

Control of Robot Directions Based on Online Hand Gestures

Mohammed A. Tawfeeq
 Computer Engineering Department,
 Faculty of Engineering,
 Mustansiriyah University, Baghdad, Iraq
 drmatawfeeq@uomustansiriyah.edu.iq

Ayam M. Abbass
 Computer Engineering Department,
 Faculty of Engineering,
 Mustansiriyah University, Baghdad, Iraq
 ayammoosen@uomustansiriyah.edu.iq

Abstract: *The evolution of wireless communication technology increases human machine interaction capabilities especially in controlling robotic systems. This paper introduces an effective wireless system in controlling the directions of a wheeled robot based on online hand gestures. The hand gesture images are captured and processed to be recognized and classified using neural network (NN). The NN is trained using extracted features to distinguish five different gestures; accordingly it produces five different signals. These signals are transmitted to control the directions of the cited robot. The main contribution of this paper is, the technique used to recognize hand gestures is required only two features, these features can be extracted in very short time using quite easy methodology, and this makes the proposed technique so suitable for online interaction. In this methodology, the preprocessed image is partitioned column-wise into two half segments; from each half one feature is extracted. This feature represents the ratio of white to black pixels of the segment histogram. The NN showed very high accuracy in recognizing all of the proposed gesture classes. The NN output signals are transmitted to the robot microcontroller wirelessly using Bluetooth. Accordingly the microcontroller guides the robot to the desired direction. The overall system showed high performance in controlling the robot movement directions.*

Index Terms— Iris recognition, Fourier descriptors, Principle component analysis, Manhattan distance, Euclidean distance, Cosine distance.

I. INTRODUCTION

Nowadays the communication techniques between humans and machines evolved significantly. These techniques tend to make ease of use especially with smart devices. One of such techniques depends on hand gestures. Hand gestures can be applied in several disciplines, such as; Human-Robot Interaction (HRI), sign language recognition, smart TV, computer games, smart house, and some controlling applications.

Considerable attention has been paid to HRI due to the technological development in hardware. This reinforced the use of hand gestures in controlling robotic systems, in which each gesture refers to a special action and can be

used to control a particular robot job. Several approaches are implemented in this field, some are based on mechanical hand gloves, accelerometers, and finger worn sensors, while others are based on image processing, geometrical measurements, statistical analysis, and classifiers. Chun Z. and Weihua S. [1] proposed an online hand gesture recognition (HGR) algorithm for a robot assisted living system. The proposed algorithm combines a neural network and hierarchical hidden Markov model (HHMM) to distinguish hand gestures. The authors use an inertial sensor worn on a finger of a human, which provides 3-D acceleration, angular velocity, magnetic data, and temperature. Raheja J. L., et al. [2] proposed an automatic hand gesture detection and recognition

system based on pattern matching using Principal Component Analysis (PCA) algorithm to control a robot arm. The authors implemented their algorithm using FPGA with a database of six small size images; each is of 60×80 pixels. Faudzi A., et al. [3] proposed a robot controlled by real-time hand gesture system to facilitate the process of Human-Robot Interaction (HRI). Their system extracts features from a preprocessed image based on bounding box and center-of-mass computation to control a robot. A glove-based technique is used to capture the hand gesture images with color filter to eliminate the undesired objects. Chaudhary A, Jagdish R. [4] introduced a method that calculates the bent finger's angle of hand gestures to control electro-mechanical robotic hand fingers. Dominio F., et al. [5] introduced a hand gesture recognition scheme using depth data and RGB images. Their method is based on segmenting the hand image into palm and finger regions, then hand's distances and curvature contour are extracted as features. The authors employed a multi-class SVM classifier to recognize the gesture. Ren Z., et al. [6] focused on building a robust part-based HGR system using Kinect sensor. They used a distance metric technique called Finger's Earth Movers Distance (FEMD) to measure the dissimilarity between hand shapes. Yao Y. and Fu Y. [7] proposed a HGR based on contour model with the use of Kinect sensor. The contour model is used to reduce gesture matching computation complexity. The proposed system allows tracking hand gesture in 3-D space. An isolated Chinese sign language gesture approach is proposed by Jiang, et al. [8]. Their approach is based on extracting feature vectors composed of trajectories and shapes of hand. A sparse dictionary algorithm is used for gesture recognition.

