A Video Surveillance Network for Airport Ground Moving Targets

Xiang Zhang[1(✉)] and Yi Qiao[2]

[1] University of Electronic Science and Technology of China, Chengdu, China
uestchero@uestc.edu.cn
[2] The Second Research Institute of the Civil Aviation Administration of China, Chengdu, China
qiaoyiqq@hotmail.com

Abstract. In this paper we describe an airport ground movement surveillance network. Airport ground videos are captured by multiple cameras, and than transmitted to the airport control center based on the optical fiber network. On the high-performance servers in the control center, various intelligent applications process video data, visualize the processing results and provide them to the air traffic controllers as a reference for airport management. Moving object detection is the foundation of many video based intelligent applications in airport surveillance. We propose detecting the moving objects in the airport ground by the use of the prior knowledge, that is, the airport ground made of cement has a gray-white color distribution. Based on this fact, firstly we use a dual-mode Gaussian distribution to fit the color distribution of the ground. Next, based on the fitted distribution we build a prior model, where pixels near the class boundary are more likely to be classified as the foreground. Finally, the prior model is used to detect moving targets within a Bayesian classification framework. Experiments are conducted on the AGVS benchmark and the results demonstrate the effectiveness of the proposed moving object detection algorithm.

Keywords: Airport ground surveillance · Moving object detection

1 Introduction

With the continuous growth of civil aviation traffic, the airport ground is getting busier and crowded. The airport must be more intelligent to face operation and capacity challenges. Compared with the common Secondary Radar or Automatic Dependent Surveillance-Broadcast (ADS-B) [1], video data contains more information, so video based intelligent surveillance is more promising in the next generation of airport management system. Since the airport ground is very broad, it needs a network of multiple cameras to cover the whole monitoring area. Furthermore, the coding and transmission of video data must be considered in such a video surveillance system. In the control center, various intelligent applications

S. W. Loke et al. (Eds.): MONAMI 2020, LNICST 338, pp. 229–237, 2020.
https://doi.org/10.1007/978-3-030-64002-6_15

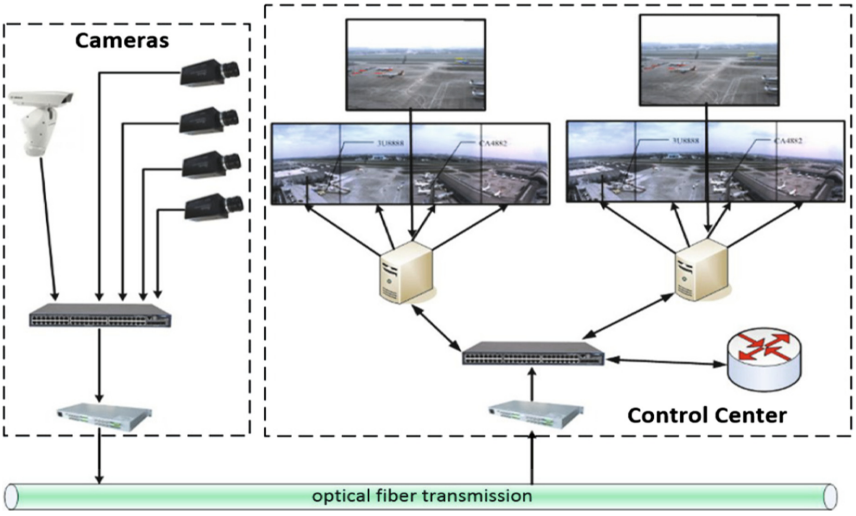

Fig. 1. Illustration of the airport ground video surveillance network.

can be developed based on video data, e.g. visual guidance and conflict alert. The foundation of most such applications is moving object detection [2], which aims to extract the moving targets on the ground from video sequences.

In this paper we present an airport ground movement surveillance network. Multiple cameras form a monitoring network covering the whole airport ground. Captured video data is transmitted to the airport control center based on a dedicated optical network. The video data format is H264. For various intelligent applications running in the control center, we propose a method of moving object detection which is specially used for ground surveillance. This method takes use of the prior knowledge in ground, that is, the airport ground made of cement has a gray-white color distribution. Based on this fact, firstly we use a dual-mode Gaussian distribution to fit the color distribution of the ground. Next, based on the fitted distribution we build a prior model, where pixels near the class boundary are more likely to be classified as the foreground. Finally, the prior model is used to detect moving targets within a Bayesian classification framework. Experiments are conducted on the AGVS benchmark [5] to evaluate the effectiveness of the proposed method.

