International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-8, June 2020

A Video Surveillance System for Unmanned Surveillance of Cantonment Boundary

Tarun Kumar, Sanjeev kr. Pippal, Aishwarya Mishra Allora Dudi, Vinod Chaudhary

Abstract*: *Since last few years, the Incidents that breach internal security and attack on the security forces are increasing day by day. These are security issues are becoming challenging to handle manually due to economical restrictions. This paper proposes an application for video surveillance to handle and monitor the intrusive incidents. The proposed application includes then human detection in no men's land around the boundary of the army cantonment. The human detection approach is proposed in this paper is developed with integration of the object detection using background subtraction, feature extraction using CNN and object classification into human and non human using SVM. The proposed approach achieves 95.6% accuracy in human detection. Application proposed in this paper is useful for unmanned surveillance of cantonment boundary.

Keywords*: *Object Detection, Classification, CNN, AlexNet, SVM.

I. INTRODUCTION

At present, internal security of any country is becoming the most challenging tasks. The intruders are targeting base camps of the securities forces. The developing countries like India, Sri Lanka etc are one of such countries where the incidents of such attacks are increasing in few years. The main causes behind these attacks are the lack of coordination between various security agencies and unavailability of state-of-art technologies for surveillance. Case Study of Pathankot Incident is discussed here in brief. On 2nd Jan 2016 attack at Pathankot airbase, there was total commotion. Causalities incurred were 4-6 attackers, 1 civilian, 7 security personnel (5 DSC, 1 IAF commando, 1 NSG). One of the reasons highlighted by Indian Express was glaring deficiencies in security which had facilitated the attackers, including trees, tall grass and shrubbery surrounding the walls of the air base [1]. Other Possible causes were:

- Dense civilian environment around IAF base.
- Complacency of security persons and systems.
- Lack of latest processes and technology and their adaptability.
- Lack of coordination.
- No use of sensors to monitor blind areas, neglect of dead grounds and non-removal of vegetation around base.
- Terrain around Air Base

Certain obvious problems of current scenario are sub-optimal resource utilization, inefficiency, human fatigue/error, lack of complete area / periphery coverage & slow response both from info dissemination and action point of view. In the developed countries, technologies are playing vital role in securities surveillance. These countries are using cameras with advance techniques for video surveillance. This surveillance includes regular traffic surveillance, border surveillance, monitoring the roads, buildings etc. The most of the surveillance techniques are based on the object detection using image processing techniques. The image based techniques are widely used in many surveillance and non surveillance applications. However, techniques based on image processing are having serious limitations of false detection, less accuracy in different environment conditions etc. The various researchers are integrating machine learning techniques to improve the limitations of the existing techniques. This paper focuses on surveillance of cantonment boundary by using the camera and image processing technique. The main objecting of this paper is to proposing an approach for detection of presence of human in surveillance area near the boundary. This because the terrorist are intruding into base camps of security agencies from the various no men's land areas near boundary. The manual patrolling of such areas are carried out periodically and continues monitoring such areas is require more economical support. The developing countries where security agencies have to monitors the security on very larger sector in the city, such continues monitoring of boundary not possible. This paper solved the issue by deploying the human detection approach on no men's land area near border. The proposed technique is integration of object detection technique and deep learning technique that detects the presence of human the surveillance zone accurately.

II. RELATED WORK

In the image or video based surveillance, the object detection techniques are mainly used to detect the presence of said object. This object may be vehicle, human, animal etc. The detection of the object is not enough for accurate surveillance, this require classification of the object also.

Revised Manuscript Received on June 30, 2020.
***** Correspondence Author
Tarun Kumar*, CSED, G. L. Bajaj Institute of Technology and Management, India. Email: ertarun123@gmail.com.
Sanjeev Kumar Pippal, CSED, G. L. Bajaj Institute of Technology and Management, India. Email: sanpippalin@gmail.com.
Aishwarya Mishra, CSED, G. L. Bajaj Institute of Technology and Management, India. Email: aishwarya.mishra1609@gmail.com.
Allora Dudi, CSED, Engineering College Bikaner, Bikaner India. Email: dudi.allora@gmail.com .
Vinod Chaudhary, CSED, Engineering College Bikaner, Bikaner India. Email: vinod.ecb@gmail.com .

