2020 2nd International Conference on Information Technology and Computer Application (ITCA) | 978-1-6654-0378-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ITCA52113.2020.00067

# A motion based object detection method

Chen Zhaoyang, Gao Haolin, Wang Kun
Zhengzhou Xinda Institute of Advanced Technology
Zhengzhou, China

***Abstract*—To improve the performance of motion detection and object detection in security monitor, a motion based object detection method is proposed. Combining motion detection, target detection is proposed. This method first improved the accuracy of motion detection, then combined the result of motion detection and object detection. The experimental results show that proposed method can effectively improve the detection of small targets, reduce the false detection and negative detection rate, and decrease the computing cost.**

***Key words: motion detection, background model, small object detection***

## I. INTRODUCTION

Object detection is an important research and application direction of computer vision. With the emergence of various target detection algorithms based on deep convolution neural network [1], it is more and more widely used in daily application. Especially in the field of security, object detection is particularly important. It can not only replace manual work to reduce cost, but also provide various intelligent analysis means to improve efficiency. However, there are some shortcomings in practical application, (1) object is too small, the detection of small target away from camera has been one of the difficulties in this field; (2) reducing missed detection and false detection, improving detection accuracy, has been the optimization goal of all kinds of detection algorithms; (3) The consumption of computing resources.Some methods requires a large amount of computing resources all the time. Recently, the requirement of computer hardware resources based on deep convolution neural network increases continuously with the rise of network complexity.

Motion detection [2] is a common and important application in the field of security, which can solve many problems in unmanned environment, and has been embedded in many cameras. However, there is a common deficiency in motion detection: high false alarm rate, unable to determine what is the object in the detected area.

Aiming at the shortcomings above, a new motion detection method is proposed, which is based on object detection. This method includes two steps. Firstly, the motion detection algorithm is used to extract the foreground target area, secondly, the convolution neural network is applied to the moving area to detect object. It can effectively improve the detection of small targets, reduce missed detection and false detection, and decrease the consumption of computing resources.

## II. RELATED WORK

### A. Motion detection

Motion detection is the detection of pixel changing regions in video sequences. Background modeling algorithm is used to create background and foreground image segment. The mainstream background modeling algorithms includes inter-frame difference method [3], Gaussian mixture model [4], background subtraction method [5], Vibe method[6] and so on. Except for frame difference method does not need to build model, other methods need modeling and real-time background update, they are more robust to the change of illumination.

*1) the inter-frame difference method*

In a continuous video sequence, the change of the continuous frame is very weak if there are no motion factors (including illumination, environment, vibration, etc.) in the scene; if there are motion factors, the the motion trace will be apparent in different image frames. The inter-frame difference method performs differential operation on two or more continuous frames in time domain, which means subtracts the value of pixel points at same position in different frames, and makes a judgement based on the absolute value of gray difference. When the absolute value exceeds a certain threshold, the pixel point can be judged as moving pixel. The two-frame difference method is shown in formula (1) below.

$$\left|I_{i,j}(t)-I_{i,j}(t-1)\right|\begin{cases}\geq T & \text{, motion point}\\ < T & \text{, non motion point}\end{cases} \quad (1)$$

In this formula, $I_{i,j}(t)$, $I_{i,j}(t-1)$ represent the pixel of *i*-th row and *j*-th column at time *t* and *t*-1 respectively，*T* is the threshold of pixel difference, which is determine according different condition.

In the result of two frame difference method, "double shadow" or "cavity" phenomenon may appear, which is harmful for post-processing. So three frame difference method based on it was proposed , as shown in formula (2) below, which effectively avoids the defect of two frame difference method.

$$\left|I_{i,j}(t)-I_{i,j}(t-1)\right|\cap\left|I_{i,j}(t+1)-I_{i,j}(t)\right|\begin{cases}\geq T & \text{, motion point}\\ < T & \text{, non motion point}\end{cases} \quad (2)$$

*2) background subtraction method*

In background subtraction method, the moving area is detected by using the difference between the current frame and the background image. The key is to establish the model of background and update real-time, in order to be robust for environmental changes. A simple way to update the background is shown in formula (3) below:

$$F(t)=(1-\alpha)*F(t-1)+\alpha*I(t) \quad (3)$$

In this formula, *F(t)* is the background image at time *t*, *F(t-1)* is the background image at time *t*-1, *I(t)* is the image of the moment *t*, $\alpha$ is the update coefficient. The absolute difference between the background and the current frame is

978-1-6654-0378-8/20/$31.00 ©2020 IEEE
DOI 10.1109/ITCA52113.2020.00067

calculated, then filtered by formula (4), the moving area can be obtained finally.

$$\left|F_{i,j}(t)-I_{i,j}(t)\right|\begin{cases}\geq T & \text{, motion point}\\ < T & \text{, non motion point}\end{cases} \quad (4)$$

### B. Object detection network

The mainstream object detection network is divided into two categories: regional recommendation and end-to-end.

