Journal of Theoretical and Applied Information Technology
30th September 2021. Vol.99. No 18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

NEURAL NETWORK ASSISTED VIDEO SURVEILLANCE FOR MONITORING HUMAN ACTIVITY

[1] ADATI ELKANAH CHAHARI, [2]ABDULKADIR IYYAKA AUDU, [3]SAMUEL NDUESO JOHN, [4]ABDULLAHI ZABAIRU, [5] ETINOSA NOMA-OSAGHAE, [6]KENNEDY OKOKPUJIE

[1]Department of General Studies Education, Federal College of Education, Yola, Adamawa State, Nigeria

[2]Department of Computer Engineering, University of Maiduguri, Borno State

[3]Department of Electrical and Electronic Engineering, Nigeria Defence Academy, Kaduna

[4]Department of Chemical Engineering, University of Maiduguri, Borno State

[5,6]Department of Electrical and Information Engineering, Ota, Ogun State, Nigeria.

E-mail: [1]eachahari@fceyola.edu.ng, [3]samuel.john@nda.edu.ng, [5]etinosa.noma-osaghae@covenantuniversity.edu.ng

ABSTRACT

A video surveillance system is a useful tool for observing and monitoring human activities in such a way that guarantees protection against risks and danger from within or from the immediate environment. The video surveillance system is already existing technology, which is simply a recording facility. In this study, the surveillance system can record video in real time and transmit it to an existing parameter that then feeds it to an intelligent approach to recognize the human activity in the recorded video. This model consists of a video surveillance system with feature vector of human beings and trained using Artificial Neural Network (ANN) algorithms (Normal and Abnormal). The model was then used to classify human activities such as hand waving, running, jumping, walking, boxing and other environment events. The pre-processing step uses a continuous stream of live AVI video format with a frame rate of 25/30 frame per seconds and a collective total number of frames as 600fps. This work consists of one normal scenario with four activities of dataset with a recognition rate of 98.5%, six abnormal activities from KTH dataset with a recognition rate of 90.8% and five abnormal activities from Weizmann dataset with a recognition rate of 83.2%. selected to evaluate the performance of the model in an indoor environment. The result obtained was 90.8% accurate.

Keywords: *Neural Network; Video Surveillance; Human Activity*

1. INTRODUCTION

The steady increase of crimes and acts of terrorism such as events bombing, assassination and fraud operations made video surveillance attracts much attention. Video surveillance is particularly interesting due to its effective and wider used in artificial intelligence, image processing [1] and computer interaction with human. The aim of surveillance system is to process, observe any change of human information/human activities for security enquiries. The surveillance system in video is divided into manual, semi-autonomous and fully-autonomous [2] Co-ordination and accuracy is the global demand in smart building due to increase in crimes and terrorism. Hence the need for sophisticated surveillance system to be develop for effective monitoring to detect human activities using machine learning approach and the system could adjust to any changes in the background. The advancement in technology had also sowed growing concern for safety worldwide [3].

Several works have proposed various methods for detection of human motion and video surveillance [3] [4] [5] [6], although these methods have successfully demonstrated the ability to detect and classify human motion in each video frames, a good number of them are highly sensitive to any slight changes in the background, thereby reducing the accuracy of the result. A little brightness in the background changes the result, a dynamic background with little illumination and slit change of pixel value occurs [6]. Therefore, there is a need to design a system that can detect and accurately classify moving object in video surveillance and yet having the ability to learn and adjust to any slight changes in the background brightness, as long as it doesn't affect the accuracy of the result. In this work, an algorithm which can detect moving object (human beings) from a given real time video frames yet learning and adjusting to any slight changes in the background is proposed. To tackle the problem; Artificial Neural Networks is used as a classifier, to oversee the background models [7] and learn its motion pattern through self-organization [8]. This can be achieved through machine learning algorithm which involves the application of Artificial Intelligent (AI) – ANN [9].

In this study, the surveillance system can record video to existing parameter that pass it to intelligent system (ANN) for recognizing the situations in real time and respond automatically. The standard of evaluating video surveillance that have been affected with background changes in the environment either in static or in dynamics depending on the various human appearance, the clothing, and the tremendous change of light and depth shadow [5]. In manual analysis video surveillance is done by human and semi-automated surveillance is partially analysed by human decision making, whereas fully automated system a video input; is analyzing and processing then treated with suspect event complex independently from intervention of any human being [2]. The video surveillance system processes start background subtraction (motion) and detection of object with method of extracting the interested area using the background constructed in detecting the image and behavior analysis, which include activity recognition or pattern description in some object.

