

A Hybrid Background Subtraction and Artificial Neural Networks for Movement Recognition in Memorizing Quran

Anton Satria Prabuwono¹

Faculty of Information Technology, Universitas Budi Luhur,
Jakarta, Indonesia

Faculty of Computing and Information Technology in
Rabigh, King Abdulaziz University, Rabigh, Saudi Arabia

Ismatul Maula², Wendi Usino³

Faculty of Information Technology
Universitas Budi Luhur, Jakarta, Indonesia

Arif Bramantoro⁴

Faculty of Computing and Information Technology in Rabigh,
King Abdulaziz University, Rabigh, Saudi Arabia

Faculty of Information Technology, Universitas Budi Luhur,
Jakarta, Indonesia

Abstract—Movement change beyond the duration of time and the variations of object appearance becomes an interesting topic for research in computer vision. Object behavior can be recognized through movement change on video. During the recognition of object behavior, the target and the trace of an object in a video must be determined in the sequence of frames. To date, the existence of object on a video has been widely used in different areas such as supervision, robotics, agriculture, health, sports, education, and traffic. This research focuses on the field of education by recognizing the movement of Quantum Maki Quran memorization through a video. The purpose of this study is to enhance the existing computer vision technique in detecting the Quantum Maki Quran memorization movement on a video. It combines the Background Subtraction method and Artificial Neural Networks; and evaluates the combination to optimize the system accuracy. Background Subtraction is used as object detection method and Back propagation in Artificial Neural Networks is used as object classification. Nine videos are obtained by three different volunteers. These nine videos are divided into six training and three testing data. The experimental result shows that the percentage of accuracy system is 91.67%. It can be concluded that there are several factors influencing the accuracy, such as video capturing factors, video improvements, the models, feature extraction and parameter definitions during the Artificial Neural Networks training.

Keywords—Movement recognition; computer vision; Quran memorization movement; background subtraction; back propagation; artificial neural networks

I. INTRODUCTION

Computer vision has been subject to excel in the world of research and development of information technology industry. The change in movement beyond the duration of time and variations in the appearance of an object becomes an interesting topic to study in today's computer vision. The movement changes of the objects are supposed to be able to be

detected and recognized by an advance system. This recognition is useful to understand the object movement behavior through a video camera. To detect moving objects on a video, the target and trace of objects must be determined in a sequence of video frames.

There are some important obstacles to consider for detecting objects on a video, such as the appearance of a background or another object that is similar to the focus of objects [1], the shadows on top of the focus of the object, and noise caused by several highlights [2]. Despite these obstacles, the object recognition technique on a video has been widely used in several fields, such as supervision [3], robotics [4], [5], agriculture [6], health [7], sports [8], education [9], and traffic [10]. This research is proposed to contribute in the field of education.

Most object movement detection techniques on a video use Background Subtraction method, the combination of it or the modified one. In total, we found that there are at least 30 references; one of them is [11] that uses a method with an optimal result of object detection. There are other methods that we found in the literature such as Feature Extraction [7], [12], Histogram of Oriented Gradients [13], [14], Neural Networks [15], [16], Face Detection Module [17], Fuzzy Inference System [18], Naïve method [9], Probabilistic method [19], Directed Acyclic Graph [20] with various results of object detection.

Object analysis means classifying objects based on their behavior. Objects can be classified by various techniques, such as matching the pattern [3], [21], system learning and Artificial Neural Networks [22], control comparison and fuzzy logic [23], and others. The Artificial Neural Networks is considered as the best technique, since it can adapt and learn better when environment changes.

To the best of our knowledge, the study that uses Background Subtraction method as object detection method and Artificial Neural Networks as object classification method is rare. There is only one related study [24] that combines these two methods. However, the study concludes that the accuracy of the proposed system is 84.6%. The result contradicts with the theoretical textbook [25] which requires the minimum accuracy of the system in Artificial Neural Networks with Back Propagation algorithm to be 90%. We argue that the increase of the accuracy may serve to fill the gap.

Quran is the holy book for Muslim. It is the source of truth and guidance to human life (Chapter *Al-Baqoroh*: 2) [26]. For this reason, Muslims has complete obligations to practice the contents of the Quran. To ease the practice, it is suggested to memorize the contents as much as possible. There are some techniques in memorizing Quran used by Huffadz, i.e. a person who memorize Quran. The memorization is believed to start from the creation of Prophet Adam (Chapter *Al-Baqoroh*: 31) [26]. Nowadays, there are many methods that can be chosen in memorizing Quran. One of the methods is the sign method which is innovated by Huffadz who emphasizes in memorizing in different ways. The famous example of sign method is the Quantum Maki. Quantum Maki uses body movements to translate the meaning of the Quran verses word by word. Due to the nature of Quantum Maki method, we believe that it can be improved by computer vision.