For example, in traffic surveillance it is necessary to recognize the vehicle after detection of it. The various classification techniques such as support vector, k-nearest classification, cascade classifiers, neural network etc. are used for such classification. There exist wide variety of techniques for object detection such as frame differencing, optical flow etc. Kumar and kushwaha [2] and Singh and Kumar [6] use background subtraction techniques for detection moving vehicle. The detection of an object requires some features to identify the desired object after background subtraction. Authors also used their techniques for security surveillance by monitoring the traffic. The Singh and Kushwaha [3] used optical flow technique for detection of human in the border surveillance. Authors used haar like features and skin color modeling for detection of human in the border. Temporal frame differencing is the simplest and fastest method, in which pixel-wise difference is computed between two consecutive frames. The moving foreground regions are determined using a threshold value [7]. Street-parking vehicles are detected using frame differencing in [8], with noise suppression. Motorcycles are detected in [7]. However, using more information is preferable; the use of three consecutive frames improves detection as in [8]. In which dual inter-frame subtraction are calculated and binarized followed by a bitwise AND to extract the moving target region. The classification of the detected object is the most important and challenging task. The accurate classification of detected object increases the overall efficiency and reliability of the system. The various classification techniques ate exists today requires training to create a working model of that classification. The features of the objects and various sample images of such objects are used to train the classifiers. In these classifiers, the support vector machine and neural network are widely used in various classification techniques. Vohra and Tiwari [9] use spatial features of objects for recognition of various engineered objects using SVM. The authors train the SVM using spatial feature and developed the model. Jagadeesh and Patil [10] uses Gaussian mixture model to train the SVM for classification of human activities in various poses. Jagtap and Kokare [11] use ANN for classification of human age by using the facial features. Sigh and Kushwha [3] uses haar like features to develop the trained model of cascade classifiers. Authors are used color model to recognize the presence of human with skin color model. All the classifiers require labeled dataset along with the knowledge of specific features. The research needs to define a method to extracts the features so that classifier can be trained. Apart from these classifiers, deep learning technique is trending in research community. This deep learning includes power of the convolution neural network in the image based classification. The CNN does not require feature dataset for training purpose instead of this generates the feature vector so that each object can be categorized on their respective categories. However, CNN requires very large dataset for training and feature extraction. The accuracy of deep learning based classification is very high. This paper proposes an approach for detection of the human in surveillance zone by using the deep learning.

III. PROPOSED WORK

This paper proposed a video surveillance system for monitoring the army cantonment boundaries. The approach is based on object detection using image processing technique, feature extraction using CNN and human classification using SVM[5]. The approach is designed to in order to improve the reliability of the system by improving the true classification rate. The proposed surveillance zone setup is illustrated in the figure 1.

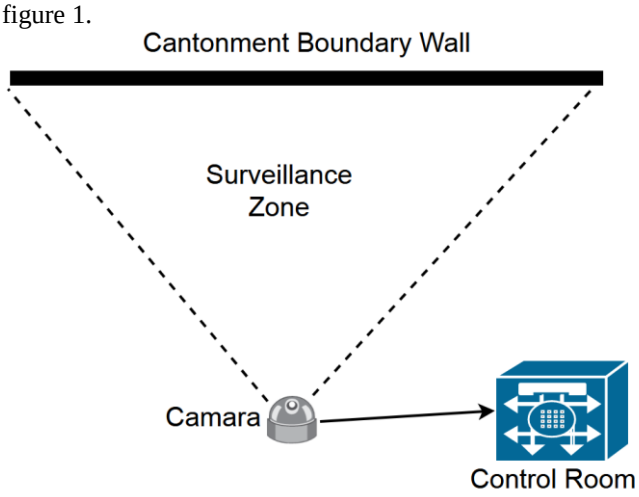

Figure 1: Surveillance Zone Setup

In figure 1, a pole mounted camera is placed towards the boundary wall of the cantonment. This camera is connected with the control room through the network. Range of the camera is the surveillance zone of the proposed system. The camera records live video of the boundary wall continuously and transfer the video stream to the control room. A control room need to place in the cantonment where the actual processing of the video is carried out. Servers process the video stream and alert the base camp in case of detection of the human in surveillance zone.

