IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10477

Preventing Cross Border Infiltration using CNN

M P Pushpalatha[1], Santosh Umesh Shet[2], Hymavathi B U[3]

Professor & HOD, Computer Science and Engineering, JSS Science and technology University, Mysore, India[1]

Student, Computer Science and Engineering, JSS Science and technology University, Mysore, India [2,3]

Abstract: Motion detection and the subsequent identification of the cause of motion plays a pivotal role in advanced surveillance systems. Cross border infiltrations are increasing rapidly and an autonomous system that identifies and classifies the cause of motion has a tremendous potential to augment the surveillance process. Deep learning methods can be used to smarten such systems to perform under any illumination setting with high accuracy. In our study we have developed a system that identifies intruders in well illuminated as well as in low or no illumination conditions. The system identifies motion in a live video feed by background subtraction and identifies the regions of interest (ROI). Subsequently, a Convolutional Neural Network (CNN) model trained using regular and thermal images classifies the regions of interest and detects if motion is caused by a human intruder. Our system detects human intruders with an accuracy of 96.13%.

Keywords: visual surveillance, background subtraction, motion detection, static background, convolutional neural networks

I. INTRODUCTION

Cross border infiltrations are posing extreme security challenges. Around the globe, instances of cross border infiltrations have seen a significant rise in recent years. India is facing extreme issues at Pakistan, China, Myanmar and Bangladesh borders; USA is facing illegal immigration and smuggling at the Mexico border; European nations, the middle-east and many other countries are facing these problems. India shares an achingly lengthy 15200 kilometres of International border with our seven neighbouring countries [1].

Failing to contain these infiltrations poses threats to the security of the country and causes socio economic damages as well. Terrorist infiltrations from across the border has led to an increase in the number of casualties of armed personnel and civilians. Infiltration has also resulted in the rise of illegal immigration which creates unrest and further exacerbates the security situations. Human trafficking and pushing contraband across the border further dent the situation. Securing such large swathes of borders that India has requires considerable manpower. The rugged terrains and extreme climates make the task of border patrolling and monitoring more challenging.

A suitable alternative to expending manpower is the use of automated systems that are highly accurate and give real time feedback of the environment to take swift corrective action to prevent colossal damage on all the fronts. Here, we propose a Convolution Neural Network based automated surveillance system that can detect intruders in any terrain and in real time, so that the armed forces can definitively know about the location and number of intruders to take further actions. The proposed system augments the security forces in the border monitoring process by alerting as soon as an intrusion is detected so that requisite action can be taken to prevent damages and casualties.

II. PROPOSED METHODOLOGY

A. *Preparing the dataset*

The input dataset consists of two sets of images and their associated labels. The first set consists of 1060 images of humans in different postures and involved in various activities [2,3,4,5]. The images are obtained in the infrared and visible light region of the electromagnetic spectrum. The second set consists of 920 images of diverse landscapes in border areas without any human presence. In Figure 1a are sample images of humans in the visible light spectrum (380nm-740nm) which are used to identify humans during daytime.

The Figure 1b are thermal image samples of humans with wavelength (9-14 micrometer). In Figure 1c are sample images of humans seen under Infrared(700nm-1mm). In Figure 2, the night vision and thermal versions of thick smoke blanketing the human behind it is presented. These images play a key role in our designed system as certain pixels highlight human presence. It is evident from this that a human intruder in the frame is clearly visible under thermal scan even though there is thick smoke.