The rest of the paper is organized as follows. The airport ground movement surveillance network is shortly described in Sect. 2. Section 3 presents details of the proposed moving object detection algorithm. Experimental results are shown in Sect. 4, followed by the conclusion in Sect. 5.

2 Proposed Surveillance Network

The airport ground movement surveillance network is shown in Fig. 1. Multiple cameras including both fixed and PTZ cameras, are installed around the airport.

Fig. 2. Spliced video in the airport ground video surveillance network.

Captured videos in the form of H264 are transmitted to the control center with a dedicated high-speed optical fiber network. In the control center, video data is processed and displayed on a large screen. Since the airport ground is very broad and the field of view of a single camera is limited, it is necessary to splice the videos based on the camera's geometry relationship. The spliced video is shown in Fig. 2. Based on the splice video, various intelligent applications can be developed. Next we introduce the proposed moving object detection algorithm, which is able to support high-precision upper applications.

3 Moving Object Detection

The most widely used strategy in moving object detection is unsupervised statistical modeling, which has been studied for more than twenty years. In recent years, the supervised moving object detection based on deep learning is rising, but this method has not met the requirements of real applications. Some statistical methods only model the background [3], while others build multi-layer models for the scene, such as bilayer modeling [4]. Our method belongs to the category of bilayer modeling. Since the prior knowledge of the airport ground is employed in our method, the proposed method is more accurate than the previous methods.

3.1 Overall Framework

First we give some notations. Let $\mathbf{Z}^t = \{\mathbf{z}_i^t\}$ be a frame at time t, each pixel $\mathbf{z}_i^t$ is a 3-tuple RGB vector, where i is the position index of pixels. For ease of notation, we may omit the superscript and subscript in case of no confusion. Let $\mathbf{L} = \{l_i\}$ be a binary silhouette map, where $l_i \in \{\mathcal{F}, \mathcal{B}\}$, with $\mathcal{F} = 1$ denoting foreground and $\mathcal{B} = 0$ referring to background. In the Bayesian classification framework, moving object detection is the Maximum Posterior Probability (MAP) estimation of $P(\mathbf{L}|\mathbf{Z})$

$$P(\mathbf{L}|\mathbf{Z}) \propto P(\mathbf{Z}|\mathbf{L})P(\mathbf{L}), \tag{1}$$

where $P(\mathbf{Z}|\mathbf{L})$ and $P(\mathbf{L})$ are the likelihood term and smoothness term, respectively. In our method, a prior model is presented and used in conjunction with

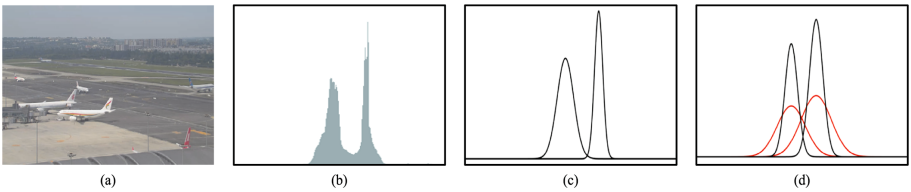

Fig. 3. (a) is one frame from sequence *S19* in the AGVS benchmark, and (b) is the gray histogram of (a). (c) is the fitted dual-model Gaussian distributions of (b). (d) shows the prior model with different σ.

the traditional bilayer model for classification. Therefore, the likelihood term in Eq. 1 consists of two terms

$$P(\mathbf{Z}|\mathbf{L}) = P_1(\mathbf{Z}|\mathbf{L})P_2(\mathbf{Z}|\mathbf{L}), \tag{2}$$

where $P_1(\mathbf{Z}|\mathbf{L})$ and $P_2(\mathbf{Z}|\mathbf{L})$ separately correspond to the prior model and the traditional bilayer mdoel. Next, we introduce each model in our method, the smoothness term and the optimization of Eq. 1.