The proposed system performs the objective different phases in the control room. These phases are as following:

A. Object Detection using Background Subtraction
B. Feature Extraction using CNN
C. Classification using SVM.

A. Object Detection using Background Subtraction

The video stream obtained from the camera is processed first by the frame extraction module. This module extracts series of frames from the video stream and stores these frames into the database. Later these frames are processed as image one by one during the background subtraction. The background subtraction requires a background image of the surveillance zone in the initial of the processing. In order to detect the presence of any object in surveillance zone, a temporal difference of extracted frame and the background image is carried out. The result of this background subtraction is and image that shows the presence of object in higher frequencies. This resultant image is then converted into binary image using thresholding operation. Now the pixels having high frequencies represents the objects and contour of the objects are extracted and then passed to CNN.

B. Feature Extraction using CNN

Detected object in surveillance zone may be any objects such as any animal, bird, vehicle or human etc. To ensure that the detected object is human, the object classification is requires. The CNN is used to classify the object.

In the proposed system object is classified into only two categories namely human and non-human. The CNN is trained for such types of images. In the proposed approach, the 1000 images of human in different postures, angles, activities and 1000 images of different objects such as animals, birds, vehicles, tree, etc. are used for training the CNN. In this paper, AlexNet [4] CNN with SVM classifier is used for feature extraction. In AlexNet, process in input RGB image of size 227x227, with 11x11 filter at the first convolution layer. The stride in this layer is set to 4 thus the dimensions of the output of this will be 55x55. In pooling layer, max pooling with stride 2 is applied by using 3x3 filters. The complete CNN network architecture is shown in figure 3.

