Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 221 (2023) 585–592

Procedia Computer Science

www.elsevier.com/locate/procedia

10th International Conference on Information Technology and Quantitative Management

Moving target detection based on background modeling and frame difference

Qiaozhi Tan[a,*], Zhongqiang Du[a], Sheng Chen[b]

[a]*Beijing Institute of Technology, Beijing 100081, China*
[b]*Wuhan Polytechnic University, Wuhan 430048, China*

Abstract

The integration of Unmanned Aerial Vehicle (UAV) technology with moving target detection has diverse applications in military reconnaissance, space remote sensing, and smart cities. Traditional motion-based target detection algorithms offer fast processing speeds but lack accuracy. Deep learning-based algorithms, while accurate for specific targets only, are complex and not suitable for resource-limited UAV platforms and lack real-time performance. Therefore, this study proposes a real-time moving target detection algorithm for UAV platforms based on traditional frame difference algorithm. The purpose of this algorithm is to improve detection accuracy, which has been hindered by the limitations of traditional algorithms caused by camera shake, background changes, and fast-moving targets. The algorithm involves rough background modeling, background updating during subsequent video image sequences, image morphology processing, and background compensation. Experimental results from multiple sets of UAV-borne video data demonstrate the algorithm's high target detection rate, low false alarm rate, and ability to detect moving targets stably in complex environments. The proposed algorithm achieves a speed of 25 FPS and a detection accuracy of 91.8%, meeting real-time and accurate detection requirements.

Keywords: Frame difference; Image registration; Dynamic background compensation.

1. Introduction

In recent years, the use of small and flexible Unmanned Aerial Vehicles (UAVs)[1] has become increasingly popular due to their ability to operate without restrictions and provide a wide range of motion, wide-angle view, and

* Corresponding author. Tel.: +86-188-0113-9006.
E-mail address: qiaozhitan@bit.edu.cn

10.1016/j.procs.2023.08.026

high resolution. As a result, UAVs are widely used in regional monitoring, traffic detection, disaster investigation, battlefield surveillance, reconnaissance, target identification, tracking[2,3] and strike assessment. The integration of moving object detection technology with UAV technology has become a popular topic in computer vision, offering new insights for both civilian and military applications, such as intelligent transportation, intelligent security, 3D reconstruction, and battlefield situational awareness[4,5,6].

Research in moving target detection algorithms[7,8] for UAV-captured video primarily focuses on detecting moving vehicles. Carnegie Mellon University's VSAM visual monitoring project [9] used high-mounted cameras to detect moving targets on the ground in 1997. ALI S developed the COCOA system [10] in 2005, which processes video images captured by manned-unmanned platforms and detects moving targets to reproduce their trajectory. Shastry A.C. used feature tracking to improve image registration accuracy in 2005 [11] and utilized the frame difference method for moving target detection, achieving a detection accuracy of only 65%. Abdelwahab proposed an airborne camera-based moving vehicle detection technology[12], which measures histogram changes around feature points to obtain the foreground and target. Dong Jing proposed a real-time moving target detection algorithm for UAV video images [13] that combines registration with the frame difference method for moving area extraction. However, the algorithm has limitations in detecting objects with fewer pixels and slower moving speeds. Peng Bo proposed a method of symmetric reconnaissance combined with block background modeling [14], achieving high detection rates but with the inability to detect targets with gray levels similar to the background. Recently, deep learning-based object detection algorithms like Faster region-based convolutional neural networks (Faster R-CNN) [15], Single-Shot Detector (SSD)[16], and You Only Look Once (YOLO) [17]have been proposed.

Various methods have been proposed for moving target detection, each with its pros and cons. The frame difference method and its variations are computationally efficient but sensitive to illumination and background changes, leading to low detection precision[18]. Background modeling methods such as the Gaussian mixture model (GMM) [19]can handle dynamic backgrounds but are less adaptable to significant scene changes. Optical flow-based methods [20] are complicated and have poor real-time performance, while deep learning-based approaches require vast quantities of data for training and substantial resource utilization during the application process, unsuitable for UAV-borne platforms due to limited resources and restricted storage capacity. To address the challenges posed by UAV platform motion and background changes, we propose a novel moving target detection algorithm that combines the benefits of the frame difference method and background modeling. Our algorithm optimizes the three-frame difference method by adopting background modeling to enhance the robustness and accuracy of the algorithm. Additionally, we improve the frame difference method by incorporating coarse and fine registration techniques, morphology operations, and statistical clustering of connected regions to ensure adaptability to static and dynamic backgrounds, reduce false alarms, and improve detection accuracy.