Fetching the right type of data for a particular use case is a daunting task as the data should have diversity in the object of interest for a model to generalize well to that data. The model should achieve translation invariance wherein it can classify the object of interest irrespective of the orientation, location, scale and brightness. Data augmentation is used to apply image transformations such as horizontal flip, width shift, feature wise standard normalization, feature wise centre and rotations to the available dataset which increases its diversity so that the model can generalize well.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10477

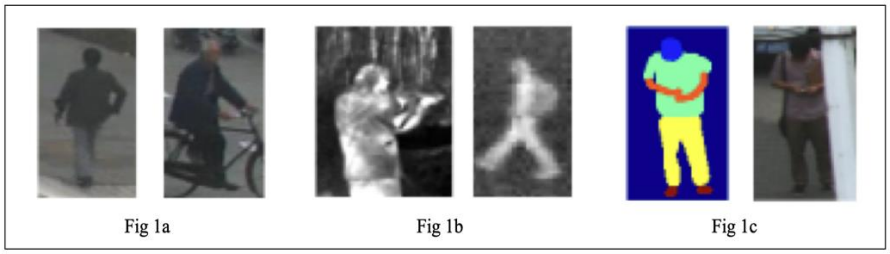


Figure 1 Shows sample images of humans in different electromagnetic spectrum.

Figure 2. Representative image showing the same frame taken with night vision and thermal camera. Source FLIR.

B. *Background Subtraction and Motion Detection*

The model chosen here is Mixture of Gaussian, supplied by the OpenCV package as MoG2[6]. The foreground obtained is a grayscale image. This grayscale image is converted to a binary image using the gray level threshold [7] value.

Our approach starts with the idea of background subtraction (BGS) algorithms which have been already proposed but BGS techniques have limitations and cannot segment images from the same color background (camouflage) and sensitivity towards low illuminated conditions. We can consider each background pixels as three different Gaussians, each of which are manually labelled as suggested by Russel and Friedman [8] as: darkest pixels can be labelled as shadow of objects, one as the road, and other with highest variance as a vehicle.

The maintenance and real time consideration are made using an incremental EM algorithm [9]. This process of modelling the recent color feature history was generalized by Stauffer and Grimson [19] of each pixel {X1,..., Xt} by a mixture of K Gaussians. Reiterating the same algorithm:

$$P(X_t) = \sum_{i=0}^{K} \omega_{i,t}.\eta\left(X_t, \mu_{i,t}, \sum_{i,t}\right)$$

where the i^{th} Gaussian has $\omega_{i,t}$ as weight associated with it, at time t and mean as $\mu_{i,t}$, parameter K is the number of distributions, and standard deviation $\sum_{i,t}$. η is a Gaussian probability density function:

$$\eta\left(X_t, \mu, \sum\right) = \sum_{i=1}^{K} \frac{1}{(2\pi)^{n/2}|\sum|^{1/2}} e^{-\frac{1}{2}(X_t-\mu)\sum^{-1}(X_t-\mu)}$$

Stauffer and Grimson[1] have assumed that each of the color components i.e RGB to be independent and have the same variance. The covariance matrix form is given by:

$$\sum_{i,t} = \sigma_{i,t}^2 I$$

An image consisting of multiple pixels each of which are characterized by a mixture of K Gaussians. Once background model is defined, various parameters are initialized. Stauffer and Grimson [19] proposed K to be set from values 3 to 5. For all the initializations such as covariance matrix, weights, and mean incremental EM algorithm is used. First foreground detection can be done followed by parameter updation. Stauffer and Grimson used, $r_j = \omega_j / \sigma_j$.

ratio to order the K Gaussians. The ordering supposes the background pixels to low variance and high weights as the background is more prominent and constant value. The first G Gaussian distributions which have threshold greater than T are retained as background distribution:

$$G = \arg\min_m \left(\sum_{i=1}^{m} \omega_{i,t} > T\right)$$

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10477

Distributions which have values less than threshold T are considered to represent foreground distribution. When new frame arrives at time t+1, a match test is made for each pixel value. A pixel is matched to a Gaussian distribution if the Mahalanobis distance which given by,

$$sqrt\left(\left(X_{t+1}-\mu_{i,t}\right)^{T}.\sum_{i,t}^{-1}\left(X_{t+1}-\mu_{i,t}\right)\right)<k\sigma_{i,t}$$