2. BACKGROUND

Many researches were conducted and applied in area of health care, life style and security application. Here is some related work among variant of human monitoring activity; [10], they proposed Back Propagation Neural Network Classifier and feature extraction, classification is performed by humans in consecutive frames. However, unable to detect and recognize people from side view. [11], Sparse Distributed LSTM Algorithm using Gaussian membership was proposed. Real-time data analysis. However, the output cannot be interpreted as a mathematical expression. Chowdhury and Tripathy (2014), Convolutionary Neural Network and High Altitude Acute Response (HAAR) classifier was proposed. Fast in processing time and increase in detection time. However, it has poor image quality. In [6], Video Surveillance is used to detected human using Silhouette-based method and Moore's algorithm was proposed. The system Successfully classified a human from a non-human. However, it is limited to understanding single human behavior and tracking. In [4], object detection of motion and tracking of video surveillance was proposed. The system provides a simple to understand method for object motion detection using background subtraction. However, background subtraction is not enough to accurately detect a moving object.

[12], K-mean clustering and Adaptive Background method was proposed. It improved the speed and efficiency of the human detection via camera and also reduced the false alarm notification. However, Hue Saturated Value depend on human position. [13], Edge Distribution of Gradient and Directional Pixels was proposed, better quality of edge distribution and accuracy. However, Complexity of the system increases with increase in number of vector size. [2] background subtraction, statistical method, temporal differencing and optical flow was proposed. Exhibit behavior understanding even after detection, tracking and classify human motion. However, Unable to accurately deal with fast moving objects. In [3], object motion detection and tracking for video surveillance was proposed. the system is efficient in object motion detection. However, the system was extremely sensitive to dynamic scene changes due to lighting and extraneous event.

In this work, we look at the problems associated to object tracking, such as background modelling, accurate classification of moving object. [6], Silhouette-Based method and Moore's algorism was proposed, human and non-human were classified successfully. However, change of the threshold value occur with distance from camera location. This work uses; background subtraction, angular Frame Difference and cascaded Classifier was proposed. Better recognition, good quality image, moving object is detected and not sensitive to dynamic and extraneous event. However, selecting of recording range was noted and create room for further research.

3. RESEARCH DESIGN

Figure 1 shows the general overview of surveillance system while figure 2 describes the system high-level architecture diagram of software and hardware components used in the gesture recognition system. As illustrated in the diagram there are four (4) Channel Digital Video Recorder

Journal of Theoretical and Applied Information Technology
30th September 2021. Vol.99. No 18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(DVR), four (4) IP Surveillance Camera's, a network switch or hub, Cat 6e Ethernet cables, Coaxial Cables, One Laptop or PC as Client and a server to host the Neural Network Service. Similarly, on the software aspect are the client application for accessing the surveillance cameras and the gesture recognition (Neural Network program) for processing. Images captured by the client application as motion is detected and fed into the neural network service as a live dataset for further processing. The flow chart (Figure 2) shows illustration of the steps of execution.

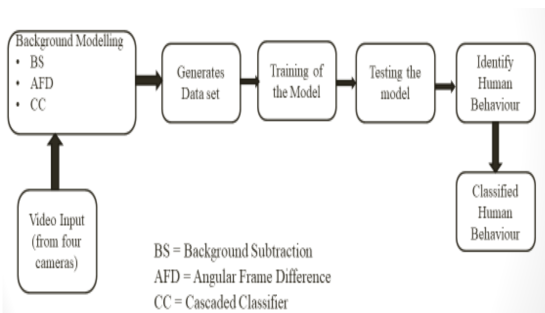

Figure 1: System Design of Neural Network Assisted Video Surveillance for Monitoring Human Activity

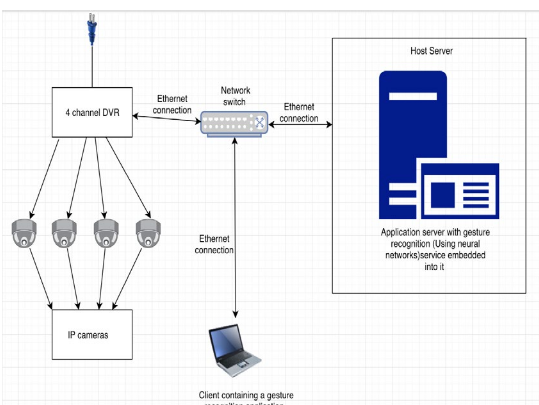

Figure 1: High Level System Architecture for Neural Network Assisted Video Surveillance for Monitoring Human Activity