2. Algorithm design and implementation

2.1. Overall framework

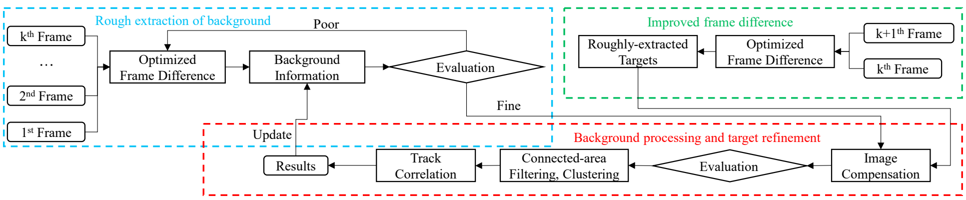


Fig. 1. The process of proposed algorithm.

The proposed algorithm, depicted in Fig. 1, is a real-time approach that integrates background modeling into an optimized three-frame difference method for UAV-based moving target detection. It follows a systematic framework consisting of several stages. Firstly, the algorithm utilizes a rough background model replacing the first two frames of the three-frame difference method. Subsequently, it employs an optimized frame difference method involving feature extraction, rough and fine matching, thresholding, filtering, and image morphology operations to obtain rough target detection information. Next, the algorithm processes the extracted background and target information, eliminating noise interference and obtaining the approximate target location through connected region screening and clustering. Lastly, it further reduces noise interference through trajectory association, updates the background information synchronously, and enhances overall detection accuracy.

The proposed algorithm improves upon the traditional three-frame difference method by enhancing the detection of small moving targets in dynamic backgrounds. By integrating background modeling and frame difference methods, the algorithm achieves enhanced robustness and accuracy for UAV-based moving target detection. Its potential for real-world applications is substantial.

2.2. Rough extraction of background

The proposed algorithm's crucial preprocessing step is the background extraction process, which employs the GMM method to extract the background from the first k frames of video data captured by the UAV platform. Because the drone-acquired video's background modeling method cannot be processed using simple mean or median methods. Therefore, the proposed algorithm uses the GMM method for initial background modeling, which estimates a Gaussian mixture distribution of the pixel values at each pixel location to model the background. The modeling principle is as follows:

$$P_b = \begin{cases} 1, & \sum_{t}^{-1}\left(I_t\left(x,y\right)-\mu_t\left(x,y\right)\right)\left(I_t\left(x,y\right)-\mu_t\left(x,y\right)\right)^T > T \\ 0, & Others \end{cases} \tag{1}$$

The preceding paragraph explains how the algorithm uses the GMM method for initializing the background model and how it determines if a pixel value corresponds to a moving target. Specifically, $I_t(x, y)$ represents the pixel value at position (x, y), and $\mu_t(x, y)$ is the mean value of the Gaussian distribution at this position. The threshold T is critical in determining if the pixel value conforms to the model and can be classified as a moving target. However, a manually set threshold may not be adaptable to changing environments, which can reduce the algorithm's accuracy. To address this issue, the algorithm incorporates an adaptive threshold calculated using the expectation-maximization (EM) algorithm. as shown in the evaluation part of Fig. 1. It's important to note that the threshold is only used for a rough background model, and the algorithm subsequently updates and optimizes the background and other information to improve accuracy. The extracted background result from Section 2.2 will replace the frame difference information of the first two frames in the initial phase of the three-frame difference method.

2.3. Improved frame difference

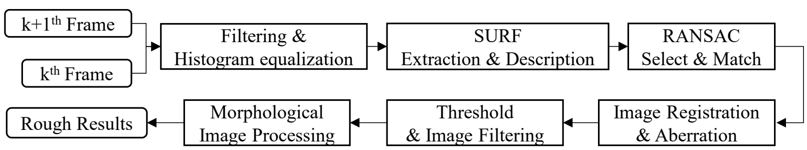


Fig. 2. The process of improved frame difference method

After extracting the background, as illustrated in Fig. 2, this proposed method introduces an improved frame difference method that overcomes the limitations of the traditional method in handling illumination interference and background disturbances, as well as its inability to adapt to dynamic background changes. The proposed method involves preprocessing operations such as filtering and equalization on the acquired video sequence images, followed by feature extraction and description using the Speeded Up Robust Features (SURF) algorithm. Abnormal data screening is then performed on the extracted feature points using the Random Sample Consensus (RANSAC) method, and an affine transformation matrix is calculated to enable image registration, allowing for dynamic compensation of the background. Finally, preliminary moving target detection is achieved through image threshold processing, filtering, morphological operations, and other techniques. The proposed method is characterized by fast calculation speed, strong anti-noise ability, and robustness.

