Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 3, Issue 2: 38-44, DOI: 10.25236/AJCIS.030206

HySamBS: A Hybrid Sample-based Background Subtraction Method

Wei Zhang[1], Lian Huang[1,*]

[1] *College of Electronic & Information Engineering, Chongqing Three Gorges University, Chongqing 404100, China*

ABSTRACT. *Background subtraction technique is the foundation of video analysis applications. Although many background subtraction methods have been proposed, it is still challenging due to the various nature of video scenes. In this paper, we propose an improved method named HySamBS based on the traditional Self-Balanced Sensitivity Segmenter. This method mainly consists of three framework: 1) modified pixel-wise adaptive feedback, 2) region-wise refinement of segmentation results, 3) frame-wise camera motion compensation. The improved feedback mechanism limits the excessive increasing of segmentation threshold. As the traditional sample-based methods are inefficient for detecting intermittent motion objects, the proposed method extra estimates the initialized reference background image via existent background samples. "Ghosts" suppression can be accelerated with identifying whether motion objects exist in the reference background image. The cancellation of neighborhood diffusion prevents motion objects from being absorbed into background samples. In addition, the effect of camera motion on foreground segmentation is further resisted by feature points matching. Finally, comprehensive evaluation results on the ChangeDetection.net dataset indicate proposed method can adapted to diverse challenging videos scenes, and the overall evaluation metric is competitive with state-of-the-art sample-based methods. The complete source code is publicly available at https://github.com/HuangLian126/HySamBS.*

KEYWORDS: *change detection, background subtraction, blinking pixel, edge similarity, motion compensation*

1. Introduction

Change detection is the fundamental task of high-level computer vision applications, such as intelligent video surveillance [1], visual object tracking [2] and action understanding [3]. Change detection can extract motion objects called "foreground" and the static objects called "background" from video sequences. In general, background subtraction (BS) is one of the most commonly used change detection techniques. BS firstly construct an initialized background model by single

ISSN 2616-5775 Vol. 3, Issue 2: 38-44, DOI: 10.25236/AJCIS.030206

frame or some historical frames. The foreground segmentation then is carried out, so each pixel of current frame can be segmented into foreground or background by comparing the difference between the background model and the current frame. In other words, foreground segmentation is binary classification procedure. Ultimately, background model updating makes the background model well adapt to scene changes. In the last two decades, BS had made significant progresses and the following are some recent surveys [4]. Due to the typical challenges such as dynamic background, illumination change, camera jitter, shadows, etc., BS still is an active field of research.

In the last few years, sample-based background subtraction methods have attracted many researchers. Barnich et al. [5] firstly proposed landmark Visual Background Extractor (ViBe), which did not fit pixels to a fixed distribution. ViBe assumes pixels whose matching counts with samples exceed the threshold are background and other pixels are foreground. Moreover, ViBe also uses random diffusion to suppress ghosts caused by intermittent object motion. However, ViBe shows bad performance in scene of dynamic background where the branches are swaying. Fixed parameters limit the potential of ViBe. Naturally, subsequent researches tend to adaptively adjust parameters. Hofmann et al. [6] proposed a new adaptive feedback framework named Pixel-Based Adaptive Segmenter (PBAS). Parameters of segmentation and updating can vary with dynamic degree of samples. Stcharles et al. [7] proposed Self-Balanced Sensitivity Segmenter (SuBSENSE). The blinking pixels are introduced to quickly control the increase or decrease of the threshold in the exponential form. So SuBSENSE yields good results in change detection. Now, supervised methods use the ground-truth of a special scene for training, and they may lead to the poor performance when the scene does not appear in training sets. It is still potential to sequentially explore unsupervised methods. Considering simplicity and accuracy, we choose the SuBSENSE method as basic framework.

This paper is organized as follow. Section 2 will firstly introduce original SuBSENSE method; then describe our method including how to estimate initialized reference background image, detect foreground, and update background samples; finally, further explain how to solve the camera jitter. Section 3 will evaluates the proposed method with other state-of-art methods from qualitative and quantitative aspects. Finally, section 4 will summarizes the method and discuss how to make better in future.

