A Dynamic Online Background Modeling Framework for Moving Object Detection from Airborne Videos

Xiaosong Lan, Shuxiao Li, Chengfei Zhu, Feimo Li and Hongxing Chang
Institute of Automation
Chinese Academy of Sciences
Beijing, China
Email:xiaosonglan@gmail.com

Abstract—Current researches on moving object detection from airborne videos are mainly based on frame difference. Though many improvements have been made on these methods, it is still difficult to extract all the moving pixels accurately. Being capable of providing more reliable motion information, background subtraction based methods have been widely used for analyzing surveillance videos captured by fixed cameras. In this paper, we design a dynamic online background modeling framework to facilitate the adoption of the available background subtraction algorithms for moving object detection from airborne videos. It can avoid accumulated stabilization errors and handle the pixels near the frame boundary well. The advantage of our framework lies in the stabilization strategies we proposed and the background model size we employed. Experimental results and analysis on the airborne videos have validated the effectiveness of the proposed framework.

1. Introduction

Detecting moving object fast and accurately from airborne videos is one challenging task in computer vision. Unlike applications with fixed cameras, we need to handle more tough problems for the aerial surveillance, such as camera motion and poor video quality caused by atmospheric turbulence. To solve these problems, much work has been done for moving object detection from airborne videos. These methods can be mainly classified into three categories.

The first category of methods are based on dense optical flow analysis [1]. Since computing dense optical flow is time-consuming, the optical flow based methods are hard to be used for real-time applications without hardware acceleration.

The second category of methods are based on frame difference, which are relatively efficient. The main drawback is their poor ability at extracting all the relevant moving pixels, leaving holes in the foreground objects. Another drawback is that they usually detect a counterpart (often called ghost) from the background regions. Though much work has been done to deal with the holes and ghost [2], [3], [4], it is still difficult to overcome their inherent shortcomings.

The third category of methods are based on background subtraction. On account of their good performance, they have been successfully applied to moving object detection from surveillance videos captured by fixed cameras. The classical methods include Gaussian Mixture Model (GMM) [5], Kernel Density Estimation (KDE) [6], Codebook [7], Visual Background Extractor (VIBE) [8] and so on. However, for aerial surveillance with moving camera setups, these methods cannot be directly utilized since they are not designed to handle the videos with moving background. There is few work which is designed based on background subtraction for motion detection from airborne videos. A. Colombari et al. [9] used a panoramic background model which was stitched and built based on the whole video. But it is not suitable for online applications. C. Yuan et al. [10] and S. W. Kim et al. [11] built the background models with the model size being the same as the image size, which can not handle the regions near the image boundaries very well. Besides, C. Yuan et al. stabilized the neighboring 90 frames to the current frame, which suffers from accumulated stabilization errors and system delay.

In this paper, we intend to solve the problem on how to make use of background subtraction methods to detect moving object from airborne videos. We proposed a dynamic online background modeling (DOBM) framework, which can facilitate the utilization of the available background subtraction algorithms. It can avoid accumulated stabilization errors and handle well the pixels near the frame boundary. To validate the effectiveness of the proposed framework, we have implemented GMM and VIBE in it and made a comparison with one state-of-the-art frame difference based method [4] for moving object detection from airborne videos. Experimental results on airborne sequences from VIVID dataset [12] show that the proposed framework achieves an excellent performance.

2. Dynamic online background modeling framework

2.1. Overview of the DOBM framework

Figure 1 shows the flowchart of the proposed framework. As shown in this figure, it includes two main modules: (1) stabilization module and (2) background subtraction module.
At first, we get the background image from the background model dynamically and conduct image registration with the current frame. Then, using the stabilization parameters obtained from the above registration, we warp the background model to the coordinate of the current frame. Having the aligned background model and the current frame, we can employ the available background subtraction algorithms for moving object detection from airborne videos.

The uniqueness of the DOBM framework lies in the following three aspects.

1. We dynamically stabilize the background model and update it with the current frame. It do not need to stabilize all frames of the video [9] or neighboring frames [10] to a same coordinate to build the background model. Instead, it only needs the current frame to update the background model. Since we estimate the parameters for the stabilization just from the current frame $c$ and the background image, instead of basing on frames in a cascade way, we can avoid the accumulation of stabilization errors. Note that the background image in frame $c$ is obtained from the background model in frame $c$ dynamically.

2. We build the background model with its size being larger than the size of original image to better handle the pixels near the frame boundary.

3. Since the background image is usually blur and does not contain moving targets, the matching strategy designed for the general video frames is not suited for such situation. We investigated different methods and proposed to utilize the tracking based matching strategy.

### 2.2. Implementation details

#### 2.2.1. Obtainment of the background image.

To align the background model to the coordinate of the current frame, we first calculate the background image from the background model.

In this paper, for GMM [5], the value of the background image located at $(i, j)$ is obtained by:

$$I_{i,j}^{bk} = \sum_{k=0}^{n} w_{i,j}^{k} \times u_{i,j}^{k}$$

where $n$ is the number of Gaussian models, $w_{i,j}^{k}$ and $u_{i,j}^{k}$ are the weight and mean value of $k^{th}$ Gaussian model.

