Vehicle Detection in UAV Aerial Video[a]

Huan Zhang[1], Cai Meng[1*], Pin Guo[2], Xilun Ding[2] and Zhaoxi Li[1]

***Abstract*—In recent years, unmanned aerial vehicle (UAV) has been increasingly applied to traffic monitoring field, especially the vehicle detection. However, there are many challenges for vehicle detection in the UAV aerial video, such as camera shake, interferential targets and a wide range of change of scene. To solve these problems, a new vehicle detection system which is suitable for detection in UAV aerial video is proposed. We use the bit plane to extract the lane surface and use the extracted lane surface to limit the detection area. We improve the Vibe algorithm so that it can be used in rapidly changing scenes. In addition, the moving target screening strategy is proposed to screen moving vehicles. This paper is the first one to introduce bit plane into detection method. Our novel detection system is another major contribution. Experiments show that our system outperforms existing detection algorithms in terms of accuracy and computation time.**

I. INTRODUCTION

In recent years, with the prosperity of the transportation industry, lane detection, vehicle detection and vehicle classification, etc. have become the most popular areas in the field of traffic monitoring [1]. At the same time, unmanned aerial vehicles (UAV), as a platform for collecting data, have been increasingly applied to traffic monitoring and other fields.

The application of UAV in vehicle detection has a series of advantages, such as flexibility, friendliness, adjustable traffic monitoring range, and on-demand image acquisition. However, there are also a series of challenges: (1) During flight, UAV could be affected by the weather. The UAV will shake and drift, so the aerial video will be swaying and distorting. (2) Due to the wide shooting range of UAV, aerial video will inevitably be introduced into the large background area, which will interfere with vehicle detection [2]. (3) UAV moves quickly and flexibly, resulting in rapid changes in background. (4) There are pedestrians and other moving interference targets. Because of these problems, the detection in aerial video needs to be robust, anti-interference and real-time.

In dynamic scenarios, motion based methods, such as the optical flow method and the motion vector based method, are expensive in computing and difficult to detect real-time. In the feature based method, the simple feature method is less robust, and the subarea method is complex, which is not suitable for real time detection. Deep learning based method, such as HDNN [3] and CNN system combined with SVM classifier [4], are used to detect in UAV images rather than aerial video. The background modeling method has faster detection speed than the feature based and motion based methods, and detect the moving target accurately.

In this paper, we study the best real-time algorithm in the current detection algorithm [20], the Vibe algorithm [14]. Many researchers have studied and optimized the algorithm. Li, et al. [5] integrated the adjacent frame difference with Vibe algorithm to remove the ghost area. The PBAS algorithm combined the advantages of the SACON algorithm and the Vibe algorithm, and optimized on these basis [6]. Ehsan Mahoor, et al. changed the pixel value in the sample set to the frame difference value, so even if the camera is jitter, their algorithm is still effective [7]. Jin and his colleges fused the improved Canny operator with the Vibe algorithm to get a more accurate foreground region [8].

Unfortunately, most of the improved Vibe algorithms are still limited to detect targets in static scenes, which is not suitable for detecting targets in moving aerial video. As mentioned, due to the large range of aerial video, the moving background area can be easily detected as the foreground target. In order to reduce the background complexity, we used lane information to restrict the detection area dynamically. Because of dramatic changes in the scene, we improved the Vibe algorithm to enable it to detect in a rapidly changing environment. There are interference targets, such as pedestrians. We proposed a series of vehicle screening criteria to screen these targets. To the author's knowledge, this paper is the first one to introduce bit plane into detection method. In addition, the design of our vehicle detection system, which is suitable to detect in aerial video, is the paper's main contribution as well.

II. DYNAMIC DETECTION AREA

The scene in UAV aerial video often changes, and the detection area should be changing accordingly. In order to reduce the background complexity and limit the shooting range, the detection area should be limited to the lane surface. The traditional fixed ROI areas are not suitable for aerial video for they are updated very slowly or not updated. Therefore, we propose our dynamic detection area using the lane information.

