TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol. 12, No. 11, November 2014, pp. 7778 ∼ 7784
DOI: 10.11591/telkomnika.v12i11.6545

Weighted Samples Based Background Modeling for the Task of Motion Detection in Video Sequences

Boubekeur Mohamed Bachir[*1], Benlefki Tarek[2], and Luo SenLin , Labidi Hocine[1]
[1]School of Information and Electronics, Beijing Institute of Technology
[1]Beijing, China 100081
[2]School of Electronics and Information Engineering, Beihang University
Beijing, China 100191
[*1]corresponding author, e-mail: msboubekeur@yahoo.fr

Abstract

In this paper, a non parametric method for background subtraction and moving object detection based on adaptive threshold using successive squared differences and including frame difference process is proposed. the presented scheme focused on the case of adaptive threshold and dependent distance calculation using a weighted estimation procedure. In contrast with the existing update procedures (First-in First-out, random pickup), We proposed an intuitive update policy to the background model based on associated decreasing weights. The presented algorithm succeeds on extracting the moving foreground with efficiency and overpasses the problematic of ghost situations. The proposed framework provides robustness to noise. Experiments show competitive results compared to existing approaches and demonstrate the applicability of the proposed scheme in a variety of video surveillance scenarios.

Keywords: *Background Subtraction, surveillance, weighted samples.*

1. Introduction

Most of static camera based monitoring systems for security purposes rely on background modeling and subtraction process for detecting and identifying moving foreground objects in the video scene, the main advantage of background subtraction techniques is that no prior knowledge on the nature of the target object to be detected is needed. The subtraction of inconsistent information existing in the background implies the retrieval of interesting foreground objects. One may easily verify the spatial consistency between neighboring pixels resulting of a high correlation between the intensity values in a tight neighborhood. The temporal information provided by the succession of frames is also a cue to detect relatively gradual or fast change in the scene. By comparing the intensity value of pixel at the same position in difference time lapses, a change if exist should be detected. One way to do is to compute the distance between the current pixel value and the background model pixel(s) followed by a comparison with a threshold. After classification, an update of the background model is necessary to ensure that the background model can learn the changes in the video scene or to learn the environment changes. In this paper we consider that the estimation of the distance metric computed to tell where the current pixel value stands should include the way that the background model is updated. In this paper we introduce a Temporal based approach for background subtraction task, using a sample background modeling with adaptive threshold. In contrast to some existing methods which consider a sampled background model, the considered set of samples is directly exploited to determine the distance metrics [1, 2], we do not assume that the background samples are equally distributed, in fact the proposed approach used associated weights to estimate the distance between the background model and the streaming frames. In addition, the update procedure in our approach neither does replace the sample of the first frame or the last frame [1], nor choose a random location to update [2]. The proposed method applied a weight related update to all the samples in the background model.

Received August 7, 2014; Revised September 14, 2014; Accepted October 2, 2014

This paper is organized as follows; in section II, a short review of background subtraction (BS) algorithms dealing with different video surveillance challenges has been presented. Section III introduces the proposed framework which consists of (i) a training phase for building the background model, (ii) a distance metric estimation and a variable threshold based decision for classifying and separating background and foreground pixels, and (iii) the proposed updating strategy adopted in the framework. Experimental results of the proposed scheme and a discussion of the issues related to noise and ghost cancellation are presented in section IV. The fifth section is devoted to the performance analysis of the proposed framework and tables a comparison between the proposed algorithm and some existing background subtraction approaches.