Murugeswari M. and Veluchamy S. [9] extracted the keypoint from a gesture image by implementing scale invariant feature transform. This keypoint is mapped into unified dimensional histogram vector depending on K-mean clustering using vector quantization. Then the gesture is recognized by multiclass SVM classifier. Ohn-Bar and Trivedi [10] developed a

real-time car vision-based system. The system classifies hand gestures by combining RGB and depth descriptor, in which, each descriptor is separately applied and compared over different modalities of RGB and depth using SVM classifier to find the optimal combination. Molchanov, et al. [11] proposed an algorithm for dynamic HGR using 3D convolutional neural networks (CNN) classifier. The CNN classifier is composed of a high-resolution subnet and a low-resolution subnet. Each subnet output class-membership probability and the final gesture classification is determined by multiplying these outputs element-wise. Jalab H. A. and Omer H. K. [12] proposed algorithm to recognize a set of four poses of hand gestures using neural network to control media player. The hand shape features extracted from a hand gesture image by regenerating a hand boundary image. Maleki B. and Ebrahimnezhad H. [13] proposed a visual mouse system based on dynamic hand gesture recognition. A white glove with different color fingertips is utilized to define 11 dynamic gestures based on mouse functions. Features are extracted from the curves of the detected trajectories of the five fingertips. A feature vector of length 672 components is created and processed. Different classifiers are applied to distinguish the gestures. Devine S., et al. [14] introduced a real time robotic arm control using hand gestures with multiple end effectors. They used the input of optical tracking device to control dual-arm robot remotely. Lamb K. and Madhe S. [15] added right and left movements to the existing electronic hospital-bed system, in which the position of the bed is controlled automatically by using hand gesture recognition. The features of the segment image are extracted by using wavelet decomposition, while the gestures are classified by using Euclidean Distance. Liu, et al. [16] constructed a 3D-based deep CNN to learn features from depth sequences. For each sequence a joined based feature vector is computed, then the output of SVM classification are fused to accomplish the recognition process.

This paper introduces the design and implementation of wireless control system that can steer the robot movement to the desired directions. The system is based on online hand gestures without the use of any gloves or sensors. A hand gesture image has been captured and preprocessed to extract the proposed features. An artificial neural network (ANN) is trained and used to classify the image gestures. The output of the neural network (NN) is converted to digital signals transmitted to a microcontroller in wireless mode to control the directions of a robot. The paper is organized as follows; the next section describes the proposed system. Section 3 introduces the methodology of hand gesture recognition. Section 4 describes the robot control circuitry, while section 5 presents the results. Finally, section 6 concludes the proposed work.

II. PROPOSED SYSTEM

There are different controls methodologies deals with robots. In this work, the proposed system controls the robot directions remotely based on online hand gestures. Online hand gesturers means the ability of using the system in harsh environments, as well as in wide range of applications. The system is designed and implemented such that it can recognize five different shapes of hand gestures using an ANN as gesture classifier. These five gesture shapes are proposed to control the directions of a wheeled robot to be moved in forward, spin left, spin right, move back, and to stop. The overall proposed system is as shown in Figure 1. It is composed of a smart phone camera used as acquisition device to capture hand gesture images. The captured image is transferred to a laptop, in which preprocessing operations are considered to extract some features. These features are used as inputs to the proposed NN classifier. The output of the classifier is converted to signals transmitted to a robot controller using Bluetooth. On the other side, a receiver is mounted to a microcontroller which is connected to a four motors robot to control the directions of the robot.

