Research on Video Detection of Object Intrusion in Substation

Yi Jiang¹, Hui Sun²*, Hanghang Zhao¹, Rui Lin¹ and Zhe Li²

¹China Southern Power Grid EHV Transmission Company, Guangzhou 510405, Guangdong Province, China
²Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

*sunhui.sd@qq.com

Abstract. The safe running of substations is an important part of power system security, as well as an important work of the power companies. The intrusion of kites, plastic trash, small animals and other objects will affect the safe running of substations. With the development of video monitoring technology, intelligent recognition technology of images and videos, video monitoring has gradually become the main monitoring means of unattended substations. This paper firstly analyses the research status and main technologies of video object recognition. To achieve the detection of foreign objects in substations, moving object detection and video object detection technologies are studied. And two outlines are proposed: using background subtraction and image classification, using YOLOv3 to detect objects in videos. In the end of the paper, YOLOv3 was tested by training a detector of small animals. The results show that the value of mAP equals to 0.7169 and the detection speed fulfills the request of real-time detection. This method could serve as reference for detecting and tracking intrusion objects in substations.

1. Introduction

The safe running of substations is an important part of power system. In the actual running of substations, there are some potential safety hazards. Foreign objects like kites and plastic trash falling on the transmission line or electrical equipment, or small animals entering the substations or power distribution rooms, will increase the probability of power accidents such as short circuit[1]. Therefore, it is important to monitor and timely clean up the foreign objects and small animals in the substation. However, traditional monitoring methods or artificial inspection could cost a lot of manpower and the effect is unsatisfactory.

With the development of smart grid, unattended intelligent substation is becoming more and more popular[2]. In recent years, with the development of encoding and decoding video technology and intelligent analysis technology for videos, video monitoring has gradually become an important part of security system in all walks of life. Intelligent fixed video monitoring could be a better choice for object intrusion detection in unattended substations.

Intelligent recognition methods of videos mainly include motion detection, video object tracking, video object recognition, behaviour recognition and other research directions. Foreign researches on video monitoring system started earlier. Domestic researchers are also active nowadays. Many enterprises, such as Texas Instruments and Hikvision, have been committed to the research and development of video object analysis system[3].
This paper is focused on the research of video recognition technology of substation object intrusion and analyses the common technologies: moving object detection algorithm and video recognition. Two detection methods which are based on background subtraction and YOLOv3 (you only look once version 3) are proposed to achieve the recognition and tracking of foreign objects in substations.

2. Moving object detection

Moving object detection is to detect the changed areas in the video sequence and extract the moving objects. According to whether the camera is fixed, moving object detection can be divided into static background and dynamic background. Because the cameras in substations are usually fixed, the background is also static. Background subtraction, optical flow and frame difference are common methods of static background moving object detection. Due to the large amount of computation and high complexity, optical flow method is difficult to meet the real-time requirements of video detection[4]. The frame difference method is to extract the contour of moving object through the difference operation of two adjacent frames in the video sequence. Its principle and application are similar to the background subtraction method. Background subtraction method is used in this paper to detect moving objects.

Background subtraction is an effective method for moving object detection. Its basic idea is to model the background and compare the current frame with the background image. The part with large difference is considered as the moving area, while the part with small difference belongs to the background area. The key of this method is background modelling and updating. Background modelling is to obtain a static region in a video, and regard it as the background. Background modelling is vulnerable to the influence of light changes, noise and shadow[5].

![Figure 1](image1.png)

(a) the recognition of moving mouse  
(b) the binary image of the moving object

Figure 1. The diagram of GMM background subtraction to recognize moving objects.

GMM (Gaussian mixture background modelling) is generally accepted as a classical method nowadays. Its key content is to assume that the grey scale distribution of pixels in the image conforms to the Gaussian distribution, and use random process to represent the change of pixel value in the video sequence, select K Gaussian Models to represent the characteristics of pixels, and compare the new pixels to be detected with the current K models. If the two match, the point is the background point, otherwise it belongs to the foreground[6].

In a segment of mouse motion video, the diagram of GMM background subtraction to identify moving objects is shown in figure 1. Because this method can only recognize the moving area, and cannot recognize its class, it is necessary to cut the moving object and recognize its class with image recognition algorithm.