This research aims at achieving system accuracy above 90% by combining two methods in order to better recognize the memorization of the Quantum Maki Quran. To achieve the goal, it is required that the distribution of data uses the hold-out method which has 70% training data and 30% test data. This method is important for stratified random sampling which randomizes data to produce proportional training and test data used in this research.

To speed up the process, however, this research is limited to the use of one chapter in Quran, namely *Al-Ikhlās* and the detected movement for memorizing Quran is in a standing position. Another limitation of this research is that the image is automatically cropped with predetermined size to standardize the process, although the consequence is that it is difficult to obtain specific parts from particular object for everyone. Moreover, the reliability of the data sets is important for movement recognition in memorizing Quran. Hence, this research uses a green color background to accommodate colorful video recognition in mp4 format.

II. THEORITICAL FRAMEWORK

Background subtraction is a method to seek particular objects in an image by comparing existing images with a background model. Modeling the background subtraction is sensitive to object motion recognition. Background subtraction detects any objects by separating background and moving foreground objects; which is computed by using the formula in

$$R(x,y) = I(x,y) - B(x,y) \tag{1}$$

Where R is the result of background separation, I is an object that is explored in position change, and B is the object background.

During the movement recognition, the Background Subtraction method is used as object detection and Artificial Neural Network is used as object classification. There are two features to represent the movement of objects during the memorization that involves body and hand movements. They are metric and eccentricity. Metric is a quantity that represents the roundness of particular object and eccentricity is the ratio calculation between two axes; i.e. major and the minor axis, in the shape of the object. In addition, the movement is recorded as a video before processed into the system.

Although the background subtraction method is considered as suitable for separating the background with moving objects, there is a possibility of another moving object or the shadow of the object that is detected as a foreground. The basic idea of this method is $|frame(n) - background| > threshold$. If there is a pixel n (the shadow of the object) that meets the equation, then the pixel is classified into a group of pixel objects, whereas the others are considered as background.

Artificial Neural Networks is a well-known artificial intelligence technique based on the mechanism of human neural networks. It is formulized by modeling human neural networks in mathematical formula with the three basic assumptions. First, the information is processed within several elementary neurons as elements. Second, the signals of information flow in and out through the path between neurons. This path is called connectors. Third, the connection between two neurons is given a weight that strengthens or weakens the signal.

The output is obtained from neuron by calculating activation function. This function; which is usually not a linear function, is computed based on the summation of the received inputs. A threshold is utilized to compare the amount of output. The structure of Artificial Neural Networks is conceptually represented in Fig. 1.

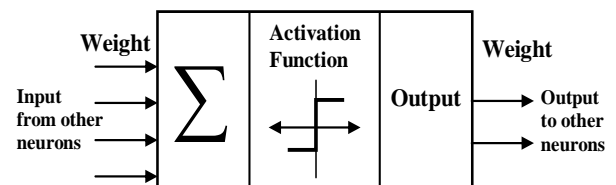


Fig. 1. Artificial Neural Networks Structure.

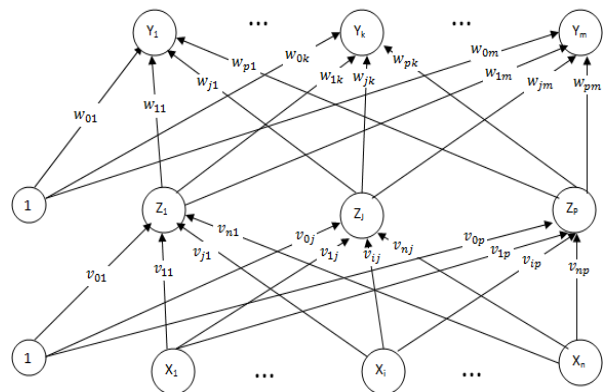


Fig. 2. Back Propagation Subtraction Process.

Back propagation is one of the algorithms in Artificial Neural Networks. This algorithm is commonly used to change the weights connected to neurons in the hidden layer. Back propagation uses an output error to change its weight in backward direction [25].

Back propagation has several units in one or more hidden layers. Fig. 2 illustrates the back propagation architecture with n input pieces (plus a bias), a hidden layer consisting of p units (plus a bias), and an output unit.

