Proc. of the 4th IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)
28-30 November 2017, Putrajaya, Malaysia

Automated Daily Human Activity Recognition for Video Surveillance Using Neural Network

Mohanad Babiker[1], Othman O. khalifa[1], Kyaw Kyaw Htike[2], Aisha Hassan[1], Muhamed Zaharadeen[1]

[1]Department of Electrical and Computer Engineering

International Islamic University Malaysia

[2]Department of Information Technology

UCSI University, Malaysia

Hanode_88@hotmail.com, khalifa@iium.edu.my

***Abstract*— surveillance video systems are gaining increasing attention in the field of computer vision due to its demands of users for the seek of security. It is promising to observe the human movement and predict such kind of sense of movements. The need arises to develop a surveillance system that capable to overcome the shortcoming of depending on the human resource to stay monitoring, observing the normal and suspect event all the time without any absent mind and to facilitate the control of huge surveillance system network. In this paper, an intelligent human activity system recognition is developed. Series of digital image processing techniques were used in each stage of the proposed system, such as background subtraction, binarization, and morphological operation. A robust neural network was built based on the human activities features database, which was extracted from the frame sequences. Multi-layer feed forward perceptron network used to classify the activities model in the dataset. The classification results show a high performance in all of the stages of training, testing and validation. Finally, these results lead to achieving a promising performance in the activity recognition rate.**

Keywords- video surveillance, Human activities recognition, intelligent system, Neural Network, Multi-layer feed forward perceptron.

I. INTRODUCTION

The recognition and the analysis of the daily human activities is an attractive area for the researchers due to it is effectiveness and wide application in image processing, sign language, artificial intelligence, and human-computer interaction. Commonly, the target of video surveillance systems processing is to monitor the behavior and the activities of the human. Also, to observe any change in the movement of the human for the security and the administrative purposes [1] [2]. In video surveillance systems, there are three main types of systems: manual, semi-autonomous and fully autonomous [3].

In manual video surveillance systems, the analysis performs by a human. In semi-automated surveillance the intervenes of the Human in video analysis and making the decision is partial while in the fully automated system the video input, the analyzing, the processing and the treating with the suspect event completely independent from any human intervention.

The basic process of surveillance system begins with background subtraction, which known as the motion and object detection and it is a process of extracting the area of interest from constructed background [4]. Then, the data of the extracted region used as an input to detect the Object and to analyzing the behavior [4]. In addition, the application of surveillance system built on the behavior analysis, which involves recognition of activity or descriptive pattern among the object [5].

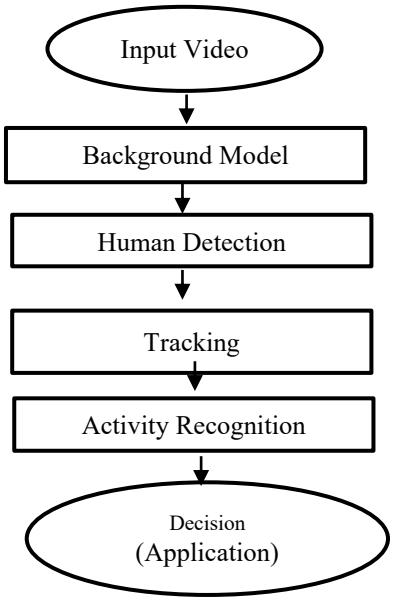

Fig 1. General Overview of Surveillance System

II. METHODOLOGY AND IMPLEMENTATION

In this work, a single static video camera used for recording the daily and usual human activities. The environment where the video sequence acquired was an indoor environment,

(which will be input to the system). Note that, there will not be any moving traffic in the background or moving leaves of trees. In addition, a single person were captured. The aim here is to test the ability of system to detect and segment a single human. The background were smooth illumination. Matlab were used in for implementation, testing and evaluation.