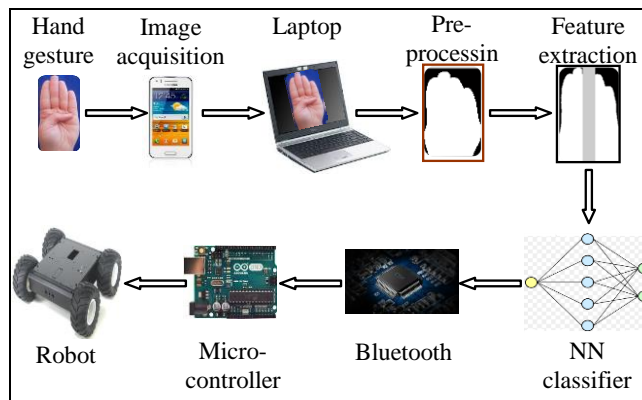


Figure 1. Overall proposed system.

III. METHODOLOGY OF HAND GESTURE RECOGNITION

In general, hand gesture techniques need to pass four process stages: image acquisition, image preprocessing, feature extraction and hand gesture recognition. In the first stage hand gesture image can be acquiesced using webcam, smart phone camera, or any external camera. In the second stage the captured image needs to be smoothed, filtered and cropped to obtain enrich data that can be used in the next stage. The third stage deals with features extraction, and the last stage uses some techniques to recognize and classifies the hand gestures. In this work the implemented HGR methodology can be described as follows:

A. Choosing hand gesture classes

The first step in this work is to decide, among the several shapes, what hand gestures need to be recognized. Five distinct gestures have been carefully chosen. Each gesture is assumed to symbolize a particular class, and then to represent a specified robot control action. Figure 2 shows the classes of the chosen gestures, each with its control action.

For each class several images have been captured, these images related to different people; males and females, for different ages. The collected images are used then to train and test the classifier neural network.

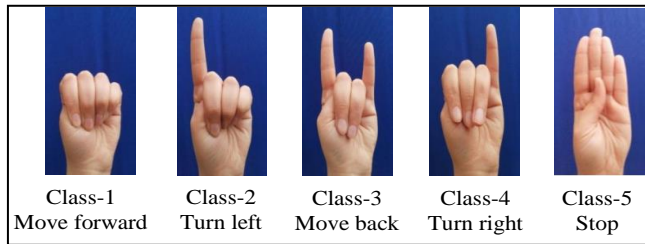


Figure 2. Hand gesture classes.

B. Image acquisition

In order to capture and process the images, a camera is required with some kind of interfacing or communication with a computer. There are several types of cameras that can be used as image acquisition devices, such as; webcam, smart phone, and standalone cameras. The problem with computer built-in webcam cameras is that it is of low resolution and has no free movement to capture the scene. The standalone cameras are in general of high cost.

Nowadays, the more available cameras are the smart phone ones. These cameras can be used online easily, thus, in this work such a camera has been used. The captured images are transferred from the smart phone to a laptop computer through USB port using (IP Webcam) software application. Noting that, it is of importance to make the color of the image background quite different than the skin color.

C. Image Preprocessing

In the preprocessing stage, several operations should be considered; isolating hand gesture from the background area, transforming the image to simple format, removing the unwanted parts, and resizing to particular dimensions. To do that, image preprocessing techniques have been used including: skin-detection, binarization, cropping the image to choose the important parts of the hand gesture, and resizing the final image to the required dimensions.

1) skin color detection

The most common method used in detecting the skin is through discrimination of its color. The pixel value in RGB color space is sensitive to the light. Since it is desired to control the robot at different environmental lighting, converting the

RGB image to another color space is required. Hence, the YCbCr model is used to separate skin color region from the background. The value of the threshold used to distinguish the skin from non-skin pixels is based on the value of Cb and Cr thresholding [17]. The condition used to detect skin region is: $(77 \leq Cb \leq 127; 133 \leq Cr \leq 170)$. If the image contains some noise or some pixel values similar to skin, then this process assumes these pixels as skin. Thus, care should be taken in capturing the hand gesture image in a different background color. The small amount of noise has no impact on the features adopted in this work.