A. The Sign Movement of Quantum Maki

The Quantum Maki is a sign movement in memorizing Quran by exploiting human body, mainly by using hand movement. This movement is developed by a husband and wife Huffadz, who successfully memorized the whole chapters of Quran, in accordance with Chapter *An-Nahl* verse 78 and Chapter *Yaasiin* verse 65 [26].

This movement is slightly different from BISINDO [27] and considerably different from SIBI [28]. The main difference is that the Quantum Maki movement does not only use hand movements, but also other limb movements in particular circumstances. For an example, to illustrate the word *al-aqdaam* is by patting the foot with both hands. Basically, the Quantum Maki is based on a movement that generally represents the meaning of the Quran verse. To illustrate the movement, several examples are captured in Fig. 3.

In general, each word in Quran verse has its own meaning and movement. However, there are several verses that are too long to memorize by using the movement. In order to facilitate the memorization, one movement is used. Additionally, to understand and remember the movement easily, general movements are repeated for other verses. For an example, the word *robby* in the phrase of *birobbil falaq* is indicated by the hand up. In any verses and chapters, the movement for the word *robby* remains the same.



Fig. 3. The use of Sign Movement of Quantum Maki.

III. METHODOLOGY

To streamline the research, a conceptual methodology is required. Fig. 4 shows the methodology of the research which is divided into two stages. The first stage is data collection which starts from interviewing the creator of the movement to get the sample data from the primary source and to validate the result of the experiments.

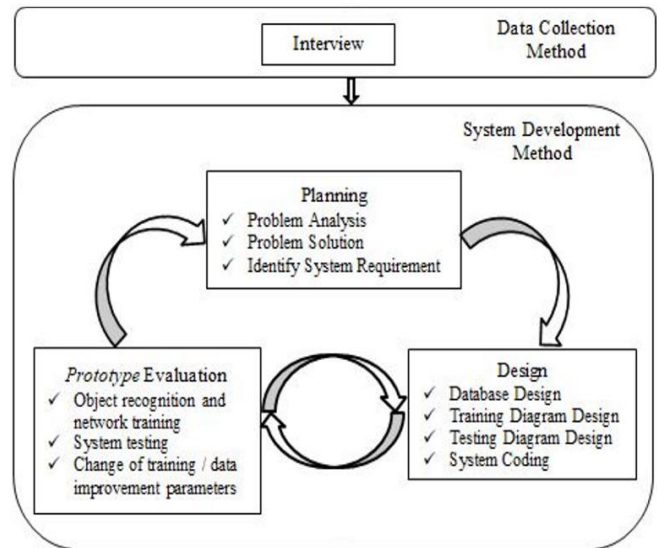


Fig. 4. Research Methodology.

The second stage is system development which emphasizes on the prototyping method to speed up the coding. To achieve the minimum value of error, the prototype is tested several times and its parameter is modified.

The training and testing diagrams are designed to detail system development. Fig. 5 shows training process diagram, while Fig. 6 shows testing process diagram. Several processes in these two diagrams are the same. For each process, the discussion is provided together with the result of its implementation to ease the understanding.

The training and testing diagrams start from video capturing which is conducted by three volunteers with a standing position. The result is inputted into the system together with its attributes, such as video dimension, frames, and duration. In background subtraction, object and background are separated through the following processes:

1) *Frame separation*: It is the process of separating each frame on a video in seconds. Frame attributes are separated and adjusted based on video duration.

2) *Image sequence*: It is the process of separation between objects and background. The selected object is used to process the segmentation, while the background is ignored. The RGB values of each frame are converted to HSV and their binary digit values are converted with XOR to eliminate negative value in truth table. The result is gray scale that is then transformed into binary digits by using the formula in

$$Binary(i,j) = \begin{cases} 1, & \text{if } Gray(i,j) > 0 \\ 0, & \text{if } Gray(i,j) < 0 \end{cases} \quad (2)$$

