Hand Gesture Detection Using Neural Networks Algorithms

N. Alnaim and M. Abbod

Abstract—**Human gesture is a form of body language usually used as a mean of communication and is very critical in human-robot interactions. Vision-based gesture recognition methods to detect hand motion are vital to support such interactions. Hand gesture recognition enables a convenient and usable interface between devices and users. In this paper, an approach is presented for hand gesture recognition based on image processing methods, namely Wavelets Transform (WT), Empirical Mode Decomposition (EMD), besides Artificial Intelligence classifier which is Artificial Neural Networks (ANN) and Convolutional Neural Network (CNN). The methods are evaluated based on many factors including execution time, accuracy, sensitivity, specificity, positive and negative predictive value, likelihood, receiver operating characteristic (ROC), area under roc curve (AUC) and root mean Square. Results indicate that WT have less execution time than EMD and CNN. In addition, CNN is more effective in extracting distinct features and classifying data accurately compared to EMD and WT.**

Index Terms—**Artificial neural networks, convolutional neural network, empirical mode decomposition, hand gesture recognition, wavelet transform.**

I. Introduction

Currently, direct contact is the dominant form of interaction between the user and the machine. The interacting channel is based on devices such as the mouse, the keyboard, the remote control, touch screen, and other direct contact methods. Human to human interaction is achieved through more natural and intuitive noncontact methods, such as sound and physical movements. The flexibility and efficiency of noncontact interaction methods has led many researchers to consider exploiting them to support the human computer interaction. Gesture is one of the most important noncontact human interaction methods and forms a substantial part of the human language. Historically, wearable data gloves were usually used to obtain the angles and positions of each joint in the user's gesture. The inconvenience and cost of a wearable sensor have limited the widespread use of such method. Gesture recognition methods based on noncontact visual inspection are currently popular due to their low cost and convenience to the user. Hand gesture is an expressive interaction method used in healthcare, education and the entertainment industry, in addition to supporting users with special needs and the elderly. Hand tracking is essential to hand gesture recognition and involves undertaking various computer vision operations including hand segmentation, detection, and tracking.

Several gesture-based techniques [1]-[3] have been developed to support human computer interaction. According to Pradipa and Kavitha [1], the main aim of gesture recognition is developing a system that can detect human actions to be used for extracting meaningful information for device control.

Hand motion can be detected using any type of camera supporting reasonable image quality. 2-D cameras such as Microsoft's Kinect, Intel's RealSense Technology and Apple's iPhone high quality camera can easily be used in detecting most hand motions on a constant surface. Video content (composed of several images) is processed in several phases including data input, pre-processing, image segmentation, feature extraction, and classification.

The objective of this study is to investigate the best method available to extract features. Deep learning techniques are evaluated including WT, EMD, and CNN comparing their classification accuracy. A hand gesture recognition system was developed based on various image processing methods. The performance of hand gesture recognition methods was evaluated using various metrics including: execution time, accuracy, sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood, negative likelihood, receiver operating characteristic, area under roc curve and root mean square. ANN was used to classify the gestures using the features extracted as inputs. Multiple training sessions were performed applying filters. A CNN deep learning tool was used to minimize previous stages.

The rest of the paper is structured as follows. A literature review of hand gesture recognition methods and techniques is provided in Section II. Related theory to gesture extraction is presented in Section III. Details on the proposed system's implementation is provided in Section IV followed in Section V by a presentation and discussion on the results obtained. Finally, conclusion and future work are discussed in Section VI.

II. Literature Review

Image processing is central to hand gesture recognition. Digital image processing is most relevant to our work where useful information related to hand gesture and movement need to be extracted from digital images. Segmentation is crucial in gesture recognition [4], [5]. Hand detection and background removal are vital to the success of the gesture recognition algorithm. In previous work, a monocular camera was used in gesture recognition algorithms to filter out the background, which can be inconvenient in a real-world setting. Most methods used in hand detection are based on Harr features, colour, context information, or even shape.

Manuscript received August 20, 2019; revised October 17, 2019.

The authors are with the Department of Electronic and Computer Engineering, Brunel University London, Uxbridge, UK, UB8 3PH (e-mail: Norah.alnaim@brunel.ac.uk, Maysam.Abbod@brunel.ac.uk).

doi: 10.18178/ijmlc.2019.9.6.873

Such methods can provide accurate performance given the successful identification of the background and the hand in the image. However, there are limitations, e.g. a hand detection method relying on the skin colour will fail if the person is wearing a glove [6].

