Balkan Journal of Electrical and Computer Engineering (BAJECE)
Volume 1 Number 1, September 2023

Intelligent Video Surveillance Systems: A Survey

Olayemi Olaniyi[1], Shefiu Ganiyuo[2] and S. J. Akam[3]

1 Department of Computer Engineering, Federal University of Technology, Minna, Nigeria, (e-mail: mikail.olaniyi@futminna.edu.ng)
2 Department of Information Technology, Federal University of Technology, Minna, Nigeria, (e-mail: shefiu.ganiyu@futminna.edu.ng)
3 Department of Computer Engineering, Federal University of Technology, Minna, Nigeria, (e-mail: sundayjames115@gmail.com)

Manuscript received Dec 20, 2022; accepted August 27, 2023.

***Abstract*— Over the years, the need for intelligent video surveillance has increased in other to enhance security and safety in the society. Government, private organization and individual need to make sure there properties are kept safe from intruders and as such intelligent video surveillance plays a key role in ensuring that this is achieved. Intelligent video surveillance is embedded with the capability of providing real time intelligent surveillance and also automatically provide analysis of video and image data without human operation. In the development of such systems, computer vision, machine learning and deep learning plays a vital role in achieving this. Therefore, this paper presents a survey on intelligent video surveillance system, overview of background concept and discussion on object detections and classification, tracking and deep networks. Also, this paper presents an efficient and faster object detection and classification techniques for intelligent video surveillance.**

Index Terms— **CNN, Deep learning, Detection, Video, Surveillance**

INTRODUCTION

The need for day to-day security cannot be over emphasized. Surveillance is a very important aspect of security and plays a vital role in ensuring that lives and properties are kept safe. In the early days, surveillance system relied on humans for its operations [1]. With the advent of video surveillance systems, governments, individuals and various organizations across society use this system to keep track of various activities for the sole aim of security and safety [2]. In today's smart cities, video surveillance is used for inventory control in retail outlets, security on corporate and educational campuses, and both security and demand monitoring on homes and rapid transit networks. Intelligent video surveillance systems are interdisciplinary systems that include electronic (sensing devices), pattern recognition and computer vision, networking, artificial intelligence, and communication [2].

Manual monitoring by human operators is an inefficient or even impractical solution because human resources are expensive and have limited capabilities. The goal of an intelligent video surveillance system is to automatically monitor people, property, and the environment without the need for human intervention [2]. As a result, this monitoring task entails automatically detecting and classifying objects (either humans or household pets), as well as performing additional analysis and taking actions. Image processing and artificial intelligence (deep learning) techniques are important in the development of intelligent video systems [3].

With advancements in deep learning, particularly Convolution Neural Network (CNN) and in computer vision applications, the accuracy of object detection and classification has improved dramatically for intelligent video surveillance. [4]. Neural network algorithms which offer state-of-the-art performance in classification and object detection are widely used in intelligent video surveillance for intrusion detection. This paper presents a review of literature for intelligent video surveillance and general overview of efficient and accurate method object detection and classification.

I. BACKGROUND OF RELATED WORK

This section provides a review of the literatures on different object detection and classification techniques as well as the advancements in neural network architectures. Recently, object detection and classification have drawn the attentions of many researchers into deep learning and its techniques. Several deep learning techniques based on CNN for real-time classification and recognition in computer vision have lately been proposed. Their performances, however, is dependent on the scenarios in which they are used [4].

A. Object detection and classification

Object classification refers to the task of assigning a label to an image. It includes the following techniques: k-nearest neighbors (KNN), Neural networks (NN), Naive Bayes classifier and CNN [2]. Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, animals, or cars) in digital images and video object. It also serves as a means of focusing attention on an object. It is a well-researched domains of which include face detection and pedestrian detection. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance. The ability to automatically detect and classify object is one of key component in intelligent video surveillance system. For a machine (computer), detecting object like human is a hard job due to wide range of possible appearance as result of changing articulated pose, clothing, lighting and background [3]. Deep convolutional neural networks (DCNNs) have proven very effective for computer

Copyright © INESEG

vision in object detection and classification tasks [6]. Furthermore, several deep learning techniques were recently proposed based CNN for real-time detection and classification in computer vision [1]. Computational complexity and object resolution requirements of CNNs limit their applicability in wide-view video surveillance settings where objects are small [4].