2. Related work

In [1], the authors presented a non-parametric kernel density estimation (KDE) for background and foreground modelling using a short term and a long term model based on the selection of N background model samples, KDE is an efficient solution to moving object extraction, however it needs a considerable computational cost. The authors in [3], classified existing background subtraction methods into recursive and non-recursive approaches, and made a comparison between simple basic methods and probabilistic modeling based approaches, their experiments showed that even basic method could produce good results, while the computational cost kept low. A fast and robust algorithm for background subtraction was proposed in [4], the authors presented a new hierarchical motion detection algorithm based on sigma-delta modulation. They have considered a conditional approach by inserting controllers into the classification and the update process. In [2], a powerful algorithm for background modeling and foreground extraction named the Visual Background Extractor (ViBe) is proposed; the algorithm adopts a sample based background modeling approach with a stochastic replacement and a spatial diffusion for the update step. Using a constant threshold value for background /foreground separation; ViBe overcomes most of background subtraction challenges. It has been argued in [5] that improved results have been found by using a threshold as the half of the standard deviation computed for all the samples in the background model over time. The authors in [6] used tow threshold values and a three successive frames for moving object detection, they affirmed that such a consideration is strongly adapted to the environment changes. In [7], the authors introduced an algorithm for moving vehicle detection using a combination of semantic and background differences, they used a limited threshold value to build the binary images, even though the results was quite impressive. The authors in [8] draw a comparison between a set of background subtraction techniques using various distance computations, in addition they have introduced a square sum of differences between RGB entries and the background frame. In this paper, we propose to use weighted squared differences between entry frames and the background model as a distance metric, as well as exploiting the weights on the update procedure. The following paragraphs present the extents of the proposed approach, and detail the steps and the choices of the adopted parameterization.

3. PROPOSED ALGORITHM: ADAPTIVE THRESHOLD

The presented algorithm consists of three phases: training and background modeling stage, foreground/background separation phase, and an update step. In the following; we detail these steps and justify the choices that were made.

3.1. Background Modeling

A non parametric background modeling strategy is adopted in the framework, the background model is considered to be a set of K frames taken during the initialization. Let:

$$BGM(x, y) = \{b_1, b_2, \ldots, b_K\} \tag{1}$$

be the collection of K background samples at location (x, y), where $b_m, m = 1, 2, \ldots, K$ are the samples collected at different times at location (x, y). Moreover, a weight given by equation (2) is associated with each background sample,

$$W_i = \frac{1/2^{i-1}}{\sum_{m=1}^{K} \frac{1}{2^{m-1}}} \tag{2}$$

Where; i is the index of the sample in the background model, the values w_i are in the range $[0, 1]$, the normalization of the weights came to ensure that regardless the number of samples chosen to build the BGM, the sum of the weights remains equal to 1. If no foreground object is present during the initialization step, the set of K pixels should represent a range of possible values for a background pixel. In such a case, the variation in intensity value over time for $BGM(x, y)$ should not be significant at all, we introduce then in equation (3) a measure $S(x, y)$ defined as the mean square of successive differences between background model samples in the same location (x, y).

$$S(x, y) = \frac{1}{K-1} \sum_{m=2}^{K} (b_m - b_{m-1})^2 \tag{3}$$

We are interested in the behavior of the pixel belonging to the background model at location (x, y), the metric $S(x, y)$ takes a values according to the degree of similarity between background model samples.

3.2. Foreground/ Background separation

In the algorithm, a decision scheme based on a distance metric estimation and a variable threshold is used for separating and classifying foreground and background pixels. The proposed foreground/ background separation scheme involves two successive tests: First, for every new pixel; a weighted distance $d(x, y)$ metric is estimated using equation (4).

$$d(x, y) = \frac{1}{K} \sum_{m=1}^{K} w_m (b_m - V^t(x, y)) \tag{4}$$

Where $V^t(x, y)$ denotes a current pixel value at location (x, y) and w_m are the weights values associated with background model samples defined in equation (2). Now given the value of the distance computed using equation (4), this calculated value is compared to the metric estimated by equation (3), if the distance is greater than $S(x, y)$; the pixel is classified a priori as a foreground, otherwise is considered a background pixel. The foreground pixels are labeled 1 and the background pixels by 0. Equation (5) shows the prior obtained binary mask.

$$M^{prioir}(x, y) = \begin{cases} foreground & if \quad d(x, y) > S(x, y) \\ background & if \quad d(x, y) < S(x, y) \end{cases} \tag{5}$$