where k is a constant threshold that equals 2.5. The following two scenarios can occur:
A pixel is classified as background if it matches with one of the above-mentioned K Gaussians, which is also a background else its classified as foreground pixel. After this classification a binary mask is obtained which separates the background from foreground and the parameters are updated wherein Mahalanobis distance is used to ascertain the match between new incoming frames.
Case 1: Consider 'm' pixels among which 'i' pixels matches K Gaussians. Considering the matched pixels, parameters are updated as:

$$\omega_{i,t+1}=(1-\alpha)\omega_{i,t}+\alpha$$

where ω is a constant learning rate.

$$\mu_{i,t+1}=(1-\rho)\mu_{i,t}+\rho X_{t+1}$$
$$\sigma_{i,t+1}^{2}=\sigma_{i,t}^{2}+\rho\left(X_{t+1}-\mu_{i,t+1}\right)\left(X_{t+1}-\mu_{i,t+1}\right)^{T}$$

where $\rho=\alpha.\eta(X_{t+1},\mu_i,\sum_i)$.
With respect to the '(m - i)' unmatched pixels; μ and Σ retain their value, the weights are updated as,

$$\omega_{j,t+1}=(1-\alpha)\omega_{j,t}$$

Case 2: If none of the 'm' pixels matches the K Gaussians, a new parameter replaces the least probable distribution K. Foreground detection is done once parameters are updated. Setting of the parameters can be found in [10][11].
For object detection and recognition, contours should be identified which is a curve joining all the continuous foreground points along the boundary, having the same color or intensity. The contours having an area greater than a predefined threshold will be the ROI for the next phase of detection. Based on the area of contour, non-significant motion (rustling of tree branches or leaves) is eliminated.

C. *Defining the model architecture and training phase*

CNN are a category of Deep Learning Networks which use the mathematical operation of convolution in at least one layer. Convolution is the operation of integrating two functions Ω and Ψ, to generate function ϕ which denotes the amount of overlap when α is shifted over by β.

$$(\phi)(t)=\int_{-\infty}^{\infty}\Omega(\tau)\,\Psi(t-\tau)d\tau$$

Here, α and β are the two input functions, * denotes the convolution operation. Convolution performs the integral product of α and β, which is reversed and shifted by amount τ.
We have developed a custom CNN architecture to recognize human intruders in the real time video feed. CNNs are highly effective in this regard and are extensively used in many computer vision applications [12]. he convolution layers in the model vary in the number of kernels used at each layer. Leaky Rectified Linear Unit [13] is used as the activation function for the convolution layers. Batch Normalization [14] is used to standardize the inputs for the subsequent hidden layers and also for achieving regularization effects. After each convolution layer, a max pooling layer [15] composed of a stride and filter size of two is used to reduce the size of the feature map by half. The output of the last convolution layer is flattened to a one-dimensional vector and fed to the fully connected layers which have Rectified Linear Unit [16] activation function. The output of the last fully connected layer is given to a single neuron to get the prediction probability of the human intruder.
The model was trained on the augmented dataset for 25 epochs using Adam optimizer [17], binary cross entropy loss function [18] and batch size of 64.

D. *Analyzing the real time video feed for human for human intruder detection*

The system gets real time video feed with static background as the input, background subtraction and motion detection are performed to continuously get the ROIs. Detecting whether the identified ROI is an infiltrator or not is the critical step. Fast processing and a high accuracy are needed to take swift corrective action to prevent infiltrations and reduce the collateral damage. The ROIs are automatically fed into the trained model as the test data and gives output labels based on classification. The model also gives the number of infiltrators identified. If the output label is human, then the alert signal is sent to the concerned authority for further action.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10477

III. Overview Of System Design

The designed system consists of static cameras positioned at desired angle to capture both thermal and normal live surveillance video feed. The Figure 3 shows internal working of the system. The video feed is continuously captured which is a sequence of frames and at any given time the two frames at t and t-1 are given as input to the system. As explained in the previous section, by frame differencing the foreground pixels are obtained which are further dilated and contour selection is done to obtain ROIs.