2. Methodology

2.1 Background Modeling

In general, there are two initialization operations for sample-based methods, such as selecting recently observed consecutive frames or repeatedly sampling from neighborhood regions of pixels in the single frame. However, using single frame for initialization may cause the situation that foreground pixels exist in background

ISSN 2616-5775 Vol. 3, Issue 2: 38-44, DOI: 10.25236/AJCIS.030206

samples. Moreover, variations of pixels among consecutive frames are often subtle. In order to estimate the initialized referenced background image, the traditional initialization method needs to be changed. We select discontinuous frames with longer sampling interval that makes the difference between the current frame and the previous frame more obvious. The background sample B is redefined as:

$$B_i(x) = \{I_{P\cdot i}^{RGB}(x), I_{P\cdot i}^{LBSP}(x)\}, \quad 1 \le i \le N, \tag{1}$$

where $I_{P\cdot i}^{RGB}$ and $I_{P\cdot i}^{LBSP}$ respectively are color and texture features of the current frame, P is sampling interval, and i is frame index. Every P frames, the current color and texture features are added to the background samples until all the samples are filled. In our method, the sampling interval is extended from 1 to P.

The initialization of sample-base methods usually ends at this stage. However, the additional reference background image that contains the original scene information of the video sequence can be completely estimated using the above background samples.

Estimation of reference background image is inspired by the Independent Multi-modal Background Subtraction (IMBS) [8]. A pixel can be observed many times at the same position, which indicates this pixel is more likely to be the background. Note that the estimation of reference background image only relies on the existing color features of samples. Firstly, setting empty initialized reference background sample RBS and corresponding similarity matching counts C:

$$RBS(x) = \{RBS_1(x), RBS_2(x), \cdots, RBS_N(x)\} \tag{2}$$

and

$$C(x) = \{C_1(x), C_2(x), \cdots, C_N(x)\}. \tag{3}$$

The first element of background samples B_1^{RGB} is inserted into the reference background samples RBS_1(x), then corresponding $C_1(x)$ = 1. Starting from the second element, B_i^{RGB} will be compared with $RBS_k(x)$, as follows:

$$RBS_k(x),\ C_k(x) = \begin{cases} RBS_k(x) = RBS_i'(x) \\ C_k(x) = C_k(x) + 1 \\ RBS_i(x) = B_i^{RGB}(x) \\ C_i(x) = 1 \end{cases} \begin{array}{l} \text{if } ||RBS_k(x) - B_i^{RGB}(x)|| \le R_{ini} \\ \\ \text{otherwise} \end{array} \tag{4}$$

with

$$RBS_k'(x) = \frac{B_i^{LBSP}(x)C_k(x) + RBS_k(x)}{C_k(x) + 1}, \tag{5}$$

where ||·|| is the Euclidean distance, parameter R_{ini} is the cluster threshold, k is the k-th index in the existing reference background samples. B_i^{RGB} will compares with k existing reference background samples one by one. If the difference between $RBS_k(x)$

ISSN 2616-5775 Vol. 3, Issue 2: 38-44, DOI: 10.25236/AJCIS.030206

and B_i^{RGB} is subtle, a fusion value then is generated to replace the original referenced background sample and corresponding counts adds 1. Otherwise, B_i^{RGB} is inserted into the *i*-th position of *RBS*(x) and the corresponding counts $C_i(x)$ is set to 1. When all background samples are processed, unimportant referenced background samples need to be filtered. Traversing the index of the maximum value in $C(x)$, then the values at corresponding index in $RBS(x)$ is treated as the initialized referenced background:

$$RBI(x) = RBS_{index}(x) \tag{6}$$

and

$$index = \arg\max\left(C(x)\right). \tag{7}$$

2.2 Foreground Segmentation

As mentioned, SuBSENSE uses feedback mechanism to control foreground segmentation, which can effectively suppress the noises. Our method use the foreground segmentation of SuBSENSE.

2.3 Background Update

SuBSENSE controls different update rate to suppress the neighborhood diffusion, but the effect is not significant. Instead of suppressing neighborhood diffuse in a complex way, we try to cancel neighborhood diffusion. Without neighborhood update, other "ghosts" suppressing method is considered.

