Chinese Journal of Aeronautics, (2013),26(5): 1211–1217

Chinese Society of Aeronautics and Astronautics
& Beihang University

Chinese Journal of Aeronautics

cja@buaa.edu.cn
www.sciencedirect.com

Moving object detection in aerial video based on spatiotemporal saliency

Shen Hao [a], Li Shuxiao [a], Zhu Chengfei [a,*], Chang Hongxing [a], Zhang Jinglan [b]

[a] *Institute of Automaton, Chinese Academy of Sciences, Beijing 100190, China*
[b] *Queensland University of Technology, Brisbane, Australia*

Received 17 September 2012; revised 17 December 2012; accepted 21 March 2013
Available online 1 August 2013

KEYWORDS

Aerial video;
Computer vision;
Object detection;
Saliency;
Unmanned aerial vehicles

Abstract In this paper, the problem of moving object detection in aerial video is addressed. While motion cues have been extensively exploited in the literature, how to use spatial information is still an open problem. To deal with this issue, we propose a novel hierarchical moving target detection method based on spatiotemporal saliency. Temporal saliency is used to get a coarse segmentation, and spatial saliency is extracted to obtain the object's appearance details in candidate motion regions. Finally, by combining temporal and spatial saliency information, we can get refined detection results. Additionally, in order to give a full description of the object distribution, spatial saliency is detected in both pixel and region levels based on local contrast. Experiments conducted on the VIVID dataset show that the proposed method is efficient and accurate.

1. Introduction

With the development of technology, unmanned aerial vehicles (UAVs) have played a vital role in modern wars and industries. Moving object detection in aerial video as the foundation of higher targets, such as tracking and object recognition, is essential for UAV intelligence. In contrast to applications with fixed cameras, such as traffic monitoring and building surveillance, aerial surveillance has the advantages of higher mobility and larger surveillance scope. Meanwhile, more challenges are involved in aerial video, such as changing background and low resolution. Therefore, much attention has been paid to moving object detection in aerial video.

Generally object detection methods can be categorized in three approaches, namely temporal-based, spatial-based, and combined approach. For moving object detection from a video, motion cue is the most reliable information, so the proposed moving object detection methods are mainly based on temporal information, such as background subtraction[1,2] frame difference,[3,4] and optical flow.[5,6] Additionally, Cao[7] proposed to use the multi-motion layer analysis in moving object detection for airborne platform. Yu[8] used the long-term motion pattern in moving vehicle detection in aerial video. However, as the lack of spatial distribution, the results for the methods based on motion cues are usually undesirable. On the other hand, the spatial-based object detection method is principally used in the domain of object detection in static images. With the development of biological vision, many

* Corresponding author. Tel.: +86 10 62550985 21.
E-mail address: chengfei.zhu@ia.ac.cn (C. Zhu).

Peer review under responsibility of Editorial Committee of CJA.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.cja.2013.07.038

researchers have shifted their attentions to saliency detection, and plenty of saliency-based object detection methods have been designed. Initially, saliency detection is mainly based on low-level features, e.g., edges, colors, and textures. Recently, many new measures have been adopted in this literature, such as region contrast,[9] patch rarities,[10] and difference in frequency domain.[11,12] In addition, Wang[13] used visual saliency in aerial video summarization. Besides, in order to give a further description for moving objects, some researchers have also tried to combine temporal and spatial information in moving object detection.[14–17] Yin[14] used a 3D Markov random field (MRF) to predict each pixel's motion likelihood and the message was passed in a 6-connected spatiotemporal neighborhood. As every pixel needs to be predicted by MRF, the computational cost is huge. Liu[15] introduced saliency in moving object detection. They developed an efficient information theoretic-based procedure for constructing an information saliency map (ISM), which was calculated from spatiotemporal volumes.

Because aerial video has the property of changing background and small objects, moving object detection is still an open problem that needs to be addressed further. As the camera is moving, it is not easy to build a reliable background. In addition, the computing resource available on a UAV platform is often limited, so the optical flow is not a suitable choice. Thus, most of the object detection methods are based on frame difference. Although motion information is very important for moving object detection, there are still several drawbacks:

(1) The detected object may be larger than its real size.
(2) There may be holes in detection results.
(3) When an object is moving slowly, its motion is unreliable.

Besides, most of the saliency detection methods are based on static images, which focus on application of image classification or recognition, so they are not suitable for moving object detection.

For the combined methods, there are also some aspects that need to be modified for moving object detection in aerial video. Firstly, most of the existing methods[15–17] are mainly aimed at applications with fixed cameras, so they are not easy to be adopted in aerial video. Secondly, calculation of pixel saliency in a whole image is time-consuming. Finally, most of the integrated spatial information is only extracted in the pixel level, so higher-level object descriptions, such as region, are missed.

In short, there are clearly three major challenges for moving object detection in aerial video: changing background, small objects, and real-time processing demand. To tackle with these problems, we propose a novel spatiotemporal saliency detection method, inspired by biological vision. Temporal and spatial saliency is adopted in moving object detection as employed by previous researchers. However, instead of calculating spatial and temporal saliency separately[15,16] we developed a hierarchical detection method. Temporal saliency is used to get a coarse segmentation, and spatial saliency is adopted to get the object's appearance details in candidate motion regions. Finally, we get refined detection results by fusing temporal and spatial saliency information. Our contributions can be summarized as follows:

(1) A novel framework for moving object detection in aerial video that combines both temporal and spatial saliency.
(2) A hierarchical saliency detection manner that can greatly reduce time cost for spatial saliency calculation.
(3) A novel spatial saliency representation method, in which spatial saliency is extracted in both pixel and region levels to give a full description of the object distribution.

2. Proposed detection algorithm

In aerial video such as the ones shown in Fig. 1, objects are usually very small; they normally show more saliency in the local region than in global image. Thus we only explore spatial saliency in candidate local regions which are obtained through temporal saliency detection. The final detection results are achieved by fusing spatial and temporal saliency information. Fig. 2 shows the flow chart of the proposed moving object detection algorithm.

Fig. 1 Sample images. Upper: original images in the VIVID dataset. Middle and bottom: segmented objects in local regions.