ROBUST AND FAST MOVING OBJECT DETECTION IN A NON-STATIONARY CAMERA VIA FOREGROUND PROBABILITY BASED SAMPLING

Kimin Yun and Jin Young Choi

Perception and Intelligence Lab
Department of Electrical and Computer Engineering, ASRI
Seoul National University, South Korea
{ykmwww, jychoi}@snu.ac.kr

ABSTRACT

This paper proposes a robust and fast scheme to detect moving objects in a non-stationary camera. The state-of-the art methods still do not give a satisfactory performance due to drastic frame changes in a non-stationary camera. To improve the robustness in performance, we additionally use the spatio-temporal properties of moving objects. We build the foreground probability map which reflects the spatio-temporal properties, then we selectively apply the detection procedure and update the background model only to the selected pixels using the foreground probability. The foreground probability is also used to refine the initial detection results to obtain a clear foreground region. We compare our scheme quantitatively and qualitatively to the state-of-the-art methods in the detection quality and speed. The experimental results show that our scheme outperforms all other compared methods.

Index Terms— Foreground probability based sampling, moving object detection, foreground, background subtraction, non-stationary camera.

1. INTRODUCTION

Finding moving objects in the scene is a fundamental problem in the research of computer vision. In this problem, it is important to achieve a computational efficiency as well as detection accuracy because the moving object detection is usually used as for a baseline function for the succeeding high level processing such as behavior analysis or event analysis [1]. Background subtraction algorithms have been proposed and shown good performances in fixed cameras [2, 3, 4, 5, 6]. However, in non-stationary cameras such as mobile or unmanned aerial vehicle (UAV) cameras, the existing methods do not work well because background is also changed by the camera movement. According to the literatures in a non-stationary camera, the existing works are categorized into three approaches: mosaic-based approach, segmentation-based approach, and compensation-based approach.

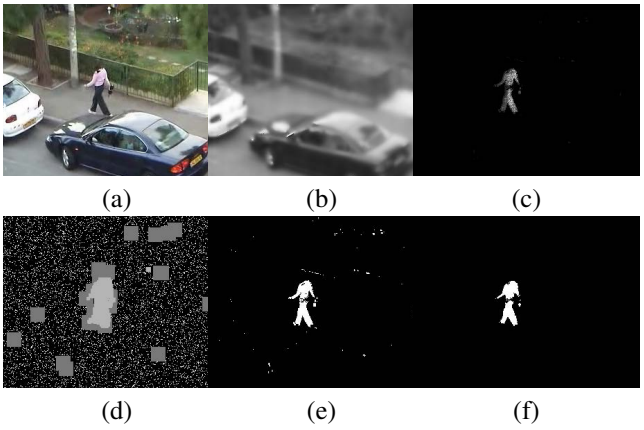


Fig. 1. Example images in the proposed procedure for moving object detection. The background model (b) is selectively updated by using the sampling map (d) which is determined by considering the foreground probability map (c). The foreground probability map is estimated from the previous detection results. The current initial foreground (e) is obtained by using the previous background model and sampling map. The final foreground (f) is fine-tuned by the foreground probability map.

The mosaic-based approach includes a procedure to make a panorama image using image registration, and then a background subtraction algorithm is applied [7, 8, 9, 10, 11, 12, 13]. In this case, the background subtraction algorithm can be directly used without modification, but image stitching errors in panorama images lead to many false detections. The second approach uses a segmentation method to separate the object motion from the camera motion. The methods in [14, 15] use the graph-cut optimization and the method in [16] uses nonparametric belief propagation(BP) to solve the Markov random field problem. Although these methods generate accurate results, they are too much slow and sensitive to parameters due to complexity of the model.

This work was partly supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No. B0101-15-0552, Development of Predictive Visual Intelligence Technology) and the Brain Korea 21 Plus Project.