B. Supervised Classification

Supervised classification is a classification of digital images where the classes of objects on Earth's surface are known priory in certain limited areas of the image (areas that are called test areas or sites). These areas fall into patterns and then rules are developed which will be extended to parts unknown in the image. Supervised classification can work with several types of algorithms, the average minimum distance algorithm (minimum distance to means), algorithm parallelepiped (Multi-level slicing) and algorithm Gaussian of maximum similarity (maximum likelihood), [7]

$$y_i(x) = Inp(w_i) - \frac{1}{2\text{Inp}|\Sigma_i|} - 1/2(x - m_i)^y \sum_i^{-1} (x - m_i) \quad (1)$$

C. Unsupervised Classification

Unsupervised classification of digital images requires the creation of groups of pixels that represent geographic features, without previously knowing what is classified, and subsequently verifying the meaning of the pixels in the digital image researched. It is based on mathematical algorithms K-means and algorithm Iterative Self Organizing Data Analysis (ISODATA), in the present study having used ISODATA algorithm [5].

$$SD_{xyc} = \sqrt{\sum_{i=1}^{n} (\mu_{ci} - X_{xyi})^2} \quad (2)$$

D. Object Tracking

Object tracking is a deep learning application in which the program takes an initial set of object detections and creates a unique identification for each of the initial detections before tracking the detected objects as they move around frames in a video [6]. Object tracking has been an interesting field of research due to its challenges and importance. It is the main aim of intelligent video surveillance system. Recently, tracking by detection methods had emerged as immediate effect of deep learning with remarkable achievements in object detection. For example, [7] used CNN features for people tracking by detection. They used a model that is based on simple Euclidean distance. The result obtain shows that simple minimum Euclidean distance association performs well compared to SNN in most scenes. Fig 1 shows an example of the result obtained.

Fig.1. People tracking by detection [8]

II. Deep learning TECHNIQUES

A. *Convolutional Neural Network*

Convolutional neural network (CNN, convNet), is a class of deep neural network, widely applicable to image analysis. It was inspired by visual system's structure. The first computational models based on this local connectivity between neurons and on hierarchically organized transformations of the image are found in Neoncognition, it describes that when neurons with the same parameters are applied on patches of the previous layer at different locations, a form of translational invariance is acquired [32] The CNN-based architecture is built up on ConvNets with several layers of convolutional filters, RELU layer and pooling algorithms [33].

$$y(x,y) * g(x,y) = \sum_{i=-\infty}^{\infty} \sum_{j=\infty}^{\infty} f(i,j) * g(x-i, y-j) \quad (3)$$

CNN consist of three main types of neural layers; (i) convolutional layers, (ii) pooling layers, and (iii) fully connected layers.

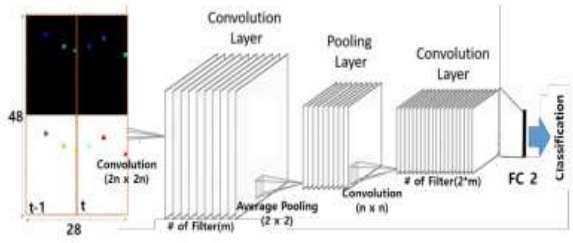

Fig.2. CNN Architecture [9].

B. *Fully Connected Layer*

Neurons in a fully connected layer have full connections to all activation in the previous layer, as their name implies. Their activation can hence be computed with a matrix multiplication followed by a bias offset. Fully connected layers eventually convert the 2D feature maps into a 1D feature vector. The derived vector either could be fed forward into a certain

Copyright © INESEG

number of categories for classification or could be considered as a feature vector for further processing [34]

$$y^{\beta} = \delta(Wy^{(d-1)} + b) \quad (4)$$

If the input to $d-1$ convolutional layer is of dimension $N \times N$ and the receptive field of units at a specific plane of convolutional layer d is of dimension $m \times m$, then the constructed feature map will be a matrix of dimensions $(N - m + 1) \times (N - m + 1)$. Specifically, the element of feature map at (i, j) location will be;

$$Y_{ij}^{\beta} = \delta(X_{ij}^{(d)} + b) \quad (5)$$

C. Convolutional Layers

Convolutional layers are considered the core building blocks of CNN architectures. The Figure 3 illustrates, convolutional layers transform the input data by using a patch of locally connecting neurons from the previous layer. The layer will compute a dot product between the region of the neurons in the input layer and the weights to which they are locally connected in the output layer.