Firstly, the edges of referenced background image, current sequence and the segmentation results respectively are extracted via Canny operator [35]. Then, $Exist(k)$ is defined as follows:

$$Exist(k) = \begin{cases} 1 & \text{if } SE_C(k) - SE_B(k) > E_{th} \\ 0 & \text{if } SE_B(k) - SE_C(k) > E_{th} \end{cases}, \tag{8}$$

where k is the k-th connected component of segmentation results, $SE_C(k)$ are pixels sums that the edges of k intersect with the edges of current frame, $SE_B(k)$ are pixels sums that the edges of k intersect with the edges of referenced background image, and E_{th} is fixed edges threshold. $Exist(k) = 1$ indicates that k exists in current sequence, and $Exist(k) = 0$ indicates that k exists in referenced background image. Moreover, two counters Com_f and Com_b are constructed:

$$Com_f(x) = \begin{cases} Com_f(x) + 1 & \text{if } F(x) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

and

ISSN 2616-5775 Vol. 3, Issue 2: 38-44, DOI: 10.25236/AJCIS.030206

$$Com_B(x)=\begin{cases}0 & \text{if } Exist(x)=1\\ Com_B(x)+1 & \text{otherwise}\end{cases}, \quad (10)$$

where Com_f is the number that pixels x are continuously classified as foreground, Com_b is the number that pixels x continuously exists in reference background image. We use formula (9) and additional formula (20) to update the samples:

$$B_r(x)=\{I^{rgb}(x), I^{lbsp}(x)\} \quad \text{if } Com_f(x)\geq T_{com} \text{ and } Com_B \geq T_{com}, \quad (11)$$

where T_{com} is the two critical values, and r is a random number. If two counters exceed the critical value simultaneously, RGB and intra-LBSP values of pixel are updated to the background sample by force without considering the classified results of pixels and whether conservative updates are implemented.

3. Results and discussion

3.1 Evaluation Dataset and Metrics

To properly evaluate the performance of our method, a benchmark dataset that contains as many video surveillance scenes as possible is needed. The landmark CDnet dataset [9] offers comprehensive video sequences in real scenes with corresponding ground-truth masks. CDnet consists of 2012 version and 2014 version that includes 11 categories: baseline, camera jitter, dynamic background, intermittent object motion, shadow, thermal, bad weather, low framerate, night videos, point-tilt-zoom and turbulence.

Besides, 7 official metrics for evaluating different methods are Recall (Re), Specificity (Sp), False Positive Rate (FPR), False Negative Rate (FNR), Percentage of Wrong Classifications (PWC), F-Measure (FM), and Precision (Pr). If the background subtraction method obtains a high Recall score without sacrificing Precision, it is considered to be a good method. FM, which is calculated jointly by the Recall and Precision, is generally accepted as a good indicator of overall quality. So FM is selected as the main evaluating metric in our evaluation process.

3.2 Parameter Setting

Some parameters need to be determined for adapting more video sequences better. All the following parameters will be discussed: sampling interval P = 4, the number of samples N = 30, LBSP similarity threshold T_r = 0.27, edge similarity threshold E_{th} = 10, the critical value of counter T_{com} = 35.

3.3 Experimental Results on CDnet

We chose SuBSENSE as the framework to be improved and mainly solves some

ISSN 2616-5775 Vol. 3, Issue 2: 38-44, DOI: 10.25236/AJCIS.030206

problems for dynamic background and intermittent object motion and camera jitter. To demonstrate the proposed framework is effective, we first compare the original SuBSENSE and HySamBS with the FM score. The completely quantitative evaluation results on the entire CDnet are given in Table 1.

For dynamic background category, The FM score of HySamBS (0.819) is higher than SuBSENSE (0.817) by 0.002. FM has been greatly improved in “boats”, “canoe” and “overpass” sequences. The suppression of dynamic noise is weakened in other scenes, such as “fountain01” and “fall” sequences. For intermittent object motion category, HySamBS makes significantly improvement, and the FM score reaches 0.7931, which increases by 20.7%. The HySamBS cancels neighborhood diffusion, and lead to balances between detection of static motion objects and the suppression of “ghosts”. These are the main reasons why FM increases for intermittent object motion category. For camera jitter category, The FM score of HySamBS (0.8207) is higher than SuBSENSE (0.8152) by 0.0055. Besides camera jitter, the “boulevard” sequence also has a strong illumination mutation, so HySamBS obtains a lower score in this sequence. It can be clearly seen that the FM score is very low for PTZ category. The HySamBS estimates the offset distance of pixels by feature point registration. The frame-wise camera motion compensation can fail if the offset distance is extremely large, and this is why the FM score is only 0.0529. On the CDnet2012 dataset, the FM score of HySamBS intuitively exceeds all other methods for intermittent object motion, camera jitter, and thermal. The average FM score of HySamBS is 0.848, which fully shows the superior performance of method. However, on the CDnet2014 dataset, the performance of HySamBS needs to be further improved since we have not considered a specific solution for PTZ sequence. On the whole CDnet dataset, the FM score can reach 0.713, and the adaptability is still acceptable.