2.3.1. Filtering and histogram equalization

This method employs median filtering to remove image noise, reduces computation by converting the image to grayscale, and enhances image details through histogram enhancement. These steps facilitate subsequent tasks as preprocessing step.

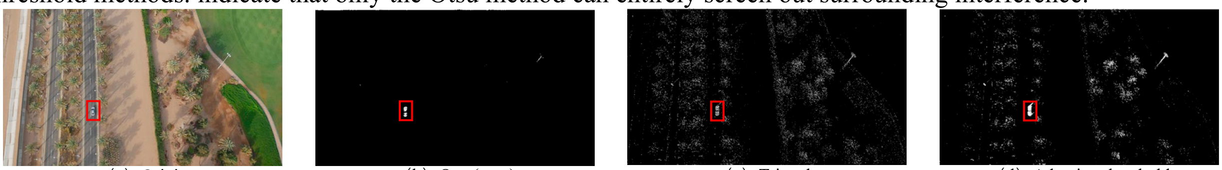
2.3.2. SURF extraction and description

The SURF algorithm is a feature detection algorithm that is known for its robustness and speed. In this step, a low Hessian threshold is typically chosen to account for factors such as camera shake and lighting. While it increases the number of recognized features and improves accuracy, a lower threshold can also slow down feature extraction. Therefore, this paper's algorithm selects the Hessian threshold based on the image size, with a threshold of 400 for images with a length and width less than 512, and 800 for larger images. If feature extraction is slow, the Hessian threshold is adjusted to 1200.

2.3.3. Threshold and morphological processing

Image registration involves calculating the optimal mapping transformation matrix for image perspective transformation, but it is not possible to achieve complete image registration correction due to calculation accuracy errors and noise. As a result, the processed image after registration may not be clean and can contain noise mixed with the target, which will interfere with target detection. To address this, it is necessary to perform threshold processing, image filtering, and morphological processing to eliminate noise and initially screen the target.

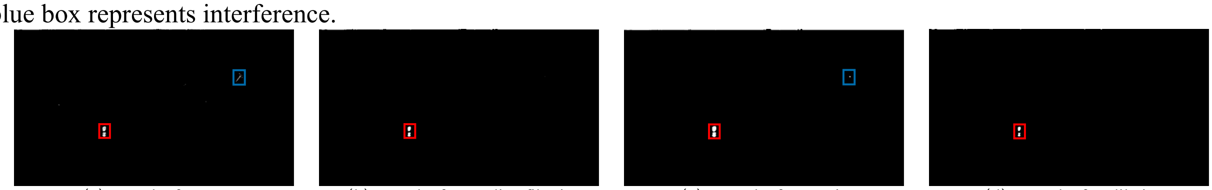
A threshold segmentation algorithm with strong adaptability and few parameters is needed to meet the algorithm's robustness and real-time requirements, given poor image registration results and plaque interference caused by inaccurate registration. Since most scenarios demand small target detection, the threshold value is crucial, and the Otsu method is chosen for its adaptability and robustness. The results shown in Fig. 3. Comparison of the results of different threshold methods. indicate that only the Otsu method can entirely screen out surrounding interference.



(a) Origin (b) Otsu(ours) (c) Triangle (d) Adaptive-threshold

Fig. 3. Comparison of the results of different threshold methods.

The resulting image may still contain patchy and point-like interferences, leading to high false alarms. To address this, a median filter is applied to process impulse noise and salt-and-pepper noise while preserving edges. Image morphology techniques, including dilation and erosion, are then used to enhance the image and process the boundary and connected regions of the target object. Fig. 4 shows the results, where the red box represents the target and the blue box represents interference.



(a) Result after Otsu (b) Result after median filtering (c) Result after erosion (d) Result after dilation

Fig. 4. Results of median filtering and image morphology processing.