Second, by analyzing the foreground mask provided by this first test, we note the presence of ghost pixels in the binary mask, especially when foreground objects are present in the scene during the initialization phase. To deal with this challenge and in order to make the framework more immune to the problem of ghost; a second test based on the computation of the difference between successive frames and making use of the metric $S(x, y)$ is added to the algorithm. A priori classified foreground pixel is finally validated as a foreground pixel in the case where only the distance between successive frames is larger than the measure $S(x, y)$, otherwise the corresponding pixel is declared as a background pixel. Equation (6) shows the obtained labeled mask.

$$M(x, y) = \begin{cases} 2 & if \quad M^{prior}(x, y) = 1 \quad and \quad Diff(x, y) > S(x, y) \\ 1 & if \quad M^{prior}(x, y) = 1 \quad and \quad Diff(x, y) < S(x, y) \\ 0 & if \quad M^{prior}(x, y) = 0 \end{cases} \tag{6}$$

and

$$Diff(x, y) = (V^t(x, y) - V^{t-1}(x, y)).(V^{t-1}(x, y) - V^0(x, y)) \tag{7}$$

Only pixels with $M(x, y) = 2$ are actually foreground pixels, two types of background pixels to be distinguished: background pixels with $M(x, y) = 0$ and those where $M(x, y) = 1$. In the equation (7), the value $V^0(x, y)$ represents the first frame, in fact in our frame work we do not make any supposition regarding the first frame, furthermore if a ghost situation occurs, it would be eliminated by the difference between frames in the equation (7).

3.3. Update phase

When coming to updating the background model, the reader can distinguish between two strategies: the blind update policy and the conservative approach strategy. In the blind update procedure, each background model pixel is updated without considering the output of the foreground/ background separation phase. The conservative approach strategy depends on the result of the classification step in a way that only classified background pixels are allowed to update the background model samples, as a consequence background samples in the locations corresponding to actually classified foreground pixels remain without change. In the framework, the proposed updating strategy is neither a blind update approach nor a conservative one. A conditional weighted conservative update strategy is presented in this paper; earlier; we have identified two types of classified background pixels in the proposed foreground/ background separation stage: pixels which are classified background from the first test, and other pixels that were first declared as a foreground and later set to background pixels using the second test. For the first type, the value of those pixels are included directly in the background model samples. Moreover for the second type and for those pixels which their corresponding frame difference is less than the S value, background model updating follows a weighted values as shown in equation (8)

$$BGM(x,y)^{new} = \begin{cases} w_m.BGM(x,y)^{old} + (1-w_m).V(x,y) & if \quad M(x,y) = 1 \\ V(x,y) & if \quad M(x,y) = 0 \end{cases} \tag{8}$$

More explicitly, at each location and for the first type of classified background pixels where $M(x, y) = 0$; all the corresponding background samples are replaced by the current pixel value $V(x, y)$. For the second type of background pixels where, $M(x, y) = 1$, the updating process is achieved by using a weighted sum between every background model sample and the current pixel value, this update can be understood as follows: the first background sample keeps w_1 of its own value and gets $(1 - w_1)$ from the current pixel value. The next background pixel keeps only w_m and gets the $(1 - w_m)$ left from the current pixel value $V(x, y)$. The increasing percentage of the current frame value in the update of other layers explains that further layer updated more impact obtained by the background model samples from the current frame. Furthermore, the metric defined by equation (3) should be updated to ensure that the threshold considered for further decision is up to date. The update process goes as mentioned on equation (9).

$$S(x,y) = \begin{cases} x & if \quad M^{prior}(x,y) = 1 \\ \frac{1}{K-1}\sum_{m=2}^{K}(b_m - b_{m-1})^2 & if \quad M(x,y) = 1 \\ S(x,y) & otherwise \end{cases} \tag{9}$$

Note that the threshold takes the value of the current pixel value when the pixel is classified as foreground object. This value has been chosen to make sure that the update for the weighted formula presented in equation (8) be effective for pixels satisfying $M(x, y) = 1$, in the case where these pixels are classified in further process as background pixels.