Feature detection is a crucial part of 2-D and 3-D image processing [7]. Before any feature extraction technique is applied, the image data is pre-processed, and different pre-processing techniques are applied to it including thresholding, binarization, and normalization. Features are then extracted and used for classification purposes. The behaviour of an image is captured based on its features. A good feature set contains attributes with high information gain and can be used to effectively classify images into different groups. A method that utilizes the symmetric properties of visual data to detect spare and stable image features was presented by Huebner and Zhang [8]. Regional features were formed by using Qualitative Symmetry operator together with quantitative symmetry range information.

Extraction and classification of local image structures are crucial to gesture recognition. Gevers *et al.* [9] proposed a method to classify the physical nature of local image structure using the geometrical and photometrical information.

To make the hand gesture recognition more accurate and thus ensuring a more natural user experience interacting with the machine interface, Bouchrika *et al.* [10], [11] applied a Wavelet Network Classifier (WNC) in a remote computer ordering application using hand gestures to place orders. Hands detection, tracking and gesture recognition techniques were applied. WNC was used for its effective classification results. An approach also proposed by Bouchrika *et al.* [12] made amendments to the Wavelet Network classification phase by making separated Wavelet Networks discriminating classes ($n - 1$) with the purpose of training each image. This resulted in less time required to complete the testing phase. The proposed Wavelet Network architecture enables quick learning and recognition of actions by avoiding unnecessary hand movements. Another hand gesture recognition approach [13] was based on wavelet enhanced image pre-processing and supervised Artificial Neural Networks (ANN).

Contour segmentation was supported in the pre-processing. Reference points were used to provide 2-D hand gesture contour images to 1D signal conversion. Wavelet decomposition was used for 1D signals. Four statistical features were extracted from the wavelet coefficients. Six hand gestures were tested. An accuracy of 97% was achieved with fast feature extraction and computation. Murthy and Jadon [14] proposed a method for hand gesture recognition using Neural Networks. It is based on supervised feed-forward neural network net-based training and back-propagation technique to classify hand gesture in ten various categories including hand pointing up, down, left, right, front, etc. Convolution Neural Networks (CNNs) were used [15] to evaluate hand gesture recognition, where depth-based hand data was employed with CNN to obtain successful training and testing results. Another CNN method was proposed [16] that uses a skin model, hand position calibration and orientation to train and test the CNN.

Hussain and Saxena [17] presented a hand gesture recognition technique based on deep transfer learning. Their work involved recognizing six static and eight dynamic hand gestures under various light conditions and backgrounds. Pang *et al.* [18] presented a technique based on deep learning, heuristic, and transfer learning. It uses a convolutional neural network tool, which can arrange a set of areas based on how to fit the target. 36 different algorithms on 50 videos were evaluated.

In this work, a hand gesture recognition system was developed to investigate and compare various methods including WT, EMD and CNN. Their classification accuracy is compared based on measuring various performance metrics.

III. Theory

Wavelets Transforms

Wavelet transforms support image processing by performing signal analysis where signal frequency differs at the end of time [19]. Wavelet transforms analysis offers accurate information on signal data in comparison to other analysis techniques. Daubechies orthogonal wavelet is used in this study. It is the known as dbN wavelets where N is the number of fading moments. The Daubechies wavelets are defined as follows:

$$\int x^n \psi(x) dx = 0, \qquad n = 0,1,\dots,K \tag{1}$$

The equation has a combination of scaling functions used to represent numerical approximations on a secured scale. The value of K is directly proportional to the orthogonality condition.

IV. Empirical Mode Decomposition

EMD is an innovative technology used in both non-stationary and non-linear data [20]. It is based on decomposing a signal into Intrinsic Mode Functions (IMF) with respect to the time domain [20], [21]. The EMD method can be compared to other analysis techniques such as Wavelet Transforms and Fourier Transforms [21]. Neuroscience experiments, seismic readings, gastro-electrograms, electrocardiograms, and sea-surface height readings are some of the data type to which the EMD technology might be applied [21]. EMD is defined as follows:

$$x(t) = \sum_{n=1}^{N} c_n(t) + r_n(t) \tag{2}$$

where r_n is the mean trend of $x(t)$, the value of c_n is the of amplitude and frequency modulated output set. The frequency decreases as the value of c_n increases.