2.4. Background processing and target refinement

2.4.1. Background compensation and update

To improve the accuracy of the algorithm, the proposed approach combines the three-frame frame difference method with background modeling. Although the frame difference method can lead to false detections in feature point determination, matching, and image stages, the algorithm refines and purifies the preliminary results to increase the detection rate. By making difference results between two adjacent frames and using background modeling, the algorithm compensates for the background and detection results, reducing false detections and completing global background compensation.

$$P_e \,\&\, P_b = P_t \tag{2}$$

The background model needs to be continuously updated during the detection process to ensure the reliability of the test results. The update formula is as follows:

$$P_{b+1} = \begin{cases} \alpha P_e + (1-\alpha) P_b & flag = 1 \\ P_e & flag = 0 \end{cases} \tag{3}$$

The symbol α in the formula represents the learning rate for updating the background, with a value range of $[0,1]$. A higher α value results in a faster learning rate and a faster update speed for the background model. The $flag$ is an indicator used to evaluate the detection results after background compensation. If there is significant interference or the target is lost in the result, the $flag$ is set to 0, which typically occurs during periods of lens shake, registration difficulty, high-speed movement of the target or camera, or target loss. These situations are typically short-lived, and the background is normally updated with $\alpha = 0.8$.

2.4.2. Target refinement

After applying global background compensation, the interference caused by background motion can be largely eliminated, and the moving target needs to be extracted from the resulting image. Following threshold segmentation, the image is binarized, and connected domains are extracted and their centroids are marked. To ensure real-time performance and reduce computational complexity, this algorithm uses 4-connected domain analysis to obtain parameters such as area, position, and centroid coordinates of each connected domain. In the purification step, the area of the connected domain serves as the threshold screening standard to identify moving targets. Given that the image registration only operates on the core part of the image, and not on the edges, a scope is defined as $Bound_{truth} = \beta \times Bound_{origin}$, where $\beta = 0.98$.

To prevent a single target from being segmented due to subsequent image processing, distance judgment is performed based on the center position parameters obtained from connected domain analysis. If the distance between two targets is less than the connected threshold, i.e., $dis(T_{c1}, T_{c2}) < d_{min-con}$, they are merged into one. Fig. 5 illustrates target detection frame aggregation. The target in the middle was separated into two parts due to image subtraction in the frame difference process. The image clustering algorithm aggregates the two parts seperated from a single target, resulting in the original target.

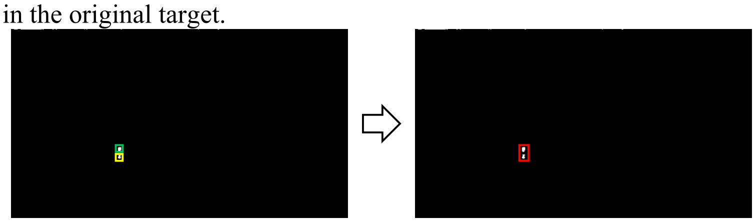


Fig. 5. Results of target refinement.

The final purification step involves associating candidate target trajectories to reduce the false alarm rate. Trajectory continuity is essential for detecting moving targets, and two conditions are considered. Firstly, the movement of a target should be continuous within a certain range, i.e., $dis(T_{c1,pre}, T_{c2,cur}) < d_{min-move}$. Secondly, interference and noise usually lack continuity, i.e., $dis(N_{c1,pre}, N_{c2,cur}) \geq d_{min-move}$. By combining these conditions, the range of motion for the central position of the candidate target can be estimated through trajectory association in continuous frame detection. This eliminates candidate targets outside the range of motion, resulting in the final moving target result.

3. Experimental results and analysis

The proposed real-time moving target detection algorithm was evaluated for its detection accuracy, real-time performance, and robustness using ground environment and moving vehicle videos captured by UAVs. The evaluation utilized the UAV123 data set and additional self-collected dataset, covering scenes such as fields, cities, and highways. The data set comprised 13,589 frames, distributed as in Fig. 6:

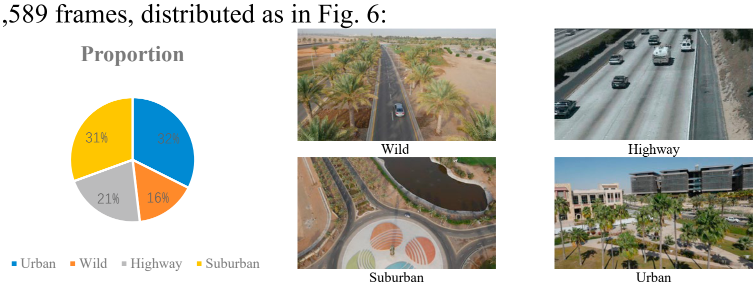


Fig. 6. Brief of experimental dataset.