The current lane detection methods include feature-based methods and model-based methods. In the model-based methods, the mathematical models of lane lines are established and optimized, such as line model, two degree curve model and hyperbolic model, etc.[9].Feature-based methods use the color, texture, margin, etc. to extract lane. Most lane detection algorithms are designed according to the specific system tasks and application environments [10] [11] [12]. These algorithms are not universal when it comes to harsh environments such as raining or nightscape. Therefore, we design a lane detection

a: Thanks to the support by NFSC with Grants no. 91748201, and by the sponsorship from the Mohamed Bin Zayed International Robotics Challenge 2017

1: School of Astronautics, Beihang University, Beijing, China.

2: School of Mechanical and Automation, Beihang University, Beijing, China

*Corresponding author: tsai@buaa.edu.cn

algorithm based on bit plane, which is composed of lane surface extraction, lane detection and lane tracking.

A. *Lane Surface Extraction*

For further lane detection, frames are usually thresholding to generate the binary images, and edges are extracted by Canny operator on the binary images. In the thresholding, the integrity of the lane surface is damaged, which will affect further lane detection. Therefore, we introduce bit planes to maintain road integrity. The bit plane describes the grayscale attribute of the image. The grayscale of a m bit image can be expressed as a polynomial:

$$P_m = a_{m-1}2^{m-1} + a_{m-2}2^{m-2} + \cdots + a_0 2^0 \qquad (1)$$

$a_i (i = 0,1, \ldots m-1)$ is the coefficient of the bit planes. An 8 bits image can be considered as 8 planes of 1 bit, including 4 high-older planes and 4 low-older planes. The high-older planes contains important information, such as the contours. The low-older contains the details of the image, as shown in Figure 1. In aerial video, the details of images change greatly, which is considered to be interference in dynamic area detection. In order to avoid small grayscale changes influencing the lane surface extraction greatly, we only use high old bit plane to get the dynamic area.

In different application scenarios, we choose bit planes according to illumination conditions. The 8th bit plane is usually used to extract the road surface. In raining or night scape, we use the 7th bit plane. Even in bad weather and night scenes, the lane surface extraction results of bit plane are still better than the Ostu results. The experimental section will introduce the details of the experiment.

On the selected bit plane, we calculate the vanishing point of the road, i.e. where the lane vanishes. By calculating the average grayscale μ_i of row i in the image (m rows and n columns), we get the vanishing point of lane.

$$\mu_i = \frac{\sum_{j=0}^{n} v_{i,j}}{n_i}, (i = 1,2, \ldots m, j = 1,2, \ldots n) \qquad (2)$$

$v_{i,j}$ represents the pixel grayscale. On the vanishing point, the average grayscale μ_i decrease sharply:

$$|\mu_{i-1} - \mu_i| > T_1 \qquad (3)$$

The threshold T_1 is set according to the specific application scenario. The pixels above the horizon are undesired and these pixels will be set to zero.

B. Lane Detection

After extracting the lane surface, the Canny operator is used to detect edges on the binary 8th bit plane images. After the probability Hough transform, the straight line can be expressed as: $\rho = x \cdot cos\theta + y \cdot sin\theta$. ρ is the distance from the line to the origin of the coordinate system. θ is the angle between the line and the x axis.

$$l = \begin{cases} l_f & \theta \epsilon (90°, 180°) \\ l_r & \theta \epsilon (0°, 90°) \end{cases} \quad if\ \theta < 0°, \theta = \theta + 180° \qquad (4)$$

According to the angle θ, the straight line would be determined whether it belongs to the left lane l_f or the right lane l_r, and the two lateral lines are collected respectively. Angle information is used to screen out the outermost lane on both sides. Finally, the outmost point P_L, P_R of the outermost lanes of the left and right sides are used as the rectangle corners to achieve the dynamic detection area. The use of rectangles instead of other curves is to facilitate lane tracking and limit computation complexity.