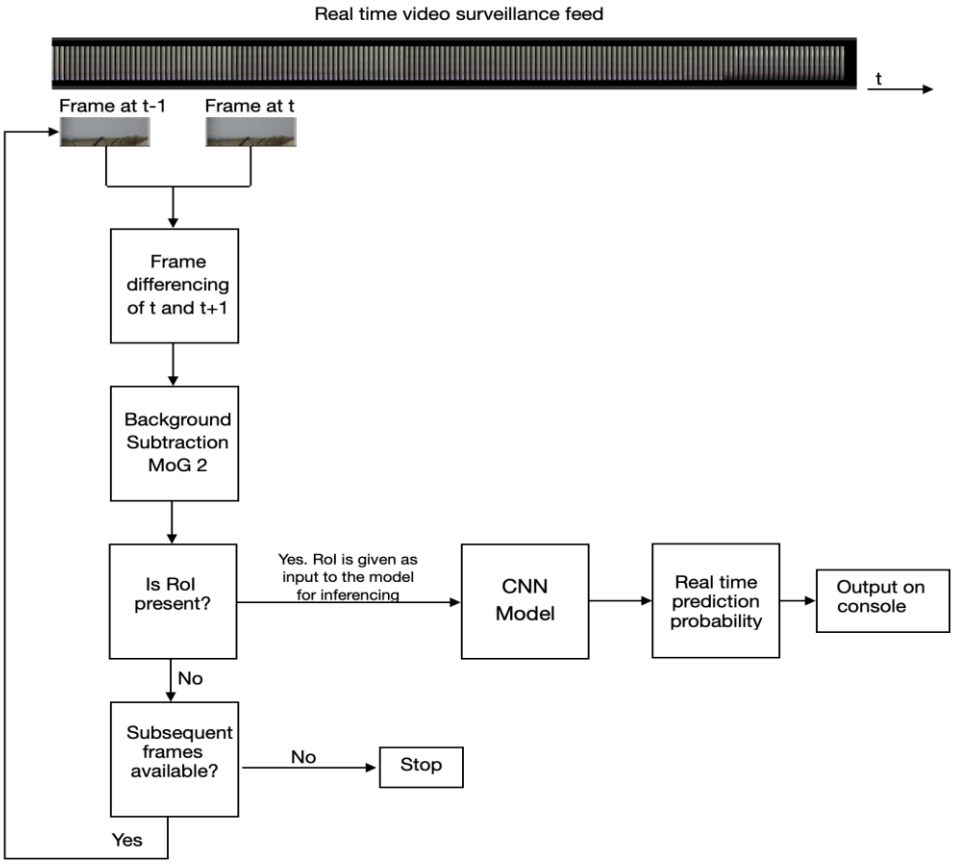


Figure 3 The figure gives an overview of the proposed surveillance system and its internal working in the form of flow diagram.

The differencing of the two frames is done to obtain objects in motion, indicated with white pixels. The white pixels upon applying MoG function gives ROIs. The obtained ROIs are inferred by the CNN to give real time prediction if a ROI is caused by humans. The output is fed to the display console as a sign of warning. If no ROIs are obtained then no inference is done by CNN. This saves unnecessary computing costs. The algorithm continues the above process until the end of the video feed wherein no more frames are available for processing.

IV. Results

Analyzing the trained model under different viewing angles and lighting conditions as show in Figure 4. In the figure 4 a) A thermal video feed from a drone is analyzed and human intruders are detected by the model. The figure 4 b) image shows a human intruder behind an object (tall grass or bushes) not seen under night vision.