The process of using GMM background subtraction and image recognition algorithm to detect small animals, kites and other intrusions is shown in figure 2.
3. Video object detection

Video object detection is actually the image object detection, that is to identify and locate the object in each frame or some frames of a video. Compared with single image, video has redundancy in time and space, which can be used to improve the detection accuracy and speed, so as to solve the blur or deformation caused by motion, shooting or partial occlusion[7]. Video object detection algorithms introducing temporal CNN (convolutional neural network) or optical flow have a large amount of calculation and high complexity. They are not suitable for real-time video detection. Single-frame object detection may be a better method. It refers to the use of frame extraction method to detect the images in the video. The algorithms used for static image object detection could be used for video’s single-frame object detection, such as R-CNN (Region-CNN) series, YOLO (you only look once) and SSD (Single Shot MultiBox Detector). YOLO and SSD are one-stage algorithms, which have faster detection speed and are suitable for video detection. In this paper, YOLOv3 is applied.

3.1. Introduction of YOLOv3

YOLOv3 is based on YOLO and YOLOv2. YOLO is a regression method. Compared with R-CNN series, YOLO no longer extracts features from multiple proposal areas, but directly predicts the coordinates and category probability of the object box from the input image, realizing end-to-end recognition, thus greatly speeding up the detection speed[8]. Compared to YOLO and YOLOv2, YOLOv3 adjusts the network structure, uses darknet-53 network for feature extraction, integrates multi-scale features, and uses logistic function instead of softmax in object classification[9]. In the premise of maintaining the speed advantage, YOLOv3 has higher prediction accuracy, especially for small object recognition.

3.2. Dataset acquisition and augmentation

Training a YOLOv3 model requires a large number of image samples containing target objects. Four approaches are used in the paper to collect images containing foreign objects like small animals and kites.

- ImageNet dataset[10]. It is the largest image recognition database in the world containing more than 20,000 categories of various animals and common items in life. Images of birds, cats, dogs and kites are included.
- Using web crawler to acquire pictures from various websites. A large number of pictures could be obtained by using a simple crawler. The main picture websites include Baidu pictures and other picture websites.
- Extracting frames containing target objects from videos. The frame number of videos on the Internet is generally about 30 FPS, that is, each second of video contains 30 pictures. A short video could provide hundreds or even thousands of pictures. Due to the large redundancy of video, we could choose to extract pictures at a certain frame interval.
- Synthesizing pictures. The application scenario of this paper is substation, but the pictures downloaded from the Internet are various, and most of the objects are large. In order to obtain pictures with substation background and small objects, we need to synthesize pictures. In this paper, image mask[11] is used for image synthesis.
Pictures acquired through the first approach have location annotations. The fourth approach could also generate the location annotations automatically through the program. Pictures acquired through the second and third approaches need to be annotated with label tools.

3.3. Cluster analysis of the dataset

YOLOv3 uses the anchor mechanism to make the generated candidate areas more close to the size of the objects in the dataset. The anchor box size is set to the size of the objects in the dataset, which could shorten the search time of the candidate box in the image and improve the detection rate of the boundary box[12]. YOLOv3 generally obtains the training set's anchor frame through clustering. The K-means algorithm is used to cluster the dataset acquired in 3.2. The number of cluster centres is set as 9. The nine anchor frames calculated by the clustering algorithm are (10,13), (40,38), (66,94), (97,192), (138,109), (204,218), (228,384), (373,326) and (395,314).

3.4. YOLOv3 model training and testing

Taking the detection of small animals as an example, about 4,000 images collected are sent to the YOLOv3 network. Objects include birds, cats, dogs and mouse. Train set, test set and validation set are divided into 8:1:1. The batch is 32 and the subdivision is 16. The learning rate is 0.001 and the decay coefficient is set to 0.0005. The maximum number of iterations is 50,000.

During the training, average loss, average IoU, recall (threshold = 0.5) and class precision curves are shown in figure 3. It could be seen that with the increase of iterations, the loss value gradually converges to 0, the IoU gradually tends to 1, and the other two curves are also near the line y = 1. This indicates that the training model has ideal convergence.