3) *Noise removal*: It is the process of removing noise on the object that is already separated from the background by using Median Filter formula, as in

$$f(x,y) = \text{median} \{g(s,t)\}_{(s,t) \in S_{xy}} \quad (3)$$

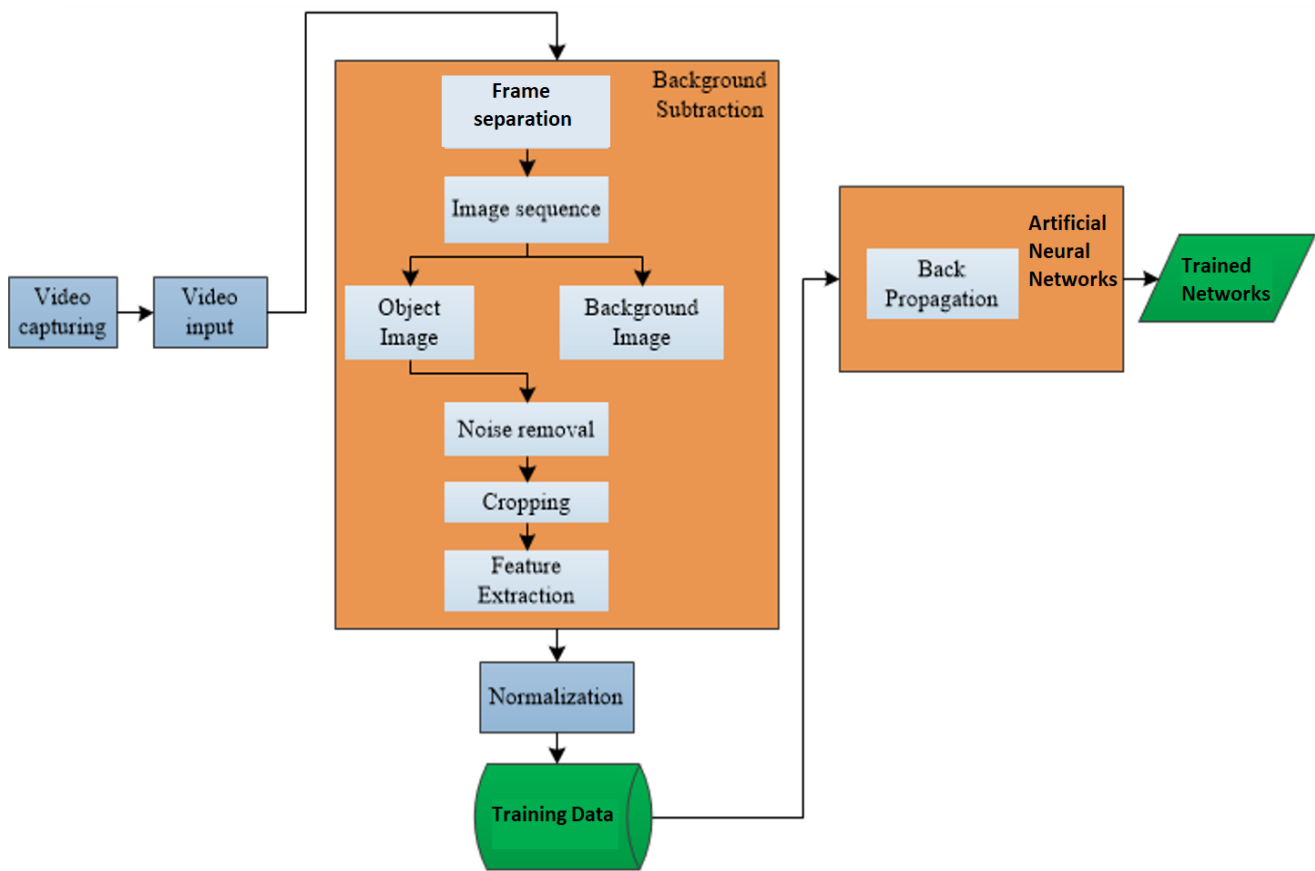


Fig. 5. Training Process Diagram.