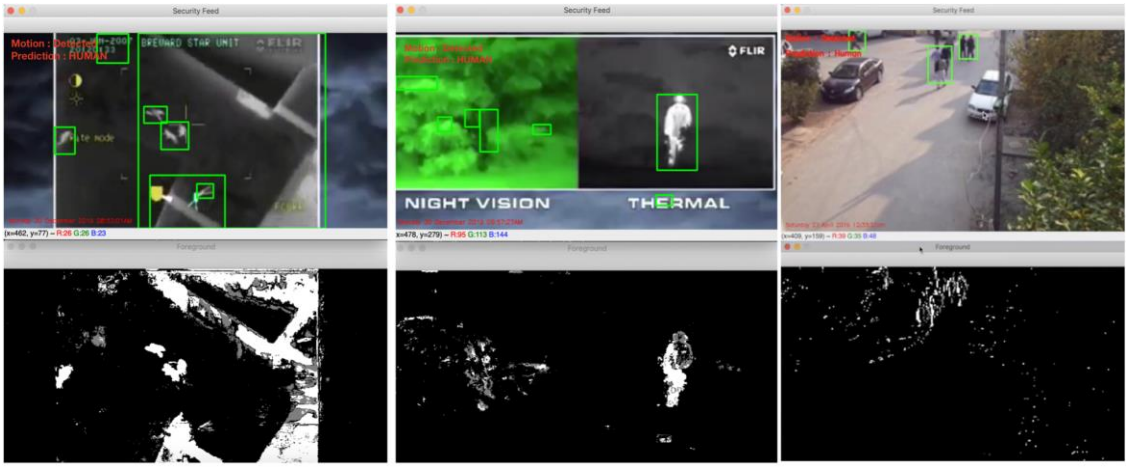
IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10477



4 a. 4 b. 4 c.

Figure 4 Shows security and foreground with thresholds feed at different lighting conditions and view.

As a result, the model draws false bounding boxes as motion is detection. When same thermal image is given as input the model is able to clearly detect the motion and draw bounding box around human intruder. The thresholds also show clear distinction between human and surrounding. Thus, we can say thermal image is more suited for surveillance during low illumination setting. The figure 4 c) shows the image where our model correctly detects humans in motion during daytime.

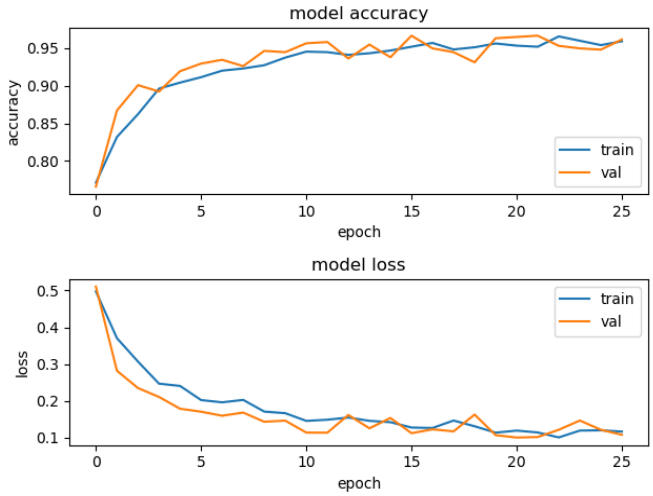


Figure 5. The plots show model accuracy and loss during training phase. The train and validation loss converge at epoch 25.

Figure 5 shows the accuracy and loss values of both the training and validation datasets over all the epochs. After training for 25 epochs, training and validation accuracies converge, demonstrating that the model has learnt and generalized well. We fine-tuned the model to settle on the optimal parameters and at 25th epoch, the model achieved a validation accuracy of 96.13% and training accuracy of 95.88% in detecting the human intruders.