3.1. Qualitative analysis of detection performance

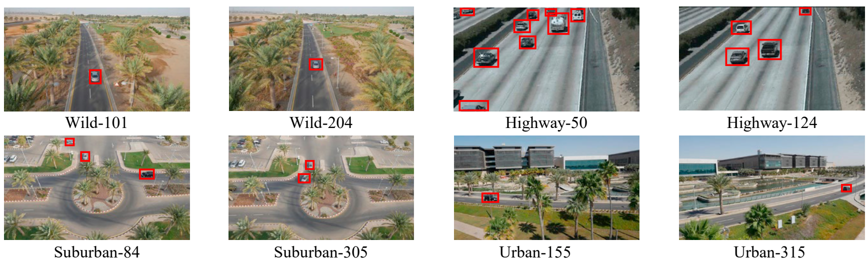


Fig. 7. Part of the moving target detection results.

Fig. 7 showcases the detection results for four classes in the test data, with the number next to each category indicating the corresponding image count. The algorithm demonstrates exceptional detection capabilities across diverse datasets, especially in multi-target scenarios with low false alarms and missed detections. This achievement is credited to the amalgamation of rough and fine target extraction in the improved frame difference method. Furthermore, the algorithm's robust detection performance on occluded targets can be attributed to the active background compensation and real-time background updates.

3.2. Quantitative analysis of detection accuracy

The algorithm's performance is quantitatively evaluated and presented in Table 1, including metrics: FP (false alarm proportion), FN (missed detection proportion), and TP (correctly detected frames proportion). The evaluation was conducted on a test platform equipped with an I7-11800H CPU, 32GB of memory, and an RTX 3070 GPU.

Table 1. Accuracy on different kinds of Datasets.

Dataset Category	FP	FN	TP
Highway	4.3%	5.5%	92.7%
Urban	5.7%	6.6%	90.6%
Suburban	4.8%	5.7%	92.5%
Wild	6.2%	4.6%	90.9%
Total	5.4%	5.6%	91.8%

The algorithm's performance was tested using all image data mentioned above, revealing varying accuracy across the four data types. Road data showed the highest accuracy due to the clean background and clear target trajectory. Urban data had the lowest accuracy due to occlusion, vegetation, and similar target interference. Suburban data had a relatively simple background, but small targets resulted in high missed detection rates. The complex background of field data with significant changes resulted in increased false detection rates.

Table 2 presents a comparison between the proposed method and the original three-frame difference. Our method exhibits superior accuracy and speed, as demonstrated in the table.

Table 2. Contrast of accuracy and speed on different kinds of methods.

Dataset Category	Accuracy	Speed (FPS)
Three-frame difference	73.2%	10.2
Three-frame difference + Canny	82.2%	17.3
Three-frame difference + ViBe	83.9%	13.2
Three-frame difference + Optical Flow	86.3%	4.4
Ours	91.8%	25.4

3.3. Detection efficiency

According to tests on the dataset used, the proposed algorithms achieve a real-time processing speed of over 25 FPS for UAV video data with the size of 1280×720. However, for larger image data, the algorithm's computational time increases. To maintain details while increasing the algorithm's running speed, images can be appropriately scaled.

3.4. Failure cases

missed detection (blue) false detection (yellow)

Fig. 8 Example of the failure cases.

In the experiment, some false detections occurred due to incomplete detection of certain targets and mutual occlusion between targets. Additionally, false detections were observed as a result of interference in complex backgrounds, as depicted in Fig. 8.

4. Conclusion

This paper presents an enhanced moving target detection algorithm for UAV-captured video. It improves upon the traditional three-frame frame-difference method by integrating background modeling. The algorithm introduces the pre-sequence frame information into the process of moving target detection in each frame, and use coherent background information for compensation to eliminates the noise and interference problems in the image, so as to improve the detection accuracy of the algorithm. At the same time, the method ensures the robustness of detection in different scenes through image registration and background cyclic updating. Finally, the algorithm adopts the space-for-time strategy, and only processes the current frame image for each frame detection, and stores and calls the information of the pre-order frame without additional processing to ensure the real-time performance of the algorithm.

The proposed algorithm meets real-time requirements with a speed of 25FPS and achieves a high accuracy rate of 91.8%. It demonstrates robustness in tests on both self-collected and public datasets, accurately detecting moving targets in most scenes while minimizing false alarms.