V. Artificial Neural Networks

ANN is a simple electronic model similar to the neural structure of the human brain. The brain functions by learning from human experiences [22], [23]. ANN is a system that processes information in a similar manner to the biological nervous system. The system is composed of enormous number of unified processing elements working together to solve certain issues [22], [23]. Specifically, it is structured for data classification or pattern recognition via learning processes [22], [23]. Self-learning and the ability to handle large data are some of the many benefits associated with using ANN. A trained neural network is regarded as an expert in the set of information which it has been given to analyse.

VI. Convolutional Neural Networks

CNN is a multi-layer neural network with a special architecture used for deep learning [18]. A CNN architecture is composed of three layers: convolutional layer, Pooling layer, and fully-connected layer. CNN is frequently used in recognizing scenes, objects, and carrying out image detection, extraction and segmentation. CNN has been significantly used in the last few years due to the following three factors: (1) it removes the necessity for feature extraction by using image processing tools and can directly learn the image data, (2) it is exceptionally good for recognition results and can be easily retrained for new recognition missions, and (3) it can be built on the pre-existing network [18].

VII. Implementation

A. Hand Gestures Input

Hand gestures represent the input to different gesture detection methods evaluated in this study. Fig. 1 illustrates ten 2-D and 3-D hand gestures with plain backgrounds. They are recorded within long distances and used in the study's experimental work.

The implementation framework illustrating the extraction and the classification steps is shown in Fig. 2. Using an iPhone 6 Plus camera with resolution 4k at 30 fps, the hand motions shown in Fig. 1 are recorded. Each recording lasts 10 seconds and the resolution of the recorded video is 3840 × 2160. The first system is created using optical flow object by estimating and displaying the optical flow of objects in the video. The length of videos is between 15 to 65 frames. Each video has a different number of frames, which depends on the first section of motion.

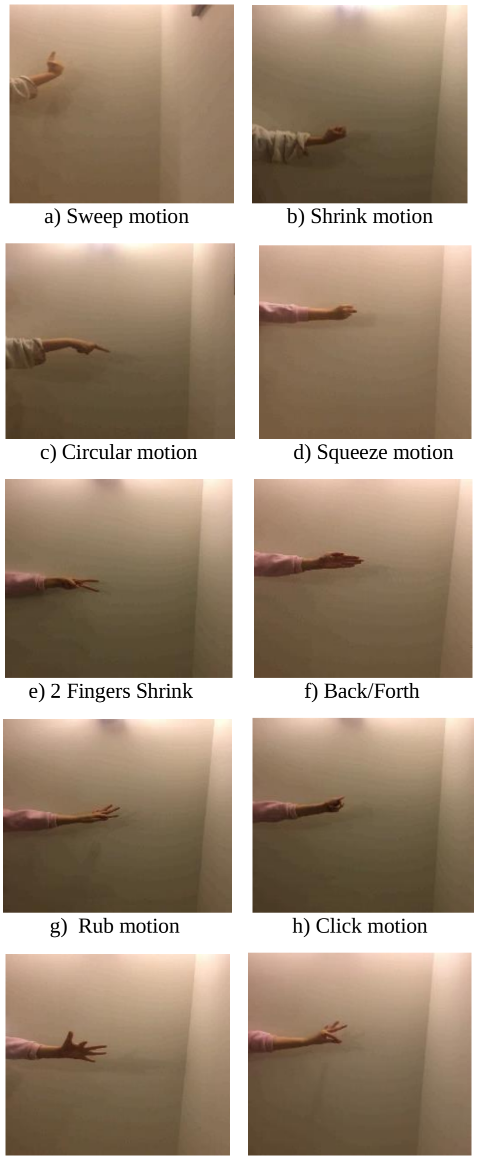
B. Computing Platform Specification

The experiment was performed using a Dell laptop XPS 15 9550 with 6th processor Generation Intel Quad Core i7, memory type DDR4 16 GB, speed 2133 MHz, 512 GB storage hard drive, 15.6-inch Ultra-HD 15.6" IPS 1920×1080 RGB Optional 3840×2160 IGZO IPS display w/Adobe RGB colour space and touch. Windows 10 (64 bits) operating system was used and the system is implemented using MATLAB R2017bV language.