For VIBE [8], we calculate the mean value of its background samples as the background image.

#### 2.2.2. Estimation of the parameters for stabilization.

After we obtain the background image, we need to find matched pairs between the current frame and the background image for the estimation of the stabilization parameters. The most commonly used method to find matched pairs between images includes three steps. (1) Extract corners, e.g. using HARRIS [13], FAST [14], from each image respectively. (2) Describe the corners with descriptors, such as SURF [15], BRIEF [16]. (3) Compare the descriptors of the two images to find the corresponding matched corners. In this paper, considering the real time requirement of airborne applications, we select FAST for corner extraction, which can extract high quality corners with a fast speed. Meanwhile we have tested both SURF and BRIEF descriptors.

However, when FAST+SURF and FAST+BRIEF are used to find the matched pairs between the background image and current image, they often fail and do not work very well. It is because the corners extracted from the current frame and the background image have different characteristics. Usually, the background image is blur and does not contain moving targets. In our experiments, we find the corners extracted from the current frame can be tracked by KLT [17] to the background image to obtain the matched pairs, which can provide more reliable and accurate matched pairs. Thus, we use KLT to obtain the matched pairs in the DOBM framework. If we fail to get enough matched pairs for the estimation of the parameters for stabilization, e.g. when the scene changes suddenly, we will reinitialize the background model.

When we get the matched pairs, RANSAC is used to estimate the parameters for stabilization.

#### 2.2.3. Background modeling.

Based on the obtained stabilization parameters, we wrap the background model to the coordinate of the current frame. Then a background subtraction method can be used to detect the moving objects. In this paper, we have implemented the most popular background subtraction methods.
subtraction algorithms GMM and VIBE in the proposed framework respectively to validate its effectiveness.

GMM models the background for each pixel by weighted mixture of Gaussians [5]. For each pixel in the current frame, if it can find a match with its corresponding Gaussian models, it will be classified as background. Otherwise, it will be classified as foreground. In this paper, the number of Gaussian models is set as 3.

VIBE models the background for each pixel by storing a set of background samples (set as 20), which is randomly updated by an elaborately designed strategy [8]. It compares a pixel to the background samples to determine whether the pixel belongs to the background or not. If the number of background samples which are similar to the pixel is more than a threshold (set as 2), this pixel is regarded as background.

Using a background subtraction algorithm, the foreground and background of the current frame can be separated. By conducting the morphological operations, connected component analysis and area thresholding on the foreground part, the detection results indicating the moving objects can be obtained. The background part of the current frame is utilized to update the background model, which will generate the new background image to get the stabilization parameters for next time stabilization iteratively.

3. Experimental results

To validate the effectiveness of the proposed DOBM framework and our designed strategies in it, experiments have been made on the public VIVID dataset [12]. We manually annotated all the moving objects of four sequences frame by frame, which are referred to as Seq1, Seq2, Seq3 and Seq4, as the ground truth. Typical frames of these four sequences are shown in Figure 2.

We have implemented GMM, VIBE (GRAY) and VIBE (RGB) in the DOBM framework for moving object detection from airborne videos. We denote these schemes as DOBM-GMM, DOBM-VIBE1 and DOBM-VIBE2 respectively.

Overlap ratio [18] is adopted for determining whether we have a successful detection or not. Usually, when the overlap ratio between the ground truth and the detected object is larger than 0.5, it is regarded as a successful detection.

In addition, PCC [8], [19] metric, which combines the true positive, false positive, true negative and false negative together, is also employed to evaluate the performance. Higher PCC value means better performance.

3.1. Analysis of the proposed strategies in the DOBM framework

To validate the effectiveness of our design in the DOBM framework, we analysed the performance of each sub-design. Three comparisons were made. (1) Stabilization between frames (Ref-1) versus stabilization between current frame and background image (Proposed). (2) The size of background model being the same as the size of the original image (Ref-2) versus $\alpha(\alpha = 1.2)$ times of the original image (Proposed). (3) Find matched pairs for stabilization by two most commonly used descriptors: SURF and BRIEF (Ref-3) versus by KLT track based strategy (Proposed).

Figure 3 shows the motion images obtained from the stabilization between frames and that from the stabilization between the current frame and background image (Proposed) at the $200^{th}$ frame on the four airborne sequences. We can see that by conducting stabilizing between the current frame and the background image got dynamically from the background model instead of between frames, we can reduce the accumulated stabilization errors.

Figure 4 illustrates why utilizing a larger model as we proposed can handle the pixels near the frame boundary better. Since pixels come in and out of the camera view could be recorded by the larger background model, it can handle the pixels in the region marked with yellow color (Figure 4 (b)) while the original size model can not (Figure 4 (a)).