A. Preprocessing and extracting the features

The basic step in the designed system is the preprocessing operation, which it is the process of transferring the image to understandable form for the computer. Series of operations based on image processing techniques and theory used to enhancing the resultant image and handled it for further process in order to extract some numeric and descriptive information from the region of interest in the video [6]. These preprocess steps can be summarized as follows [6] [7]:

- Determining the background frame: in this paper, the video was recorded with static background model. Therefore, we set the first frame as a reference frame to use in further process.
- Background subtraction: which it is an operation for subtracting the background frame (reference frame) from the foreground image frame. Then, the subtracted object, which is the human body, handled for further process.
- Threshold of the image using Otsu method. This is important to reduce the intra class variance of the white and black pixels.
- The resultant image is an image with black and white color the background is represented by black color while the object will be in white color.
- The 2D median filter will perform to reduce the noise and distortion in the extracted image. The median filter is a nonlinear digital filtering technique, this operation often used to remove noise from an image or signal.
- The dilation and the erosion, which the background subtraction process enhanced by the morphological operation causes it.
- Perform the blob analysis operation to calculate the features of the white pixels, which are, represent the human body. These features known as bounding box, which is representing in four values (x, y, height, length). The rectangular box around the human body, area, and centroid, which is the center of gravity of the human will saved for next process.

After performing the preprocessing operation, a sheet of features database was built from the extracted information from the clear binary picture. For illustration, from the previous process, we calculated and analyzed the images to extract the values of the bounding box, centroid, and the area. For each kind of activities, these values have the same series of values, which will differ from the values of the other activities.

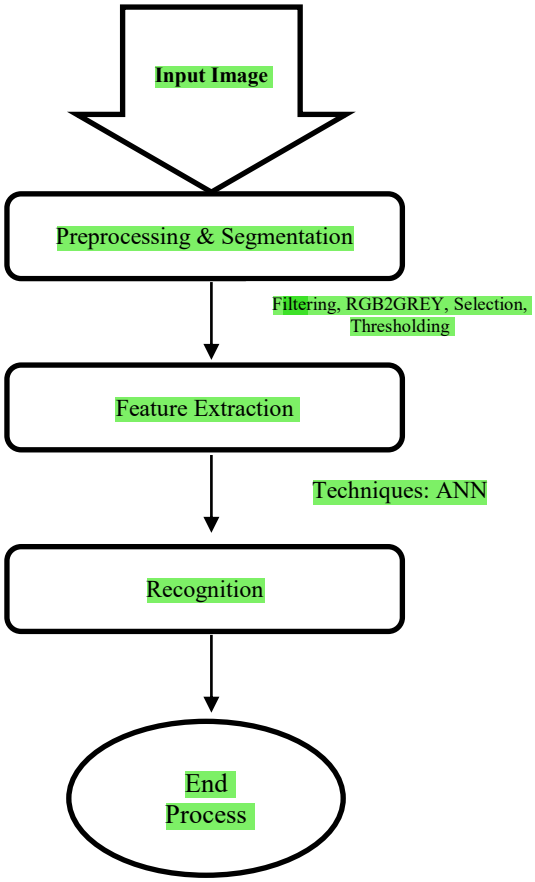

Figure. 2 Flow Chart of the Preprocessing Steps

B. Construct the neural network

After preparing the database sheet, a multilayer perceptron feeds forward neural network used to train the designed system by mapping the input data and connected it with a numeric relation with specific output through a number of neurons nodes. In each of 5 activities, 1000 samples were collected for each activity. The samples were divided into training samples (70% of samples), testing samples (15% of samples), and validation samples (15% of the samples). In training process, the information of the extracted features took as an input for the network. The target data is the information refers to the designated class in the input.

After training the system and recognize from it efficiency we test the system by the image before running the whole system in the video because what is work in the image surely can also work on the sequence of images which it is the content of any

video. Lastly, we used the graphical user interface provided in matlab to monitor the whole processes starting from background subtraction and next preprocess until the final detection tracking and recognition. Lastly, all of these processes performed using matlab 2014, and for final monitoring and observing for the whole process in one time graphical user interface was used for this. Figure 3 below shows the flow of the designed system.