Figure 3. Average loss, average IoU, recall (threshold = 0.5) and class precision curves.
Table 1. AP values of each class

| Class | AP     |
|-------|--------|
| Bird  | 0.5566 |
| Cat   | 0.8459 |
| Class | 0.7782 |
| Mouse | 0.6868 |

The model is then tested with the test set, and the result is shown in table 1. Cats perform the best and birds the worst. And mAP (mean Average Precision) equals to 0.7169 as an average. When the detection is accelerated by NVIDIA 2080Ti GPU, the detection speed is about 61 FPS. And the detection speed is 8 FPS without GPU.

Using a video of the distribution room containing a moving mouse and static background to test the model, frames extracted from the recognized video are shown in figure 4 (a), (b), (c). Figure 4 (d) is the mouse's motion track, which is drawn by recording the coordinates of the mouse. It could be used to predict the mouse’s motion range and find its nest location.

![Figure 4](image)

**Figure 4.** A video of moving mouse is detected using YOLOv3: figure (a), (b), (c) are frames extracted from the recognized video; figure (d) shows the trajectory of the moving mouse.

4. Summary and outlook

This paper analyses the researches and main technologies of intelligent video monitoring and video object recognition technology nowadays. Two outlines are proposed to achieve the detection of intrusion objects in substations: using background subtraction and image classification, using YOLOv3.

The second method based on YOLOv3 works well after test:

- The value of mAP equals to 0.7169 after training with the dataset of four classes of small animals acquired from the four approaches.
The detection speed could fulfill the request of real-time detection when some frames are extracted to be detected at a certain frame interval. The trajectory of small animals can be drawn and saved according to the detection, and served to predict the animals’ motion range or nest location.

In the future, the research of video object detection will be focused on these aspects:

- Improving YOLOv3’s network structure to improve the recognition accuracy;
- Using light networks and compress the model file to improve the detection speed;
- Exploring the application of object tracking and multi-object tracking in the video.

References

[1] Lin X 2017 Analysis on measures of small animals defense in substation and discussion on improvement Technology Innovation and Application vol 33 (Heilongjiang: Harbin) p 65

[2] Wang Q 2014 The research and design of unattended intelligent substation of safety monitoring system Advanced Materials Research vol 912, ed Wen-Pei S and Jimmy C.M. Kao p 1469

[3] Niu X 2018 Recognition of moving objects in video surveillance of intelligent substation (Shenyang: Shenyang Agriculture University) p 3-4

[4] Baker S, Scharstein D, Lewis J P, Roth S, Black M J and Szeliski R A 2007 A Database and Evaluation Methodology for Optical Flow International Journal of Computer Vision 92.1 p 1-31

[5] Sengar S S and Mukhopadhyay S 2019 Moving object detection using statistical background subtraction in wavelet compressed domain Multimedia Tools and Applications vol 79 (12) (Netherlands) p 5919-5940

[6] Li G, He X, Zhang S and Gao M 2011 Improved moving objects detection method based on GMM Application Research of Computers vol 028(012) (Chengdu) p 4738-4741

[7] Kang K, Li H, Yan J, Zeng X, Yang B, Xiao T, Zhang C, Wang Z, Wang R, Wang X and Ouyang W 2017 T-CNN: Tubelets With Convolutional Neural Networks for Object Detection From Videos IEEE Transactions on Circuits and Systems for Video Technology vol 28(10) p 2896-2907

[8] Redmon J, Divvala S, Girshick R and Farhadi A 2016 You only look once: Unified, real-time object detection Proc. of the IEEE Conference on Computer Vision and Pattern Recognition p 779-788

[9] Redmon J and Farhadi A 2018 YOLOv3: An incremental improvement arXiv preprint arXiv:1804.02767 [cs]

[10] Deng J, Dong W, Socher R, Li L J, Li K and Li F F 2009 ImageNet: A large-scale hierarchical image database IEEE Computer Society Conference on Computer Vision and Pattern Recognition p 248-255

[11] Wibedi D, Misra I and Hebert M 2017 Cut, paste and learn: Surprisingly easy synthesis for instance detection Proc. of the IEEE International Conference on Computer Vision p 1301-1310

[12] Ren S, He K, Girshick R and Sun J 2015 Faster r-cnn: Towards real-time object detection with region proposal networks IEEE Transactions on Pattern Analysis & Machine Intelligence vol 39(6) p 1137-1149