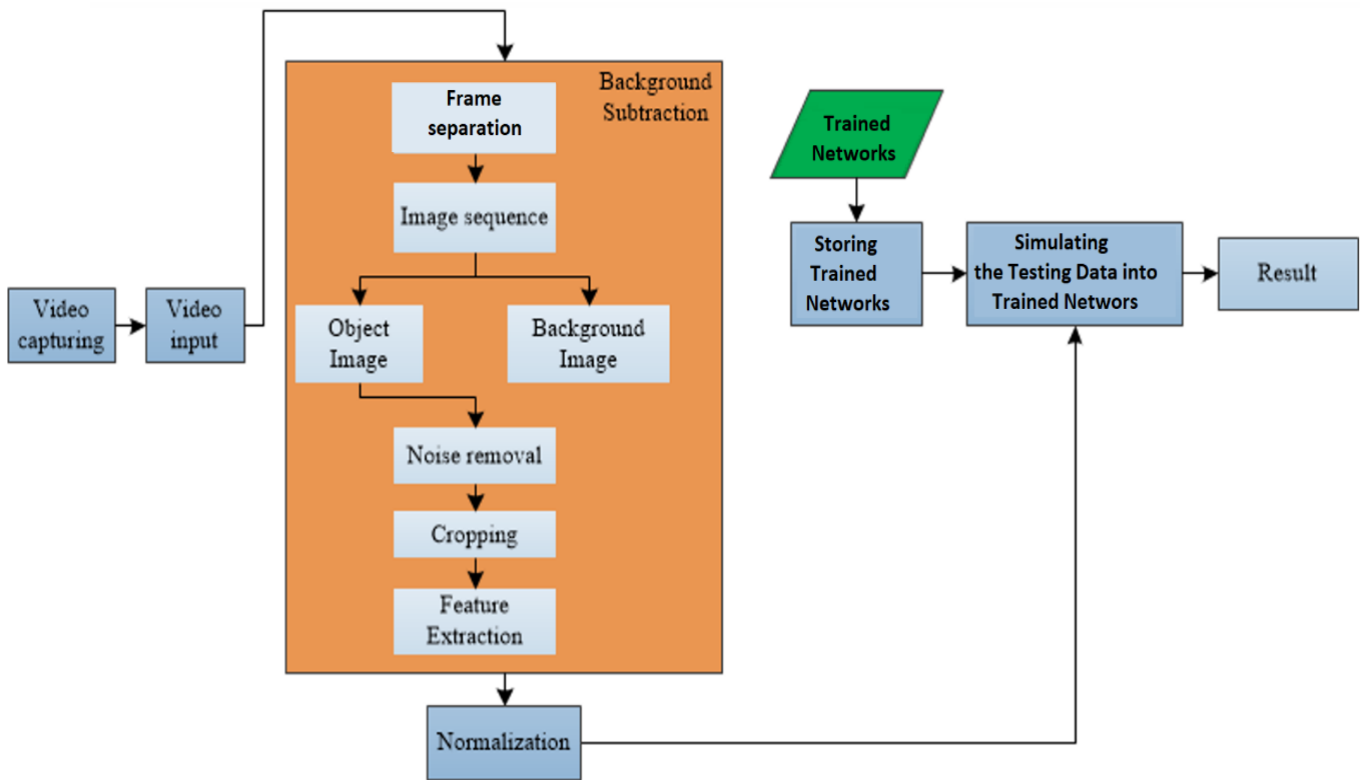


Fig. 6. Testing Process Diagram.

4) *Cropping*: It is an automatic process of taking specific parts from particular object. This research uses an arbitrary cropping size of 170x170 pixels to automate and speed up the whole process.

5) *Feature extraction*: It is a process of taking features from particular object. There are two features used in this paper: Metric and Eccentricity; which are also used in [29]. The mathematical equation for the Metric feature extraction is represented in

$$M = \frac{4\pi * A}{P^2} \quad (4)$$

Where M is metric value, A is object area, and P is object size.

The equation of Eccentricity feature extraction is in

$$e = \sqrt{1 - \left(\frac{b}{a}\right)^2} \quad (5)$$

Where e is eccentricity value, a is length of major ellipse foci, and b is length of minor ellipse foci.

The sample result of Background Subtraction processing is illustrated in Fig. 7.

Normalization is the process of normalizing the value of object features. The average value of the two features are selected and calculated by using the formula in

$$\bar{X}(m) = \frac{\sum fe(m)}{N} \quad (6)$$

Where m is a feature (*metric* and *eccentricity*), $\bar{X}(m)$ is a feature value in each frame, $\sum fe(m)$ is the number of feature value in each frame, and N is the number of frame.

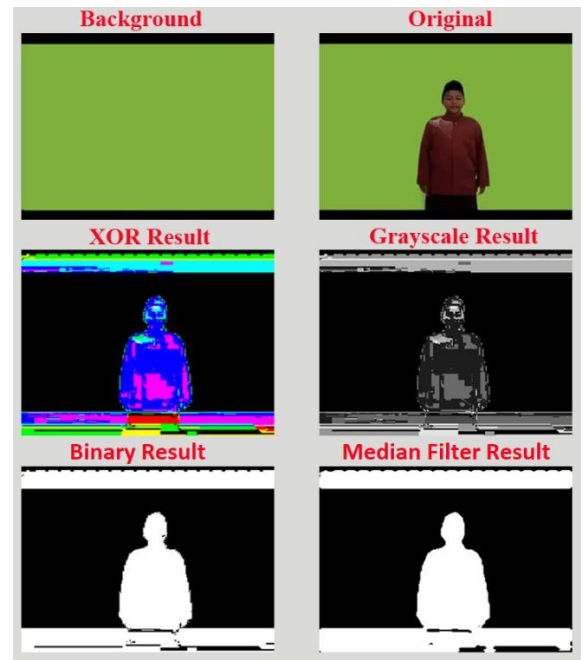


Fig. 7. Background Subtraction Output in One Frame.