3.2 Prior Model

The prior model is based on the prior knowledge in the ground, that is, miss-detections of moving object detection in ground surveillance are mainly caused by the camouflage phenomenon, which means the foreground and background share similar color distributions. Furthermore, the camouflage in ground surveillance occurs between the gray-white airport ground and airplanes dominated by white. In other words, the gray-white areas have a higher probability of camouflage occurrence. We compute the gray histogram of Fig. 3a, as shown in Fig. 3b. Figure 3a is from the AGVS benchmark, which is a special dataset for the research of airport ground surveillance. We can see that there are two significant peaks in the gray histogram. Because the airport ground is the main area in all videos of AGVS, we can infer that the two peaks in each histogram correspond to the gray and white areas of the airport ground, respectively. based on the least square method, we use a bimodal Gaussian distribution to fit the gray histogram

$$p_h(\mathbf{z}_i) = \sum_{r=1}^{2} \hat{\omega}_r \mathcal{N}(\mathbf{z}_i; \hat{\mu}_r, \hat{\sigma}_r), \tag{3}$$

where $\hat{\omega}_r$ is the weight of each Gaussian component, and $\hat{\mu}_r$ and $\hat{\sigma}_r$ are the mean and standard deviation of the ith Gaussian component.

We compute the dual-model Gaussian distribution of Fig. 3b, as shown in Fig. 3c. It can be seen that this bimodal model does reflect the probability of camouflage occurring in different gray pixels to a certain extent, because the largest camouflage probability is at the gray and white pixels of the ground, and the larger the gray difference from the ground pixel, the smaller the probability

of camouflage occurrence. Next, we define the prior model as the conditional probability of σ

$$p_h(\mathbf{z}_i|\sigma) = \sum_{r=1}^{2} \hat{\omega}_r \mathcal{N}(\mathbf{z}_i; \hat{\mu}_r, \sigma), \tag{4}$$

where $\hat{\omega}_r$ and $\hat{\mu}_r$ are the same as in Eq. 3, σ is a variable parameter, and the two Gaussian components have the same σ. The prior model with different σ is shown in Fig. 3d. We can see that when σ is small, the camouflage probability of different pixels varies greatly (black curve), while when σ is large, the camouflage probability difference of pixels becomes small (red curve). When σ is infinite, the camouflage probability of all pixels is equal, which means this prior model does not work.

We have known that in the case of camouflage, classification tends to the background. The purpose of designing the prior model is to force the classification to the moving object when camouflage occurs, thus compensating the algorithm for the background bias. This is achieved by $P_1(\mathbf{Z}|\mathbf{L})$ in Eq. 2. Let $\alpha = \max[p_h(\mathbf{z}_i)]$ and β be the scale factor, first we define two probability items based on the prior model, $\beta(1 + p_h(\mathbf{z}_i|\sigma)/\alpha)$ and $\beta(1 - p_h(\mathbf{z}_i|\sigma)/\alpha)$. The scale factor β determines the absolute difference between the two probability terms. Next we construct $P_1(\mathbf{Z}|\mathbf{L})$ with the two probability items as

$$P_1(\mathbf{Z}|\mathbf{L}) = \prod_i [\beta(1 + \frac{p_h(\mathbf{z}_i|\sigma)}{\alpha})]^{l_n} [\beta(1 - \frac{p_h(\mathbf{z}_i|\sigma)}{\alpha})]^{1-l_n}. \tag{5}$$

In the framework of MAP, the meaning of Eq. 5 is that, if $\beta(1 + p_h(\mathbf{z}_i|\sigma)/\alpha)$ is larger than $\beta(1 - p_h(\mathbf{z}_i|\sigma)/\alpha)$, $\mathbf{z}_i$ is more likely to be classified as foreground. It can be seen that Eq. 5 does achieve our goal that the greater the camouflage probability, the higher the probability of being classified as foreground.

3.3 Traditional Bilayer Model

The traditional bilayer model is used in conjunction with the prior model for classification. We construct the bilayer model with the same manner as in [6], where a non-parametric model is separately kept for the background and foreground. Assuming the foreground probability and background probability based on the bilayer model are $p_f(\mathbf{z}_i)$ and $p_b(\mathbf{z}_i)$, respectively, the $P_2(\mathbf{Z}|\mathbf{L})$ in Eq. 2 is constructed as

$$P_2(\mathbf{Z}|\mathbf{L}) = \prod_i p_f(\mathbf{z}_i)^{l_n} p_b(\mathbf{z}_i)^{1-l_n}. \tag{6}$$

3.4 Smoothness Term and Optimization

The smoothness term allows to enforce the spatial coherence of pixel labels into the classification process. We use the Potts smoothness term as in [7]. Now we have all the probability terms in Eq. 1, and the MAP estimation of $P(\mathbf{L}|\mathbf{Z})$ is the final result of moving object detection. We convert the MAP estimation to an energy minimization problem. There have been many graph cut methods for energy minimization, and we use the α-expansion method in [8].

