Vehicle detection base on Road Segmentation for surveillance satellites video

Yun Meng¹, Xiaoyong Wang¹,²*, Shuai Yuan¹

¹College of Technology and Communication Kaifeng Henan, China
²Henan Financial University, China
wangxiaoyong@hafu.edu.cn

Abstract. With the development of data acquisition ability, the LEO satellites can work with a surveillance camera on board by focusing a particular area for dozens of seconds, so it drives us to develop some applications based on this data. In this paper, a data preparation method named DERS (Density Estimation based Road Segmentation) is proposed to divide the road from background to reduce the worse impact from relative-motion and then improve the precession and recall for vehicles for surveillance satellite video sequence. DERS has three main steps: frame difference, morphology, pyramid iteration, segmentation. Our experiments with the satellite videos from SkySat-1, and compared between Gaussian Mixture Model (GMM) and GMM with Ders shows that the precision rate and recall has improved of the traditional method with Ders, the precision of GMM with Ders has been improved from 64.42% to 88.71, the recall of GMM with Ders is from 76.21% to 74.16%, and the F-Score has been improved from 69.82% to 80.79%.

1. Introduction
There are about 70%-80% of the information obtained by humans comes from vision, so the images play an indispensable role in information interaction between humans and computers [1]. High-quality and high-resolution images always have more information and details, and with more and more LEO high-resolution commercial staring satellites launched in the past decades, more and more researchers begin to focus on developing some application base on this data. However, equipment on bord cannot be as excellent as it on earth due to the limited space and energy supply of the satellite, so the image captured on the satellite always has a poor resolution and it led to the vehicle always with unobvious constant, what’s more, the relative motion from satellites posture adjustment in order to start at the region. So, the background always moving also, so the vehicle detection has much challenge [2-4]. Tao Yang [5] has proposed a method base on the road segmentation by GPS, and although it can reach a satisfied performance, it can stop work for a sequence which GPS information cannot be achieved. Xiaoying Jin [6] has proposed a method for vehicle detection also with road segmentation, and in this paper, the road segmentation with textures, characters training for the resolution is 0.6m-1.0m, so the precision cannot reach a good level because the satellite video cannot always accompany with very high-resolution.

So, due to the limitation on the satellite video and the disadvantages of previous researchers [3-7], in this paper, a new method DERS for road segmentation was proposed to reach a better precision level.
2. Methodology

DERS is a road segmentation method for vehicle detection for satellite surveillance video, which has three basic phases: Frame Difference, Target map building, morphology, pyramid. Frame Difference is a traditional method for moving target detection with good precision level, but it always with many noises spot.

Frame Difference

 Targets 

 Map 

 Morphology 

 Pyramid 

 Road Map 

 &

 Final Result 

 Targets

Figure 1 Ders

There are many road segmentation methods, but almost all of them depend on the other requirements. DERS is a method base on real video sequence but not GPS draw labels, texture training etc. in advance. What is more, some researchers aim at Deep-Learning to divide the road and background, but complex background always cannot provide enough features for their work.

Segmentation preparation is a data preparation for road segmentation. In order to statistic the density of the vehicle target, FD (Frame Difference) is used to detect the moving target in advance. A serial of images can be pack into a group that detects the vehicles in a group as the Figure 2 shown.

F1 F2 F3

D1 D2

T1 ...

Figure 2 Groups

Three frame difference method can be used to estimate the vehicles density. The \( F_n \) is the frames from the satellites video sequence, and \( D_1 \) is the absolute difference result from \( F_1 \) and \( F_2 \), and the \( T_1 \) is the overlap result from \( D_1 \) and \( D_2 \). So, the data preparation can be described as the Eq (1) shown.

\[
D_n = |F_n - F_{n-1}|
\]

\[
T_n = D_n \cap D_{n-1}
\]

(1)

(2)

where the \( F_n \) is the frame from video with index \( n \), and the \( D_n \) is the detection mask with frame difference method, \( T_n \) is the vehicle mask.

After the data preparation for the video sequence with frame difference, the layout of the road can be described easier by the vehicle’s density. Unfortunately, the road description is very blur, because buildings may be covered some noise from relative motion. So, it is necessary to process the image with morphological methodology to remove the small noise of the mask.

After that, in order to more easily to separate the road in frame, a queue was set to save the target at different time, as the Figure 3 shown, a few frames merged together with a to increase density.
It is fortunately that vehicle cluster is always smaller cluster than things on earth surface because the movements of car always independent while the ground-things always move together into a spots group because motion of static-things on earth is always from relative motion with little change. So, the vehicles and ground-things can be distinguished within a certain margin through a threshold value as Figure 4 shown. (a) is the real motion, and (b) is abstractly motion.

Fortunately, the vehicle always has less points (vehicle less than 3m can be saved, so it has potential that vehicle will be deleted which is more than 3m), a threshold value was set to filter clusters, and it can remove some ghost clusters as Figure 5 shown.
Pyramid is an image sequence, each image in which is composed of a low-pass filter and a second sampling sample of its precursor \[6\]. In this paper, in order to build road map from discrete cluster, the image is down-sampled in the first to remove the cluster in low density area. And then, the image can be up-sampling to connect the cluster to build the map as (b) in Figure 6 shown.