Table 1 Complete results for proposed HySAMBS and original SuBSENSE on CDnet dataset.

	Re	*Sp*	*FPR*	*FNR*	*PWC*	*Pr*	*FM*
baseline	0.9620	0.9979	0.0021	0.0380	0.3512	0.9338	0.9472
camera jitter	0.8774	0.9851	0.0149	0.1226	1.9200	0.7848	0.8207
dynamic bgk.	0.8786	0.9975	0.0025	0.1214	0.3791	0.8006	0.8190
inter. obj. mo.	0.7912	0.9926	0.0074	0.2088	2.0906	0.8388	0.7931
shadow	0.9472	0.9902	0.0098	0.0528	1.1603	0.8196	0.8759
thermal	0.8695	0.9853	0.0147	0.1305	1.9156	0.8032	0.8325
bad weather	0.8298	0.9974	0.0026	0.1702	0.5966	0.8131	0.8167
low framerate	0.8261	0.9910	0.0090	0.1739	1.4282	0.6008	0.6420
night videos	0.6517	0.9717	0.0283	0.3483	3.5301	0.3619	0.4416
PTZ	0.8423	0.6760	0.3240	0.1577	32.2264	0.0278	0.0532
turbulence	0.8692	0.9992	0.0008	0.1308	0.1617	0.7903	0.8217
overall (2012)	0.8877	0.9914	0.0086	0.1123	1.3028	0.8301	0.8481
overall (2014)	0.8038	0.9271	0.0729	0.1962	7.5886	0.5188	0.5550
overall (2012+2014)	0.8496	0.9622	0.0378	0.1504	4.1600	0.6886	0.7149

ISSN 2616-5775 Vol. 3, Issue 2: 38-44, DOI: 10.25236/AJCIS.030206

4. Conclusion

The proposed method mainly focuses on some problems for intermittent object motion, dynamic background and camera jitter. Improved blinking pixel is suitable for this case that motion objects pass through regions of the dynamic background. The ability to suppress periodic noise remains to be improved. Instead of neighborhood update, the extra reference background image can be the convenient way that effectively eliminates "ghosts" produced by intermittent object motion and ensures the complete foreground objects as well. Note that some scenes, such as with hard shadow, complex background texture features, and similar color features between background and foreground, have negative effect on the use of edge similarity. Although our method can overcome local illumination changes, the ability to tolerate global illumination needs to be further improved. In the future, we will continue to consider the unsupervised method, and use the appropriate semantic segmentation and optical flow estimation to solve the *PTZ* sequence from temporal-spatial information.

References

[1] B. Tian, B.T. Morris and M. Tang (2017). Hierarchical and networked vehicle surveillance in ITS: a survey. IEEE Transaction on Intelligent Transportation Systems, vol.16, no.1, p.25-48.

[2] K.M. Abughalieh, B.H. Sababha and N.A.Rawashdeh (2018). A video-based object detection and tracking system for weight sensitive UAVs. Multimedia Tools & Applications, vol.78, no.1, p.9149-9167.

[3] K.P. Chou, M. Prasad and D. Wu (2018). Robust feature-based automated multi-view human action recognition system. IEEE Access vol.6, p.15283–15296.

[4] T. Bouwmans (2014). Traditional and recent approaches in background modeling for foreground detection: An overview. Computer Science Review, vol.11, p.31-66.

[5] O. Barnich, M.V. Droogenbroeck (2011). ViBe: A universal background subtraction algorithm for video sequences. IEEE Transactions on Image Processing. vol.20, p.1709-1724.

[6] M. Hofmann, M.; P. Tiefenbacher and G. Rigoll (2012). Background segmentation with feedback: The Pixel-Based Adaptive Segmenter. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, p.38-43.

[7] P. St-Charles, G. Bilodeau and R. Bergevin (2015). SuBSENSE: A universal change detection method with local adaptive sensitivity. IEEE Transactions on Image Processing, vol.24, p.359-373.

[8] D. Bloisi, D, L. Iocchi (2012). Independent multimodal background subtraction. International Conference on Computational Modeling of Objects Presented in Images: Fundamentals, Methods and Applications. 2012, p.39-44.

[9] N. Goyette, P.M. Jodoin and F. Porikli (2012). Changede-tection.net: A new change detection benchmark dataset. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, p.1-8.