A novel object detection technique for dynamic scene

Xiu Li, Liansheng Chen, Zhixiong Yang, Wanlu Zhang
Shenzhen Key Laboratory of Information Science and Technology
Graduate School at Shenzhen, Tsinghua University
Shenzhen, China
cls13@mails.tsinghua.edu.cn

***Abstract*—Object detection in video streams plays an important role in many computer vision applications and background subtraction is the most popular method which compares each new frame to a model of the background. However, the static background is practically impossible and dynamic background makes the perfect object detection very difficult. In the paper, based on the visual background extraction (ViBe) algorithm, we present a new method dealing with this problem. A new update mechanism is present to get more robust model for the dynamic regions. We describe our method in full details and compare it to other background subtraction techniques. Our experimental results show that our proposed method outperformed several state-of-the-art object detection approaches. Moreover, it can process 60 frames per second on an Intel i5 3.1GHz CPU with C++ code.**

I. Introduction

In many image processing and computer vision applications, accurate object extraction from video stream is of great importance for consecutive tracking, recognition, and behavior analysis. Background subtraction, which compares an observed image with the preserved background model, has become the mainstream technique for accurate foreground extraction.

In the simplest case, a static background frame can be compared to the current frame and pixels with high deviation are classified as foreground. However, it is rarely the case and even the stationary camera does not mean a static background due to different kinds of complicated environments, such as water wave, spring, swaying leaves, et al. [6], which leads to many false alarms. The camera jitter [3] makes the situation even worse.

The movement of dynamic background leads to the pixel value failing to be compared with corresponding model and thus results in false alarms. However because the displacement in consecutive frames is usually very small, we may find corresponding model in a very small neighborhood, which we named as neighborhood match. If there exists a successful match, we replace two samples in model with the value, which we call sharp update. If we conduct these two operations for all pixels, there will be many false negatives and foreground information will pollute the background model. As a result, such operations just work in dynamic background region.

The rest of paper is organized as follows. In section II, we present related background subtraction methods. In section III, a detailed description of our algorithm will be provided. Then the experimental results of the proposed method in several videos are given in Section IV. Finally, we conclude our work and talk about the future work in Section V.

II. Related Work

Background subtraction has become the mainstream technique in object detection. Its main idea is to create and maintain a model of the scene without objects and then to detect the foreground by comparing the current frame with the estimated background. The advantage behind this concept is that no prior knowledge is required to detect the object as long as their appearance differs enough from the background (i.e. they are not camouflaged [3]). A multitude of algorithms and methods for background modeling have been developed, such as Gaussian mixture model (GMM) [14], kernel density estimate (KDE) [5], Codebook (CB) [8], Visual background Extraction (ViBe) [2] and many improved versions [16], [11], [12], [13] based these methods. Excellent survey papers can be found in [10], [9], [4], [3].

Due to the dynamic nature of real-world scenes, there are inevitably many false alarms and thus numerous methods have been presented to handle with this issue. Some statistical models [14], [5] are used to represent the multi-modal essence of dynamic background. In fact, these models are merely able to represent very small background movement. Some hold the idea that the model converges too slowly and design some algorithms [7], [13] to speed up the update process in the dynamic region. But there is still some delay in model. Some instead leave the problem to be solved in regularization step [11].

III. The Proposed Method

As stated in the previous section, based on ViBe [2], we propose a new approach to handle with the problem of dynamic scene. We present a mechanism of neighborhood match and sharp update to remove false alarms and speed up update process in the dynamic region, which is described in detail in Section III-A. Some details about post-process and

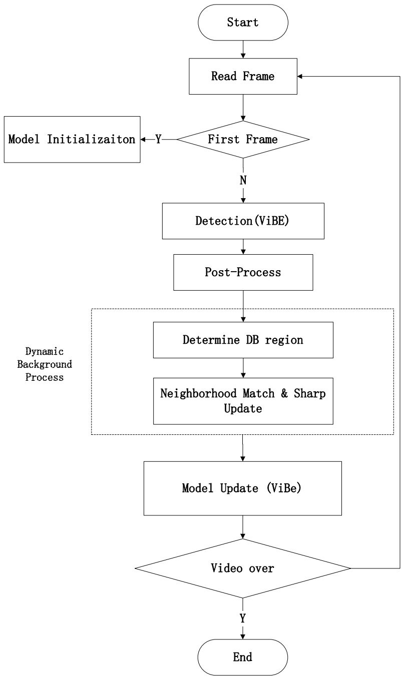

Fig. 1. Overall framework of our method

locating dynamic region is given in Section III-B. An overall framework is provided in Figure 1.