C. Lane Tracking

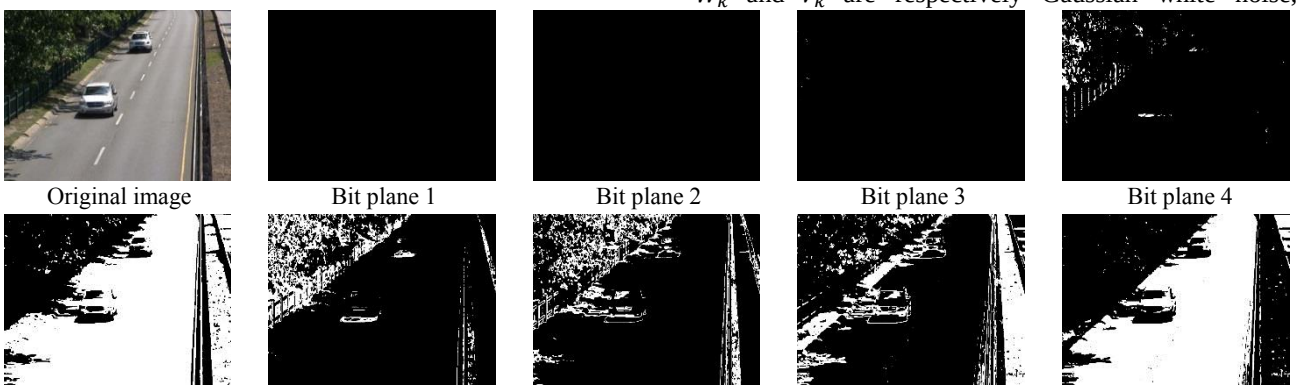
Due to the jitter and mutation in aerial video, the dynamic detection areas between adjacent frames will change greatly. Therefore, the object located at the edge of the lane can be easily detected as a vehicle. Here, we use Kalman filtering to limit area variance [13]. Traditionally, Kalman filter is used to track ρ and θ of lanes [21]. In our method, we track the outermost point of the lane. Take the left lane for example. Kalman filter uses previous state and current measurement to predict the optimal state. Lane tracking system can be described by linear stochastic differential equation:

$$X_k = AX_{k-1} + BU_k + W_k \quad Z_k = HX_k + V_k \qquad (5)$$

The position and its variants are defined as the state vector X, and the position is defined as the measurement vector Z:

$$\mathrm{X} = [X_{Pl}\ Y_{Pl}\ X_{Pl}'\ Y_{Pl}']^T \quad \mathrm{Z} = [X_{Pl}\ Y_{Pl}]^T \qquad (6)$$

This 4×4 state updating matrix A is set to the unit matrix. The measurement matrix H is set to the identity matrix. Both W_k and V_k are respectively Gaussian white noise, i.e.



Original image | Bit plane 1 | Bit plane 2 | Bit plane 3 | Bit plane 4

Otsu threshold | Bit plane 5 | Bit plane 6 | Bit plane 7 | Bit plane 8

Figure 1. The binarization results for lane surface extraction. The top left is the original image. The bottom left is the binary image by the Otsu threshold. The others are the binary images by bit planes, ranging from bit plane 1 to bit plane 8. The high-older bit planes contains the main information of lane and the low-older bit planes contains the details.

$W_k \sim N(0, Q)$, $V_k \sim N(0, R)$. The Kalman filter is used to track the outmost points of lanes as follows.

The current state is predicted by using the optimization result of the former state $X_{k-1|k-1}$. There is no control, so the U_k is set to 0. Q represents the noise matrix of the system, which is set to the identity matrix in the application. The measurement noise matrix R is set as an identity matrix. The specific calculation steps are as follows:

$$X_{k|k-1} = AX_{k-1|k-1} + W_k \tag{7}$$

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q \tag{8}$$

$$X_{k|k} = X_{k|k-1} + K_k(Z_k - HX_{k|k-1}) \tag{9}$$

$$K_k = P_{k|k-1}H^T / (HP_{k|k-1}H^T + R) \tag{10}$$

$$P_{k|k} = (I - K_k H)\, P_{k|k-1} \tag{11}$$

Through these steps, lane tracking is completed by the predicting of the outmost points of the lanes. Since each Kalman prediction requires only one iteration, the time cost is negligible. By designing the dynamic detection area, the background is effectively eliminated. The complexity of vehicle detection is reduced.

III. Improved Vibe Algorithm

To the best of our knowledge, Vibe algorithm is used for the video surveillance in static background. When applied to the detection of dynamic scenes, there will be slow background updates and ghost problems [14] [15]. Therefore, we improved the Vibe algorithm and applied it to the dynamic scenes.