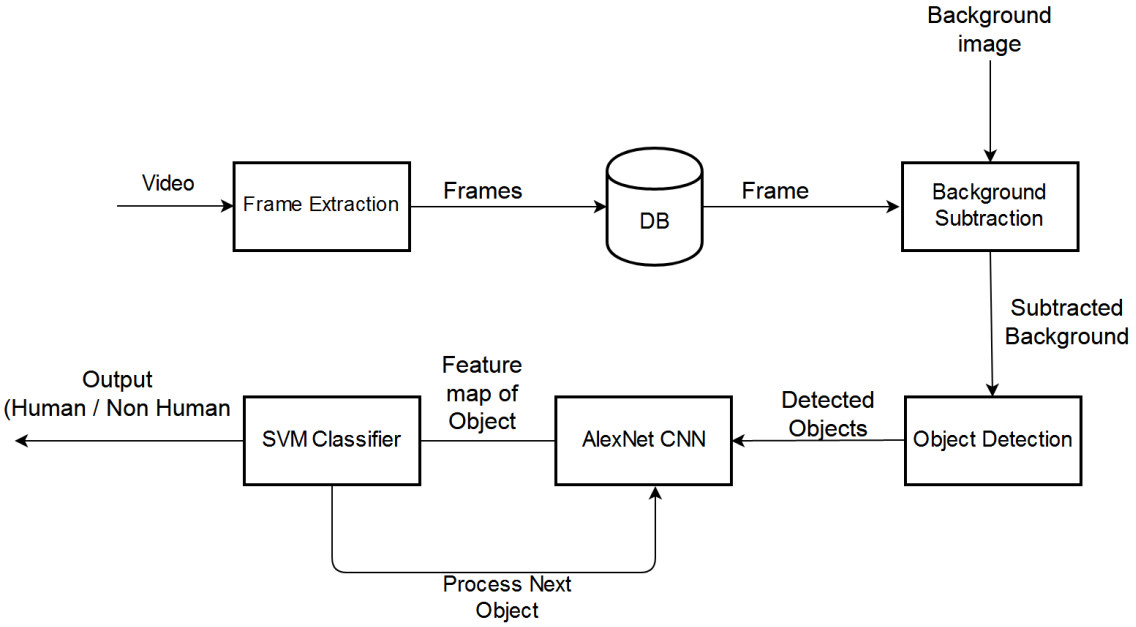

Figure 2. Workflow of Proposed System

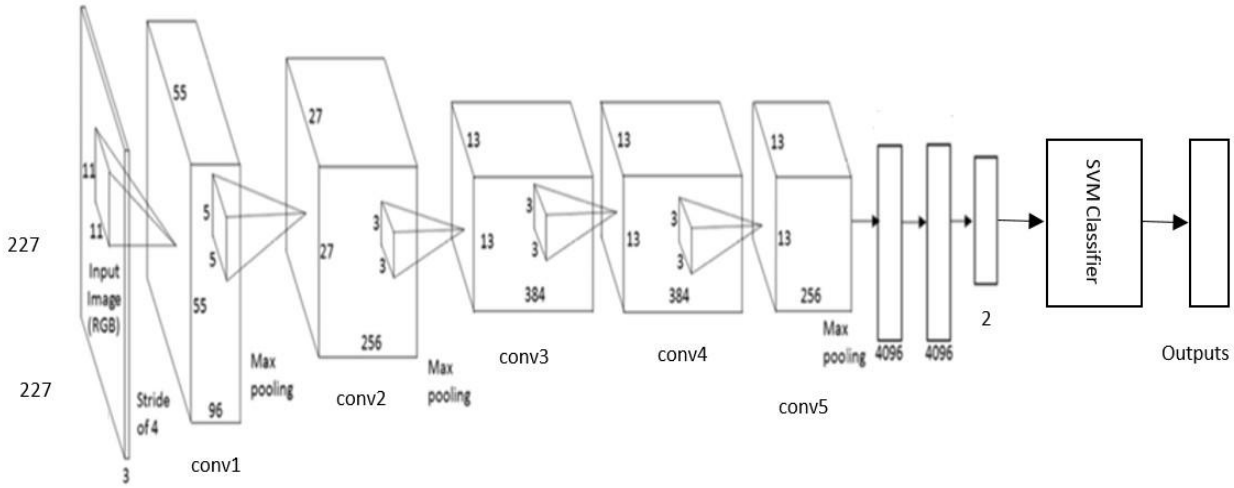

Figure 3. Traditional architecture of AlexNet CNN with SVM classifier

The last layer of this CNN is fully connected layer with size of 4096x1. Output of this layer will be a feature vector for the defined classes such as human and not human. This feature vector is used to train the SVM classifier so that the SVM classifier can recognize the presence of human in the surveillance zone.

SVM classifiers are the most efficient classifiers for binary classification. In the proposed system, the SVM is trained only for the two classes. An algorithm for the proposed system is proposed.

Algorithm: Human Detection
Input: Frame as F
Procedure Human_detection (Frame F_i)
Read the input frame F_i as F
Read the background image as B
Sub = F -B
Sub_Gray = RGB_to_Gray(Sub)
Sub_Bin = Threshold(Sub_Gray)
Detected_objects [ ] = Find_contours(Sub_Gray)
While Detected_objects has next
Read countour region from F as ob
Ob = Resize(ob, 227,227)
Feature_vector = CNN_Net (ob)
Classfication_result = SVM_Classifier (Feature_vector)
If Classification_result == Human then

Alert base camp
End if
End while
i++;
Call Human_detection (Frame F_i)
End Procedure

IV. RESULTS AND ANALYSIS

The performance of the proposed approach is tested on the dataset containing 2000 images of humans and non human objects. AlexNet CNN reduces the top-5 error from 26% to 15.3%. Hence, in this paper we tested the accuracy of the SVM classifiers. This is beause, CNN only performs the action of features detection. The final outcome is the result of the SVM classifier. The support vector machine based on linear kernel function is used for classification. To train the SVM, four training datasets are designed. The first training set contains features of the 100 images of the human and 100 images of the non humans. In the second training set, the numbers of samples are increased for both the categories. In human category, 200 samples and in non human category 200 samples are used. Third training set includes 300 samples of each category and fourth training set includes 500 samples each category. Accuracy of the linear SVM corresponding to all four training datasets is given in table 1. To obtain the stable model, SVM is trained for different kernel size.