A sliding window with \(n \times m\) was set to build the road mask. The \(E\) is the mean value for the window, and the \(\delta\) is variance of the window.

\[
E = \frac{1}{n \times m} \sum_{i=0}^{n} \sum_{j=0}^{m} P(i,j)
\]

\[
\delta = \sum_{i=0}^{n} \sum_{j=0}^{m} (P(i,j) - E)^2
\]

\[
Cov = \alpha E + \beta \delta
\]

where the \(P(i,j)\) is the intensity of the point \((i,j)\), \(\alpha\) and \(\beta\) are tow adaptive parameters to modify the importance of \(E\) and \(\delta\). \(Cov\) is the confidence level of the window, and we can distinguish the weather it is a road region with the Eq (6):

\[
M = \begin{cases} 
1 & \text{Cov } \geq T \\
0 & \text{Cov } < T
\end{cases}
\]

where \(T\) is the threshold value of \(Cov\), and compared whether this points are a part of road region. \(M\) is the road region mask, and the region can be marked as road if the \(Cov\) of the window is larger than \(T\) as (c) in Figure 6 shown. And then, after some frames being learning, the road can be segmented with the road map as (d) in Figure 6 shown, and the road can be built more exactly with frames processed.

Therefore, only the pixel of targets from new frame in the road map can be regard as vehicle while other targets are regard as ghost vehicle to improve precession.

\[
T_r = T_n \cap T_m
\]

where \(T_n\) is the new targets in new frames, and \(T_m\) is road map learned from past frames, \(T_r\) is the target detected in this frame.

3. Experiments

In this section, an opened SkySat-1 video data was selected for evaluating Ders which was launched by Skybox on July 8th, 2014 at Baikonur, Kazakhstan [12,5], and the video has 30fps in full motion during 90 seconds. The view of SkySat-1 is 2km by 1km, and GSD is 1.1m. This video was captured by Skybox at Burj Khalifa on April 9th, 2014. Our experiments finished with OpenCV 4.5 by C++ 17 on a simple laptop with the guide of section 2 step by step, and the detection result can be shown in Figure 7.
At first, 10 frames are bunched together for road segmentation, and size of slide window was set as $8 \times 8, \alpha = 0.5, \beta = 0.5$, and the cluster composed of off-line points can be dilated with the template size of $3 \times 3$ to an integrated target, and the small noise (less than $1 \times 1$, it has potential to lose some targets) can be eroded with the template size of $2 \times 2$. Then, with 10 frames added together, after 420 frames, the road map can be shown in Figure 7.

A pyramid method is used to down-sampling some times and then up-sampling reverse, and the road can be segmented from the background with sliding box easier. The marked frame can be shown in (b) of Figure 8.

As Figure 9 shown, vehicles detected from relative-motion are distributed over the road, and comparison between GMM and GMM with Ders can be seen at Figure 10.
In order to quantify evaluation purpose, standard evaluation measurements are used to describe Ders: true positive (TP), false positive (FP), false negatives (FN), precision (Pre), recall (Rec), and F-score (F-S) are used in this paper $^{[1]}$, which can be found in Eq (8)-(10).

$$Pr_e = \frac{1}{n} \sum_{i=1}^{n} \frac{TP_i}{TP_i + FP_i}$$ (8)

$$Rec = \frac{1}{n} \sum_{i=1}^{n} \frac{TP_i}{TP_i + FN_i}$$ (9)

$$F - S = \frac{(1 + \alpha^2) \times Rec \times Pr_e}{\alpha^2 \times (Rec + Pr_e)}$$ (10)

where, $n$ is the size of group (and it was set 5), $\alpha$ is a const threshold value 1 in this paper. And, the detection result can be shown in Figure 10.

From Figure 11, the precision of GMM with Ders has been improved from 64.42% to 88.71, the recall of GMM with Ders is from 76.21% to 74.16%, and the F-Score has been improved from 69.82% to 80.79%. So, it is clearly that the F-S has improved, although Ders has potential to lose some target while removing noises.

4. Conclusion

In this paper, a road segmentation method has been proposed for MOD (Moving Target Detection). Firstly, frame difference method was used for road segmentation, and then filter the targets from road map from continually learning of the video sequence. From experiments between GMM no Ders and GMM with Ders shows that the F-Score has improved to a better level, it means Ders+GMM has combined score and better than GMM only. However, the number of TP has less than GMM. It means a few targets may be loosen while Ders is working, because there have many threshold values in Ders,
and this value let Ders cannot sperate the road and background reach the acme. So, Ders cannot support with targets on few and scattered road.

**Acknowledgements**

This paper is supported by Key Project in Science and Technology of Henan Province (NO.212102210411).

**References**

[1] Wang X, Li F, Ma J, et al. A New Parallel Scheduling Algorithm Based on MPI[C]. IEEE International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, 2018.

[2] Zhang J, Jia X, Hu J. Local region proposing for frame-based vehicle detection in satellite videos[J]. Remote Sensing, 2019, 11(20): 2372.

[3] Gill N K, Sharma A. Vehicle detection from satellite images in digital image processing[J]. International Journal of Computational Intelligence Research, 2017, 13(5): 697-705.

[4] Wang X, Li F, Xin L, et al. Moving targets detection for satellite-based surveillance video[C]//IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019: 5492-5495.

[5] Yang T, Wang X, Yao B, et al. Small Vehicle Detection in a Satellite Video of an Urban Area[J]. Sensors, 2016, 16(9).

[6] Jin X, Davis C H. Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks[J]. Image and Vision Computing, 2007, 25(9): 1422-1431.

[7] Karathanasopoulos A, Dunis C L, Khalil S, et al. Modelling, forecasting and trading with a new sliding window approach: the crack spread example[J]. Quantitative Finance, 2016, 16(12): 1875-1886.

[8] Zeng D, Zhao F, Ge S, et al. Fast cascade face detection with pyramid network[J]. Pattern Recognition Letters, 2019, 119: 180-186.

[9] Mansour A, Hassan A, Hussein W M, et al. Automated vehicle detection in satellite images using deep learning[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2019, 610(1): 012027.