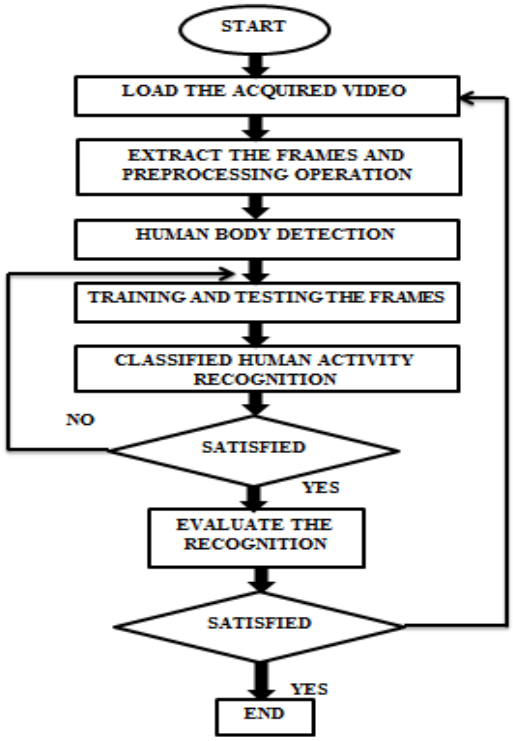

Figure. 3 Flow Chart of The designed system

C. *Evaluation criteria*

The main goal of this paper is to measure the accuracy of the designed system by observing the validity of the system to recognize the human activity. This known as recognition rate, which it can be analyzed through the terminology of the confusion matrix. Static error rate or the confusion matrix is one of the supervised learning terminologies that used to measure the performance of classification model [7]. This classification performed based on the number of correctly predicted from the classification set model and the number of wrongly predicted from the same classification set model [8]. This relation can obtain simply from the equation [8]:

$$Recognetion\ Rate(\%) = \frac{\sum_{i=1}^{N} ci}{N} \times 100\% \quad (1)$$

Where ci is the number of the corrected classified in the dataset samples, N is the total number of samples. Many statistical model and analysis program can present a very robust result and powerfully it has the capability to analyze and classify a huge and complex set of database [9] [10].

III. RESULT AND ANALYSIS

Through the simulation the designed system able to recognize five types of activities which are walking, sitting, boxing, hand waving, and laying the resultant binary image displayed by dimension (128*128). Then image handheld to next steps in the system to extract the database. The videos were used to extract the initial data is different from the data used in when we test the system.

Figure 4 shows some samples of the image preprocess steps. As it can be seen, mostly the process gives a clear binary image.

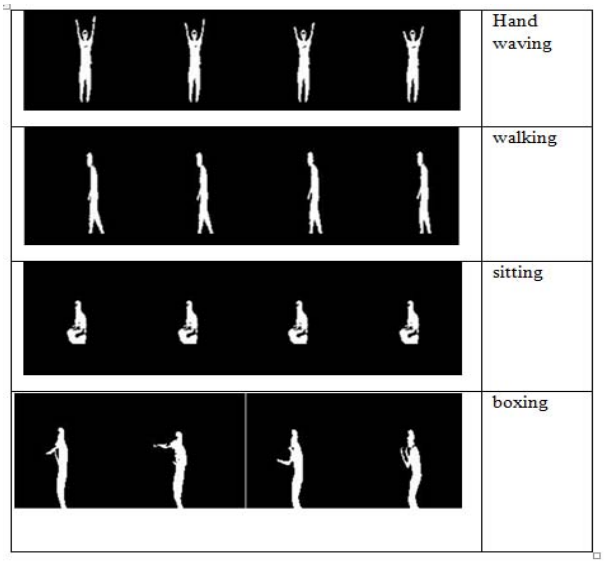

Figure. 4 samples of preprocessed images

A. *Result of the performance of the classification*

The tables below shows the confusion matrix of the number of corrected classify out of number of uncorrected classify with overall all accuracy in the stages of training testing and validation.