Fig. 4. Left to right: two frames in *S1* and *S21*, ground truth, KDE [12], KNN [10], Bodids [6] and the proposed method, respectively.

Fig. 5. Left to right: two frames in *S12* and *S20*, ground truth, SOBS [15], PBAS [14], CodeBook [11] and the proposed method, respectively.

4 Experimental Results

Comparative experiments are shown in this section. Ten state-of-the-art unsupervised algorithms with available public codes are chosen for comparison: SuBSENSE [9], KNN [10], Bodids [6], CodeBook [11], KDE [12], ViBe [13], PBAS [14], SOBS [15], GMM [16], FGMM [17]. The AGVS dataset is used as the benchmark dataset. Next, we separately describe the experimental setting, qualitative analysis and quantitative analysis.

4.1 Experimental Setting

There are several parameters in the proposed method, e.g. $\hat{\omega}_1$, $\hat{\omega}_2$, $\hat{\mu}_1$, $\hat{\mu}_2$, α, β and σ. Among them, $\hat{\omega}_1$, $\hat{\omega}_2$, $\hat{\mu}_1$, $\hat{\mu}_2$ and α are learned from the gray histogram during the operation of the algorithm. Therefore, two parameter, β and σ, needs to be tuned in our method. An empirical way is adopted in our method to set the two parameters, that is, optional values of one parameter are tested while with the other parameter fixed, and we choose the one resulting in the best detection result by visual observation. With such a method, we set the two parameters as $\beta = 0.11$ and $\sigma = 3$.

4.2 Qualitative Analysis

The detection results of some algorithms are shown from in Fig. 4 and Fig. 5. Mild camouflage is shown in Fig. 4, where the white-dominated airplanes are

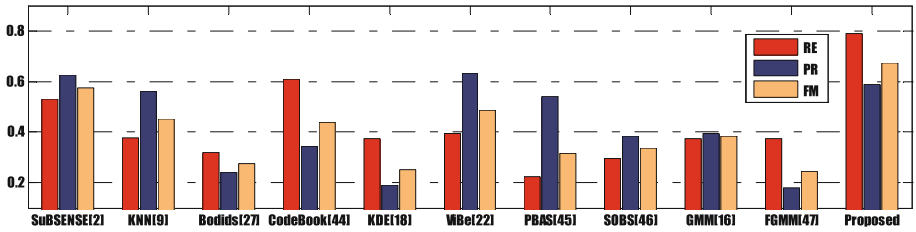

Fig. 6. Detection performance of 11 algorithms on AGVS.

moving on the gray area of the ground. We can see that the detection results by the proposed method are the most complete. The camouflage in Fig. 5 is more severe than in Fig. 4. The aircraft in Fig. 5 are all white, and the ground area they pass through is also bright white. Therefore, the detection defect of each comparison algorithm in Fig. 5 is very serious. The proposed method is effective for this severe camouflage, and its detection results are still the most complete. In fact, in addition to camouflage, Fig. 4 and Fig. 5 contain other challenges, e.g. shadow. It can be seen that the presented method is invalid for shadow.

4.3 Quantitative Analysis

We choose *RE*, *PR* and *F-measure* (*FM*) for quantitative comparison of detection accuracy:

$$RE = \frac{TP}{TP + FN}, \quad PR = \frac{TP}{TP + FP}$$

and

$$FM = \frac{2 \times RE \times PR}{RE + PR},$$

where *TP*, *FP* and *FN* are the numbers of true positives, false positives and false negatives, respectively. Higher *RE* means less detection defects, and higher *PR* indicates less detection noises. *FM* is the comprehensive result of *RE* and *PR*. When both *RE* and *PR* are close to 1, *FM* is also close to 1. When either *RE* or *PR* deteriorates, *FM* will also decrease.