C. Implementing Wavelet Transform with ANN

The system is implemented using the db8 WT tool following the steps outlined below:

1. Read each video using a video reader function.
2. Create an optical flow object that spreads the object velocities in an image.
3. Estimate and display the optical flow of objects in the video.
4. Divide a video into certain frames; each frame contains 8 IMFs.
5. Apply the *appcoef2* function which is used to compute an approximation coefficient of 2-D signals.
6. Extract each level using the *wrcoef* function to reconstruct the coefficients of each level in the video.
7. The execution time of WT is estimated only once.
8. The image data is trained and tested using a Neural Network system. The NN has 20 hidden neurons in a single hidden layer to train data and it stops when the error is reached in 20 epochs. When more hidden layers are added the depth of the neural network is increased, the neural network model becomes a deep learning model. Thus, A single layer is selected for this experiment.
9. The execution times of image data training and testing are also calculated.



a) Sweep motion b) Shrink motion

c) Circular motion d) Squeeze motion

e) 2 Fingers Shrink f) Back/Forth

g) Rub motion h) Click motion

i) Dance motion j) Pinch motion

Fig. 1. Hand gestures used in the study.

D. Empirical Mode Decomposition with Neural Network Implementation

The implementation of EMD is similar to WT with the addition of *reshape* function that returns the M-by-N matrix whose elements takes column-wise from X. The function used is *ceemd* representing a noise improved data analysis algorithm.

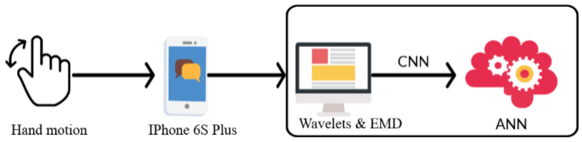


Fig. 2. The implementation framework.

E. Convolutional Neural Network Implementation

CNN forms an integral part of deep learning, as it is used to train data without using any image processing tool. In our experiment, a new directory is created for each video. Ten images are generated to transfer the image frame RGB to grey and resize it to 48×27 from the original image size . All videos have 70 frames. The image's data is split into training and testing datasets. The CNN topology is created in seven layers with each layer having the following functionality and size: *ImageInputLayer* Input size [48,27,1], *Convolution2-DLayer* Filter size [5,5], *ReLULayer* (Rectified Linear Unit), *MaxPooling2-DLayer* Pool size [2,2], *FullyConnectedLayer* Input size [auto] and Output size [10], *SoftmaxLayer* and *ClassificationOutputLayer* Output size [auto]. The hyperparameters of the CNN are generated inside the training options function. The value of max epochs parameter is set to 200 epochs.

F. Parameters for Comparison

The performance of WT, EMD and CCN algorithms was compared using number of parameters. This includes execution time, that is the duration taken by the software to process the given task. Sensitivity measures the percentage of positives which are properly identified. Specificity is a measure of the false positive rate. The PPV and NPV are the percentages of positive and negative results in diagnostic and statistics tests which also describe the true positive and true negative results. The LR+ and LR− are known measures in diagnostic accuracy. Area under ROC curve (AUC) is the typical technique to measure the accuracy of predictive models.

VIII. Results

The experiments were executed ten times to obtain the mean of ten-hand motions. Two different training and testing were presented and compared to find the best mini gestures detection tool. Training accuracy is achieved by implementing a model on the training data and determining the accuracy of the algorithm.

Fig. 3 and Fig. 4 show the signal extracted features using IMF method for 10 different gestures in WT and EMD techniques respectively. The IMF function is applied under two conditions. The first condition involves the entire data, where the number of extrema and the number of zero crossings are equal or vary at most by one. The second condition is that the mean of the envelope explained by the local maxima or the envelope clarified by the local minima has value zero [24]. The extracted features are each assigned a class and fed to ANN for training.

In the IMF graphs shown in Fig. 3 and Fig. 4, the X axis represents time (in microseconds) and the Y axis represents frequency with 8 signals of IMF (levels). The speed of motion starts from 0 microsecond till the end of time with a stable signal rate. Examining the IMFs for WT shown in Fig. 3, there is a substantial fluctuation in the blueline in the IMF graphs for different hand gestures. In addition, there is notable variation in the red and light blue signals in all graphs. Other signals exhibit steadier paths.