IV. EVALUATION AND DISCUSSION

Artificial Neural Networks is the process of training and dataset classification which leads to be recognized by the system. Back Propagation is the algorithm that is used for training. There are nine videos available in this research. Six videos are used as training data and three videos used as testing data. The training data that have been processed are shown in Table I. The training parameters of the Artificial Neural Network used in this research are shown in Table II.

TABLE I. INPUT VALUE, TARGET DATA AND TRAINING DATA

Data	Target	1	2	3	4
1	Metric	0.5707734	0.4466702	0.5800725	0.5115093
	Eccentricity	0.8088200	0.7554581	0.7944091	0.7990654
2	Metric	0.5794906	0.4797728	0.5774935	0.5159322
	Eccentricity	0.8088360	0.7408384	0.7875225	0.7850985
3	Metric	0.5381068	0.4471097	0.5383233	0.4761612
	Eccentricity	0.7825329	0.7473362	0.7395582	0.7372278
4	Metric	0.5968912	0.4525911	0.5808415	0.5124966
	Eccentricity	0.8125884	0.7575972	0.7982806	0.7947707
5	Metric	0.5757907	0.4296351	0.5949379	0.5151999
	Eccentricity	0.8064739	0.7075989	0.7908770	0.7864578
6	Metric	0.5393527	0.4505028	0.5338850	0.4326231
	Eccentricity	0.7798373	0.7489395	0.7397976	0.7162581
Attribute	Verse	One	Two	Three	Four
	Meaning	Say, "He is Allah, [who is] One,	Allah, the Eternal Refuge.	He neither begets nor is born,	Nor is there to Him any equivalent

TABLE III. PARAMETERS OF BACK PROPAGATION TESTING

Parameter Name	Values
Hidden Layer	24
Looping (Epoch)	11500
Learning Rate	0.05
Error Value	0

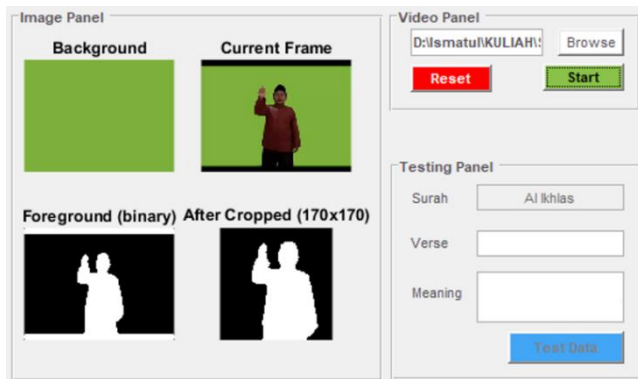


Fig. 8. Trial Process Interface.

The trial process has the same steps as the training process has. However, the normalized data are required to be simulated to the network. Fig. 8 shows the interface of the developed system to accommodate the trial process. The result is divided into current frame, background, foreground in binary, and after cropped image. This interface enables users to evaluate the trial result effectively. The trial results of other six testing data are available in Table III.

All testing data for each movement are tested into a network that has been previously trained. The trial was done 12 times with 11 movements were correctly detected and one movement was incorrectly detected. In other words, the percentage of accuracy system in detecting the Quantum Maki

Quran memorization movement is 91.67%. The accuracy percentage is calculated by using the equation in

$$accuracy = \frac{\text{amount of correct test data}}{\text{amount of test data}} \times 100\% \quad (7)$$

The equation is generally used by other existing works. Hence, it is not required to replicate and compare the experiment from other existing works. This is also because the implementation of this research is different from others in term of its unique requirement and process.

V. CONCLUSION

Detecting the movement during Quantum Maki Quran memorization has three main processes, namely the introduction of objects (Background Subtraction), the training of Artificial Neural Networks (Back Propagation) and the trial process.

The trial process is used to seek the accuracy of the system in recognizing the movement of Quran memorization on the data demonstrated by three different volunteers. The system accuracy results in 91.67% that fulfils the minimum requirement for Artificial Neural Networks. Several factors are considered as influencing the accuracy, such as video capturing, models demonstration, video editing, feature extraction, and Artificial Neural Networks parameters.