Average *RE*, *PR* and *FM* of all methods on the AGVS dataset are shown in Fig. 6. PTZ videos are not considered, because unsupervised algorithms cannot deal with this problem. We can see that the proposed method has the best *RE* on all sequences, and compared with other algorithms, the *RE* is much improved. This means the defect detection due to camouflage is greatly improved by our method. Particularly, the presented method does not have the best *PR* on most sequences, which indicates the increase of false positives compared with other algorithms. As shown in Fig. 4, increased false positives are mainly the missclassified shadowed pixels. This problem may be resolved by using shadow removal as post-processing. Regarding with the computational time, our method can achieve the processing speed of 15 FPS on Intel i7 processor, which meets the requirements of practical applications.

5 Conclusion

An airport ground movement surveillance network was introduced in this paper. This system consisted of three parts, multiple cameras, optical fiber transmission and control center. There was multiple intelligent applications in the control center and moving object detection was the basis of most such applications. A new moving object detection algorithm was proposed by the use of the prior knowledge of airport ground. This method included a prior model and the traditional bilayer model, and the final classification was achieved via energy minimization. Experimental results on the AGVS benchmark demonstrated the effectiveness of the proposed method.

Acknowledgement. This work was supported by National Natural Science Foundation of China (U1733111, U19A2052), Key R&D projects in Sichuan Province (2020YFG0037), and Sichuan Science and Technology Achievement Transformation project (2020ZHCG0015).

References

1. Yang, H., Zhou, Q., Yao, M., Lu, R., Li, H., Zhang, X.: A practical and compatible cryptographic solution to ADS-B security. IEEE Internet Things J. **6**(2), 3322–3334 (2019)
2. Roy, S.M., Ghosh, A.: Foreground segmentation using adaptive 3 pahse background model. IEEE Trans. Intell. Transp. Syst. (2020, forthcoming)
3. Zhong, Z., Zhang, B., Lu, G., Zhao, Y., Xu, Y.: An adaptive background modeling method for foreground segmentation. IEEE Trans. Intell. Transp. Syst. **18**(5), 1109–1121 (2017)
4. Hao, J.Y., Li, C., Kim, Z., Xiong, Z.: Spatio-temporal traffic scene modeling for object motion detection. IEEE Trans. Intell. Transp. Syst. **14**(1), 295–302 (2013)
5. http://www.agvs-caac.com/
6. Sheikh, Y., Shah, M.: Bayesian modeling of dynamic scenes for object detection. IEEE Trans. Pattern Anal. Mach. Intell. **27**(11), 1778–1792 (2005)
7. Zabih, R., Kolmogorv, V.: Spatially coherent clustering using graph cuts. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2004)
8. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **23**(11), 1222–1239 (2001)
9. Charles, P.L.S., Bilodeau, G.A., Bergevin, R.: SuBSENSE: a universal change detection method with local adaptive sensitivity. IEEE Trans. Image Process. **24**(1), 359–373 (2015)
10. Zivkovic, Z., Heijden, F.V.D.: Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern Recogn. Lett. **27**(7), 773–780 (2006)
11. Kim, K., Chalidabhongse, T.H., Harwood, D., Davis, L.: Real-time foreground-background segmentation using codebook model. Real-Time Imaging **11**, 172–185 (2005)
12. Elgammal, A., Duraiswami, R., Harwood, D., Davis, L.S.: Background and foreground modeling using non-parametric kernel density estimation for visual surveillance. Proc. IEEE **90**(7), 1151–1163 (2002)

13. Barnich, O., Droogenbroeck, M.V.: Vibe: a universal background subtraction algorithm for video sequences. IEEE Trans. Image Process. **20**(6), 1709–1724 (2011)
14. Hofmann, M., Tiefenbacher, P., Rigoll, G.: Background segmentation with feedback: the pixel-based adaptive segmenter. In: Proceedings of IEEE Workshop on Change Detection (2012)
15. Maddalena, L., Petrosino, A.: The SOBS algorithm: what are the limits? In: Proceedings of IEEE Workshop on Change Detection (2012)
16. Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 747–757 (2000)
17. El Baf, F., Bouwmans, T., Vachon, B.: Type-2 fuzzy mixture of gaussians model: application to background modeling. In: Bebis, G., et al. (eds.) ISVC 2008. LNCS, vol. 5358, pp. 772–781. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-89639-5_74