A. Neighborhood Match and Sharp Update

In outdoor environments with fluctuating background, the false alarms derive from two sources. First, there are false alarms due to random noise which should be homogeneous over the entire image. Second, there are false alarms due to small movements in the scene background that are not represented in the background model. This can occur, for example a tree branch moves further than it did during model generation. Also small camera displacements due to wind load are common in outdoor surveillance and cause many false alarms. This kind of false detection is usually spatially clustered in the image and it is not easy to eliminate using morphology or noise filtering because these operations might also affect small and occluded targets. Our method aims to suppress the false detections due to small and unmodelled movements in the scene background.

In ViBe, the pixels are only compared with their own models and the ones similar with at least two samples in the model are considered as background, having a chance of one sixteenth to make its model and the model of neighborhood updated with its current value. In this way, only pixels sharing similar appearance with the samples in the model can be incorporated in to the model and thus in most cases the model just contains samples from only one object. Although there is a probability that the model contains samples from different objects due to the fact that two objects both similar with the same object may be very different, it takes a very long time to collect enough samples and requires special condition that the pixel values vary very slowly, such as slow illumination. In addition, with the mechanism of propagation, updating the model of its neighborhood with its own pixel value, the model in the edge of objects may have the chance of containing samples from different objects but collecting enough samples takes a long time. The models in the region of dynamic background should always be multi-modal and thus the algorithm of ViBe fails to handle with the problem of dynamic scene. Specifically, the false alarms due to the small movements of background objects prevent the moved background objects incorporated in to the models of new pixel locations and further the unchanged models will all the same lead to false alarms, which is a drop-dead halt.

If some part of the background moves to occupy a new pixel, but it was not part of that pixel, then it will be detected as a foreground object. However, this object will have a high probability to be a part of the background distribution at its original pixel. Assuming that only a small displacement can occur between consecutive frames, we decide if a detected pixel is caused by a background object that has moved by comparing the detected pixel value with models in the neighborhood and finding out some matched models. We name the procedure of finding matched models in the neighborhood as neighborhood match which can break the drop-dead halt.

The object of dynamic background does not stay too long on one spot and collecting enough samples (typically 2 samples) to represent the moving object needs about the time of 32 frames in the situation of the object being there. If the object is always moving, it will take more time to collect enough samples. Moreover, if the enough samples fail to be collected immediately, some false alarms of background objects will also fail to be found out and removed, which furthermore prevent the background objects incorporated in to the models of new pixel locations. In a word, although the mechanism of neighborhood match break the drop-dead halt, the slow update mechanism may lead to a vicious circle mentioned above. In order to break the vicious circle, for the false alarms of moving background object, we replace two samples in the model of new pixel location with two copies of the current pixel value, which is named as sharp update. In this way, the moving background objects are able to be incorporated in to the models of new pixel locations immediately and further movements can also be detected. As a result, a more robust and reliable model is obtained and less false alarms will be detected which is beneficial to post process including median filtering and morphological operations, which be described in following section. As is shown in Figure 2, we presented several results from ViBe and our method, and it is obvious that much less false alarms are detected in our results, which demonstrates the advantage of our model.

B. Further Details

In addition, note that the operations of neighborhood match and sharp update is only conducted in the region of dynamic background which is indicated in a 2D map of pixel level, named as blink map. In [15], for each pixel, the previous updating mask (prior to any modification) and a map with the blinking level is stored. This level is determined as follows. For a pixel, if the current updating label is different from the previous updating label, then the blinking level is increased by 15 (the blinking level being kept within the [0,150] interval), otherwise the level is decreased by 1. A pixel is considered as blinking if its level is larger or equal to 30. Directly detecting blinking pixels following the steps above can be inconvenient since the borders of moving foreground objects would also be included in the result. Consequently, we find blinking pixels just in the region of background which is indicated in the post-processed and dilated updating mask.

For post-process, we first remove noise with median filter with a large size, which may lead to some false negatives. A close morphological operation, finding contours and convex hull for each contours are followed to determine a region where there is no noise. A combination both results with and operation will remove most noise and lead to very few false negatives. In fact, it is a fusion of pixel-level information and block-level information. A close operation with large size aims to get closed contour and then holes in these contours are filled. Twice dilation morphological operation aims to ensure all true positive are in the result. Experimental performance show the result region usually contains all foreground region. A basic ViBe detection is conducted again in the result region with very small R to get almost the true positives and very few false positives. A further dilated version is passed to find the dynamic region.