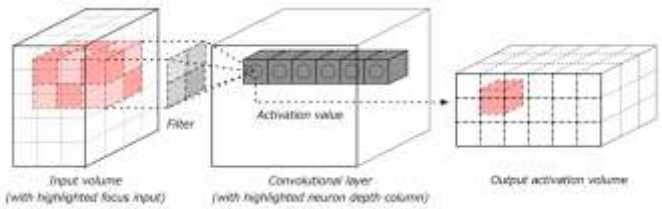

Fig.3. Convolution layer with input and output [10].

D. Pooling Layer

After the convolutional layer, a new layer called a pooling layer is introduced. After a nonlinearity (ReLU) has been applied to the feature maps produced by a convolutional layer. The largest pool is used for maximum pooling, The goal of maximum pooling is to reduce the size of an input image by down sampling it [35].

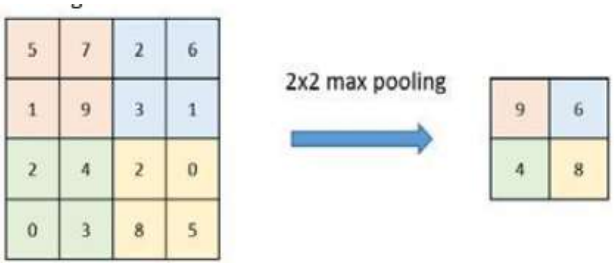

Fig.3. Pooling layer [35].

E. Long Short-Term Memory networks (LSTM)

The LSTM departed from typical neuron-based neural network architectures and instead introduced the concept of a memory cell. The memory cell can retain its value for a short or long time as a function of its inputs, which allows the cell to remember what's important and not just its last computed value.

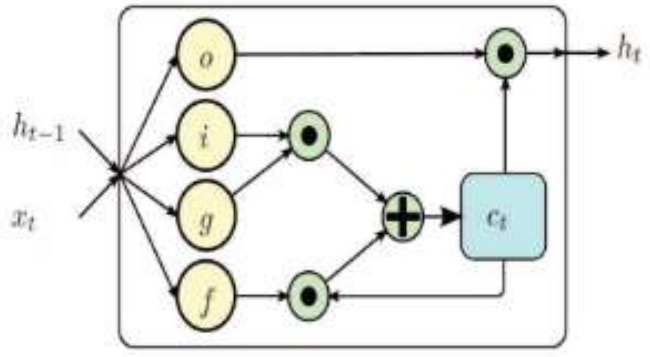

Fig.4. LSTM networks [12].

F. R Self-organized maps (SOM)

Self-organized map (SOM) was invented by Dr. Teuvo Kohonen in 1982 and was popularly known as the Kohonen map. SOM is an unsupervised neural network that creates clusters of the input data set by reducing the dimensionality of the input. SOMs vary from the traditional artificial neural network in quite a few ways (Jones, 2017).

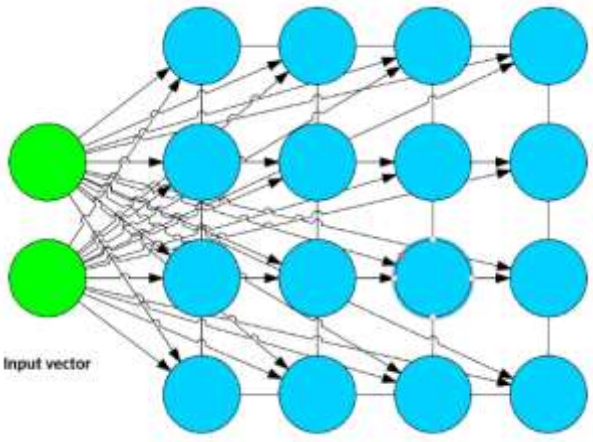

Fig.5. Self-organized maps [11].

III. Review of Related Works

Several researches have been carried out on intelligent video surveillance and object detection, classification, and tracking. Authors in [13] developed an object detector using multi regional convolutional neural network (RCNN) detectors. Object detection and tracking which plays a very important role in surveillance for traffic control, counting and public security field. The system adopted multi-detector model based on faster RCNN, a combination of CNN and Amulet is use to extract the raw feature from image, region proposal network (RPN) is use to predict the expected Region of interest (ROI). Multiple detection is used to detect the image. However, the system performed better when detection only vehicles.