Table- I: Accuracy of SVM with different kernel Size

S. No.	Training Set	No. of Samples		Accuracy			Training Time in Sec
		Humans	Non-Humans	K=1	K=3	K=5	
1	1	100	100	0.79	0.80	0.78	8.103
2	2	200	200	0.90	0.98	0.86	9.056
3	3	300	300	0.97	0.97	0.98	11.044
4	4	500	500	0.94	0.93	0.96	13.567

Table- II: Classification Accuracy of proposed approach

S. No.	Test Data Set	Number of samples		Accuracy (%)	
		Human	Non-human	Human	Non- human
1	First	100	100	97.0	99
2	Second	100	100	91.0	100
3	Third	200	200	98.5	100
4	Fourth	200	200	94.6	99
5	Fifth	100	100	97.0	100

During the training, accuracy of the SVM model is the highest at (k=3) in training set third. We adopt this model for the proposed approach.

The final accuracy of the complete model is also analyzed. In this analysis, the performance of the proposed system is tested on images where position and posture of the human is varies with random pattern. To do so, the test dataset is divided into five different sets. In first set, all images of humans are included where the human is sitting in different location in the surveillance zone. In the second set, the images in which human is lying down in the surveillance zone are includes. The third set contains the images in which human is present in random posture like standing, hands up etc. In fourth set images in which human is trying to jump the boundary wall are included and in last set the images of humans with mask on their faces are included. The accuracy of the proposed system on different test dataset is given in table II.

The average accuracy of the proposed system in classification of the human is 95.6%. The non human objects are classified with almost 100% accuracy. Figure 4 illustrate the performance of the proposed system in different datasets. The accuracy of the system is down in the second dataset this is because the human in the images is lying down and features of this posture are not classified accurately by the SVM classifiers. In overall, system is enough and capable to detect the presence of human in no men's land zone.

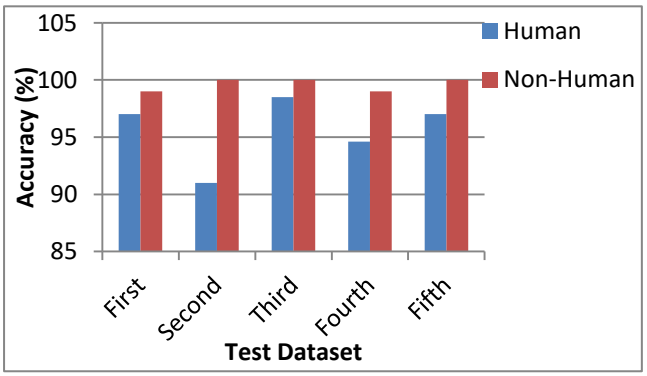

Figure 5. Accuracy of proposed system in different test datasets

A. Comparison

The approach proposed in this paper is compared with human detection approach based proposed in [3]. The average accuracy in human detection using skin color modeling is 67.5% and accuracy of human detection using cascade classifier is 89%.

The proposed approach improves the overall accuracy in human detection by 6% of the approach proposed in [3].

V. CONCLUSION

This paper proposes an application of object detection and machine learning in video surveillance. This proposed application is designed for surveillance of cantonment boundary for intrusive entry of the human. The approach combines the traditional image processing based object detection and CNN based machine learning for the video surveillance. The proposed approach detects the presence of human in the surveillance zone with 95.6% accuracy. Proposed approach is also better than the existing and leading approach for human detection by 6%. The application proposed in this paper may play revolutionary changes in internal security of the cantonment. The proposed application does not require manpower to operate. Hence, proposed approach is beneficial for developing countries such as India.