2) Binarization

The output image from the skin detection process is a 3D matrix. It is easier to deal with a 2D binary image than a 3D one. There are many methods in image processing that can obtain a binary image from a 3D image. One of such methods is to use a fixed threshold. But before converting the image to binary form it is required first to convert it to a grayscale. The binarization in this step depends on the threshold values used to distinguish the skin from non-skin pixels mentioned in the previous section, in which, every pixel value lies within the threshold range is assumed to be skin pixel and coded as 1, otherwise it coded as 0.

3) Image cropping and resizing

The binary image may contain some unwanted areas. These areas can be removed by cropping the image. Since the aim of the system is to work online, thus it is required to propose an efficient and fast function that can deal with this process. The easiest way to do this job is by considering a function that detect and remove each row and/or column with no information. Using such function can solve the problem of the extra area but at the same time creates a new problem. The problem is that, some gesture classes should be considered with their related non-informatics upper area. Also, the hands of people are not of equal sizes, as well as the hand size varies by gender. To treat this problem, it is found empirically (by measuring the dimensions of different hands) that

the width of the palm in its largest zone to the length of the hand (palm and fingers measured from the wrist) is in general having approximately the ratio of (15:28). Thus, the proposed function is modified to crop each non-informatic column in the hand gesture image, while removing only the non-informatic rows that lies above this ratio.

The cropped images are then resized to a dimension of 150×280 pixel. Figure 3 shows an example of a hand gesture image with its preprocessed versions.

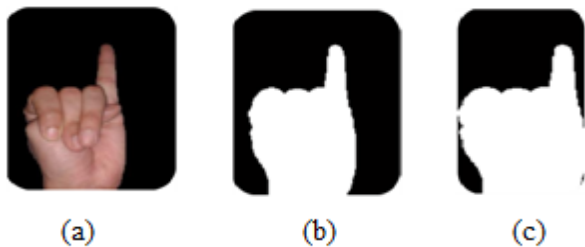


Figure 3. Hand gesture image preprocessing: (a) colored image, (b) binary image, (c) cropped and resized image.

4) Feature extraction

Work online means that everything must be achieved in a very short time. According to this fact, it is required to extract the less sufficient number of features with the shortest calculations time.

In this stage only two representative features of each hand gesture are obtained using the following simple shape analysis algorithm:

- a) The preprocessed image is partitioned vertically into two equally width segments left L and right R, as shown in Figure 4.

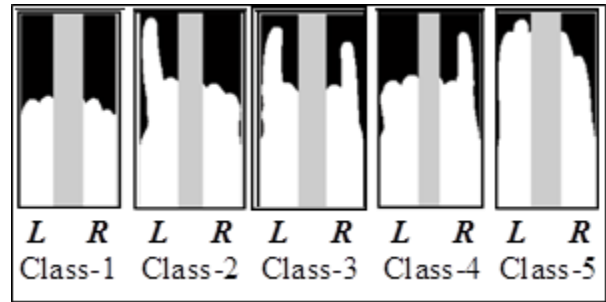


Figure 4. Hand gesture image segmentation.

- b) Compute the number of white pixel's histogram w and the number of black pixel's histogram b for each of the both segments. In each segment, if one finger or more is extended, then the number of white pixels is increased while the number of black pixels is decreased and vice versa. Notice that, the gesture classes are chosen carefully so that there is only one case in which two successive extended fingers can appear in any segments. This prevents any confusion between the gesture classes.
- c) The two extracted features are the ratios of the number of white pixels w to the number of black pixels b in each segment.

$$f_R = \frac{w_R}{b_R} \quad f_L = \frac{w_L}{b_L} \quad (1)$$

Where, f_L is the left segment feature, and f_R is the right segment feature. These two features are used to train and to test the proposed neural network.