4. EXPERIMENTAL RESULTS

The proposed framework was first implemented on MATLAB and later on Visual Studio C++ to test its real time performance. Our experiments were conducted on an I7 CPU with 2.2 GHz, the Change Detection dataset introduced by [9] and publicly available on www.changedetection.net has been used. This data set contains a variety of video sequences including most of the challenges that usually face background subtraction algorithms. A set of video sequences have been chosen to test the performance of the proposed approach, and to experiment our algorithm with various challenges presented in this dataset in its 2012 version. The number of initials frames chosen to initialize the background model is considered K = 10 in all our conducted experiments. Figure 1 shows the results of the proposed algorithm using frames from the baseline category.Figure 2 illustrates the results of the proposed algorithm applied to the case of dynamic background presented as the Canoe sequence. In order to challenge our algorithm with ghost situation, we have considered the highway sequence with the 860^{th} frame as the initial frame. Figure 3 shows the obtained results compared to the adaptive mixture of Gaussian introduced by Zivkovic in [10]. Experiments show competitive results compared to existing approaches and demonstrate the applicability of the proposed framework in a variety of video surveillance scenarios. In the following, we will discuss the performance of the proposed scheme, and make a comparison with some existing methods and algorithms dealing with background subtraction.

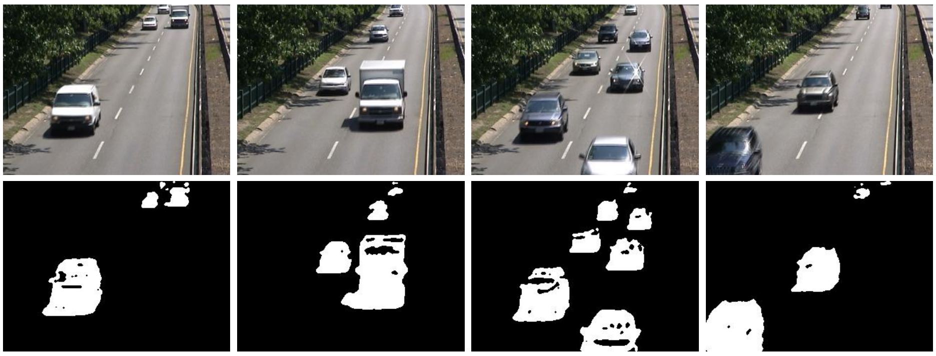

Figure 1. Proposed method for Baseline category: frames from the highway sequence.

5. EVALUATION AND DISCUSSION

In most of the literature, the performance of Background/Foreground classification is estimated by considering a reference as ground truth, and computing several metrics.The authors in [9] presented seven (07) metrics to assess the efficiency of motion segmentation algorithms, through the computation of TP, TN, FP, and FN that are respectively true positive count, true negative count, false positive count, and the false negative count. The seven metric presented in [9] are the following:

$$Recall(RE) = \frac{TP}{TP + FN} \tag{10}$$

$$Specicity(SP) = \frac{TN}{TN + FP} \tag{11}$$

$$False\ Positive\ Rate(FPR) = \frac{FP}{FP + TN} \tag{12}$$

$$False\ Negative\ Rate(FNR) = \frac{FN}{TN + FP} \tag{13}$$

$$Percentage\ of\ Wrong\ Classification(PWC) = 100.\frac{FN + FP}{TP + FN + FP + TN} \tag{14}$$

$$Precision(PR) = \frac{TP}{TP + FP} \tag{15}$$

$$F - measure = 2\frac{Pr.Re}{Pr + Re} \tag{16}$$

In order to evaluate the proposed scheme, we have compared the obtained results with those publicly available on the change detection website (Table 2). We have computed the seven

Table 1. The performance of the proposed algorithm.