Human activities recognition system:

a. Recording of the video
b. Detection and tracking process
c. Constructing of the neural network
d. Testing and deployment of Human Activities

3.1 Recording the Video

The camera system (acquiring) recorded video to simulate the real video surveillance system in the first step. Where four HD cameras were used to record the scene and stored in DVR Decode. An indoor environment was used to acquire sequence of video in the primary scenario, outdoor in the other scenarios and movement traffic in the background or moving object in the background was avoided. Therefore, single person was captured at a time to test the system performance on detection and segmentation of single human. The background was smooth illumination. Microsoft Visual studio was used for running the program for implementation, testing and evaluation.

3.2 Detection and Tracking Process

In detection and tracking, neural network is contracted through collection of data which is considered human activity recognition system with core steps for the simulation. The collection of data is preprocessing step, which defines process of transferring the image to understandable form on the machine after operations of enhancing and processing of extracting some numeric and descriptive information from the region of interest in the videos.

The second data collected is the analysis operation in pre-processing step to calculate features of the white pixels represented by the human body. The third step builds the database of the artificial neural network in extracted fractures, by using the collected data from the blob analysis, the human body was detected and the tracking in each frame was performed based on the extracted features which are used in the coding.

3.3 Constructing the Artificial Neural Network

ANN is an adaptive system that changes the structure based on external or internal information that passes through the network during the learning phase by mapping the input data and connect it with numerical relation that specify output through adjustment of the neurons of the hidden layer nodes. The samples are training samples with 70% and 30% for testing. The graphical user interface aimed to visualize the process from background subtraction and next preprocess until the final detection, tracking and recognition is achieved and display on graphic user interface. Tables 1 show the main algorithms and parameter to build the neural network.

Journal of Theoretical and Applied Information Technology
30th September 2021. Vol.99. No 18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Artificial Neural Network used for Proposed Video Surveillance System (Tutorials Point, 2017).

Number of the input	For sce1= 4, sce2= 16, sce3= 16
Number of out put	For sce1= 4, sce2= 6, sce3= 9
Number of hidden layers	10
Type of neural network	multilayer perceptron feeds forward
Training algorithm	Scaled conjugate gradient
Algorithm to evaluate the performance	Cross entropy

3.4 Testing and Development

ANN The system required equipment to deploy for the video surveillance to work properly which involve hardware components shown in Table 2 and software components shown in Table 3.

Table 2: Properties of Hardware used for Neural Network Assisted Video Surveillance System.

Equipment	Specification
Laptop Computer	Intel Core i5 8th Gen. 2.5-2.6Ghz 12Ghz RAM USB 4.0, VGA/HDMI Port.
Closed Circuit Camera	Camera diameter (71 mm), overall height (71 mm), camera thickness (30 mm) Video/Motion Sensing Capability Up to 30 frames per second
Digital Video Recorder (DVR)	Multiple Video Channels (In/Out) Wifi/3G Connectivity Internal Storage (HDD) Micro-USB (optional for power) Up to 1080p HD resolution
Siemens Cables/Coaxial Cables	Several Meters/Yards
Laptop Computer	Intel Core i5 8th Gen. 2.5-2.6Ghz 12Ghz RAM USB 4.0, VGA/HDMI Port.

Table 3: Properties of Software used for Neural Network Assisted Video Surveillance System.

Software	Specification
Microsoft Visual Studio, Math Lab.	Software development Platform
Dataset	UCI (University of California In-vane)
Python	Programming and Scripting

4. RESULTS AND DISCUSSION

The pre-processing steps for the acquired activities in an indoor environment were produced in two section of three scenarios of scenario one Normal and scenario two and three Abnormal. AVI video format, the frame rate is 25/30 fps, continue scene recording and are the properties use in these video.