5) Hand gesture recognition

In this stage hand gestures need to be classified based on the extracted features. A multi-layer feed forward neural network with back propagation training algorithm (MLBP) is considered. The network architecture is 2-5-5. This means input layer with two units, one hidden layer of five neurons and output layer of five neurons. Each input data presents to the network

is a vector consisting of two elements f_L and f_R . The five hidden neurons are chosen experimentally to get the best results. Finally the output layer is assumed to have five neurons to represent the five gesture classes. The output is designed to produce a single positive value and four negative values in each time, in which the positive output represents the corresponding gesture class. Thus to get such output, the activation function used in the hidden layer and in the output layer is the hyperbolic tangent function. This function bounded the output to be within $[-1, 1]$. The output from this network is passed to a single layer competitive subnet based on the rule of the winner take all, which means that, when the competition is completed only one neuron will have a nonzero signal of binary 1 value such that it can easily be transmitted as a control signal. The architecture of the proposed network is shown in Figure 5.

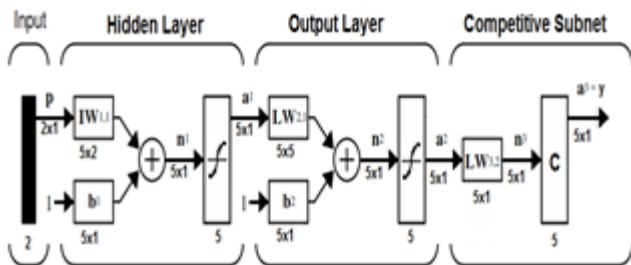


Figure 5. Neural network architecture.

D. Robot Control

In general robots can be categorized into two types, static robot and mobile robot. Static robot is commonly used in industrial applications, while the mobile type is used in enormous fields, such as, agriculture, surveillance, inspection hospitals, transportation tasks, etc. In this work a wheeled rover robot is implemented. This type of robots is easy to be controlled and can be considered as an experimental model for robotic systems. Rover robot needs wireless connection to ensure free movement such as Wi-Fi, Bluetooth, etc. The following subsections describe the main techniques and equipment used to control the rover robot.

1) Bluetooth

Bluetooth is a wireless communication mean which is used to transmit data and voice over short distances based on high speed radio waves. The core specification of Bluetooth (version 2) recommends a range of not less than 10 meters [18]. HC-05 is a Bluetooth module uses Serial Port Protocol (SPP). This module is fully qualified Bluetooth V2.0 + EDR (Enhanced Data Rate) 3 Mbps Modulation with 2.4 GHz radio transceiver and baseband. HC-05 module uses Adaptive Frequency Hopping Feature (AFH). Its size is of $12.7 \text{ mm} \times 27 \text{ mm}$ [19]. HC-05 can be interfaced easily with Arduino board (Uno) as shown in Figure 6.

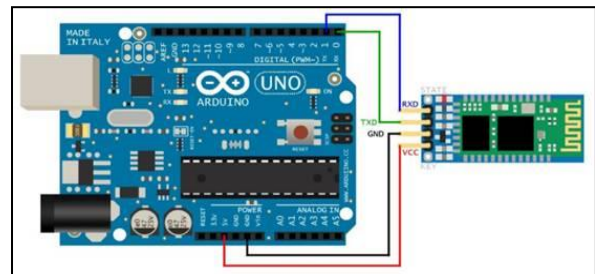


Figure 6. Arduino/Uno with HC-05.

2) Microcontroller

In this work, Arduino /Uno have been selected as a microcontroller due to its functionality and small size board in addition it can easily be connected to a computer using USB port. As the interfacing with HC-05 is configured, the microcontroller Arduino can be programmed for sending and receiving data using IDE software. After that, Arduino works independently. Arduino / Uno has 14 input/output digital pins (6 of which can be used as output with PWM), 6 pins used for analogue inputs and quartz crystal with 16 MHZ.

3) Rover 5 Robot

Rover 5 comes with complete motor set (encoder and driver). It has 4-independent dc motors with 7.2 V rated voltage and speed 1 Km / 1 hr, L293D motor driver and 1000 pulses at 3 rotations quadrature encoder for each motor [20].