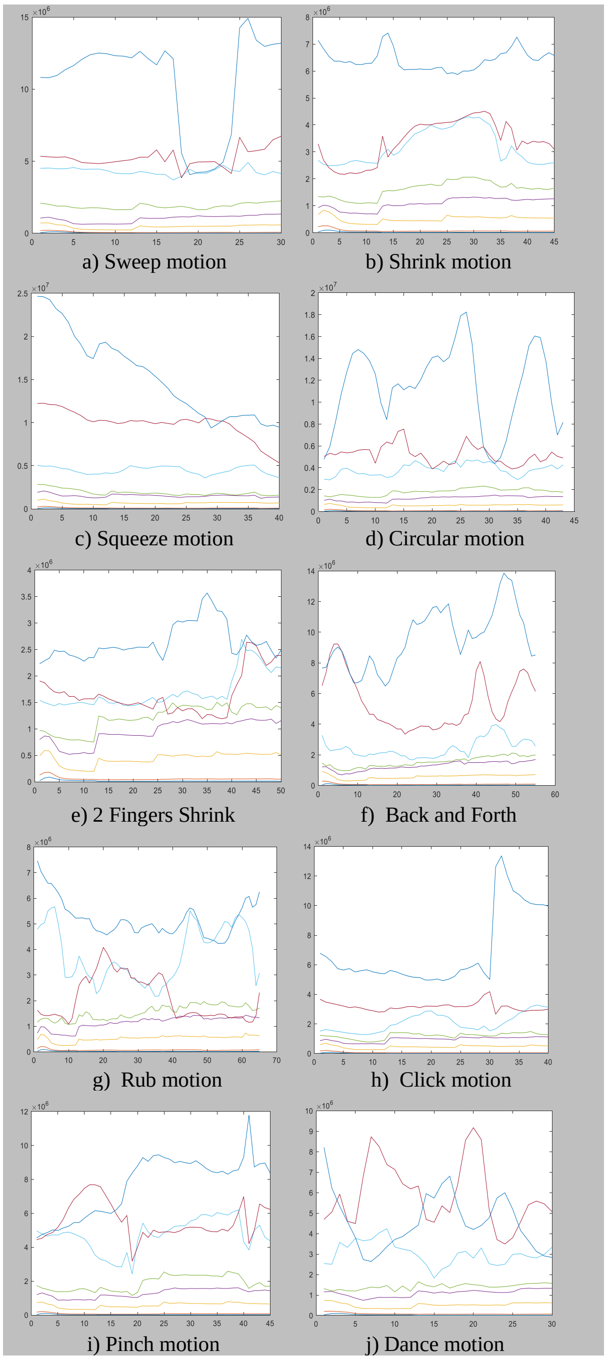


Fig. 3. IMF for 10 different motions using WT.

For the IMF graphs for EMD shown in Fig. 4, signals for different gestures are generally steadier compared to WT IMF graphs shown in Fig. 3. Slight fluctuations can be noticed in the blue signal, especially for the back and forth hand gesture. Minimum variation can be seen in the path of the red signal for all hand gestures. All other signals show steady lines for different gestures.

A summary of the values acquired for various parameters in training mode is listed in Table I. It can be noticed that the execution time of WT is less than that of EMD and CNN. The accuracy result of CNN is better than WT and EMD. The value of sensitivity in CNN is higher than WT and EMD. Specificity in CNN is the highest followed by EMD and the lowest result was recorded for WT. The PPV and NPV of WT is lower than EMD and CNN. The best value for LR+ and LR− are recorded for CNN. For RMS, the value of EMD and CNN are lower than WT. Finally, The AUC is 0.90 for WT, 99 for EMD, and 1 for CNN.