REFERENCES

1. http://indianexpress.com/article/india/india-news-india/pathankot-terror-attack-something-seriously-wrong-with-our-security-establishment-parliament-panel/ (URL)
2. T. Kumar, and D. S. Kushwaha, "Traffic surveillance and speed limit violation detection system", *J. Intell. Fuzzy Syst*. 32, 3761–3773 (2017).
3. D. K. Singh and D. S. Kushwaha, "Automatic Intruder Combat System: A way to Smart Border Surveillance", *Def. Sci. J*. 67, 50 (2017).
4. M.Z. Alom, T.M. Taha, C. Yakopcic, S. Westberg, P. Sidike,M. S. Nasrin, B. C. Van Esesn, A. A. Awwal, V. K. Asari, "The history began from AlexNet: a comprehensive survey on deep learning approaches*"*. *arXiv preprint* arXiv:1803.01164. 2018 Mar 3.
5. T. Joachims, "Making large-scale SVM learning practical", *Technical report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen*, Universität Dortmund; 1998.
6. N. Singh and T. Kumar, "An Improved Intelligent Transportation System: An Approach for Bilingual License Plate Recognition." In *Information and Communication Technology for Intelligent Systems* 2019 (pp. 29-38). Springer, Singapore.
7. K. Park, D. Lee, and Y. Park, "Video-based detection of street parking violation", in *Proc. International Conference on Image Processing, Computer Vision, and Pattern Recognition(IPCV)*, 2007, pp. 152-156.
8. N. Buch, S. Velastin, and J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 920-939, 2011.
9. R. Vohra and K. C. Tiwari, "Spatial shape feature descriptors in classification of engineered objects using high spatial resolution remote sensing data", *Evolving Systems*. 2019:1-4.
10. B. Jagadeesh and C. M. Patil, "Video Based Human Activity Detection, Recognition and Classification of actions using SVM", *Transactions on Machine Learning and Artificial Intelligence*. 2019 Jan 5;6(6):22.
11. J. Jagtap and M. Kokare, "Human age classification using facial skin aging features and artificial neural network". *Cognitive Systems Research.* 2016 Dec 1;40:116-28.

AUTHORS PROFILE

Tarun Kumar received Ph.D. degree in Computer Science and Engineering at Motilal Nehru National Institute of Technology Allahabad, Allahabad, India. At present he is working as Associate Professor in Computer Science and Engineering Department of G L Bajaj Institute of Management & Technology. His research interests include image processing, Cloud Computing and Blockchain Technology. He has published more than 14 research articles in various SCI and Scopus indexed journals and conferences.

Sanjeev Kumar Pippal received Ph.D. and Masters degree in Computer Science and Engineering at Motilal Nehru National Institute of Technology Allahabad, Allahabad, India. At present, is working as Head of Department in Computer Science and Engineering at G L Bajaj Institute of Management & Technology. His area of interest includes Cloud, AI and Sustainable Engineering.

Aishwarya Mishra is an Assistant Professor in Department of CSE at G L Bajaj Institute of Technology and Management, Greater Noida. She holds B. Tech (2017) & M. Tech. (2019) degrees from MMMUT, Gorakhpur. She is now doing research on Cloud, AI , and Sustainable Engineering.

Allora Dudi has completed her B. Tech. degree in honours from College of Engineering and Technology, Bikaner. During the course of degree she did her Oracle certification OCA and worked on Linux environment too. Throughout the degree she had also been studying network security and server implementation. She is currently pursuing her M. Tech in Computer Science from Engineering College, Bikaner. She has also received a scholarship from college for M. Tech. She has published many research papers previously and presently she is working on paper related to video surveillance system. This video surveillance system is based on handling of intrusive incidents..

Vinod Chaudhary is working as Assistnat Professor in Computer Science and Engineering Department of Government Engineering College, Bikaner (India). He received his Master of Engineering from NTTTR, Chandigarh. He is reviewer of various international journals and author of various publications. His research interest includes image processing and software engineering.