Pedestrian Detection for Transformer Substation Based on Gaussian Mixture Model and YOLO

Qiwei Peng,Wang Luo, Gongyi Hong, Min Feng, Yuan Xia, Lei YuXiaolong Hao, Xu Wang, Mingxuan Li
NARI Group Corporation (State Grid Electric Power Research Institute)
State Grid Xinjiang Electric Power Science Research Institute
Nanjing, China Xinjiang, China
pengqiwei, luowang , honggongyi, fengmin, xiayuan ,yulei5,haoxiaolong@sgepri.sgcc.com.cn

***Abstract*—Safety is a core requirement of the transformer substation where is dangerous due to high voltage. It requires to detect pedestrians efficiently based on the surveillance video near the substation to ensure the safety of the pedestrian and the device. This paper presents a new method of pedestrian detection for transformer substation based on Gaussian Mixture Model (GMM) and YOLO. We use GMM to model the background and detect the pedestrians preliminarily, at the same time, YOLO, a kind of detection method based on convolution neural network (CNN), is also applied for pedestrian detection. Through combining the two results with different weights, the network outputs a better detection result. Our extensive experiments show that the work reaches 20% higher than the single method.**

***Keywords*-GMM; YOLO; pedestrian detection.**

I. INTRODUCTION

In recent years, the monitoring devices has increased rapidly to monitor the pedestrians around the substation [1]. In order to achieve an intelligent surveillance, by contrast, the pedestrian detection method [2] for transformer substation is still needed to be improved. This task suffers from the environment of the transformer substation because of the illumination, the clutter background and scale of the pedestrian in the video.

Since the monitor device is stationary, we can model the background and make use of the background information. In the work of [3], several Gaussian models are proposed to compute the probability distribution of different objects. We can accomplish the object-detection [4], [5] according to the divers probability distribution. It has a good effect on generating the background [6] and the primary object detection result.

YOLO[7], a new and effective method to detect objects based on regression instead of classifying, can be used of pedestrian detection. Since the image only get through the network just for once and the network can output all the detect results, the unified architecture of base YOLO model can process images at $45f/s$ in real time. YOLO has an outstanding detect result especially for big objects because the training images from the PASCAL, has a large pedestrian object.

In our work, we first use GMM method to model the background and distinguish the foreground in the initial few frames. The background model will be updated for every 25 new frames. Then we can detect the pedestrians use GMM in the following frames through comparing the difference between the background and the foreground. Since the background has relatively little change, we can get a good result s_1 of moving object detection at this step. Then compute the parameter of s_1 including the width (w), height (h) of the result, and the proportion (p) between the result and the full image.

The YOLO detection method will be used after the initial background generation and get a result s_2, s_2 also includes the following parameters: (w), (h) and (p). Through comparing IoU p withs_1 and s_2, we choose different weight λ to balance the final result $\lambda s_1 + (1-\lambda)s_2$. This paper is constructed as follows. In Sec 2, we review previous related works on pedestrian detection, GMM, and YOLO. In Sec 3, the details of proposed method are introduced. And sec 4 describes the experimental setup and presents the results of experiments. We conclude our work in Sec 5 at last.

II. RELATED WORK

Pedestrian detection is a long-term academic problem, especially the development of computer vision. Here, we focus on reviewing the related works on background generation and the deep convolution network.

A. Background generation

In order to detect objects in images or videos efficiently, background generation is a traditional way to distinguish ROI (region of interest) with background. The common method of background generation is based on texture and color features of pixels, such as ViBe [8]. Through comparing the feature of sample pixels and neighbor pixels to accomplish background generation. Gaussian mixture model is a kind of method based on statistics. It uses the K (substantially $3-5$) Gaussian models to characterize the features of each pixel in the image, the matching will be implemented between the current Gaussian mixture model and the update

978-1-5090-0768-4/16 $31.00 © 2016 IEEE
DOI 10.1109/IHMSC.2016.130

model, if the pixel matches successful, it is determined to be background pixel, otherwise the foreground.

B. Deep convolution network

Nowadays, convolution network has been used to object detection and image classification [9]. The network learns the features and the classifier automatic through training millions of parameters in this structure. YOLO is also a kind of deep convolution network inspired by the GoogleNet [10]. This model is implemented as a convolutional neural network and evaluated on the Pascal VOC detection dataset. The full name of YOLO is You Only Look Once, which means the image just get through the network just for once and finish the detection task.