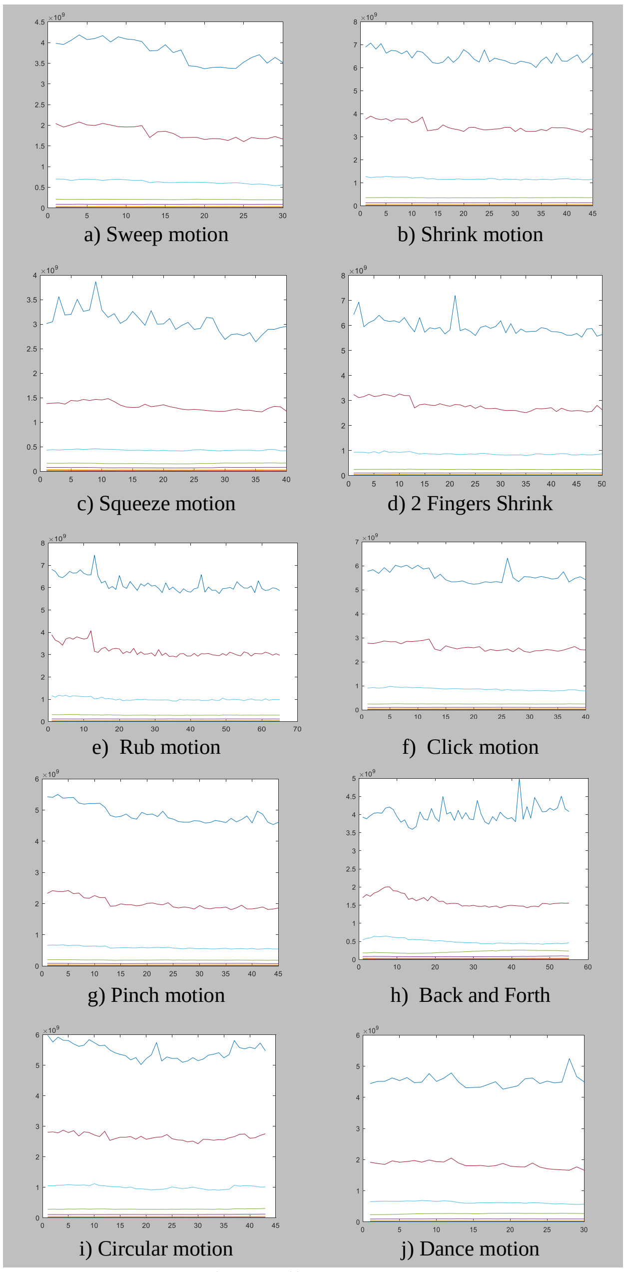


Fig. 4. IMF for 10 different motions using EMD.

The parameter values of CNN are constant for all categories. Its execution time is approximate 636 minutes, a substantially long and unacceptable duration to train the system using only ten hand movement pictures (which are used in the experiment). Positive Likelihood (LR+) of EMD is higher indicating more accuracy compared to WT and CNN. Overall, CNN has the best values in most parameters when training was performed except for execution time.

Comparative performance values for the three methods are listed in Table II. CNN's execution time is higher than WT and EMD. For accuracy, WT achieved a lower value compared to EMD and CNN. Accuracy results of CNN outperformed WT and EMD. Similarly, CNN has a higher sensitivity value compared to WT and EMD. Specificity in WT is lower than EMD and CNN. EMD and CNN have higher PPV and NPV values than WT. For LR+ and LR−, CNN values are higher than WT and EMD. The value of RMS for EMD is the lowest while WT has the highest value. Lastly, The AUC is 0.93 for WT, 1 for EMD, and 1 for CNN.

As in the training phase, the duration of CNN execution took similar time i.e. 636 minutes, an impractical timing given that only ten images were tested. It is notable that CNN has a significantly low value of 1 for the Positive Likelihood (LR+) compared to WT (19.29) and EMD (20.81).

TABLE I: COMPARISON BETWEEN WT, EMD AND CNN IN TRAINING MODE

	WL	EMD	CNN
Exe Time ± SD (sec)	6.0755 ± 1.243	9.171 ± 2.329	636.433± 113.922
Accuracy ± SD	0.440 ± 0.068	0.909± 0.051	1± 0
Sensitivity ± SD	0.433 ± 0.167	0.803 ± 0.151	1± 0
Specificity ± SD	0.968± 0.0126	0.996 ± 0.004	1± 0
Positive Predictive Value (PPV)	0.490 ±0.158	0.968±0.041	1± 0
Negative Predictive Value (NPV)	0.958 ±0.013	0.983 ±0.010	1± 0
Positive Likelihood (LR+)	16.783 ±13.312	129.406±158.592	1± 0
Negative Likelihood (LR−)	0.599 ±0.196	0.237 ±0.155	1± 0
RMS ± SD	1.223±0.129	0.371±0.053	1± 0
AUC ± SD	0.901±0.038	0.999±0.001	1± 0