Similarly, [14] designed and implemented an intelligent surveillance system with smartphone enabled. The system

Copyright © INESEG

uses a passive infrared (PIR) sensor and a microcontroller (MCU) attached to a smartphone through the MCU for motion detection. When motion is detected, video is captured and the footage is sent to the user via short message service (SMS). The surveillance record is stored in a cloud and the link to the record is also sent to the user via email. The developed system ensures efficient use of memory by storing the record in a cloud. It is cost-effective and also offers efficient energy use as the camera is only activated when motion is detected by the PIR sensor. However, the developed system cannot efficiently differentiate radiation changes between humans, household pets, or other animals.

Also, [15] developed an intelligent surveillance system for low-cost convolutional neural network (CNN) design. The developed system makes use of hardware accelerator known as Neural Compute Stick (NCS) with ROCK64 for high-speed calculation of images. A lightweight MobileNet network is use to extract the features and the classify images. The authors used the NCS to load a single shot multibox detector (SSD) network for human detection. Also, the Darknet architecture of You Only Look Once (YOLO) is used for extraction and classification of images and combine with SSD to create bounding box for the region of interest of the detected images. A simple mail transfer protocol was used to send email to deliver the detected object. However, the system has low human recognition accuracy and therefore cannot be used for other intelligent surveillance applications

Furthermore, [16] developed enhanced background subtraction algorithm for smart surveillance system using adaptive gaussian mixture technique. The smart system can efficiently detect motion and detect object by means of background subtraction with illumination change. However, the developed system cannot different between human and home pet.

In addition, [17] developed a Real-Time Action Detection in Video Surveillance using Sub Action Descriptor with Multi-CNN. The system presented a novel real-world surveillance video dataset and a new approach to real-time action detection in video surveillance system. The joint space of the sub-action descriptor was not considered. Also, more powerful temporal feature methods, such as a skin-color MHI or optical flow, and other deep architectures of CNNs are not considered.

Similarly, [10] developed an activity recognition using temporal optical flow convolutional features and multi multiplayer LSTM. The activity recognition framework for industrial systems proposed with a trained map CNN model help to select only the salient region that are activated for persons in the video frame which reduce verboseness and ambiguity of information in video frame. However, the system focuses on the activity recognition of a single person in a video frame.

Surveillance video analysis for store-base using deep learning techniques proposed by [36]. A skeleton recognition algorithm is adopted in place of object detection algorithm to conquer occlusion problem for gathering sufficient customer information and realizing crowd counting and density map drawing. For human tracking and counting, multiple human tracking algorithm and human re-identification (ReID) technology are adopted. However, the system was only trained to track only human object in the area of surveillance

Also, [3] developed people tracking system Using CNN Features. They represented each person with 4096 Faster-RCNN feature vectors, and the Euclidean distance method was used to calculate the distance between two feature vectors of each input pair. A pair is considered the same person if their Euclidean distance is a minimum. This is due to the assumption that convolutional features of similar objects generated by Faster-RCNN should be quite similar compared to features of dissimilar objects.

Furthermore, [19] suggested A General-Purpose Intelligent Surveillance System for Mobile Devices using Deep Learning. The developed system was divided into two: a detection and a classification module. The detection module combined background subtraction techniques, optical flow and recursively estimated density. The classification module is based on a convolutional neural network (CNN) used to classify objects into one of the seven predefined categories using a pre-trained CNN. However, the dataset is enormous for the targeted mobile device and so it become a problem to process in a real time.

In addition, [20] designed and developed an Edge Intelligence-Assisted Smoke Detection in Foggy Surveillance Environments. The system was developed using the architecture of convolutional neural network (CNN) for detecting smoke in video streams. Pre-trained MobileNet model was trained on ImageNet dataset which focus on trying to achieve accuracy and eliminating rate of false alarm in Foggy Surveillance Environments. However, the method can only be applied for smoke detection.

Similarly, [22] presented the use of Adaboost and CNN in crowded surveillance environment for people counting based on head detection. In this system three module were used to achieve people counting. The module includes: Two off-line training and one online detection stage. The first off-line training, Adaboost algorithm is adopted to learn a fast-cascaded head detector with Histogram of oriented gradients (HOG) feature. In the other off-line training, a CNN is trained using a new dataset gotten from the detection result in applying the cascaded head detector to the original dataset. Then in the online detection stage, the cascaded head detector is applied to the test image to get head proposal then they post-process. However, the method used is prone to uniform noise. Weak classifiers being too weak can lead to low margins and overfitting

In addition, [23] developed a Smart Surveillance as an Edge Network Service: from Harr-Cascade, SVM to a Lightweight CNN. The system uses histogram Oriented Gradient (HOG) and Support vector Machine (SVM) algorithm for fast and

Copyright © INESEG

accurate human detection. The system also uses Harr cascade, Harr-like feature made up of three shapes: two rectangular features, three rectangular features and four rectangular features alongside Lightweight CNN as the classifier trained with keras dataset. However, the Harr cascade has high rate of complexity, result obtain are highly tricky. The Haar Cascaded, HOG+SVM, GoogleNet and L-CNN required a lot of processing power and can make the system quite slow for video surveillance.