Table. 1the Classification Result In Training Stage

Activity	Total No Of Samples	No Of Correct Classify	No Of Wrong Classify	Recognition Rate %
walking	718	709	9	98.7
laying	870	687	192	78.3
Hand waving	524	512	12	97.7
boxing	678	678	0	100
Sitting down	703	703	0	100
Over all	3493	3289	213	94

Table 2 The Classification Result in Testing Stage

Activity	Total No Of Samples	No Of Correct Classify	No Of Wrong Classify	Recognition Rate
walking	125	121	4	96.8
laying	198	157	41	79.3
Hand waving	120	115	5	95.8
boxing	164	164	0	100
Sitting down	143	143	0	100
Over all	750	700	50	93.3

Table 3 The Classification Result in Validation Stage

Activity	Total No Of Samples	No Of Correct Classify	No Of Wrong Classify	Recognition Rate
walking	142	141	1	99.3
laying	192	156	36	81.3
Hand waving	104	101	3	97.1
boxing	158	158	0	100
Sitting down	154	154	0	100
Over all	750	710	40	94.7

TableI-III shows the performance result and the recognition rate in each stage of classification in the neural network. For more illustration, the whole dataset of the samples divided into training, testing and validation samples. The number of samples in each is 70%, 15%, 15% respectively from the total number, which is 5000 samples. In table 1, it can be seen that 709 of the samples are correctly classified as walking activity. This correspondence to 20.3% from all total of 3500 samples. While 9 samples are incorrectly classified as walking activity. This correspondence to 0.3% from all samples of the training data set. In addition, out of 718 samples, 98.7% was classified correctly, as walking activity and 1.3% incorrectly wrong classified. Same goes to the rest of the activities in the other stages. Note that all of these calculations based on the equation (1). Overall, the resulting shows a highly satisfactory level of accuracy in each stage. This due to basic calculated features from the initial images. Figures below show the comparison of the accuracy of the activities recognition in each level of classification