4) Motor Controller

The used robot motor controller unit is a four channel DC motor controller board with encoder support, specifically designed and manufactured by "Dagu HiTech Electronic" for Rover 5 chassis [21]. Each channel is connected to a motor and supplied with 4.5-12V dc voltage. And each motor has current sensing which determine whether the motor has excessive load or stalled.

In this work the output of the classifier is forwarded using MATLAB code to transmit the signal through HC-05 module, then the preloaded Arduino program uses this signal to control the motors direction of the robot. Figure 7 shows the interfaces of Rover 5, Arduino, motor controller board and HC-05 module.

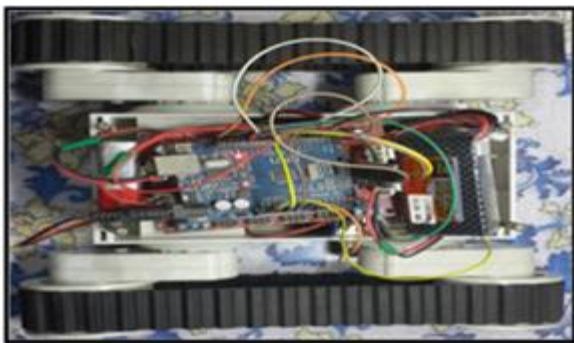


Figure 7. Rover 5 robot interfacing.

IV. RESULTS

The accuracy of HGR system mainly depends on the ability of the classifier used for distinguishing each gesture which in turn is highly relies on the goodness of the extracted features to represent each gesture. In this work about 35 different hand gesture snapshot images have been used to train the proposed MLBP

neural network. From each image two features is extracted, as a result a vector of size (2×35) is used to train the network. Each vector presents the network with its corresponding target class. Figure 8 shows the distribution of the training extracted features with their classifications.

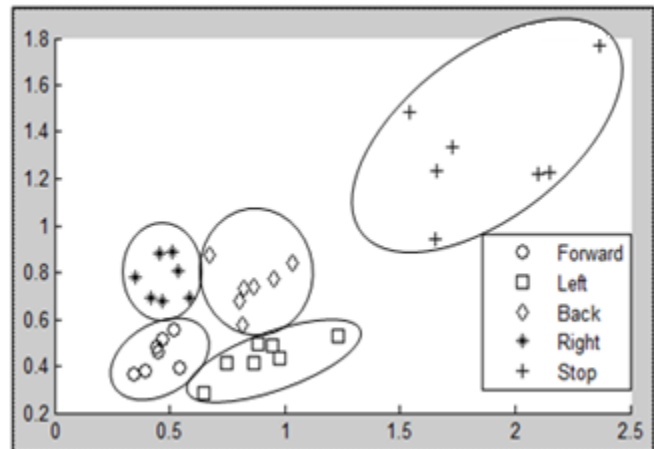


Figure 8. Training data classification.

In this Figure, it is very clear that the extracted features for each class can be grouped as a separate cluster. This makes the data of each class significantly differ than the others. Best training performance for the proposed NN showed that the sum squared error (sse) is about 5×10^{-7} , and the output of the network is highly closed to the target values, in which the linear regression equation between the output y and the target t is:

$$y = t + 2 \times 10^{-6} \quad (2)$$

In addition, it is found that the correlation coefficient R between all the outputs and the targets values is equal to 1.

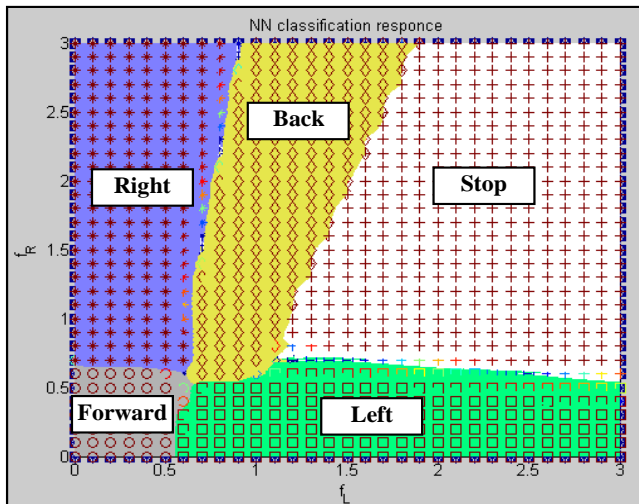


Figure 9. Neural network classification

After the training process is completed, the neural network has been tested with more than 20 online different snapshot images for the proposed hand gestures. The network is perfectly recognized all these gestures, and accordingly the robot is controlled to move in the optimized direction.