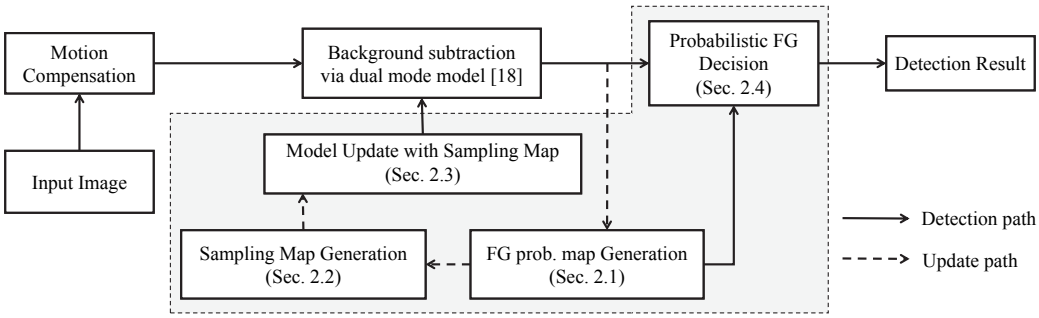


Fig. 2. Overall scheme of the proposed method, where the shaded parts are newly added.

Lastly, the compensation-based approach ccompensates the camera movement compensation to fit the previous model to the current image. Because an accurate estimation of camera motion is intractable and time-consuming, most of compensation-based algorithms use a simple camera model like the affine or projective camera model. As a result, many false detections occur at image boundary due to inaccurate estimation of camera movement. To reduce false detections with low computation, Kim *et al.* [17] proposed a spatio-temporal background modeling and Yi *et al.* [18] proposed grid-based modeling. In [19], they used feature clustering instead of pixel. While these works [17, 18, 19] reduced many false detections and achieved real-time performances, they also lose a true object region as a side effect and still show unsatisfactory performance in illumination changing environments.

In this paper, we propose a new scheme to improve the robustness of the state-of-the-art compensation-based method [18], reducing the loss of true object region and the false detections in illumination changes as well as maintaining real-time performance. Our main idea is to use the spatio-temporal properties of moving objects. The proposed scheme is realized by a novel sampling strategy based on the foreground probability using the spatio-temporal properties. From the assumption that the objects move smoothly in consecutive frames, we predict the next positions of objects. To keep the computational efficiency in the prediction, we just use the foreground probability that the objects are likely to appear at the spatial and temporal neighbors instead of accurate velocity estimation. Through this foreground probability, we can distinguish actual objects and false detections as well as reduce the search space to find the actual positions of objects.

Based on the concept of selective attention [20] for background subtraction in a stationary camera, we learn the spatio-temporal properties of objects. From the assumption that the objects appear at the neighbors of the previous detections, we build the foreground probability map(Fig. 1(e)). Then, we restrict the search space using the sampling map(Fig. 1(c)) obtained from the foreground probability, and detect the moving objects(Fig. 1(d)). Lastly, we refine the object region using foreground probability(Fig. 1(f)), and update the background model and the next foreground probability. In the experiment, we present the comparisons of our method to the state-of-the-art works in both detection quality and computational loads.

2. PROPOSED METHOD

Our approach is based on dual model background subtraction method (MCDin5.8ms) [18], an efficient method that accounts the imperfect estimation of camera movements. Fig. 2 depicts the overall scheme of the proposed method. The motion compensation and dual model background subtraction are adopted from the baseline [18]. Unlike the baseline, the background model is selectively updated by using the sampling map (details are described in Sec. 2.3). The sampling map is determined by considering the foreground probability map (in Sec. 2.2). The foreground probability map is estimated from the previous detection results (in Sec. 2.1). The current initial foreground is obtained by using the previous background model and sampling map. The final foreground is fine-tuned by using the foreground probability map (in Sec. 2.4).

2.1. Foreground Probability Map

To build a foreground probability map, our assumption is that objects movements are smooth spatially and temporally. Likewise, Chang *et al.* [20] define three properties of foreground pixels: temporal, spatial, and frequency properties. Frequency property is used to remove the inconsistent pixels which are changing periodically. In case of a non-stationary camera, however, it is hard to use this property because false detections are also consistent like true detections. In this paper, we adopt temporal and spatial properties among three properties to express our assumption of moving objects.