In Vibe, only one frame is needed to build the background model. By using adjacent pixels with spatial consistency, N samples are randomly selected from the eight neighborhoods as their sample set. The sample set of the whole image pixel is the initial background model. The sample set of any pixel $M(x, y, t)$ in the image is $M(x, y, t) = \{v_1, v_2, \dots, v_i, \dots v_N\}$, $v_i (i = 1,2, \dots N)$ is the pixel value of the sample set point.

The pixel $M_1(x, y, t)$ in the current frame is defined as the center of a circle, and the R is defined as the radius of the circle recorded as $SR(M_1)$. Compare $M_1(x, y, t)$ with the pixel of the same position in the previous frame M. If the intersection point is greater than the threshold T_{Vibe}: $\#\{SR(M_1) \cap M\} > T_{Vibe}$, the current point $M_1(x, y, t)$ is the background point, and enter the background update step.

When the current pixel is determined as a background pixel, it has a probability of $1/\varphi$ to update itself (φ is usually set to16), i.e., a pixel value v_i is randomly selected from the N samples to be updated. It also has a probability of $1/\varphi$ to update the neighbor pixels. The probability of sample point remained in the sample set decreases exponential over time. The background model can be updated quickly and effectively. When a pixel is continuously determined as foreground pixel of multi frames, it is updated to background pixel. That is to prevent the phenomenon of deadlock.

A. Improvement

In static environment, the background is almost unchanged, and the Vibe algorithm is robust. However, when the UAV moves quickly, the camera jittering, the background changes drastically, and the amount of the foreground increases sharply. The ghost appears, and the false objects are eliminated very slowly. We use morphological open operation to eliminate the noise and then we use morphological close operation to fill the holes. At the same time, the previous scene model is not suitable for the current scene, so we set different update rates of $1/\varphi$ based on the complexity of the scene to speed up the elimination of false targets. In the current frame, the number of foreground targets is defined as $\sum_t^i$:

$$1/\varphi = \begin{cases} 1/16, & \sum_t^i \le 10 \\ 1/4, & 10 < \sum_t^i < 20 \\ 1/2, & \sum_t^i \ge 20 \end{cases} \tag{12}$$

As shown in Table 2, the scene updates faster with the increase of $1/\varphi$.

IV. Vehicle Screening Method

After detection in dynamic detection area, a series of moving targets will appear. Therefore, we filter the moving targets to get vehicles. In aerial video, the widely used vehicle screening features, such as color, texture, shadow, texture, symmetry, are easily affected by illumination and other factors. Therefore, we provide our screening criteria: scale criteria, shape criteria, and direction criteria.

A. Scale Criteria

Vehicle detection is usually carried out at a fixed scale, and the scale characteristics of vehicles are generally ignored. The size of moving target in aerial video is related to the flight height of UAV. When the flight altitude is low, the moving target will be correspondingly large, and the smaller foreground target is the unwanted object, and vice versa. Therefore, UAV flight height can be used to determine the size of vehicle. Let $V(x, y, t)$ represent the current moving target area. V_L and V_H are low threshold and high threshold of the vehicle area. H is defined as the flight height. V_L and V_H are the functions of H. We use pinhole camera models to build the projection models.

The corresponding relation between the point coordinates in the real world coordinate system $P(x, y, z)$ and the image pixel coordinates system (u, v, d) (d is depth data) can be described as:

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = C \cdot \left(R \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} + T \right) \quad d = z \cdot s \tag{13}$$

$$C = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{14}$$

R and T are the camera attitudes. R represents the rotation matrix and T represents the translation matrix. C represents the internal parameter matrix of the camera. s refers to the scale factor of depth. f_x, f_y represent the camera focal length on the x and y axis in the camera coordinate system. c_x, c_y refer to the camera aperture center. f_x, f_y, c_x, c_y could be

obtained by camera calibration. Since the pod can be self-stabilized, we set the pod to always look down, and the optical axis of the pod is perpendicular to the ground. Therefore, the camera does not rotate and translate. R is set to a unit matrix I, and T is set to zero. The above formula can be simplified to:

$$u = x \cdot f_x/z + c_x \quad v = y \cdot f_y/z + c_y \tag{15}$$

We use the laser on the UAV to obtain the distance z between the pod and the ground, which is similar to the flight height H. Utilizing actual vehicle width and height, the vehicle size threshold V_L and V_H in the image can be obtained. If the size of current target is:

$$V \ll V_L \text{ or} V \gg V_H \quad V_L = f(H), V_{H=}f(H) \tag{16}$$

It wouldn't be the candidate vehicle targets. Remove current targets. The threshold V_L and V_H are set according to the specific application scenarios.