TABLE II: COMPARISON BETWEEN WT, EMD AND CNN IN TESTING MODE

	WL	EMD	CNN
Exe Time ± SD (sec)	0.158 ±0.027	0.195 ±0.053	636.433 ±113.922
Accuracy ± SD	0.437±0.079	0.915 ±0.049	1± 0
Sensitivity ± SD	0.389±0.262	0.769±0.211	1± 0
Specificity ± SD	0.970±0.0162	0.995±0.007	1± 0
Positive Predictive Value (PPV)	0.496 ±0.254	0.979±0.044	1± 0
Negative Predictive Value (NPV)	0.955±0.012	0.981±0.014	1± 0
Positive Likelihood (LR+)	19.292±18.316	20.81 ±44.378	1± 0
Negative Likelihood (LR−)	0.622±0.197	0.254±0.215	1± 0
RMS ± SD	1.204 ±0.122	0.354±0.073	1± 0
AUC ± SD	0.937±0.030	1.00 ± 0	1± 0

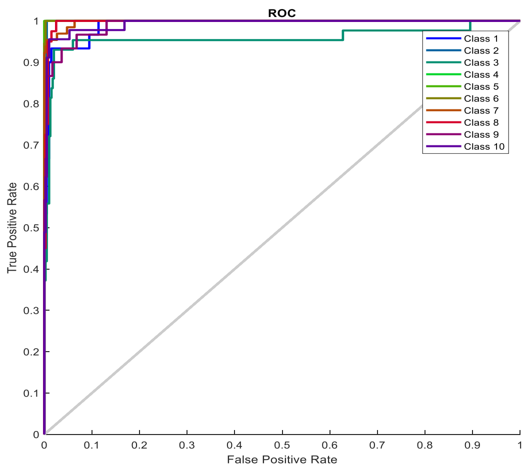


Fig. 5. ROC for 10 different classes in WT.

Fig. 5 and Fig. 6 show the Receiver Operating Characteristic ROC curve which is applied in binary classification to learn the output of a classifier. There are two strategies of ROC to be drawn for multiclass curve: One VS. One and One VS. Multi, with the latter being used in this study. According to the WT and EMD graphs, the 10 classes had 10 ROC curves reached the upper left corner which are

100% True Positive Rate (Sensitivity) and 100% False Positive Rate (1−Specificity). The ROC curve of EMD is extremely near to the upper left corner compared to WT.

CNN provides a better accuracy when compared with WT and EMD. However, CNN's duration of execution is substantially high. WT and CNN memory usage is lower than EMD.

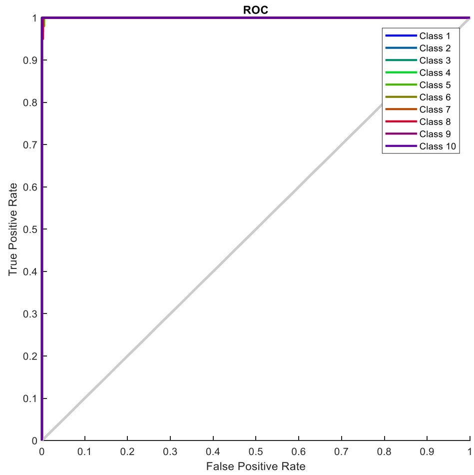


Fig. 6. ROC for 10 different classes in EMD.

IX. Conclusions and Future Work

Hand gesture recognition is essential to support a natural HCI experience. The most important aspects of gesture recognition are segmentation, detection, and tracking. In this study, a system has been created for hand motion detection using WT and EMD for features extraction. Classification is supported using ANN and CNN. Ten 2-D and 3-D motion images with plain backgrounds and recorded within long distances were used. Experiments were performed to compare the performance of various methods using number of measures. Results showed that CNN provides better accuracy compared to WT and EMD. However, its computational requirements are relatively high. Memory usage of WT and CNN was lower than EMD. In future work, the number of motions will be extended using a 3-D Holoscopic imaging system.

Acknowledgements

The first author would like to show his appreciation in terms of the financial support received from the Ministry of Higher Education in Saudi Arabia.

References

[1] R. Pradipa and S. Kavitha, "Hand gesture recognition-analysis of various techniques, methods and their algorithms," in *Proc. International Conference on Innovations in Engineering and Technology*, Tamil Nadu, India, 2014, pp. 2003-2010.

[2] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Computing Surveys*, vol. 38, no. 4, 2006.

[3] F. Mahmoudi and M. Parviz, "Visual hand tracking algorithms," IEEE *Gemometric Modeling and Imaging - New Trends*, London, UK, 2006.

[4] N. M. Zaitoun and M. J. Aqel, "Survey on image segmentation techniques," *Procedia Computer Science*, vol. 65, pp. 797–806, 2015.