Temporal property M_T is defined as a recent history of the foreground at each pixel position as

$$M_T^t(n) = (1 - \alpha_T)M_T^{t-1}(n) + \alpha_T D^t(n), \qquad (1)$$

where t is time index and α_T is temporal learning rate. $D^t(n)$ is binary detection map which means that $D^t(n) = 1$ if pixel n belongs to foreground and $D^t(n) = 0$ if pixel n belongs to background at time t.

Spatial property measure the coherency of nearby pixels of foreground as

$$M_S^t(n) = (1-\alpha_S)M_S^{t-1}(n) + \alpha_S \frac{1}{w^2} \sum_{i \in N(n)} D^t(i), \quad (2)$$

where α_S is spatial learning rate, $N(n)$ denotes a spatial neighborhood around pixel n, and w^2 is the area of neighborhood. Then, the foreground probability $P_{FG}^t(n)$ is defined as multiplication of temporal and spatial properties, *i.e.*,

$$P_{FG}^t(n) = M_T^t(n) \times M_S^t(n). \quad (3)$$

2.2. Sampling Map Generation

Because we learn the temporal and spatial properties of foreground, the additional computational loads are inevitable. To keep the efficiency even in the additional loads, we try to restrict the search space based on the foreground probability without loss of detection performance. According to the attentional sampling [20], we extract the candidate pixel positions to run the background subtraction and model update. If a sampled position has a high foreground probability, we also extract the neighbor pixels where the neighborhood area is proportional to the foreground probability. Also, we can extract the positions randomly as 5% of entire pixels to detect the newly appeared objects. See [20] for details.

2.3. Model Update with Sampling Map

We adopt the dual model background subtraction method [18] as a baseline and modify the updating part by utilizing the sampling map. Yi *et al.* [18] built a grid unit model(*i.e.*, 4×4 region is modeled by dual models), which reduces the false detections because spatially adjacent pixels share the mean and variance. The mean $\mu^t(i)$ and variance $\sigma^t(i)$ of a grid i at time t are updated by the weight sum of previous model $\{\mu^{t-1}(i), \sigma^{t-1}(i)\}$ and current observation $\{m^t(i), v^t(i)\}$ as

$$\mu^t(i) = (1-\alpha_A^{t-1})\mu^{t-1}(i) + \alpha_A^{t-1} m^t(i), \quad (4)$$

$$\sigma^t(i) = (1-\alpha_A^{t-1})\sigma^{t-1}(i) + \alpha_A^{t-1} v^t(i), \quad (5)$$

where α_A^{t-1} is time-varying learning rate at time $t-1$.

In our scheme, background subtraction is applied to only a small portion selected by the sampling map. In addition, we modify the updating rules considering the selected pixels. When a grid contains selected pixels, the mean and variance observation of the model on the corresponding to the grid, $m^t(i)$ and $v^t(i)$ are calculated as

$$m^t(i) = \frac{1}{|\mathbf{G_s}(i)|} \sum_{j \in \mathbf{G_s}(i)} I^t(j), \quad (6)$$

$$v^t(i) = \max_{j \in \mathbf{G_s}(i)} (\mu^t(i) - I^t(j))^2 \quad (7)$$

where i, j, I^t denote grid index, pixel index, and intensity map of image at time t respectively, whereas $\mathbf{G_s}(i)$ denotes the group of selected pixels in the i-th grid. In other words, we calculate the mean and variance observations by using only the selected pixels in a grid.

On the other hands, when a grid does not contain any selected pixels, we keep the mean unchanged and initialize the variance to a high value. If the camera is static, we can just keep the previous model, but, in case of non-stationary camera, we get many false detections when the previous models are kept. Because pixel intensity changes drastically in a non-stationary camera due to rapid illumination change, we initialize the variance to a high value for a fast model adaptation.