The region recommendation method is also called two-step method. Firstly, a series of candidate frames containing objects are generated by the algorithm, or the recommended regions, then the candidate frames are classified by convolution neural network. This kind of algorithm has advantages in accuracy and positioning accuracy, such as R-CNN[7], Fast R-CNN[8], Faster R-CNN[9].

The end-to-end method is called one-stage method. Because these methods directly regress the border of the target without generating candidate boxes, and they have advantages in running speed, such as YOLO[10], SSD[11] and so on.

## III. Motion Based Object Detection Method

In the application of security, both motion detection and object detection can achieve the detection of moving targets, but have drawbacks. The disadvantage of motion detection is that there are more false positives and it is uncertain what target is in the detected area. The deficiency of object detection includes (1) hard to detect small object, (2) missed and false detection often occurs, (3) high computing cost.

A new detection method is proposed, which the improved motion detection combined with object detection. The working flow chart is shown in figure 1. The video is decoded and input into the system. If there is a moving object, the corresponding motion area is passed into the convolution neural network of the object detection module for object determination, otherwise the object detection is skipped directly.

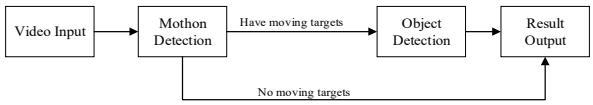



Figure 1. The flow chart of proposed method

### A. Improved motion detection

Motion detection depends on the result of foreground detection, and this kind of algorithm may not detect the whole object completely. To solve the problem, the proposed method processes the result of foreground detection to achieve better results. The concrete steps are as follows:

(1) Find the relationship between all foreground blocks, and if the distance between the two blocks is less than the threshold TH_Distance, mark each other as relational pairs;

(2) Traversing each foreground block and merging each relational pair, that is, merging all the foreground blocks into a whole block;

(3) Determining whether there is a cross-over between all merged foreground blocks, if so, merge them.

According to the above algorithm, the process result is compared shown in figure 2 below. The above image shows the result of motion detection, and there are many moving blocks for the same object, the below image show the result of proposed method, the objects detected in its result are more completed.

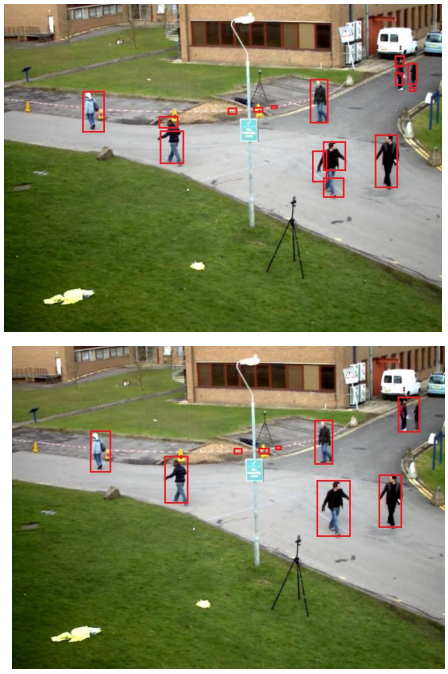


Figure 2. The comparison of motion detection

### B. training of object detection model

According to the input size(416 Pixel*416Pixel), the foreground block were resized and processed. An dataset containing 50000 images were collected, which contains manual cutting images at required size, and object zone image extracted from motion detection. Then they were labeled by LabelImg. Finally, the training parameters was initialized, and the detection network was trained, and adjust the parameters according according to the training results until the detection result achieve requirement.

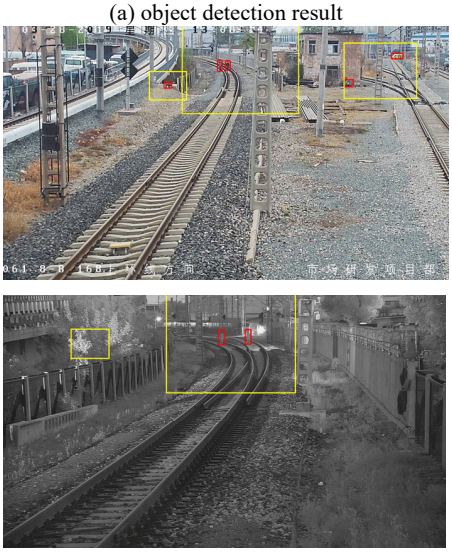
## IV. Experiments and Results

The experimental environment is 32G memory, E5-2620 CPU, Nvidia GPU TITAN-xp, VS2015, OpenCV3.2. The test data is videos recorded inside the railway guard bar containing objects such as pedestrians, the resolution of video is 1920*1080.

The detection contains daytime detection and night detection. Before detection, image enhancement is used to process night images to improve the accuracy of detection.