B. Shape Criteria

It is generally believed that the shape of the target vehicle is rectangular. Therefore, the ratio of target length l_t^i and target width w_t^i can be used to filter targets:

$$r = l_t^i/w_t^i \tag{17}$$

When $r > 6$ or $r < 0.5$, it is not a candidate vehicle target.

C. Direction Criteria

In the lane where there is no crossing, the vehicles don't move back and forth, so we propose the direction criteria: the motion direction of the candidate vehicle is monotonous. We introduce cross products to determine the direction. The algorithm is illustrated as follows:

Algorithm The direction criteria

1: Take the target t of frame i as $P_t^i(x_t^i, y_t^i)$ ($i = 1,2,3 \ldots \ldots N-2$, N is the amount of frames);

2: Take the same moving target t of adjacent frames as $P_t^{i+1}(x_t^{i+1}, y_t^{i+1})$ and $P_t^{i+2}(x_t^{i+2}, y_t^{i+2})$ calculate the cross product: $cp_i = |\overrightarrow{P_t^i P_t^{i+1}} \times \overrightarrow{P_t^{i+1} P_t^{i+2}}|$ and record the result;

3: If cp_i is different from the sign of posterior cross product cp_{i+1}, the direction of movement are different. The target isn't considered as vehicle.

4: Repeat for the frames.

V. EXPERIMENT

We carried out several experiments to evaluate our vehicle detection system on aerial video and changedetection.net (CDnet) [21]. CDnet is one of the most widely used dataset for moving object detection. Experiment results are as follows.

A. UAV Platform

The UAV is designed and assembled independently by us. We use the DJI S900 UAV frame and 16000mAh, 6S LiPo battery. Flight control system consists of the STM32F427VI and TLC algorithm [16] [17] [18] [19]. The IMU unit uses the Xsens MTI-G-700. IMU and GPS collect data and calculate the necessary speed, position, angular velocity and attitude of UAV, so as to provide necessary data for flight control. The XBee module realizes real-time data exchange between ground station and UAV. The distance between UAV and moving target is obtained by using SF11 laser altimeter. The image acquisition device is a double axle pod, and the image processing algorithm is completed on the airborne PC104. The overall UAV system and parameters are shown in Figure 2 and Table 1.

Table 1. Hardware system parameters.

Frame Weight	Diagonal Wheelbase	Max Distance of Propeller
3300g	900mm	1200mm
Load	Height	Wind Resistance
2000g	480mm	F-5
Working Temperature		Max Speed
-10 °C ~ +40 °C		10m/s
Max Flight Time		
25min（16000mAh/6s/1900g）		

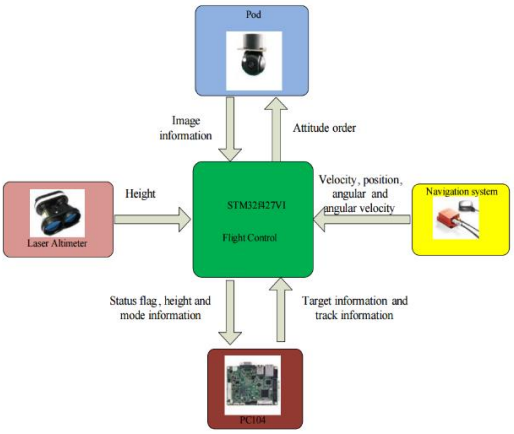


Figure 2. UAV physical map and UAV hardware diagram.