III. Proposed Method

In this section, we will introduce our pedestrian detection model. This system is merged by YOLO and Gaussian mixture model. YOLO is a system of object detection and is proposed by using a single neural network to predict the probabilities of several classes. This design has the advantage of maintaining real-time speed with high accuracy. The disadvantage of YOLO is that its constraint limits the number of nearby objects which can be predicted, especially small objects. In the proposed method, we use GMM to detect the object with small size, and get a significant boost in performance. This model utilizes the advantages of both. The process of integration will be described in detail.

Since there is no existing dataset for transformer substation, we build a new dataset using surveillance videos of transformer substation. We obtain training images from video frames, in which there are persons of small and large scales.

A. GMM

In order to improve the performance of detection when the foreground has a very small proportion in the background. We add Gaussian mixture model to background subtraction. Gaussian mixture model which models the value of each single pixel is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMM is represented as

$$p(x_t) = \sum_{i=1} w_{i,t} \times \eta(x_t, \mu_{i,t}, \tau_{i,t}) \tag{1}$$

$$\eta(x_t, \mu_{i,t}, \tau_{i,t}) = \frac{1}{|\tau_{i,t}|^{1/2}} e^{-\frac{1}{2}(x_t-\mu_{i,t})^T \tau_{i,t}^{-1}(x_t-\mu_{i,t})} \tag{2}$$

$$\tau_{i,t} = \delta_{i,t}^2 I \tag{3}$$

Where, k is the number of single Gaussian function, $w_{i,t}$ is the weight of the i-th single Gaussian function, $\eta(x_t, \mu_{i,t}, \tau_{i,t})$ represents the probability density, $\tau_{i,t}$ represents the means of single Gaussian function.

GMM is an the adaptive mixture model and it is updated as follows

$$\omega_{k,t} = (1-\alpha)\omega_{k,t-1} + \alpha(M_{k,t}) \tag{4}$$

Where α is the square of the learning rate, $M_{k,t}$ is 1 for the matched model and 0 for others and $\omega_{k,t}$ is the weight of time T.

All of the pixels are classified based on their pixel values, and will be marked as foreground when the pixel values dont match each background pixel.

B. YOLO

Our object detection system is based on YOLO. The YOLO partitioned the image into multiple grids and predict the bounding box of interested object for each grid. Since the original YOLO is trained using Pascal VOC dataset, in which the scale of person is large, directly using pre-trained YOLO to detect the person of small scale is not applicable. We fine-tune the YOLO using the dataset collected from the video of transformer substation.

Our system adopts the same method to divide the input image into a $S \times S$ grid. One grid cell must be detected if there is an object in this cell. Each grid cell predicts B bounding boxes and confidence scores. Confidence is defined as $P_r(Object) * IOU_{pred}^{truth}$, it reflects how confident the box B contains an object. If there is no object existed, the confidence scores set be zero. Where, the intersection over union (IOU) is calculated by using the predicted mask and the ground truth.

We have 5 values to predictions: x, y, w, h, and confidence in each bounding box. The (x, y, w, h) represent the rectangle box containing an object. The confidence represents the probability of the existence of object. At the same time, we predicts a conditional class probabilities in a grid cell containing an object belong to the i-th class $P_r(Class_i|Object)$. We only predict one class in every grid cell. We obtain a class-specific confidence scores for each box, these scores describe the probability of one class appearing in the box.

C. The mixed pedestrian detection system

In our detection system, YOLO detection is mixed with Gaussian mixture model. An overview of the mixed detection system is shown in Figure 1.

First, we calculate the bounding box b_1 after obtaining the mask of foreground in one given image by background subtraction. Second, we use YOLO detection method to calculate the bounding box b_2 and probability s_2 of pedestrian in the same image. Third, we calculate the coincidence rate s_1 of b_1 and b_2. Finally, the probability of pedestrian detection is obtained by the mixture function represented as

$$S = s_1\lambda + s_2(1-\lambda) \tag{5}$$

Where λ is the weight set empirically.