Also, [24] implemented a video structured description technology-based intelligence analysis of surveillance videos for public security applications. A pre-trained CNN architecture was adapted for tracking and re-identification of people and they analyze the result with CUHK03 dataset. Finally provided both manually cropped images and automatically detected bounding boxes with DPM detector, which respectively contains 13,164 images of 1360 pedestrians captured by six surveillance cameras. However, the summarization of the video stream is limited to 18 fps processing rate and cannot be combine with spectrum sensing technologies for smarter surveillance.

Furthermore, [11] presented an Efficient CNN based summarization of surveillance videos for resource-constrained devices. The study investigated deep features for shot segmentation and intelligently divide the video stream into meaningful shots. Deep features were extracted from two consecutive frames to determine whether the underlying frames belong to the same or different shot. Features were extracted from the fully connected layer (FC7) of CNN model which is trained using MobileNet architecture (version 2) on ImageNet dataset. However, the summarization of video stream is limited to 18 fps processing rate and cannot be combine with spectrum sensing technologies for smarter surveillance.

In addition, [25] developed a Kernel ELM and CNN based Facial Age Estimation. They introduced a two-level system for apparent age estimation from facial images. Then first classify samples into overlapping age groups. Within each group, the apparent age is estimated with local repressors, whose outputs are then fused for the final estimate. They use a deformable parts model-based face detector, and features from a pre-trained deep convolutional network. Kernel extreme learning machines are used for classification. However, the system cannot handle real and apparent age estimation task, and only uses a pre- trained convolutional neural network and cannot train a convolutional neural network by itself

Also, [26] designed and developed a surveillance system using CNN for face recognition with object, human and face detection. They developed a surveillance system using convolutional neural network (CNN). Region of object they considered in an entire image is picked by object detection and discriminate whether the area is human or not human or human face and then analyze his movement if the detected object is a human. However, among several frames, there are successful and unsuccessful ones. It is difficult to judge why object disappear.

Similarly, [27] implemented a vegetable category recognition system using Deep Neural Network. They implemented a caffe framework based on convolutional neural network (CNN) for the system classification and used Deep Neural Network (DNN) for the vegetable category recognition. However, as the number of iterations increases the performance of the system decreases.

Also, [28] presented an adaptive feature learning CNN for behavior recognition in crowd scene. A 3D scale convolutional neural network (3DSCNN) is implemented on crowd video scene, the 3D-CNN was used in a large-scale supervised crowd dataset which optimized convolutional architectures settings. The outcomes from 3DS-CNN captured information related to objects, scenes, and actions in a video, making them useful for different applications that do not fine tune the architectural setup. However, the system cannot handle action recognition with available dataset with variation in temporal and scale information.

In addition, [29] designed HOG-CNN based on real time face recognition. They used HOG-CNN model to recognize faces. Histogram of Oriented Gradient (HOG) was used as the feature extractor, also for detecting all the faces in the image and the CNN is used as the training algorithm for classifying the images. However, the system made consideration to only face detection of humans.

Furthermore, [30] designed and implemented an engineering vehicles detection based on faster R-CNN for power grid surveillance. CNN methods were divided into two categories, one is the two-stage methods based on region proposal and the other is the one-stage methods based on regression. The feature extraction part of these methods is done by the convolutional neural network. Some methods are based on region proposal such as R-CNN, Fast R-CNN and SPPnet, which adopt selective search algorithm to generate candidate boxes. Also, YOLO was used as the topmost feature map to predict confidences and bounding boxes for all categories over a fixed grid. SSD detects multiple categories by a single evaluation of the input image. However, the architecture used to train the system cannot handle complex application scenes.