B. Lane Surface Extraction Experiment

As shown in Figure 1, the result of lane surface extraction based on bit plane is superior to the Ostu binarization result on the highway dataset (1700 frames). In order to verify the performance of the lane surface extraction method based on the bit plane in different environments, we test on traffic datasets (1570 frames) and fluidHighway dataset (1364 frames). Even if the scene of these datasets is night and the image quality is terrible, the bit plane can still extract the relatively complete lane surface.

C. Dynamic Detection Area Experiment

The dynamic detection area results are shown in Figure 3. Aerial video was taken in 5m, 10m, and 20m. The black line represents the outmost lanes. The blue rectangle is the detection area obtained from the outmost points, and the purple rectangle is the final detection area after Kalman filtering. It can be seen that the detection area can be effectively reduced.

D. Improved Vibe Algorithm Experiment

The results of the improved Vibe algorithm are shown in Table 2. With the change of background pixel update rate, background removal becomes faster and more effective. The improved algorithm can be applied to dynamic environment.

Figure 3: The black straight lines are detected by the probability Hough transform and selected as the outermost lanes. The blue box is the dynamic detection area fitted by the outermost lanes. The purple box is the filtered dynamic detection area obtained by Kalman filter.

Table 2. Frame number for updating with different φ value.

Update rate $1/\varphi$	1/16	1/8	1/4	1/2
Required frame amount for updating	100	70	60	50

E. Aerial Video Experiment

We compare our algorithm with Vibe algorithm, GMM algorithm and PBAS algorithm, because the Vibe is the basis of our algorithm, the GMM is frequently cited and utilized. The PBAS is nearly the best detection algorithms in recent years [20]. We add simple morphological operation to the original Vibe and GMM to eliminate noise. The experiment results are shown in Figure 4.

The images in Figure 4 are the Vibe results, GMM results, PBAS results and our system results. The average time cost per frame for the four algorithm is 23 milliseconds, 40 milliseconds, 42 milliseconds and 30 milliseconds. Our detection system is superior to other algorithms. After combining the dynamic detection area, the improved Vibe algorithm and the vehicle screening strategy, the vehicle detection system can detect the moving vehicles accurately in the dynamic large-scale environment.

F. CDnet Experiment

We test our method on the twoPositionPTZCam dataset, which belongs to the PTZ category in CDnet. As proposed in [21], detection algorithms have the lowest performance on the PTZ category due to the camera jitter, and the camera jitter is inevitable in aerial video. The main advantage of this comparison is that we can easily compare our method with many most advanced methods which have been ranked according to the following measures. The evaluation used by the CDnet is in Table 3 and the results are in Table 4.

Table 3. Evaluation.

TP (True Positive)	Pixel number labeled as foreground correctly.
FP (False Positive)	Pixel number labeled as foreground incorrectly.
FN (False Negative)	Pixel number labeled as background incorrectly.
TN (True Negative)	Pixel number labeled as background correctly.
Re (Recall)	**TP / (TP + FN)**
Sp (Specificity)	**TN / (TN + FP)**
FPR	**FP / (FP + TN)**
FNR	**FN / (TP + FN)**
Precision	**TP / (TP + FP)**
PWC	**100 * (FN + FP) / (TP + FN + FP + TN)**
F-Measure	**(2 * Precision * Re) / (Precision + Re)**
Average ranking (used to rank by CDnet)	**(Re + Sp + FPR + FNR +PWC + F-Measure + Precision) / 7**

VI. CONCLUSION

In this paper, an UAV airborne vehicle detection system is proposed. We use the bit plane to limit dynamic detection area, reduce the background complexity and improve detection efficiency. Then the Vibe algorithm is improved so that it can be applied to dynamic scenes. Finally, the foreground targets are filtered to get the moving vehicles. Experiments show that the vehicle detection system is accurate, robust and real-time.