Table 1. The average computational loads of each algorithm

Methods	Time per frame	frame/sec.
Generalized BP [16]	35.3s	0.028 fps
ViBe [6] w. motion comp.	11.23ms	89.05fps
MCD NP [17]	16.08ms	62 fps
MCD in 5.8ms [18]	5.74ms	174 fps
Proposed	4.80ms	208 fps

2.4. Probabilistic Foreground Decision

When the foreground decision relies on only the background, many false detections occur due to illumination change and inaccurate estimation of camera movement as shown in Fig. 1(e). However, we can refine the foreground using foreground probability in Sec. 2.1. First, we multiply the foreground probability map to the initial foreground obtained by the background subtraction. We can determine the detection map by a simple thresholding method to the multiplied map. However, in this case, foreground regions include inner holes and noisy detection regions. To cope with this problem, we use the watershed algorithm [21] which effectively segments the foreground regions. We cut the foreground probability map to a high threshold, and then apply the watershed algorithm with the seed points remaining after thresholding. This refinement reduces false detections and fills the foreground clearly with low computation.

3. EXPERIMENTS.

We compared our method to the state-of-the-art methods: segmentation-based method [16] and compensation-based methods [6, 17, 18]. For [6], we added the motion compensation for non-stationary camera as shown in the authors websites.[1]

Fig. 3 and Fig. 4 show quantitative results and the qualitative results of the compared methods. As shown in Fig. 3, our method shows the best performances except the case of *Cycle* sequence. Because the *Cycle* sequence has complex camera motion, compensation-based methods including ours might yield false detections. General BS [16], as a segmentation-based method, does not assume the specific camera model,

[1] http://www2.ulg.ac.be/telecom/research/vibe/

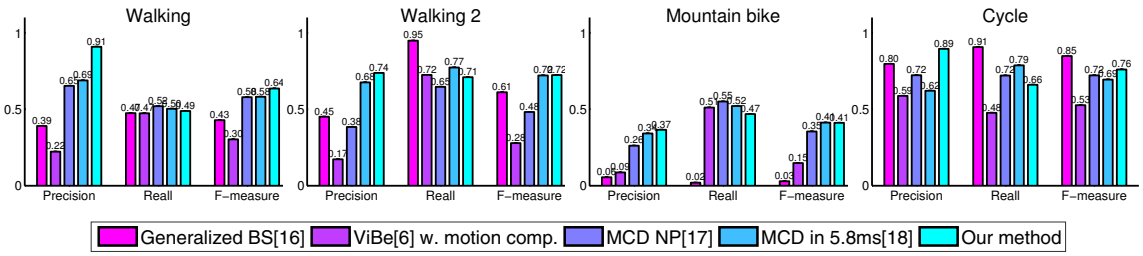
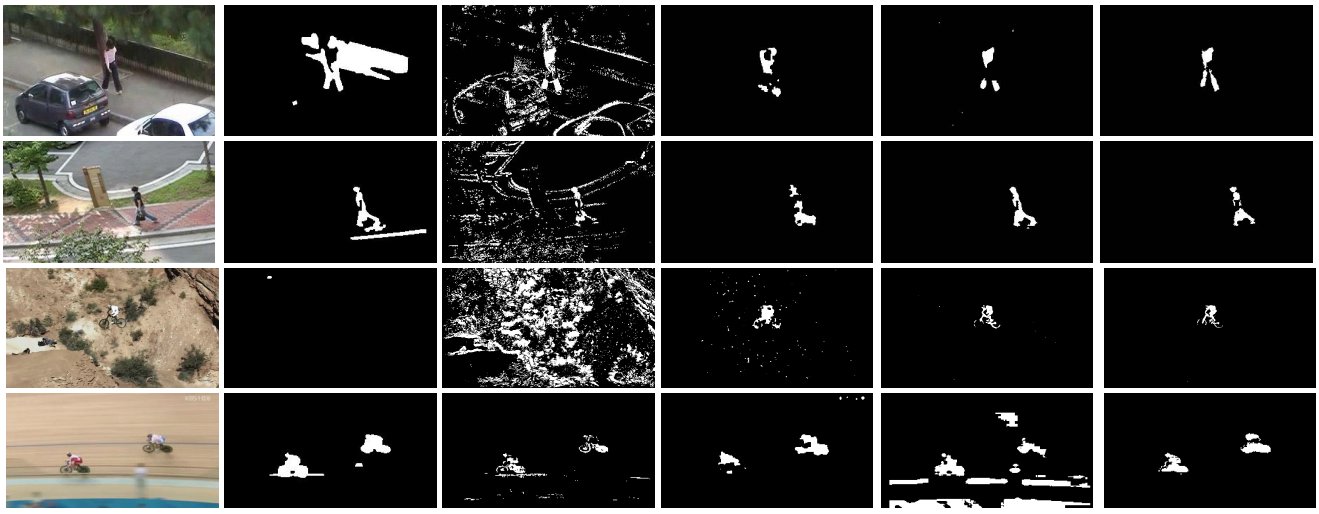