The future works include more detected object movements in memorizing the Quantum Maki Quran. It is also suggested to use other methods such as Gaussian Filter to detect the movement in memorizing the Quantum Maki Quran as comparison, to use other extraction features that are adapted to the complexity of the Quantum Maki Quran memorization movement, to compare the accuracy of the Background Subtraction method with other methods, to use other algorithms of training to classify the movement of objects in memorizing the Quantum Maki Quran, to group similar movements to identify better. There is also an expectation to include more chapters in Quran to improve the reliability of the system.

TABLE IV. TESTING DATA TRIAL RESULT

No	Video Name	Movement	Input		Trial Output	Result
			Metric	Eccentricity		
1	Faris-capture-1.mp4	1	0.5807734	0.8088200	0.9470	verse-1
		2	0.4466702	0.7554581	2.0296	verse-2
		3	0.5800725	0.7944091	3.1292	verse-3
		4	0.5115093	0.7990654	3.9514	verse-4
2	Mizar-capture-1.mp4	1	0.5794906	0.8088360	0.8730	verse-1
		2	0.4797728	0.7408384	2.1135	verse-2
		3	0.5774935	0.7875225	3.0652	verse-3
		4	0.5159322	0.7850985	3.9589	verse-4
3	Azzam-capture-1.mp4	1	0.5381068	0.7825329	0.7156	verse-1
		2	0.4471097	0.7473362	1.8845	verse-2
		3	0.5383233	0.7395582	3.1824	verse-3
		4	0.4761612	0.7372278	2.0850	verse-2

ACKNOWLEDGMENT

Thanks to Universitas Budi Luhur Indonesia for providing financial assistance to publish the research outcomes.