[5] N. Dhanachandra and Y. J. Chanu, "A Survey on image segmentation methods using clustering techniques," *European Journal of Engineering Research and Science*, vol. 2 no. 15, 2017.

[6] M. Rao, P. Kumar, and A. Prasad, "Implementation of real time image processing system with FPGA and DSP," in *Proc. IEEE International Conference on Microelectronics, Computing and Communications*, Durgapur, India, 2016.

[7] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2010.

[8] K. Huebner and J. Zhang, "Stable symmetric feature detection and classification in panoramic robot vision systems," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.

[9] T. Gevers, S. Voortman, and F. Aldershoff, "Color feature detection and classification by learning," in *Proc. IEEE International Conference on Image Processing*, 2005.

[10] T. Bouchrika, M. Zaied, O. Jemai and C. B. Amar, "Ordering computers by hand gestures recognition based on wavelet networks" in *Proc. 2nd International Conference on Communications, Computing and Control Applications*, Merseilles, France, 2012.

[11] T. Bouchrika, M. Zaied, O. Jemai, and C. Ben Amar, "Neural solutions to interact with computers by hand gesture recognition," *Multimedia Tools and Applications*, vol. 72, no. 3, pp. 2949-2975, 2013.

[12] T. Bouchrika, O. Jemai, M. Zaied, and C. Amar, "Rapid and efficient hand gestures recognizer based on classes discriminator wavelet networks," *Multimedia Tools and Applications*, vol. 77, no. 5, pp. 5995-6016, 2017.

[13] X. Fu, J. Lu, T. Zhang, C. Bonair, and M. L. Coats, "Wavelet enhanced image preprocessing and neural networks for hand gesture recognition," in *Proc. IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, Chengdu, China, 2015, pp. 838-843.

[14] G. R. Murthy and R. S. Jadon, "A review of vision based hand gestures recognition," *International Journal of Information Technology and Knowledge Management*, vol. 2, pp. 405-410, 2009.

[15] J. Pyo, S. Ji, S. You, and T. Kuc, "Depth-based hand recognition using convolutional neural networks," in *Proc. 13th International Conference on Ubiquitous Robots and Ambient Intelligence*, Xi'an, China, 2016, pp. 225-227.

[16] H. Lin, M. Hsu and W. Chen, "Human hand gesture recognition using a convolution neural network," *IEEE*.

[17] S. Hussain, R. Saxena, J. A. Khan, and H. Shin, "Hand gesture recognition using deep learning," in *Proc. International SoC Design Conference (ISOCC)*, Seoul, South Korea, 2017, pp. 48-49.

[18] S. Pang, J. del Coz, Z. Yu, O. Luaces, and J. Díez, "Deep learning to frame objects for visual target tracking," *Engineering Applications of Artificial Intelligence*, vol 65, pp. 406-420, 2017.

[19] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia, PA, SIAM, 1992.

[20] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903-995, 1998.

[21] M. Lambert, A. Engroff, M. Dyer and B. Byer. *Empirical Mode Decomposition*. [Online]. Available: https://www.clear.rice.edu/elec301/Projects02/empiricalMode

[22] I. Aleksander and J. Taylor, *Artificial Neural Networks*, vol. 2, Amsterdam, North-Holland, 1992.

[23] B. Fritzke, "Growing cell structures — A self-organizing network in k-dimensions," *Artificial Neural Networks*, pp. 1051–1056, 1992.

[24] C. Junsheng, Y. Dejie, and Y. Yu, "Research on the Intrinsic Mode Function (IMF) criterion in EMD method," *Mechanical Systems and Signal Processing*, vol. 20, pp. 817-824, 2016.

Norah M. Alnaim is a faculty at the Computer Science Department in Imam Abdulrahman Bin Faisal University in Saudi Arabia. She had a master of science in computer information systems in 2012 from St. Mary's University, San Antonio, TX, USA. She is currently a PhD student in the Department of Electronic and Computer Engineering, Brunel London University, London, UK. She is working on gesture recognition and artificial intelligence field.

Maysam Abbod is a reader in intelligent system in the Department of Electronic and Computer Engineering at Brunel University London. Dr Abbod received BSc in electronic engineering in 1987, PhD in control engineering from Sheffield University in 1992. He has published 117 refereed journal papers, 16 chapters in edited books and 122 refereed conference papers. His main research interests are in intelligent systems for modelling, control and optimisation