In order to check the response of the NN for different data sets, experimental tests have been made using a grid of virtual data which represents the extracted features within the range $[0 - 3; 0 - 3]$. The results showed the capability of the NN to response perfectly in group of clusters proportional to the proposed classes of the training data clusters as shown in Figure 9.

In online applications, analyzing and reducing the execution time is very important. The time required for preprocessing the image is proportional to the size of the captured image. In this work, it is found that, in average, the total time required for preprocessing an image of size (280×200) pixel is about 43 msec, and the time required for segmenting the preprocessed image and extracting its features is less than 0.5 msec. While the required time for classification using the trained neural network is about 2 msec. In total, the required execution time is about 45.5 msec as shown in Figure 10.

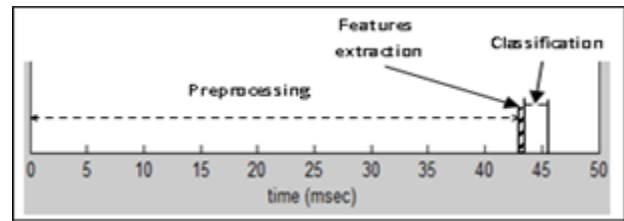


Figure 10. Execution time.

Comparing the results of the proposed system with the results of each of HHMM system introduced by [1], pattern matching using PCA algorithm introduced by [2], and FEMD introduced by [6], showed that, the performance index of the neural network of HHMM is about 2.7×10^{-6} with accuracy of less than 85%, and the accuracy in [2] is about 95%, while in FEMD the mean accuracy is less than 94% with mean running time of more than 4 sec. in contrast, the performance index for the network of the system proposed in this work is about 5×10^{-7} which is met the training goal with classification accuracy of about 99%, and average running time of about 43 msec.

V. CONCLUSIONS

This work introduces the design and implementation of a robot direction control system based on online hand gestures. Five different shapes of hand gestures have been chosen carefully to be processed and recognized using only two features extracted simply in very short time. The idea of extracting the features based on segmenting the preprocessed hand gesture image in column-wise needs only two halves. Two extracted features represent the ratio of white pixels to the black pixels of each half are efficient to train and to test the proposed NN. Arduino control the motors of Rover 5 robot efficiently using 4 channel DC motor controller unit. The control system leads easily to steer Rover 5 robot to move forward, spin left, spin right, move back, and to stop. The proposed NN showed very high accuracy in recognizing and classifying all of the test gestures. As a result, the robot direction is controlled perfectly.

In this work, the required time to extract the features is about 0.5 msec. and the total execution time required from the moment of capturing the gesture image up to transmitting the output signal of the classifier is about 45.5 msec. which is considered as a reasonable time for online system.

As a suggestion for future work, the proposed system can be extended to deal with videos rather than snapshots, and to preprocess the hand gestures for any background color. This may increase the complexity of the system but simple and smart algorithm can be used to isolate the hand gesture from the background such as edge detection technique with skin color detection.