Similarly, [31] presented an implementation of Machine Learning for Gender Detection using CNN on Raspberry Pi Platform. The implementation of the system is based on the architecture of convolutional neural network (CNN) and solution permits users to extract some relevant information from the visual data containing image labelling, face and landmarks detection, optical character recognition (OCR). Also REST API was used to interact with Google's cloud vision platform. The real-time implementation of the hardware as well as software solution was done on a Raspberry Pi 3 model B+ board with Pi Camera module. However, the system cannot identify human from their movement as well as their facial properties. From the foregoing, the ability of intelligent

Copyright © INESEG

video surveillance systems were to detect and distinguish between human intruder from home pet in the area of surveillance with accuracy and improved speed but with false alarm and failure to notify the user with the right intruder detected. This paper tries to tackle such limitations presented by [8], [2], [9], [10] by making use of raspberry pi and faster object detection and classification technique to improve video surveillance system.

IV. Proposed Design and Methodology

A. Methodology

To implement the proposed system for intelligent video surveillance, Faster R-CNN architecture will be considered the architecture is fast, accurate and suitable for detecting and classifying humans [37] object and home pets [35]. The Faster R-CNN architecture is divided into two modules: The Region Proposal Network (RPN) and a Fast R-CNN Detector. The RPN and the Fast R-CNN detector share the same convolutional layers. Faster R-CNN, by consequence, could be considered as a single and a unified network for object detection. To generate high quality object proposal, a highly descriptive feature extractor in the convolutional layers can be used. The Fast R-CNN detector uses many regions of interest (ROIs) as input. Then, the ROI pooling layer extracts a feature vector for each ROI. This feature vector will constitute the input for a classifier formed by a series of fully connected (FC) layers.

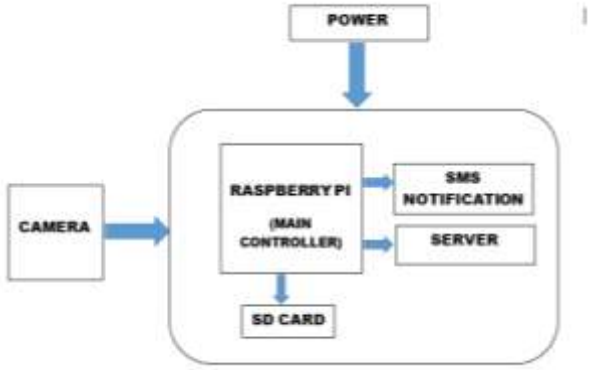

Fig.6. Proposed system block diagram

The embedded system of the proposed system, include Raspberry Pi 3B 8MP camera module, Raspberry Pi 3B as the main controller for all the object detection and programming for the whole system, SMS notification, buzzer (Alarm notification) and power supply.

The system comprises of five basic components

- Raspberry Pi 3
- Raspberry Pi 3 8MP Camera module
- SMS notification
- Power supply
- Raspberry microSD card.

The Raspberry Pi 3B is a basic module for processing images/videos, executing object detection on acquired video frames to detect objects. The Board has ARM cortex A53 clocked at 1.2GHz, 4000MHz Video Core IV multimedia GPU, 1Gb memory, power supply, HDMI, USB ports and other features.

The camera module takes in video stream then the raspberry pi 3B module implement the object detection on the captured frames. Figure 7 shows the flow diagram of the proposed system;

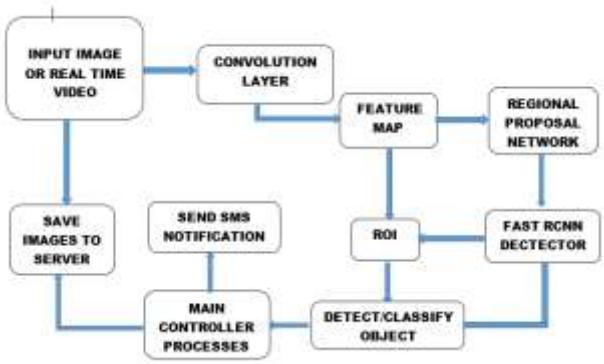

Fig.7. Flow diagram of the proposed system

The electronic components to be used are; Raspberry pi 3B, PiCamera module SMS notification and power supply. The camera and the power supply all the required inputs to the Raspberry pi 3B, while the SMS notification act as the output for the system. Once frames/video are acquired from the camera and fed into the Raspberry pi 3B controller, the image is being processed using the faster regional convolutional neural network as stored within the programmed Raspberry pi 3B.

