circuits and Systems (ISSCS) (2011)
11. Yoon, H.S., Soh, J., Bae, Y.J., Seung Yang, H.: Hand gesture recognition using combined features of location, angle and velocity. *Pattern Recognit.* **34**(7), 1491–1501 (2001). doi:[10.1016/S0031-3203\(00\)00096-0](https://doi.org/10.1016/S0031-3203(00)00096-0)
12. Freeman, W.T., Roth, M.: Orientation histograms for hand gesture recognition. Paper Presented at the International Workshop on Automatic Face and Gesture Recognition (1995)
13. Amin, M.A., Hong, Y.: Sign language finger alphabet recognition from gabor-PCA representation of hand gestures. Paper Presented at the Machine Learning and Cybernetics, 2007 International Conference on (2007)
14. Chen, F.S., Fu, C.M., Huang, C.L.: Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image Vis. Comput.* **21**(8), 745–758 (2003). doi:[10.1016/S0262-8856\(03\)00070-2](https://doi.org/10.1016/S0262-8856(03)00070-2)
15. Francke, H., Ruiz-del-Solar, J., Verschae, R.: Real-time hand gesture detection and recognition using boosted classifiers and active learning. *Advances in Image and Video Technology*, pp. 533–547. Springer (2007)
16. Rubaai, A., Kotaru, R., Kankam, M.D.: Parallel computation of continually on-line trained neural networks for identification and control of induction motors. Paper Presented at the Industry Applications Conference, 1999. Thirty-Fourth IAS Annual Meeting. Conference Record of the 1999 IEEE (1999)
17. Haitham, S., Sameem, A.: Gesture feature extraction for static gesture recognition. *Arab. J. Sci. Eng.* **38**, 3350–3351 (2013)
18. Hasan, H., Abdul Kareem, S.: Fingerprint image enhancement and recognition algorithms: a survey. *Neural Comput. Appl.* **23**, 1606–1608 (2013)
19. Hasan, H., Abdul-Kareem, S.: Static hand gesture recognition using neural networks. *Artif. Intell. Rev.* (2012). doi:[10.1007/s10462-011-9303-1](https://doi.org/10.1007/s10462-011-9303-1)
20. Hasan, H., Abdul-Kareem, S.: Human-computer interaction using vision based hand gesture recognition systems: a survey. *Neural Comput. Appl.* (2013). doi:[10.1007/s00521-013-1481-0](https://doi.org/10.1007/s00521-013-1481-0)
21. Kim, T.K., Cipolla, R.: Gesture recognition under small sample size. *Computer Vision ACCV*, pp. 335–344. Springer (2007)
22. Krueger, M.W.: *Artificial Reality II*, vol. 10. Addison-Wesley, Reading (1991)
23. Shet, V.D., Shiv, V., Prasad, N., Elgammal, A., Yacoob, Y., Davis, L.S.: Multi-cue exemplar-based nonparametric model for gesture recognition. Paper presented at the ICVGIP (2004)
24. Ionescu, B., Coquin, D., Lambert, P., Buzuloiu, V.: Dynamic hand gesture recognition using the skeleton of the hand. *EURASIP J. Appl. Signal Process.* **2005**, 2101–2109 (2005)
25. Swapna, B., Pravin, F., Rajiv, V.D.: Hand gesture recognition system for numbers using thresholding. *Comput. Intell. Inf. Technol.* **250**, 782–786 (2011)
26. Triesch, J., von der Malsburg, C.: Robust classification of hand postures against complex backgrounds. Paper Presented at the Automatic Face and Gesture Recognition, 1996. Proceedings of the Second International Conference on (1996)
27. Licsr, A., Szirnyi, T.: Hand-gesture based film restoration. Paper presented at the PRIS (2002)
28. Chang, C.C., Chen, J.J., Tai, W.K., Han, C.C.: New approach for static gesture recognition. *J. Inf. Sci. Eng.* **22**(5), 1047–1057 (2006)
29. Huang, D.Y., Hu, W.C., Chang, S.H.: Vision-based hand gesture recognition using PCA+gabor filters and SVM. Paper Presented at the Intelligent Information Hiding and Multimedia Signal Processing, 2009. IHH-MSP'09. Fifth International Conference on (2009)
30. Meng, H., Furo, S., Jinxi, Z.: Hidden Markov models based dynamic hand gesture recognition with incremental learning method. Paper Presented at the Neural Networks (IJCNN), 2014 International Joint Conference on (2014)
31. Chao, H., Meng, M.Q., Liu, P.X., Xiang, W.: Visual gesture recognition for human-machine interface of robot teleoperation. Paper Presented at the Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on (2003)
32. Liu, Y., Zhang, P.: An Automatic hand gesture recognition system based on Viola–Jones method and SVMs. Paper Presented at the Computer Science and Engineering, 2009. WCSE'09. Second International Workshop on (2009)
33. Bekir, C.: Hand Gesture Recognition. (Master thesis), Dokuz Eylul University (2012)
34. Zhou, R., Junsong, Y., Jingjing, M., Zhengyou, Z.: Robust part-based hand gesture recognition using kinect sensor. *IEEE Trans. Multimed.* **15**(5), 1110–1120 (2013). doi:[10.1109/TMM.2013.2246148](https://doi.org/10.1109/TMM.2013.2246148)
35. McConnell, R.K.: US Patent, No. 4567610 (1986)
36. Khan, R.Z., Ibraheem, N.A.: Hand gesture recognition: a Literature. *IJAIA* **03**(04), 169 (2012)
37. Oprisescu, S., Rasche, C., Bochao, S.: Automatic static hand gesture recognition using ToF cameras. Paper Presented at the Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European (2012)
38. Athitsos, V., Sclaroff, S.: Boosting nearest neighbor classifiers for multiclass recognition. Paper Presented at the Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on (2005)
39. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. *Comput. Vis. Image Underst.* **106**(1), 59–70 (2007)
40. Kanan, C., Cottrell, G.: Robust classification of objects, faces, and flowers using natural image statistics. Paper Presented at the Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (2010)
41. Guodong, G., Dyer, C.R.: Learning from examples in the small sample case: face expression recognition. *IEEE Trans. Syst. Man Cybern. B Cybern.* **35**(3), 477–488 (2005). doi:[10.1109/TSMCB.2005.846658](https://doi.org/10.1109/TSMCB.2005.846658)
42. Abu-Mostafa, Y.S., Psaltis, D.: Recognitive aspects of moment invariants. *Pattern Anal. Mach. Intell. IEEE Trans.* **6**(6), 698–706 (1984). doi:[10.1109/TPAMI.1984.4767594](https://doi.org/10.1109/TPAMI.1984.4767594)
43. Du Ming, T.: Boundary-based corner detection using neural networks. *Pattern Recognit.* **30**(1), 85–97 (1997). doi:[10.1016/S0031-3203\(96\)00057-X](https://doi.org/10.1016/S0031-3203(96)00057-X)
44. Ankit, C.: Recent Trends in Hand Gesture Recognition, vol. 3. Science Gate Publishing P.C. (2015)