Fig. 3. Quantitative results of each sequences using pixel-wise *precision*, *recall*, F-measure.



(a) input image (b) General BS [16] (c) ViBe [6] (d) MCD NP [17] (e) MCDin5.8ms [18] (f) Our method

Fig. 4. Qualitative results on several images from different sequences. From top to bottom: *Walking*, *Walking 2*, *Mountain bike*, and *Cycle*. First columns are input images, and the other columns show the results of the compared methods: (b) General BS [16], (c) ViBe [6] with motion compensation, (d) MCD NP [17], (e) MCDin5.8ms [18], and (f) Our method.

so they can remove false detections and shows the best performance in the *Cycle* sequence. However, the resulting detection region contains large neighbor backgrounds, like the case of *Walking* and *Walking 2* sequences. Moreover, General BS sometimes miss the object completely as shown in the *Mountain bike* sequence when a foreground is not distinguished from a complex background. In [6] as shown in Fig. 4(c), many false detections arise in the image edge because they do not consider the inaccurate estimation of camera movements. Non-panoramic moving object detection in moving camera(MCD NP) [17] produces an incomplete foreground with inner hole and noise in Fig. 4(d). Though our method is based on the MCDin5.8ms [18], our method detects the objects clearly without foreground missing(*Walking* sequence) and drastic noise(*Cycle* sequence) unlike the result in [18].

We measure the computation loads of the compared methods on Intel Core i5-3570 3.4GHz PC with 320 $\times$ 240 image without parallel processing. As a result, Table 1 shows the run-time comparisons using average computation time. Generalized BP [16] takes about 30 seconds to proceed one frame, moreover, it also needs the optical flow calculation. Our method is the fastest algorithm among the compared methods including the baseline [18] owing to the proposed sampling method. We uploaded a supplementary video to Youtube to illustrate the distinctive comparison on the compared methods.[2].

4. CONCLUSIONS

We proposed a new scheme to improve the robustness of moving object detection in a non-stationary camera. To reduce the loss of true objects and the false detections, we used spatio-temporal properties of moving objects. From the spatio-temporal properties, we built a foreground probability map and generated a sampling map which selects the candidate pixels to find the actual objects. We applied the background subtraction and model update to only the selected pixels. Lastly, we refined the foreground to reduce false detections and fill the foreground hole clearly using the foreground probability. In the experiments, our method outperformed the state-of-the-art methods in the detection quality and speed.

[2]http://youtu.be/2UOu4OuBYUs

5. REFERENCES

[1] In Su Kim, Hong Seok Choi, Kwang Moo Yi, Jin Young Choi, and Seong G Kong, "Intelligent visual surveillance — A survey," *International Journal of Control, Automation and Systems*, vol. 8, no. 5, pp. 926–939, Oct. 2010.

[2] Chris Stauffer and W Eric L Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition (CVPR)*, 1999.