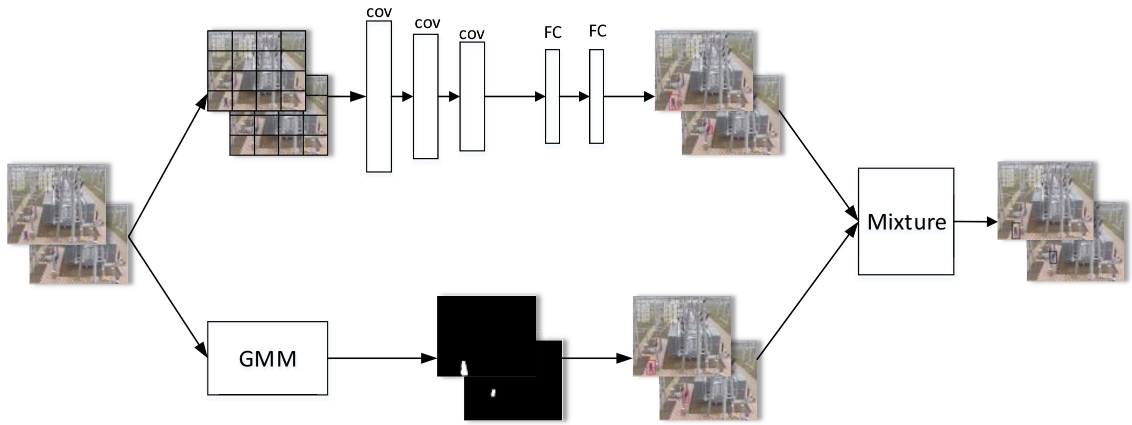

Figure 1. The overview of the mixed detection system.The above flow is the YOLO detection system and the beneath flow is the GMM detection system.The whole figure shows the complete mixed detection system.

IV. Experiment

In this section, we introduce the implementation details and our experimental setup. Then we show the results of pedestrian detection and analyze the method we proposed.

A. Background modeling

Gaussian mixture model (GMM) is used for moving object detection based on background modeling in our work. While the pedestrian is too small in the video frame, detections depended on YOLO may miss lots of small objects. In contrary, using GMM to obtain the foreground objects can solve this problem to a certain extent. In our work, Gaussian mixture model is weighted by 5 Gaussian models and updated every 25 frames. The foreground detection results are showed in Figure 2.

B. YOLO

As a good performed method in object detection, YOLO processes the images at about 30 frames per second with GPU and achieves a real-time detection in our experiment. But YOLO imposes strong scale limitation. If there exists a large variation of scale, the detection results will be deviated. This problem will be solved through combining the GMM and YOLO. We divide the input image into a 7×7 grid in YOLO model. Each grid cell predicts 2 bounding boxes and their related confidence scores.

We investigate the detection performance of YOLO. We show the recall with the variation of IOU in Table I. It can be seen that the performance of directly using YOLO is not satisfactory. The reason is perhaps that the person in the video is very small, which is difficult to be detected using YOLO.

Table I
The detection performance of YOLO

IOU	Recall	Precision
0.3	71.52%	50.02%
0.4	63.19%	45.96%
0.5	43.75%	31.82%
0.6	30.55%	22.22%
0.7	14.58%	10.61%

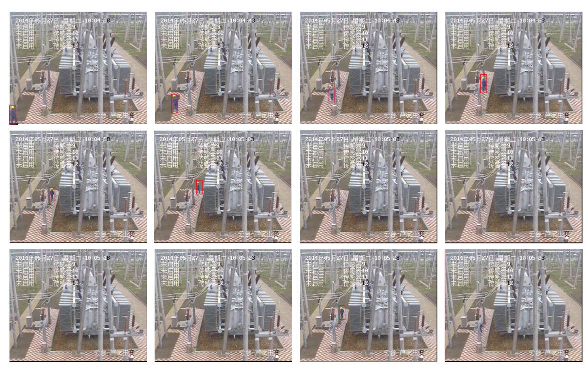

Figure 3. Pedestrian detection results of mixed detection system.