This study has certain limitations. Acceleration processors such as GPUs were not considered to further enhance the algorithm's running speed, and the rapid movement of small moving objects may invalidate the trajectory correlation part, challenging real-time and stable detection of moving targets under the UAV platform. These limitations warrant further research in the future.

References

[1] Y. Han, H. Liu, Y. Wang and C. Liu. (2022) "A Comprehensive Review for Typical Applications Based Upon Unmanned Aerial Vehicle Platform," IEEE *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **15**: 9654-9666.

[2] Y. Han, C. Deng, B. Zhao and D. Tao. (2019)"State-Aware Anti-Drift Object Tracking," IEEE Transactions on Image Processing, 28(8): 4075-4086

[3] Han, Y, Wang, H, Zhang, Z, and Wang, Z. (2020). “Boundary-aware vehicle tracking upon UAV.” Electronics Letters, 56(17):873-876.

[4] Deren, Li, and Ming, Li. (2014) “Research advance and application prospect of unmanned aerial vehicle remote sensing system.” Geomatics and Information Science of Wuhan University 39 (5): 505–513.

[5] Reinartz, P, Krau, T, Ptzsch, M, Runge, H, and Zuev, S. (2006) “Traffic monitoring with serial images from airborne cameras.” ISPRS Workshop, High-Resolution Earth Imaging for Geospatial Information 61(3-4): 149-158.

[6] Pollard, T. and Antone, M. (2012). “Detecting and tracking all moving objects in wide-area aerial video.” Proceedings of 2012 Computer Vision and Pattern Recognition Workshops.

[7] Shi, Hao, Zhonghao Fang, Yupei Wang, and Liang Chen. (2022). "An Adaptive Sample Assignment Strategy Based on Feature Enhancement for Ship Detection in SAR Images" *Remote Sensing* **14(9)**: 2238.

[8] Shi, H., He, C., Li, J., Chen, L., and Wang, Y. (2022). “An improved anchor-free SAR ship detection algorithm based on brain-inspired attention mechanism.” Frontiers in neuroscience.

[9] Collins, R. T. (2000) “A system for video surveillance and monitoring: VSAM final report.” Carnegie: Carnegie Mellon University.

[10] Saad, A. (2006). “Cocoa-tracking in aerial imagery, airborne intelligence.” Surveillance Reconnaissance Systems & Applications III, 6209(1): 105-114.

[11] Shastry, A. C., and Schowengerdt, R. A. (2005) “Airborne video registration and traffic-flow parameter estimation.” IEEE Transactions on Intelligent Transportation Systems, 6(4): 391-405.

[12] Abdelwahab, M. A., and Abdelwahab, M. M. (2016) “A Novel Algorithm for Vehicle Detection and Tracking in Airborne Videos.” Proceedings of IEEE International Symposium on Multimedia.

[13] Dong, J, Fu, D, and Yang, X. (2013). “Real-time moving object detection and tracking by using UAV videos.” Journal of Applied Optics 34(2): 255-259.

[14] Peng, B, Cai, X, Zhang, Y, and Li, S. (2017) “Automatic vehicle detection from UAV videos based on symmetrical frame difference and background block modeling.” Journal of Southeast University, 47(4): 685-690.

[15] Ren, S, He, K, Girshick, R, and Sun, J. (2017) “Faster R-CNN: Towards real-time object detection with region proposal networks.” IEEE Transactions on Pattern Analysis & Machine Intelligence 39(6): 1137-1149.

[16] Redmon, J, Divvala, S, Girshick, R, and Farhadi, A. (2016) “You Only Look Once: Unified, Real-Time Object Detection.” Computer Vision & Pattern Recognition, IEEE.

[17] Liu W, Anguelov D, Erhan D. (2016) “Single Shot MultiBox Detector.” European Conference on Computer Vision, Amsterdam, Netherlands.

[18] Cheng Ai L, Huang X, Li X Y. (2017) “Review on the algorithm of moving target detection.” Information & Communication.

[19] Mushawwir, L. A, and Supriana, I. (2016) “Gaussian mixture model and spatial-temporal evaluation for object detection and tracking in video surveillance system.” International Conference on Information Technology Systems and Innovation, IEEE.

[20] Charif, Z. (2011) “Fast modified Horn & Schunck method for the estimation of optical flow fields.” Signal Processing Systems, IEEE.