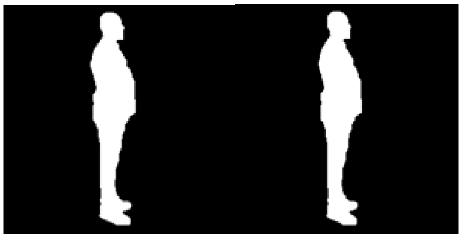

Figure 3: Pre-processing Normal Scenarios (standing)

4.1 Scenario One

In scenario one which is normal activity one (1) is used in different line of view with four cameras, provided in KTH dataset [8]in an outdoor environment. There are certain standard properties that are used in these videos, which include live stream AVI video format, the frames rate is 25fps and the total number of frames is 599, of which the results are been observed and calculated. Figure 3, show samples from the preprocessing steps. Figure 4 shows examples of the recognized activity.

Table 4: Recognition Accuracy (Standing Scenario).

Activity	Recognition Accuracy (%)
Standing front	98.0
Standing Left	100
Standing Right	100
Standing Back	96.0
Mean	98.5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 4: Sample of recognized activity for standing scenario

4.2 Scenario Two

Six (6) activities are used in this scenario which are abnormal, provided in KTH dataset [8] in an outdoor environment. The standard properties for these scenarios are used of continue scene video format, with frame rate of 25 frame per second and total frame of 599 that is used for results observation. We train a multi-class support vector machine using one –against-all method using Microsoft visual studio 2019. Figure 5. shows samples from the preprocessing steps.

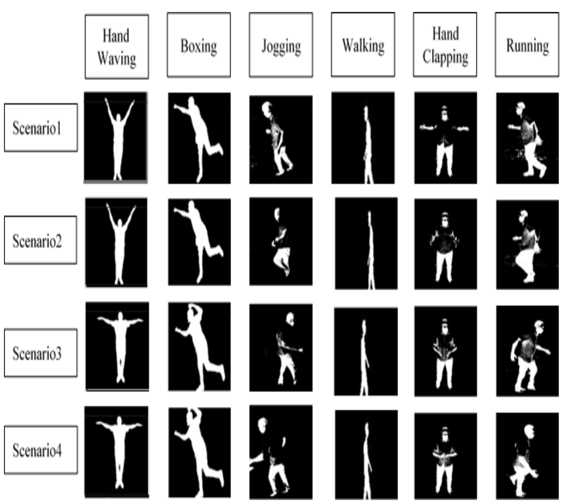

Figure 5: Pre-processing other Scenarios.

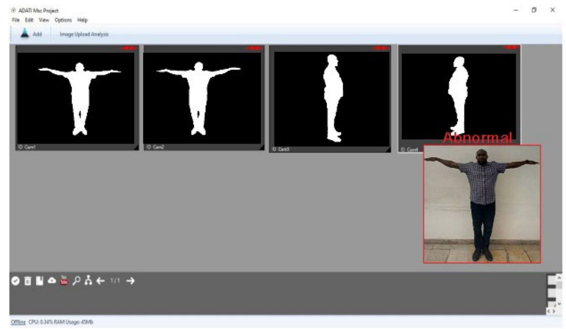

Figure 6: Sample of recognized activity for other scenario

The neural network [9] plays an important role in this aspect of accuracy and the recognition process when samples are being collected, each activity will be handled by neural network for classification and learning of different machine designated activities patterns including abnormal patterns as shown in Figure 6. Table 5 shows percentage recognition rate of training and testing samples.

Table 5: Recognition Accuracy (other Scenarios).

Activity	Recognition Accuracy (%)
Boxing	98.6
Hand clapping	85.6
Hand waving	97.8
Walking	85.7
Running	87.5
Jogging	89.6
Mean	90.8

However, in this scenario, the total number of samples was 418 and the total number of training sample was 70% while 30% for testing, and the accuracy of the classification percentage of all samples tested was 90.8%. Table 6 shows the details of a number of training and testing samples.

4.3 Scenario Three

Scenario 3, involves the algorithm of color segmentation which was extracted from the background of the Weizmann dataset (Figure 7) [8] and the extracted features.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 7: Weizmann Dataset Used for Scenario Three

Table 6: Recognition Accuracy (Weizmann Dataset).

Activity	**Recognition Accuracy (%)**
Bending	85
Jacking	84
Jumping	82
P jumping	92
Running	75
Side Walking	80
Skipping	70
Walking	83
One Wave Waving	90
Two Wave Waving	91
Mean	83.2

After training the system for recognition in neural network, its efficiency was tested. The system image was tested before running the whole system in Microsoft Visual Studio 2019 and it was discovered that what worked in the image system worked on sequence of images, with consistency in any video surveillance system.