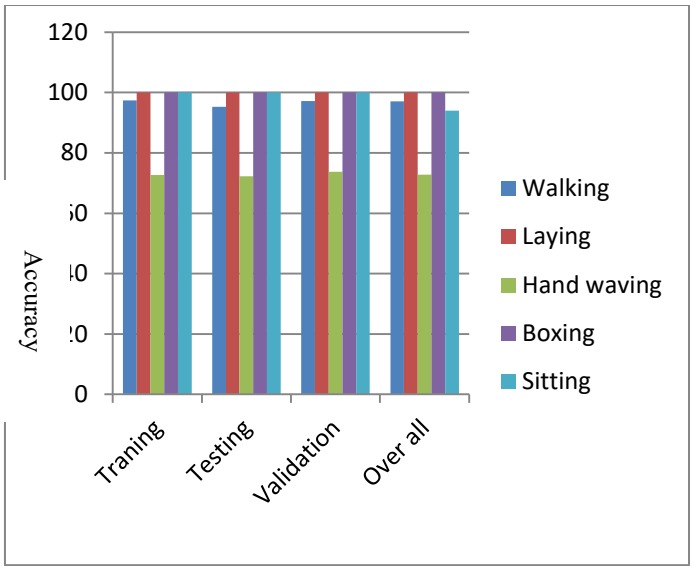

Figure. 5 The performance of the classification

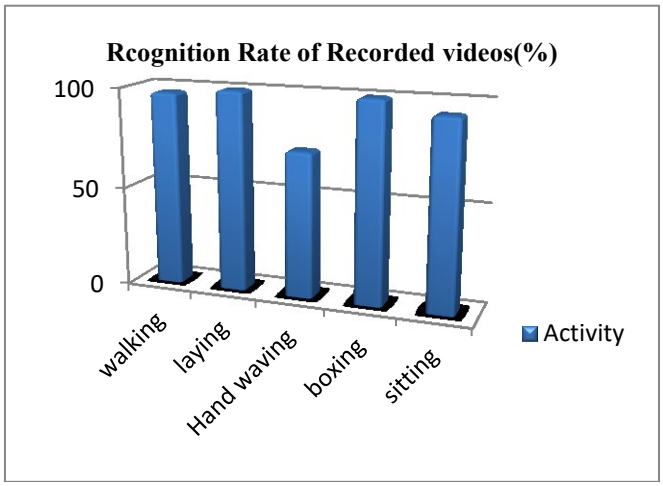

Figure. 6 The Recognition Rate of the Activities

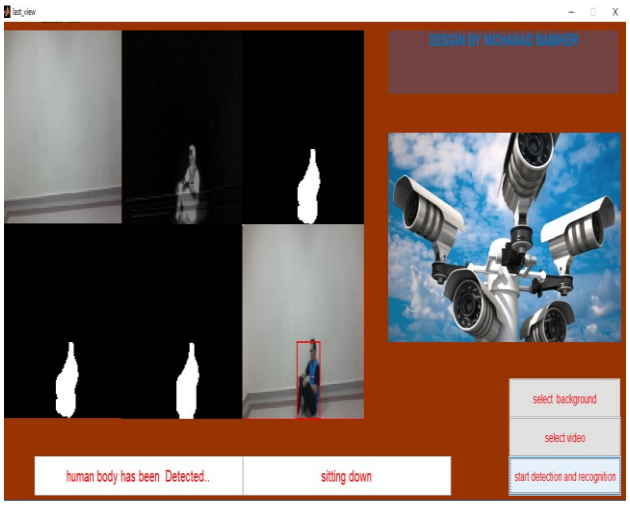

Figure. 7 Recognized Activity Displayed By GUI

IV. CONCLUSION AND FUTURE WORK

In this paper, an intelligent video surveillance system to detect and track the human body and to recognize the type of the daily human movement and activity is introduced. The designed system capable to work in an indoor environment with a single and static camera. The basic features of the human movement were extracted through series of image processing operations. the results showed a satisfactory level of accuracy in all stages of the training testing and validation. The overall recognition rate for the designed system was 94% For future work, the designed system can be tested in more complex activity which has a high similarity in the human shape and movement. In addition, it's recommended to testify the system on challenging video datasets such as KTH, and Weizmann video dataset.

REFERENCES

[1] Hossen, Muhammad Kamal, and Sabrina Hoque Tuli. "A surveillance system based on motion detection and motion estimation using optical flow." In Informatics, Electronics and Vision (ICIEV), 2016 5th International Conference on, pp. 646-651. IEEE, 2016.

[2] Htike, Kyaw Kyaw, and Othman O. Khalifa. "Human Posture Recognition: Methodology and Implementation." Journal of Electrical Engineering & Technology 10, no. 4 (2015): 1910-1914.

[3] Htike, Kyaw Kyaw, Othman O. Khalifa, Huda Adibah Mohd Ramli, and Mohammad AM Abushariah. "Human activity recognition for video surveillance using sequences of postures." In e-Technologies and Networks for Development (ICeND), 2014 Third International Conference on, pp. 79-82. IEEE, 2014.

[4] Kaur, Rajvir, and Sonit Singh. "Background modelling, detection and tracking of human in video surveillance system." In Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH), 2014 Innovative Applications of, pp. 54-58. IEEE, 2014.

[5] Meng, Binghao, Lu Zhang, Fan Jin, Lu Yang, Hong Cheng, and Qian Wang. "Abnormal Events Detection Using Deep Networks for Video Surveillance." In International Conference on Cognitive Systems and Signal Processing, pp. 197-204. Springer, Singapore, 2016.

[6] Peng, Qiwei, Gongyi Hong, Min Feng, Yuan Xia, Lei Yu, Xu Wang2and, and Yang Li. "Off-position detection based on convolutional neural network." (2016).

[7] Sapana, Miss, K Mishra, and K S Bhagat. 2015. “A Survey on Human Motion Detection and Surveillance.” International Journal of Advanced Research in Electronics and Communication Engineering 4 (4).

[8] Wang, Haixia, and Li Shi. "Foreground model for background subtraction with blind updating." In Signal and Image Processing (ICSIP), IEEE International Conference on, pp. 74-78. IEEE, 2016.

[9] Wang, Wen-June, Jun-Wei Chang, Shih-Fu Haung, and Rong-Jyue Wang. "Human posture recognition based on images captured by the Kinect sensor." International Journal of Advanced Robotic Systems 13, no. 2 (2016): 54.

[10] Zhang, Shugang, Zhiqiang Wei, Jie Nie, Lei Huang, Shuang Wang, and Zhen Li. "A Review on Human Activity Recognition Using Vision-Based Method." Journal of Healthcare Engineering2017 (2017).