Figure 5 demonstrates why we choose KLT for finding the matched pairs between current frame and background image. When we used the corner matching methods mentioned in section 2.2.2, the corners extracted in the background image and current frame have different characteristic. Thus, it is difficult to find enough well matched pairs,
as shown in Figure 5 (a). By using KLT to track the corners from the current frame to the background image, more correctly matched pairs can be obtained as shown in Figure 5 (b). We count the total mismatch times (the number of matched pairs used for RANSAC is less than 30) for these methods on all airborne sequences and show the results in Table 1, which also indicate the superiority of KLT.

![Figure 5](image)

**Figure 5.** All corners (marked as blue points) and correctly matched pairs (connected by green lines). (a) shows the corners extracted from the current frame (left) and the background image (right) while (b) show the corners extracted from the current frame (left) and the tracked corners on the background image (right).

We also calculate the average PCCs on all airborne sequences for comparison. The results were shown in Table 2. We can observe that all these strategies contribute to the good performance of the DOBM framework, being consistent with our analyses above. Note that, the larger size model do not work with GMM, since the background model built with GMM has poor quality which brings more negative effects.

### Table 1. Total Mismatch Times of SURF, BRIEF and KLT on All Airborne Sequences.

<table>
<thead>
<tr>
<th>Mismatch Times</th>
<th>VIBE1</th>
<th>VIBE2</th>
<th>GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SURF</td>
<td>93</td>
<td>79</td>
<td>91</td>
</tr>
<tr>
<td>BRIEF</td>
<td>6</td>
<td>2</td>
<td>65</td>
</tr>
<tr>
<td>KLT</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

### Table 2. Performance Comparisons about the Proposed DOBM Framework. REF-X indicates we replace our corresponding sub-design in DOBM by REF-X.

<table>
<thead>
<tr>
<th>PCC (%)</th>
<th>VIBE1</th>
<th>VIBE2</th>
<th>GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref-1</td>
<td>44.75</td>
<td>44.71</td>
<td>40.05</td>
</tr>
<tr>
<td>Ref-2</td>
<td>76.54</td>
<td>77.48</td>
<td>60.78</td>
</tr>
<tr>
<td>Ref-3 (SURF)</td>
<td>71.10</td>
<td>71.19</td>
<td>51.11</td>
</tr>
<tr>
<td>Proposed</td>
<td>76.92</td>
<td>77.49</td>
<td>59.46</td>
</tr>
</tbody>
</table>

### 3.2. Performance of the proposed DOBM framework

Table 3 shows the PCCs of DOBM-VIBE1, DOBM-VIBE2, DOBM-GMM and the method in [4] on four sequences. From the table we can find DOBM-VIBE1, DOBM-VIBE2 and DOBM-GMM all have better performance than the frame difference based method [4], even though the method in [4] has combined spatial saliency and tracking information together. This validates the effectiveness of the DOBM framework.

![Figure 6](image)

**Figure 6.** The PCCs of DOBM-VIBE1, DOBM-VIBE2, DOBM-GMM and the method in [4] on four airborne sequences.

**Table 3. Performance of the Proposed DOBM Framework.**

<table>
<thead>
<tr>
<th>PCC (%)</th>
<th>Method in [4]</th>
<th>DOBM-VIBE1</th>
<th>DOBM-VIBE2</th>
<th>DOBM-GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq1</td>
<td>66.69</td>
<td>97.09</td>
<td>97.35</td>
<td>70.75</td>
</tr>
<tr>
<td>Seq2</td>
<td>59.46</td>
<td>66.57</td>
<td>65.71</td>
<td>55.38</td>
</tr>
<tr>
<td>Seq3</td>
<td>51.87</td>
<td>77.14</td>
<td>76.32</td>
<td>61.43</td>
</tr>
<tr>
<td>Seq4</td>
<td>30.46</td>
<td>66.26</td>
<td>70.56</td>
<td>50.31</td>
</tr>
<tr>
<td>Average</td>
<td>52.12</td>
<td>76.92</td>
<td>77.49</td>
<td>59.46</td>
</tr>
</tbody>
</table>
H. Yalcin, M. Hebert, R. Collins, and M. Black, “A flow-based framework. A lot of background subtraction algorithms originally designed for applications with fixed camera setups can be easily applied for applications with moving camera setups. Thanks to the stabilization strategies and larger background model size proposed in the DOBM framework, it can avoid the accumulation of the stabilization errors and handle pixels near the image boundaries well, resulting in high performance. Experimental results on different airborne sequences show that the proposed DOBM framework has a decent performance. This is also indicated in Table 3.

4. Conclusions

In this paper, we proposed a general dynamic online background modeling framework (DOBM) for moving object detection from airborne videos. Based on the DOBM framework, a lot of background subtraction algorithms originally designed for applications with fixed camera setups can be easily applied for applications with moving camera setups. Thanks to the stabilization strategies and larger background model size proposed in the DOBM framework, it can avoid the accumulation of the stabilization errors and handle pixels near the image boundaries well, resulting in high performance. Experimental results on different airborne sequences show that the proposed DOBM framework achieves excellent performance.

In the future, we will explore more accurate and robust stabilization algorithms to further improve the DOBM framework.

Acknowledgments

This work is supported by National Natural Science Foundation of China (61175032 and 61302154).

References