V. CONCLUSION

In this paper, an autonomous system that uses CNN to identify intruders along the international borders has been proposed. The designed system exploits the strengths of CNN for identification of intruders in various real time environments with a high degree of accuracy and fast inferencing. The model can be further trained with data specific to the area of system deployment and adopted seamlessly in the different types of terrains that are present in our country: from the arid, barren lands of Rajasthan to the frigid snows capes in Siachen. The system gives accurate predictions in both day and night environments - as it has been trained under most scenarios so that both daylight video footage and thermal video footage can be used as the input. We plan to introduce further capabilities to identify drones in the visual

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10477

range of the camera and other unidentified objects. Thus, the system we have developed significantly reduces the human effort that has to be expended.

REFERENCES

[1]. Wikipedia, geography of India https://en.wikipedia.org/wiki/Geography_of_India#cite_note-DBM-1
[2]. Yet Another Computer Vision Index To Datasets (YACVID) https://riemenschneider.hayko.at/vision/dataset/index.php?filter=+pedestrian
[3]. FLIR - Thermal Imaging and Night Vision https://www.flir.in
[4]. OTCBVS Benchmark Dataset Collection http://vcipl-okstate.org/pbvs/bench/
[5]. LSI Far Infrared Pedestrian Dataset https://e-archivo.uc3m.es/handle/10016/17370
[6]. OpenCV- BackgroundSubtractorMOG2 https://docs.opencv.org/3.4/d7/d7b/classcv_1_1BackgroundSubtractorMOG2.html
[7]. Wikipedia - Thresholding (image processing) https://en.wikipedia.org/wiki/Thresholding_(image_processing)
[8]. Friedman N, Russell S. Image Segmentation in Video Sequences: A Probabilistic Approach. Proceedings Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI 1997), 1997; 175 -181.
[9]. Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. J Royal Statistical Society, Series B (Methodological) 1977; 39(1): 1-38.
[10]. Atev S, Masoud O, Papanikolopoulos N. Practical mixtures of gaussians with brightness monitoring. IEEE Conf on Intt Transportation Systems, Proceedings (ITS 2004),2004; 423-428.
[11]. Zang Q, Klette R. Parameter analysis for Mixture of Gaussians. CITR Technical Report 188, Auckland University, 2006.
[12]. Salman Khan; Hossein Rahmani; Syed Afaq Ali Shah; Mohammed Bennamoun; Gerard Medioni; Sven Dickinson, "A Guide to Convolutional Neural Networks for Computer Vision," in A Guide to Convolutional Neural Networks for Computer Vision , Morgan & Claypool, 2018
[13]. LeakyReLU Layer https://keras.io/api/layers/activation_layers/leaky_relu
[14]. S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In ICML, 2015.
[15]. Max-pooling https://computersciencewiki.org/index.php/Max-pooling_/_Pooling
[16]. Rectified Linear Unit https://keras.io/api/layers/activation_layers/relu/
[17]. Adam Optimizer: A Method for Stochastic Optimization by Diederik P. Kingma, Jimmy Ba https://arxiv.org/abs/1412.6980
[18]. Binary Crossentropy class https://keras.io/api/losses/probabilistic_losses/#binarycrossentropy-class
[19]. Stauffer C, Grimson W. Adaptive background mixture models for real-time tracking. Proc IEEE Conf on Comp Vision and Patt Recog (CVPR 1999) 1999; 246-252.

BIOGRAPHY

Dr. M P Pushpalatha is currently the Professor and HOD of Computer Science and Engineering at JSS Science and Technology University, Mysore. She has over three decades of teaching experience and her research mainly focuses on machine learning and healthcare informatics with particular emphasis on the applications of healthcare technology with socially relevant issues.

Santosh Umesh Shet is a Bachelors of Engineering graduate in Computer Science at JSS Science and Technology University, Mysore. His research work includes computer vision, interpretability on medical imaging, and advanced driver assistant systems.

Hymavathi B U is a Bachelors of Engineering graduate in Computer Science at JSS Science and Technology University, Mysore. Her research work Computer Vision and application of AI in healthcare sector.