V. Conclusion

Security remains a major concern to everyone, everybody wants to be protected from being attacked, and means to prevent this has been a challenge over the years. A lot of solution has already been put in place to tackle insecurity. This study intends to provide improvement on already existing motion and object detection techniques. The anticipated system shall intelligently detect object and by means of SMS notification sends alert the user the right action to be taken. The system proposed in this study is also cost effective and thereby an average citizen can afford, in other to enhance security in home and place of work.

References

[1] Y. Kurylyak, “A Real-Time Motion Detection for Video Surveillance System,” no. September, pp. 386–389, 2009.
[2] S. Ibrahim, “A comprehensive review on intelligent surveillance systems,” vol. 1, pp. 7–14, 2016.
[3] A. A. Shafie, F. Hafizhelmi, and K. Zaman, “Smart Video Surveillance System,” no. October 2018, 2010.
[4] B. Benjdira, T. Khursheed, A. Koubaa, A. Ammar, and K. Ouni, “Car Detection using Unmanned Aerial Vehicles : Comparison between Faster R-CNN and YOLOv3,” 2019 1st Int. Conf. Unmanned Veh. Syst., pp. 1–6, 2019.
[5] A. Deshmukh, S. Deshmukh, A. Zalte, K. Gaware, and P. S. S. Deore, “Intelligent video surveillance system using cnn,” no. 05, pp. 483–486, 2020.
[6] F. Bousetouane and B. Morris, “Fast CNN Surveillance Pipeline for Fine-Grained Vessel Classi fi cation and Detection in Maritime Scenarios,” no. August, pp. 242–248, 2016.