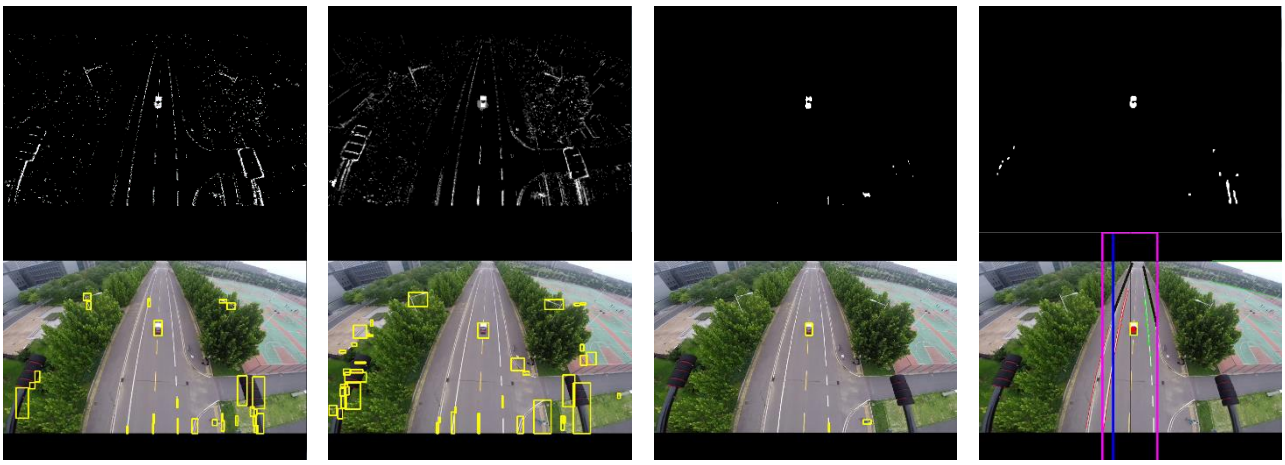


Figure 4. The images from left to right are the detection result of the Vibe algorithm, the GMM algorithm result, the PBAS algorithm result and our system result.

Table 3. Test results on the twoPositionPTZCam dataset

Method	Re	Sp	FPR	FNR	PWC	F-Measure	Average ranking
C-EFIC [22]	0.8686	0.8947	0.1053	0.1314	10.5973	0.6207	0.6144
Our method	0.6353	0.9929	0.0071	0.3647	1.1176	0.5666	0.5993
Multimode Background Subtraction [23]	0.5973	0.9963	0.0037	0.4027	0.5850	0.5520	0.5400
EFIC [24]	0.9177	0.9217	0.0783	0.0823	7.8707	0.5842	0.5282
Multimode Background Subtraction V0 (MBS V0) [25]	0.5770	0.9945	0.0055	0.4230	0.7821	0.5118	0.4988
PAWCS [26]	0.6976	0.9912	0.0088	0.3024	1.1162	0.4615	0.4725
IUTIS-5 [27]	0.6749	0.9902	0.0098	0.3251	1.2166	0.4282	0.3833
AAPSA [28]	0.5128	0.9638	0.0362	0.4872	3.9676	0.3302	0.3772

Therefore, the system has a broad application prospect. Because of the outdoor working environment, improving the robustness of lighting is a good direction for future work.

REFERENCES

[1] Hadi R A, Sulong G, George L E. Vehicle detection and tracking techniques: a concise review [J].Signal & Image Processing, 2014, 5(1):1-12.

[2] Siam, M, R. Elsayed, and M. Elhelw. On-board multiple target detection and tracking on camera-equipped aerial vehicles. IEEE International Conference on Robotics and Biomimetics, 2012:2399-2405.

[3] Chen, X, Xiang S, Liu C L, et al. Vehicle detection in satellite images by hybrid deep convolutional neural networks [J]. IEEE Geoscience and remote sensing letters, 2014, 11(10): 1797-1801.

[4] Ammour N, Alhichri H, Bazi Y, et al. Deep learning approach for car detection in UAV imagery [J]. Remote Sensing, 2017, 9(4): 312.

[5] Li Y, Chen W, Jiang R. The integration adjacent frame difference of improved ViBe for foreground object detection[C]//2011 7th International Conference on Wireless Communications, Networking and Mobile Computing. 2011.

[6] Hofmann M, Tiefenbacher P, Rigoll G. Background segmentation with feedback: The pixel-based adaptive segmenter[C]//IEEE Computer Vision and Pattern Recognition Workshops (CVPRW). 2012: 38-43.

[7] Mahoor E, Maghsoumi H, Asemani D. An improved motion detection algorithm using ViBe[C]//IEEE International Conference on Computing, Communication and Networking Technologies (ICCCNT). 2015: 1-5.