C. Mixed detection model

Considering the above two situations, we combine two detection methods as a mixed detection model. The probability is calculated by the formula 5. In our experiment, weight is set to 0.7. If the object is small, there is a real probability to be undetected it by YOLO. So we need to give GMM detection a greater weight to ensure that the object is detected. If the object is big, both two detection methods can achieve a great performance, weight distribution can be very random. So we finally set the parameter w to 0.7, and get satisfied results. Some experimental results are showed in

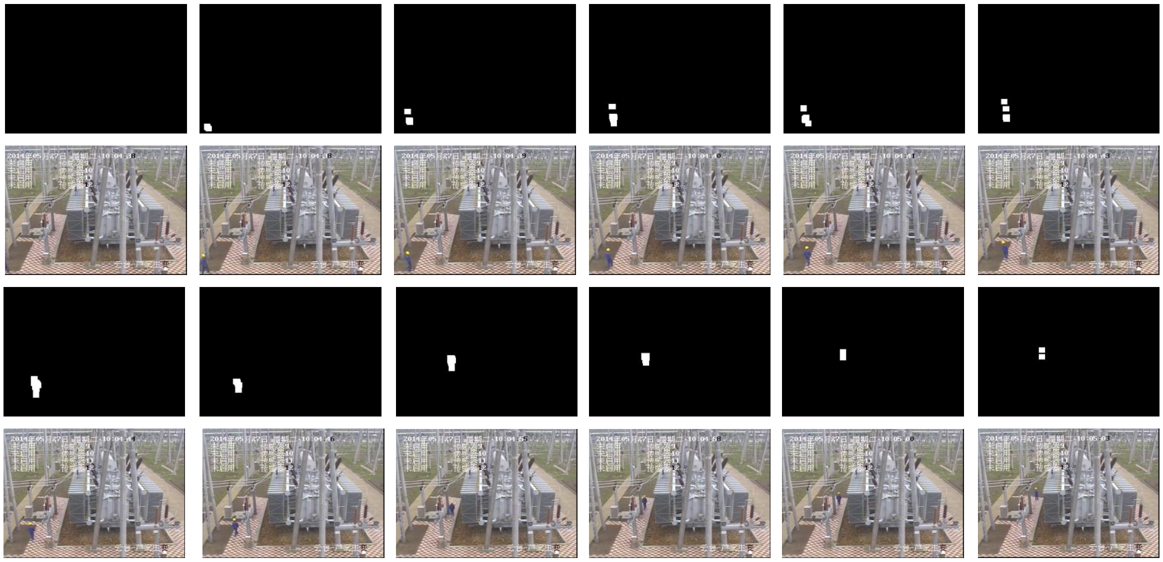

Figure 2. The white pixels of the masks in the first and third lines present the foreground objects which are detected from the background. The images in the second and forth lines are frames from the detected video.

Table II
THE FIRST LINE SHOWS THE DETECTION RESULT OF GMM, THE SECOND LINE SHOWS THE DETECTION RESULT OF YOLO AND THE THIRD LINE SHOWS THE DETECTION RESULT OF MIXED DETECTION.

	accuracy	false rate	miss rate
GMM	68.3%	20.2%	11.9%
YOLO	70.9%	6.8%	22.3%
Mixed detection	92.7%	3.4%	3.9%

Figure 3. The detection accuracy of the proposed method is greatly improved compared with the previous two methods. Table II shows the comparison of the accuracies.

V. CONCLUSION

In this paper, we propose a method for pedestrian detection. The mixed detection system consists of YOLO pedestrian detection and GMM foreground detection. The experimental results indicate that it is an effective and robust detection method. In the future, we are committed to seek a more rapid approach.

REFERENCES

[1] J. Deng and J. Zhu, "Research on pedestrian detection algorithms based on video," in *Computer Design and Applications (ICCDA), 2010 International Conference on*, vol. 5. IEEE, 2010, pp. V5–474.

[2] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 4, pp. 743–761, 2012.

[3] C. E. Rasmussen, "The infinite gaussian mixture model." in *NIPS*, vol. 12, 1999, pp. 554–560.

[4] M. Dixit, N. Rasiwasia, and N. Vasconcelos, "Adapted gaussian models for image classification," in *CVPR*, 2011.

[5] J. Li, C. Shao, W. Xu, and C. Dong, "Real-time pedestrian detection based on improved gaussian mixture model," in *Measuring Technology and Mechatronics Automation, 2009. ICMTMA'09. International Conference on*, vol. 3. IEEE, 2009, pp. 269–272.

[6] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 2. IEEE, 2004, pp. 28–31.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *arXiv preprint arXiv:1506.02640*, 2015.

[8] O. Barnich and M. Van Droogenbroeck, "Vibe: a universal background subtraction algorithm for video sequences." *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 20, no. 6, pp. 1709–1724, 2011.

[9] I. S. A. Krizhevsky and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *Eprint Arxiv*, 2014.