Copyright © INESEG

[7] H. M. Valentin and M. V. Boldea, “Using mathematical algorithms for classification of LANDSAT 8 satellite images,” no. March, pp. 1–6, 2015, doi: 10.1063/1.4912899.
[8] D. Chahyati, M. I. Fanany, and A. M. Arymurthy, “ScienceDirect ScienceDirect Tracking People by Detection Using CNN Features,” Procedia Comput. Sci., vol. 124, pp. 167–172, 2018, doi: 10.1016/j.procs.2017.12.143.
[9] H. C. Shin and J. Y. Lee, “Pedestrian Video Data Abstraction and Classification for Surveillance System,” 9th Int. Conf. Inf. Commun. Technol. Converg. ICT Converg. Powered by Smart Intell. ICTC 2018, pp. 1476–1478, 2018, doi: 10.1109/ICTC.2018.8539426.
[10] A. Ullah, K. Muhammad, J. Del Ser, S. W. Baik, and V. Albuquerque, “Activity Recognition using Temporal Optical Flow Convolutional Features and Multi-Layer LSTM,” IEEE Trans. Ind. Electron., vol. PP, no. c, p. 1, 2018, doi: 10.1109/TIE.2018.2881943.
[11] T. Hussain, K. Muhammad, A. Ullah, Z. Cao, S. W. Baik, and V. H. C. De Albuquerque, “Cloud-assisted multiview video summarization using CNN and bidirectional LSTM,” IEEE Trans. Ind. Informatics, vol. 16, no. 1, pp. 77–86, 2020, doi: 10.1109/TII.2019.2929228.
[12] I. I. Conference and E. Workshops, “Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW) 2017 10-14 July 2017,” no. July, pp. 585–590, 2017.
[13] W. Tan, “Object Detection with Multi-RCNN Detectors,” pp. 193–197.
[14] A. H. Sanoob, J. Roselin, and P. Latha, “Smartphone Enabled Intelligent Surveillance System,” no. c, pp. 1–7, 2015, doi: 10.1109/JSEN.2015.2501407.
[15] L. W. Yang and C. Y. Su, “Low-cost CNN Design for Intelligent Surveillance System,” 2018 Int. Conf. Syst. Sci. Eng., pp. 1–4, doi: 10.1109/ICSSE.2018.8520133.
[16] O. M. Olaniyi, J. A. Bala, S. O. Ganiyu, and P. E. Wisdom, “A Systematic Review of Background Subtraction Algorithms for Smart Surveillance System,” vol. 8, no. 1, pp. 35–54, 2020.
[17] C. Jin, S. Li, and H. Kim, “Real-Time Action Detection in Video Surveillance using Sub-Action Descriptor with Multi-CNN,” pp. 1–29.
[18] Q. Xu, W. Zheng, X. Liu, and P. Jing, “Deep Learning Technique Based Surveillance Video Analysis for the Store,” Appl. Artif. Intell., vol. 34, no. 14, pp. 1055–1073, 2020, doi: 10.1080/08839514.2020.1784611.
[19] A. Antoniou, “A General Purpose Intelligent Surveillance System For Mobile Devices using Deep Learning,” pp. 2879–2886, 2016.
[20] K. Muhammad, S. Khan, S. Member, and V. Palade, “Edge Intelligence-Assisted Smoke Detection in,” IEEE Trans. Ind. Informatics, vol. PP, no. c, p. 1, 2019, doi: 10.1109/TII.2019.2915592.
[21] S. Hargude and M. T. It, “i-surveillance : Intelligent Surveillance System Using Background Subtraction Technique,” vol. 1.
[22] C. Gao, P. Li, Y. Zhang, J. Liu, and L. Wang, “Author ’ s Accepted Manuscript People counting based on head detection combining environment Reference : To appear in : Neurocomputing,” Neurocomputing, 2016, doi: 10.1016/j.neucom.2016.01.097.
[23] S. Y. Nikouei, Y. Chen, S. Song, R. Xu, B. Y. Choi, and T. Faughnan, “Smart surveillance as an edge network service: From harr-cascade, SVM to a Lightweight CNN,” Proc. - 4th IEEE Int. Conf. Collab. Internet Comput. CIC 2018, pp. 256–265, 2018, doi: 10.1109/CIC.2018.00042.
[24] Z. Xu, C. Hu, and L. Mei, “Video structured description technology based intelligence analysis of surveillance videos for public security applications,” 2015, doi: 10.1007/s11042-015-3112-5.
[25] H. Kaya, H. Dibeklio, and A. A. Salah, “Kernel ELM and CNN based Facial Age Estimation,” pp. 80–86.
[26] Y. Byeon and S. Pan, “A Surveillance System Using CNN for Face Recognition with Object , Human and Face Detection,” pp. 975–984, doi: 10.1007/978-981-10-0557-2.
[27] Y. Sakai, T. Oda, M. Ikeda, and L. Barolli, “A Vegetable Category Recognition System Using Deep Neural Network,” 2016, doi: 10.1109/IMIS.2016.84.
[28] A. N. Shuaibu, A. S. Malik, and I. Faye, “Adaptive Feature Learning CNN for Behavior Recognition in Crowd Scene,” pp. 357–361, 2017.
[29] H. Ahamed, I. Alam, and M. Islam, “HOG-CNNBasedRealTimeFaceRecognition,” 2018 Int. Conf. Adv. Electr. Electron. Eng., pp. 1–4, 2018.
[30] X. Xiang, N. Lv, X. Guo, S. Wang, and A. El Saddik, “Engineering vehicles detection based on modified faster R-CNN for power grid surveillance,” Sensors (Switzerland), vol. 18, no. 7, 2018, doi: 10.3390/s18072258.
[31] M. H. Gauswami, “Implementation of Machine Learning for Gender Detection using CNN on Raspberry Pi Platform,” 2018 2nd Int. Conf. Inven. Syst. Control, no. Icisc, pp. 608–613, 2018.
[32] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep Learning for Computer Vision : A Brief Review,” vol. 2018, 2018.
[33] P. Wang et al., “Regional Detection of Traffic Congestion Using in a Large-Scale Surveillance System via Deep Residual TrafficNet,” vol. 6, 2018, doi: 10.1109/ACCESS.2018.2879809s.
[34] H. C. Shin and J. Y. Lee, “Pedestrian Video Data Abstraction and Classification for Surveillance System,” 9th Int. Conf. Inf. Commun. Technol. Converg. ICT Converg. Powered by Smart Intell. ICTC 2018, pp. 1476–1478, 2018, doi: 10.1109/ICTC.2018.8539426
[35] R. Article, “ANIMAL DETECTION USING DEEP LEARNING ALGORITHM,” vol. 7, no. 1, pp. 434–439, 2020.
[36] Q. Xu, W. Zheng, X. Liu, and P. Jing, “Deep Learning Technique Based Surveillance Video Analysis for the Store,” Appl. Artif. Intell., vol. 34, no. 14, pp. 1055–1073, 2020, doi: 10.1080/08839514.2020.1784611.
[37] H. Jiang and E. Learned-miller, “Face Detection with the Faster R-CNN,” 2017, doi: 10.1109/FG.2017.82

Copyright © INESEG