REFERENCES

- [1] Chun Z., and Weihua S. "Online hand gesture recognition using neural network based segmentation," in IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, USA, pp. 2415-2420, 2009.
- [2] Raheja J.L., Shyam R., Kumar U., and Prasad P. B., "Real-time robotic hand control using hand gestures," in Second International Conference on Machine Learning and Computing (ICMLC), Bangalore, India, pp. 12-16, 2010.
- [3] Faudzi A, Ali M, Azman M, and Ismail Z. "Real-time hand gestures system for mobile robots control," in ELSEVIER International Symposium on Robotics and Intelligent Sensors (IRIS) Procedia Engineering, Kuching, Sarawak, Malaysia, ELSEVIER, vol. 41, pp. 798–804, 2012.
- [4] Chaudhary A, and Jagdish R., "Bent fingers' angle calculation using supervised ANN to control electro-mechanical robotic hand," *Comput Electr Eng*, vol. 39, no. 2, pp.560-570, 2013.
- [5] Dominio F., Donadeo M., Marin G., Zanuttigh P., and Cortelazzo G. M., "Hand gesture recognition with depth data," in ACM/IEEE Proceedings of the 4th ACM/IEEE International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream, Barcelona, Spain, ARTEMIS '13, pp. 9–16, 2013.
- [6] Ren Z., Meng J., and Zhang Z., "Robust part-based hand gesture recognition using kinect sensor," *IEEE T MULTIMEDIA*, vol. 15, no. 5, pp. 1110-1120, 2013.
- [7] Yao Y., and Fu Y., "Contour model based hand gesture recognition using kinect sensor," *IEEE T CIRC SYST VID*, vol. 24, no. 11, pp. 1935-1944, 2014.
- [8] Jiang Y., Tao J., Ye W., Wang W., and Ye Z., "An isolated sign language recognition system using RGB-D sensor with sparse coding," in IEEE Computational Science and Engineering (CSE) IEEE 17th International Conference, Chengdu, China, pp. 21-26, 2014.
- [9] Murugeswari M., and Veluchamy S., "Hand gesture recognition system for real-time application," in IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), Ramanathapuram, India, pp. 1220-1225, 2014.
- [10] Ohn-Bar E., and Trivedi M. M., "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations," *IEEE T INTELL TRANSP*, vol. 15, no. 6, pp. 2368-2377, 2014.
- [11] Molchanov P., Gupta S., Kim K., and Kautz J., "Hand gesture recognition with 3d convolutional neural networks," in IEEE Computer Vision and Pattern Recognition Workshops Conference (CVPRW), Boston, Massachusetts, USA, pp 1-7, 2015.
- [12] Hamid A. J., and Herman K. O. Human, "computer interface using hand gesture recognition based on neural network," in IEEE 5th National Symposium on Information Technology: Towards New Smart World (NSITNSW), Riyadh, Saudi Arabia, pp 1-6, 2015.
- [13] Maleki B., and Ebrahimnezhad H., "Intelligent visual mouse system based on hand pose trajectory recognition in video

- sequences,” *Multimedia Syst*, vol. 21, no. 6, pp. 581-601, 2015.
- [14] Devine S., Rafferty K., and Ferguson S., “Real time robotic arm control using hand gestures with multiple end effectors,” in *IEEE UKACC 11th International Conference on Control (CONTROL)*, Belfast, United Kingdom, pp.719-723, 2016.
- [15] Lamb K., and Madhe S., “Hand gesture recognition based bed position control for disabled patients,” in *IEEE Advances in Signal Processing Conference (CASP)*, Pune, India, pp. 170-174, 2016.
- [16] Liu Z., Zhang C., and Tian Y., “3D-based deep convolutional neural network for action recognition with depth sequences,” *Image vision comput*, vol. 55, part 2, pp. 93-100, 2016.
- [17] David J. Rios-Soria, Satu E. S., and Sara E. Garza-Villarreal., “Hand gesture recognition using computer-vision techniques,” in *WSCG 21st International Conference on Computer Graphics, Visualization and Computer Vision*, Czech Republic, pp. 1-8, 2013.
- [18] Bluetooth SIG. Bluetooth specification Version 2.0 + EDR. Bluetooth®, 2004.
- [19] Electronica 60 Norte. Datasheet Bluetooth to serial port module HC05. Yucatan, Mexico, 2017.
- [20] Sparkfun Electronics. Rover 5 datasheet. Colorado, USA.
- [21] JAMECO electronics. Motor controller 4 channel 4.5a 4.5-12v- for rover 5 chassis.