[8] Jin D, Zhu S, Sun X, et al. Fusing Canny operator with vibe algorithm for target detection[C]// IEEE Control and Decision Conference (CCDC), 2016 Chinese. 2016: 119-123.

[9] Xining Y, Jianmin D, Dezhi G, et al. Research on Lane Detection Based on Improved Hough Transform [J]. Computer Measurement & Control, 2010, 2: 019.

[10] McCall J C, Trivedi M M. Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation [J]. IEEE transactions on intelligent transportation systems, 2006, 7(1): 20-37.

[11] Mammeri A, Boukerche A, Tang Z. A real-time lane marking localization, tracking and communication system [J]. Computer Communications, 2016, 73: 132-143.

[12] Rathinam S, Kim Z W, Sengupta R. Vision-based monitoring of locally linear structures using an unmanned aerial vehicle[J]. Journal of Infrastructure Systems, 2008, 14(1): 52-63.

[13] Mammeri A, Boukerche A, Lu G. Lane detection and tracking system based on the MSER algorithm, hough transform and kalman filter[C]//Proceedings of the 17th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems. ACM, 2014: 259-266.

[14] Barnich O, Van Droogenbroeck M. ViBe: A universal background subtraction algorithm for video sequences [J]. IEEE Transactions on Image processing, 2011, 20(6): 1709-1724.

[15] Sun, S, Qin Y, Xianbing M A. ViBe foreground detection algorithm and its improvement with morphology post-processing for outdoor scene [J]. Computer Engineering & Applications, 2013, 49(10):159-162.

[16] Li Z, Meng C, Zhou F, et al. Fast Vision-based Autonomous Detection of Moving Cooperative Target for UAV Landing. Journal of Field Robotics, 2018, DOI: 10.1002/rob.21815.

[17] Ding X, Guo P, Xu K, et al. A Review of Aerial Manipulation of Small-scale Rotorcraft Unmanned Robotic Systems [J]. Chinese Journal of Aeronautics, 2018.

[18] Ding X, Yu Y. Motion planning and stabilization control of a multipropeller multifunction aerial robot [J]. IEEE/ASME Transactions on Mechatronics, 2013, 18(2): 645-656.

[19] Ding X, Yu Y, Zhu J J. Trajectory linearization tracking control for dynamics of a multi-propeller and multifunction aerial robot-MMAR[C]//IEEE International Conference on Robotics and Automation (ICRA). 2011: 757-762.

[20] Sobral A, Vacavant A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos [J]. Computer Vision and Image Understanding, 2014, 122: 4-21.

[21] Wang Y, Jodoin P M, Porikli F, et al. CDnet 2014: An expanded change detection benchmark dataset[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops. (CVPRW). 2014: 387-394.

[22] Wang Y, Luo Z, Jodoin P M. Interactive deep learning method for segmenting moving objects [J]. Pattern Recognition Letters, 2017, 96: 66-75.

[23] Sajid H, Cheung S C S. Universal multimode background subtraction [J]. IEEE Transactions on Image Processing. 2017, 26(7): 3249-3260.

[24] Allebosch G, Deboeverie F, Veelaert P, et al. EFIC: Edge based foreground background segmentation and interior classification for dynamic camera viewpoints[C]//International Conference on Advanced Concepts for Intelligent Vision Systems. Springer, Cham, 2015: 130-141.

[25] Sajid H, Cheung S C S. Background subtraction for static & moving camera[C]// IEEE International Conference on Image Processing (ICIP). 2015: 4530-4534.

[26] St-Charles P L, Bilodeau G A, Bergevin R. A self-adjusting approach to change detection based on background word consensus[C]// IEEE Winter Conference on Applications of Computer Vision (WACV). 2015: 990-997.

[27] Bianco S, Ciocca G, Schettini R. How far can you get by combining change detection algorithms? [C]//International Conference on Image Analysis and Processing. Springer, Cham, 2017: 96-107.

[28] Ramírez-Alonso G, Chacón-Murguía M I. Auto-adaptive parallel SOM architecture with a modular analysis for dynamic object segmentation in videos [J]. Neurocomputing, 2016, 175: 990-1000.