IV. Experimental Results

The proposed method is evaluated using the dataset and evaluation metrics in CVPR 2012 Change Detection challenge (CDnet 2012) [6].One fixed set of parameters is used for all sequences and some of them inherit from that in [2], [15]. All experiments run on an Intel i5-3450 processor with 3.10GHz, 16GB RAM and Win7 OS. The algorithm is implemented by C++ codes, and its average processing speed is about 60 fps for the image size of 320x240.

A. Performance for dynamic scene

In the CDnet 2012 dataset, the video sequences of dynamic background (DB) and camera jitter (CJ) typically share the problem of dynamic scene. Some of these video sequences share the problem of shadow, highlight, illumination change and camouflage. As a result, to demonstrate the contributions of our method, we compare our method with several state-of-the-art methods in the video sequences of badminton in CJ, canoe, fall, fountain01 and overpass in DB, which do not share the problems of shadow, highlight, illumination change and camouflage. These state-of-the-art methods include ViBe [2] on which our method is based, SuBSENSE [13] which is the best one in all improved versions for ViBe, CDet [1] which performs best for the video sequences in DB, GPRMF [1] which performs best for the video sequences in CJ. Detailed statistical data is provided in Table I and the data in bold is the best in each row.

It is obvious that our method performs better than the other five methods for dealing with the problem of dynamic scene. Our performance is better than the other methods on most of the video sequences and performs very little worse than the best one in a few of video sequences and moreover all of the overall measurement is better than the other methods.

B. Overall performance and discussion

We compare our method with methods mentioned in section IV-A using the metrics of recall, precision, F-Measure and time cost. Table II displays the statistical data. It is obvious that our method performs much better than ViBe and increases the first three metrics by more than 10%. Although SuBSENSE and CDet obtain better performance but the gaps are not very large. In fact, these three methods design some tricks to deal with the problem of shadow, highlight, illumination change and camouflage, which are the reasons of these gaps. But these tricks meanwhile introduce some extra time cost and thus our method runs faster than the three methods.

V. Conclusion

In this paper, we presented a ViBe-based background subtraction algorithm, which deals with the problem of dynamic scene and static object very well. Neighborhood match can remove the false alarms of dynamic scene and sharp update can speed up the update process, make the model more robust and reliable.

As mentioned above, in our algorithm no specific mechanisms are designed to deal with the shadow, highlight, illumination change and camouflage. We believe a good representation of pixel will be the answer and try to find out the appropriate representation to deal with these issues in the future.

Acknowledgment

This research was partly supported by National Natural Science Foundation of China (NSFC, Project No.: 71171121/61033005) and National 863 High Technology Research and Development Program of China. (863 Project No.: 2012AA09A408).

References

[1] changedetection website. http://changedetection.net/.

[2] O. Barnich and M. Van Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *Image Processing, IEEE Transactions on*, 20(6):1709–1724, 2011.

[3] T. Bouwmans. Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11:31–66, 2014.

[4] S. Brutzer, B. Höferlin, and G. Heidemann. Evaluation of background subtraction techniques for video surveillance. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1937–1944. IEEE, 2011.