4. CONCLUSION

This work elucidates the core knowledge of human activity recognition in video surveillance system approach that implements an automated human body detection, tracking and activity recognition system using neural network.

The designed system was evaluated, in visual studio that used methods of background subtraction to extract the background of the image. The experiment was group into three (3) scenarios; scenario one, normal and scenario two and scenario three abnormal. Scenario one consist of one activity viewed in four directions (standing front, standing left, standing right, standing back) that can be detected with high accuracy of 98.5%. Scenario two contain six activities namely (hand waving, jogging, walking, and running) which is KTH dataset video with 90.8% accuracy. Scenario three used the algorithm of segmentation in extracting background from the Weizmann dataset and also extract the features based on the bound box, then the centroid and thereby observing the Spatial Temporal Interest Point (STIP) to overcome the similarity in both scenarios, one and two activities with 83.2%. The overall result shows an acceptable level of accurate rate of 90.8% of the rate recognition and present the major limitations is the small size of the database used, that endeavor to complete human activity recognition system designed. The STIP features at the end are better in recognizing the human activity rather than the basic features of binary large object analysis because it is associated with distinguish capability with similar activities in a meaningful accuracy. Furthermore, the corresponding samples of different people tend to cluster and remain true when database is with promising not only for tracking but for recognition of human activity.

5. ACKNOWLEDGEMENT

The authors acknowledge the financial support of Covenant University.

REFRENCES:

[1] Culibrk, D., Oliver, M., Socek, D., Kalva, H., & Furht, B. (2007). A neural network approach to Bayesian background modeling for video object segmentation. IEEE International Conference. 1(1), 2007.

[2] Sapana, K.Mishra, and Kanchan, S. Bhagat (2015). A Survey on Human Motion Detection and Surveillance. International Journal of Advanced Research in Electronics and Communication Engineering, 4(4), 1044-1048.

[3] Sahasri, M., and Gireesh, C. (2017). Object Motion Detection and Tracking for Video Surveillance. International Journal of Engineering Trends and Technology, (12), 161-164.

[4] Sapana, K. Mishra, and Kanchan, S. Bhagat (2014). Human Motion Detection and Video Surveillance Using MATLAB. International Journal of Scientific Engineering and Research, 2347-3878.

[5] Mishra P.K., Saroha G.P., (2016). A study on video surveillance system for object detection and tracking. 3rd International conference on computing for sustainable global development, New Delhi, India, 1(3).

[6] 9. MohanRaj, T., and Nandini, A. (2014). Human Detection from Video Surveillance Using Silhouette-Based Method. International Journal of Advanced Information and Communication Technology, 5(5), 2348-9928.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[7] 1. Bouwmans T. (2014). Traditional and Recent Approaches in Background Modeling for Foreground Detection: An Overview Preprint submitted to Elsevier April 2014 pp. 2-10.

[8] Ke, S.-R., Thuc, H., Lee, Y.-J., Hwang, J.-N., Yoo, J.-H., and Choi, K.-H. (2013). A Review on Video-Based Human Activity Recognition. Computers, 2(2), 88-131. https://doi.org/10.3390/computers2020088.

[9] D. R. Kruse, "Scholarpedia,"[Online] Available:www.scholarpedia.org/article/fuzzy_neural_network. [Accessed 22 12 2017].

[10] Shilbayeh, N.F., and Al-Qudah, G.A. (2008). Face Detection System Based on MLP Neural Network. Recent Advances in Neural Networks, Fuzzy Systems & Evolutionary Computing. 238 – 243.

[11] Nwiabu, N.D., and Igbudu, K.E. (2017). Oil and Gas Pipeline Monitoring using Artificial Neural Network. International Journal of Computer Applications. 180(1), 8887 – 0975.

[12] Parmar, P., and Pandi, G.S. (2015). Performance Analysis and Augmentation of K-means Clustering, Based approach for Human Detection in Videos. International Journal of Engineering Development and Research, 3(2), 2321- 9939.

[13] Vishwakarma, D.K., Dhiman, A., Maheshwari, R., & Kapoor, R. (2015). Human Motion Analysis by Fusion of Silhouette Orientation and Shape Features. International Journal of Security and its Applications, 6, 2.