[3] Y Sheikh and M Shah, "Bayesian object detection in dynamic scenes," in *Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 74–79.

[4] Dar-Shyang Lee, "Effective Gaussian Mixture Learning for Video Background Subtraction.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 827–832, 2005.

[5] Teresa Ko, Stefano Soatto, and Deborah Estrin, "Warping background subtraction.," in *Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 1331–1338.

[6] Olivier Barnich and Marc Van Droogenbroeck, "ViBe: A Universal Background Subtraction Algorithm for Video Sequences.," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709–1724, 2011.

[7] Constant Guillot, Maxime Taron, Patrick Sayd, Quoc Cuong Pham, Christophe Tilmant, and Jean-Marc Lavest, "Background subtraction adapted to PTZ cameras by keypoint density estimation," in *British Machine Vision Conference (BMVC)*, 2006, pp. 34.1–34.10.

[8] Lionel Robinault, Stéphane Bres, and Serge Miguet, "Real Time Foreground Object Detection using PTZ Camera.," in *International Conference on Computer Vision Theory and Applications (VISAPP)*, 2009, pp. 609–614.

[9] Kiran S Bhat, Mahesh Saptharishi, and Pradeep K Khosla, "Motion Detection and Segmentation using Image Mosaics.," in *IEEE International Conference on Multimedia and Expo*, 2000, pp. 1577–1580.

[10] Rita Cucchiara, Andrea Prati, and Roberto Vezzani, "Advanced video surveillance with pan tilt zoom cameras," in *Proc. of the 6th IEEE International Workshop on Visual Surveillance on ECCV*, 2006.

[11] Eric Hayman and Jan-Olof Eklundh, "Statistical Background Subtraction for a Mobile Observer.," in *International Conference on Computer Vision (ICCV)*, 2003, pp. 67–74.

[12] Anurag Mittal and Daniel P Huttenlocher, "Scene Modeling for Wide Area Surveillance and Image Synthesis.," in *Computer Vision and Pattern Recognition (CVPR)*. 2000, pp. 2160–2167, IEEE Comput. Soc.

[13] Ying Ren, Chin-Seng Chua, and Yeong-Khing Ho, "Motion Detection with Non-stationary Background.," in *ICIAP*. 2001, pp. 78–83, IEEE Comput. Soc.

[14] J Xiao and M Shah, "Motion layer extraction in the presence of occlusion using graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, Dec. 2005.

[15] T Schoenemann and D Cremers, "High resolution motion layer decomposition using dual-space graph cuts," in *Computer Vision and Pattern Recognition (CVPR)*, 2008.

[16] S Kwak, T Lim, W Nam, B Han, and J H Han, "Generalized background subtraction based on hybrid inference by belief propagation and Bayesian filtering," in *International Conference on Computer Vision (ICCV)*, 2011, pp. 2174–2181.

[17] Soo Wan Kim, Kimin Yun, Kwang Moo Yi, Sun Jung Kim, and Jin Young Choi, "Detection of moving objects with a moving camera using non-panoramic background model," *Machine Vision and Applications*, Oct. 2012.

[18] Kwang Moo Yi, Kimin Yun, Soo Wan Kim, Hyung Jin Chang, Hawook Jeong, and Jin Young Choi, "Detection of Moving Objects with Non-stationary Cameras in 5.8ms: Bringing Motion Detection to Your Mobile Device," in *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2013.

[19] Jiman Kim, Xiaofei Wang, Hai Wang, Chunsheng Zhu, and Daijin Kim, "Fast moving object detection with non-stationary background," *Multimedia tools and applications*, vol. 67, no. 1, 2013.

[20] Hyung Jin Chang, Hawook Jeong, and Jin Young Choi, "Active attentional sampling for speed-up of background subtraction," in *Computer Vision and Pattern Recognition (CVPR)*, 2012.

[21] Fernand Meyer, "Topographic distance and watershed lines," *Signal Processing*, vol. 38, no. 1, pp. 113–125, July 1994.