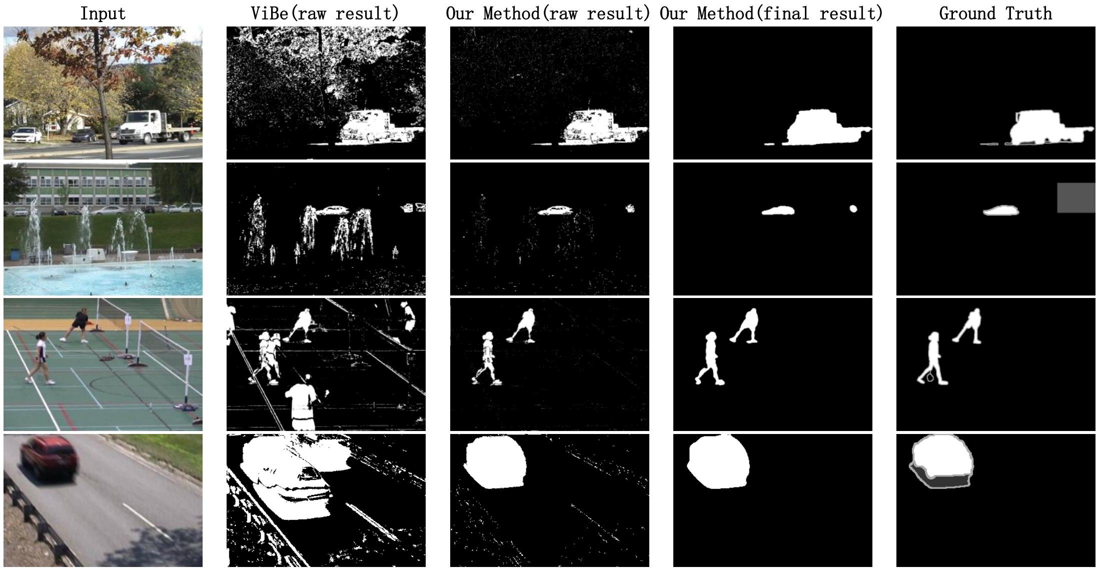

Fig. 2. Comparison with ViBe (the threshold R = 10)

TABLE I
METHOD COMPARISON ON PERFORMANCE FOR DYNAMIC SCENE

category	video name	method measure	ViBe	SuBSENSE	CDet	GPRMF	Our Method
dynamic scene	badminton	recall	0.799	0.922	**0.971**	0.823	0.965
		precision	0.576	0.843	0.851	**0.964**	0.937
		F-Measure	0.669	0.881	0.907	0.888	**0.951**
	canoe	recall	0.858	0.659	**0.966**	0.945	0.95
		precision	0.855	**0.993**	0.974	0.978	0.957
		F-Measure	0.856	0.792	**0.97**	0.961	0.953
	fall	recall	0.789	0.857	**0.998**	0.969	0.988
		precision	0.267	0.876	0.887	0.199	**0.942**
		F-Measure	0.399	0.866	0.926	0.331	**0.964**
	fountain02	recall	0.802	0.923	**0.95**	0.884	0.945
		precision	0.862	**0.966**	0.948	0.937	0.965
		F-Measure	0.831	0.944	0.949	0.91	**0.955**
	overpass	recall	0.763	0.785	0.882	**0.98**	**0.98**
		precision	0.851	0.943	0.919	0.891	**0.97**
		F-Measure	0.805	0.857	0.9	0.934	**0.975**
	overall	recall	0.802	0.829	0.953	0.92	**0.966**
		precision	0.682	0.924	0.916	0.794	**0.954**
		F-Measure	0.712	0.868	0.93	0.805	**0.96**

TABLE II
METHOD COMPARISON ON OVERALL PERFORMANCE (MS/FRAME)

	ViBe	SuBSENSE	CDet	GPRMF	Our Method
recall	0.682	0.828	**0.903**	0.837	0.846
precision	0.736	0.857	0.84	0.814	0.854
F-Measure	0.668	0.826	**0.861**	0.794	0.818
time cost (ms/frame)	**2 (C++)**	22 (C++)	25 (C++)	unknown	16 (C++)

[5] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *Computer VisionECCV 2000*, pages 751–767. Springer, 2000.

[6] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. Changedetection. net: A new change detection benchmark dataset. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 1–8. IEEE, 2012.

[7] M. Hofmann, P. Tiefenbacher, and G. Rigoll. Background segmentation with feedback: The pixel-based adaptive segmenter. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 38–43. IEEE, 2012.

[8] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground–background segmentation using codebook model. *Real-time imaging*, 11(3):172–185, 2005.

[9] D. H. Parks and S. S. Fels. Evaluation of background subtraction algorithms with post-processing. In *Advanced Video and Signal Based Surveillance, 2008. AVSS'08. IEEE Fifth International Conference on*, pages 192–199. IEEE, 2008.

[10] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: a systematic survey. *Image Processing, IEEE Transactions on*, 14(3):294–307, 2005.

[11] Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(11):1778–1792, 2005.

[12] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin. A self-adjusting approach to change detection based on background word consensus. In *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, pages 990–997. IEEE, 2015.

[13] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin. Subsense: A universal change detection method with local adaptive sensitivity. *Image Processing, IEEE Transactions on*, 24(1):359–373, 2015.

[14] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2. IEEE, 1999.

[15] M. Van Droogenbroeck and O. Paquot. Background subtraction: Experiments and improvements for vibe. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 32–37. IEEE, 2012.

[16] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 28–31. IEEE, 2004.