Two video frames in day and night are tested as following. The first scenario contains five objectives, person x2、dog x1、truck x1、bicycle x1, the second scenario contains two objects, person x2.

(a) object detection result



(b) The detection result of proposed method

Figure 3. The comparison of object detection result

The yellow boxes in figure 3 represents the moving area, and the red box contain the detected object. The results shows that the detection results of proposed method perform better in both recall and precision. The reason is that the proportion of tiny object in the original large image(1920*1080) is very small, they must normalized to a small size (416*416) before input into the convolution neural network, which leads to fewer features of the original tiny object. Then the network can not detect the target. Motion detection provides only a region contains motion information, and the proportion of small targets in this region is larger than that in the original map, so it does not lose too much information in normalization. Therefore, it can reduce false positives and increase the detection number of small objects.

To compare the accuracy of motion detection, object detection and motion based object detection method, the missed detection rate and the false detection rate are selected as the performance indexes in this experiment. The 150 videos clips containing pedestrian and other objects inside railway guardrails were selected for testing. In these test videos, there are 568 moving targets to be detected. Supposing M represents missed detection rate, F represents false detection, in the following formulas (5) and (6), A represents the number of targets to be detected, B represents the test video sample number is 150, C represents the actual number of targets to be detected that have not been detected, D represents the number of video samples containing false detection.

$$M = \frac{C}{A} \tag{5}$$

$$F = \frac{D}{B} \tag{6}$$

The results in experiment are showed in table 1.

TABLE 1. THE ACCURACY OF THREE METHOD

| | C | D | M | F |
|---|---|---|---|---|
| Motion detection | 16 | 51 | 2.8% | 34.0% |
| Object detection | 117 | 5 | 20.6% | 3.3% |
| Proposed method | 23 | 2 | 4.0% | 1.3% |

It is found that false detection more often happened for the swinging leafs nearby the railway guardrail in motion detection, and miss detection often happened for small and object in images such as trains, fast moving dogs. By combining motion detection with object detection, the false detection rate and missed detection rate can both reduced.

Because there is not always motion inside of the railway guardrail, so the object detection module will not run all the time. Only when the motion detection module returned motion zone, the object detection module detects the current image. This can reduce the occupation of GPU resources. The CPU and GPU utilization rate of the three algorithms are recorded 10 times and averaged, the average occupancy rate of the two algorithms are shown in table 2 below. Because of the reduction of the utilization rate of computing resources, more video can be processed in parallel.

TABLE 2. THE AVERAGE CPU AND GPU UTILIZATION

| | Average CPU utilization | Average GPU utilization |
|---|---|---|
| Motion detection | 26.2% | 0 |
| Object detection | 15.4% | 46% |
| Proposed method | 23% | 31.2% |

## V. CONCLUSION

Aiming at the ineffective motion detection of security monitor, a motion based object detection method is proposed. Firstly, the background is modeled to detect the moving area, secondly, the moving area is calculated by the convolution neural network of object detection. Finally, the object category and bounding box are obtained. This method can effectively improve the detection of small objects, reduce the false alarm rate and decrease computing cost.

## ACKNOWLEDGMENT

This work was financially supported by the Big Data Industry Development Pilot Demonstration Project of MIIT in 2018 and Major Scientific and Technological Innovation Projects of Zhengzhou (188PCXZX773)

## REFERENCES

[1] Liu Zhicheng, Zhu Yongxin, Wang Hui, et al. Multi target real-time detection based on convolutional neural network [J]. Computer engineering and design, 2019,40 (04): 1085-1090.

[2] Liu Min, Wu Minghu, Wang Juan. Improved method of motion detection in dynamic scene [J]. Computer Engineering, 2017,43 (05): 290-293.

[3] Kang Jie, Li Xiaojing. Moving target detection based on mean background and three frame difference [J]. Journal of Shaanxi University of science and technology, 2018,36 (01): 148-153.

[4] Liu Wei, Hao Xiaoli, LV Jinin. Efficient moving target detection based on Adaptive Gaussian mixture modeling [J]. Journal of Image and Graphics, 2020,25 (01): 113-125.

[5] Liang Shuo. Research on moving target detection and tracking algorithm based on background subtraction [D]. Xi'an Petroleum University, 2019.

[6] Gao Jianfu, Chen Jian. Moving target detection method based on improved vibe algorithm [J]. Computer applications, 2017,37 (S2): 99-102.

[7] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[8] Girshick R .Fast R-CNN[J].arXiv preprint arXiv: 1504.08083,2015.

[9] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada, 2015: 91–99.

[10] Redmon J, Divvala S, Girshick R, et al. You only lookonce: Unified, real-time object detection[J]. arXiv preprintarXiv: 1506.02640, 2015.

[11] Liu Wei, Anguelov D, Erhan D, et al . SSD: Single shotmultibox detector[J]. arXiv preprint arXiv: 1512.02325,2015.