REFERENCES

- [1] M. Tiwari, "A Review of Detection and Tracking of Object from Image and Video Sequences," *Int. J. Comput. Intell. Res.*, vol. Volume 13, pp. 745–765, 2017.
- [2] R. kishor Gupta, "Object Detection and Tracking in Video Image," National Institute of Technology Rourkela, 2014.
- [3] J. Qu, Z. Liu, and W. He, "Video Pedestrian Detection Based on Orthogonal Scene Motion Pattern," *Math. Probl. Eng.*, vol. 2014, no. 1, p. 8, 2014.
- [4] H.-B. Zhang, F. Guo, M. Zhang, Y. Lin, and T.-C. Hsiao, "Human motion correction and representation method from motion camera," *J. Eng.*, vol. 2017, no. 7, pp. 370–375, 2017.
- [5] A. S. Prabuwno, M. A. Burhanuddin, and S. M. Said, "Autonomous contour tracking using staircase method for industrial robot. In Control, Automation," *Robotics and Vision, 2008. ICARCV 2008. 10th International Conference on*, pp. 2272-2276. IEEE, 2008.
- [6] P. Christiansen, L. N. Nielsen, K. A. Steen, R. N. Jørgensen, and H. Karstoft, "DeepAnomaly: Combining Background Subtraction and Deep Learning for Detecting Obstacles and Anomalies in an Agricultural Field," *Sensors (Switzerland)*, vol. 16, no. 11, p. 21, 2016.
- [7] X. Yang, X. Niu, Z. Liu, Y. Zhao, G. Zhang, W. Liang, and W. J. Li, "Accurate extraction of the self-rotational speed for cells in an electrokinetics force field by an image matching algorithm," *Micromachines*, vol. 8, no. 9, p. 17, 2017.
- [8] H. Tian, and W. X. Jun, "Sensors & Transducers Study on the Detection of Moving Target in the Mining Method Based on Hybrid Algorithm for Sports Video Analysis," vol. 181, no. 10, pp. 16–22, 2014.
- [9] G. N. Rao, P. Jagadeeswara Rao, R. Duvvuru, S. Bendalam, and R. Gemechu, "An Enhanced Real-Time Forest Fire Assessment Algorithm Based on Video by Using Texture Analysis," *Perspect. Sci.*, vol. 8, pp. 618–620, 2016.
- [10] N. Yaghoobi Ershadi and J. M. Menéndez, "Vehicle Tracking and Counting System in Dusty Weather with Vibrating Camera Conditions," *J. Sensors*, vol. 2017, p. 9, 2017.
- [11] I. Vujović and I. Kuzmanić, "Wavelet Energy and The Usefulness of its Powers in Motion Detection," *Adv. Electr. Comput. Eng.*, vol. 17, no. 2, pp. 61–70, 2017.
- [12] G. Yao, T. Lei, J. Zhong, P. Jiang, and W. Jia, "Comparative Evaluation of Background Subtraction Algorithms in Remote Scene Videos Captured by MWIR Sensors," *Sensors*, vol. 17, no. 9, 2017.
- [13] H. Tian, and W. X. Jun, "Sensors & Transducers Study on the Detection of Moving Target in the Mining Method Based on Hybrid Algorithm for Sports Video Analysis," *Sensors & Transducers*, vol. 181, no. 10, pp. 16–22, 2014.
- [14] A. P. Dias, M. F. Silva, and R. Guedes, "Identification of Marks on Tires Using Artificial Vision for Quality," *International Journal for Quality Research*, 2015.
- [15] R. Tapu, B. Mocanu, and T. Zaharia, "DEEP-SEE: Joint Object Detection, Tracking and Recognition with Application to Visually Impaired Navigational Assistance," *Sensors*, vol. 17, no. 11, p. 2473, 2017.
- [16] N. Shu, Z. Gao, X. Chen, and H. Liu, "Computational Model of Primary Visual Cortex Combining Visual Attention for Action Recognition," *PLoS one*, vol. 10, no. 7, 2015.
- [17] A. K. Skelin, T. G. Supuk, and M. Bonkovic, "Visual Peoplemeter: A Vision-Based Television Audience Measurement System," *Adv. Electr. Comput. Eng.*, vol. 14, no. 4, pp. 73–80, 2014.
- [18] J. Ren, Y. Chen, L. Xin, and J. Shi, "Lane Detection in Video-Based Intelligent Transportation Monitoring via Fast Extracting and Clustering of Vehicle Motion Trajectories," *Math. Probl. Eng.*, vol. 2014, p. 12, 2014.
- [19] R. I. Hammoud, C. S. Sahin, E. P. Blasch, B. J. Rhodes, and T. Wang, "Automatic Association of Chats and Video Tracks for Activity Learning and Recognition in Aerial Video Surveillance," *Sensors (Switzerland)*, vol. 14, no. 10, pp. 19843–19860, 2014.
- [20] S. Mangale, R. Tambe, and M. Khambete, "Object Detection and Tracking in Thermal Video Using Directed Acyclic Graph (Dag)," *ICTACT J. Image Video Process.*, vol. 8, no. 1, pp. 1566–1574, 2017.
- [21] N. A. Mailani, and V. Amrizal, "Comparative Analysis of the Accuracy of Backpropagation and Learning Vector Quantisation for Pattern Recognition of Hijaiyah Letters," *IEEE 6th Int. Conf. Inf. Commun. Technol. Muslim World*, 2016.
- [22] A. A. Mol, L. S. Martins-Filho, J. D. S. da Silva, and R. Rocha, "Efficiency parameters estimation in gemstones cut design using artificial neural networks," *Comput. Mater. Sci.*, vol. 38, no. 4, pp. 727–736, 2005.
- [23] A. H. Aghamohammadi, A. S. Prabuwno, S. Sahran, and M. Mogharrebi, "Solar cell panel crack detection using particle swarm optimization algorithm," *Pattern Analysis and Intelligent Robotics (ICPAIR), 2011 International Conference on*, vol. 1, pp. 160-164. IEEE, 2011.
- [24] E. Khoramshahi, J. Hietaoja, A. Valros, J. Yun, and M. Pastell, "Real-time recognition of sows in video: A supervised approach," *Inf. Process. Agric.*, vol. 1, no. 1, pp. 73–81, 2014.
- [25] R. J. Schalkoff, *Artificial Neural Networks*. New York: McGraw-Hill, 1997.
- [26] M. T. U. D. Al-Hilali, and M. M. Khan, *Translation of the Meanings of the Noble Quran in the English Language*. King Fahd Complex for the Printing of the Holy Qur'an, 2018.
- [27] T. Handhika, R. I. M. Zen, D. P. Lestari, and I. Sari. "Gesture recognition for Indonesian Sign Language (BISINDO)," *Journal of Physics: Conference Series*, vol. 1028, no. 1, p. 012173. IOP Publishing, 2018.
- [28] A. Rahagiyanto, A. Basuki, and R. Sigit, "Moment Invariant Features Extraction for Hand Gesture Recognition of Sign Language based on SIBI," *EMITTER International Journal of Engineering Technology*, vol. 5, no. 1, pp.119-138, 2017.
- [29] J. J. Athanasiou, and P. Suresh, "Systematic Survey on Object Tracking Methods in Video," *Int. J. Adv. Res. Comput. Eng. Technol.*, p. 24, 2012.