Sequence	RE	SP	FPR	FNR	PWC	PR	F-Measure
Highway	0,9340	0,9918	0,0082	0,0660	1,1581	0,8783	0,9053
Office	0,8056	0,9989	0,0011	0,1944	1,4431	0,9821	0,8852
Pedestrian	0,8966	0,9977	0,0023	0,1034	0,3251	0,7977	0,8443
PETS 2006	0,8368	0,9966	0,0034	0,1632	0,5509	0,7625	0,7979
Canoe	0.8014	0.9957	0.0043	0.1986	0.0114	0.8778	0.8379
Over pass	0.7188	0.9969	0.0031	0.2812	0.0069	0.7584	0.7381

metrics stated earlier. The proposed framework achieved a percentage of missed classification less than 1% for the category of baseline, however it shows some lack of precision when tested with sequences including dynamic background. By analyzing the obtained results we believe that the proposed framework is very suitable for video based traffic monitoring systems.
The proposed method has shown robustness against ghost situation due to the conditional instantaneous update of the background model, and to the computation of the distance metric which involved decreasing weights from the BGM.

The table 1, shows the results obtained when applying the presented algorithm on the category of baseline with the high sequences. The proposed algorithm has some defects when considering the other categories (camera jitters, Dynamic background).

6. CONCLUSION AND FURTHER WORK

In this paper; we have presented a framework for background subtraction using an adaptive threshold presented as the mean square of successive differences for the background model samples, and a weighted distance computation for estimating the distance between incoming frames and the background model. In contrast with existing update strategies, we have introduced a weighted update policy to the background model based on associated weights to ensure the accuracy in the distance estimation step. The proposed framework achieved comparative results when considering surveillance application sequences and traffic monitoring systems videos. The presented framework deals with problem of ghost and noisy scene, as well as the gradual slow illumination change. In further work we seek to improve the presented approach to fit an extensive range of challenges: dynamic scene and bad weather situations.

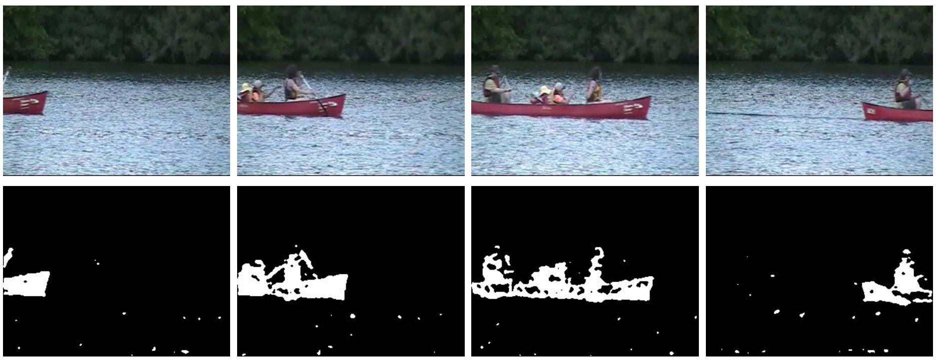

Figure 2. Proposed method for Dynamic Background Category: frames from the canoe sequence.

References

[1] Elgammal. A. et al., “Non-parametric model for background subtraction,” *In Computer VisionECCV 2000*, vol. -, pp. 751–767, 2000.

Table 2. Comparison between the proposed method and some of the existing background subtraction algorithms

Method	RE	SP	FPR	FNR	PWC	PR	F-Measure
Proposed	**0.8683**	0.9963	0.0037	**0.1317**	**0.8693**	0.8582	0.8551
Mahalanobis distance [8]	0.3154	0.9991	0.0009	0.6846	2.8698	0.4642	0.9270
GMM/Zivkovic [10]	0.8085	0.9972	0.0028	0.1915	1.3298	0.8382	0.8993
Euclidean distance [8]	0.8385	0.9955	0.0045	0.1615	1.0260	0.8720	0.9114

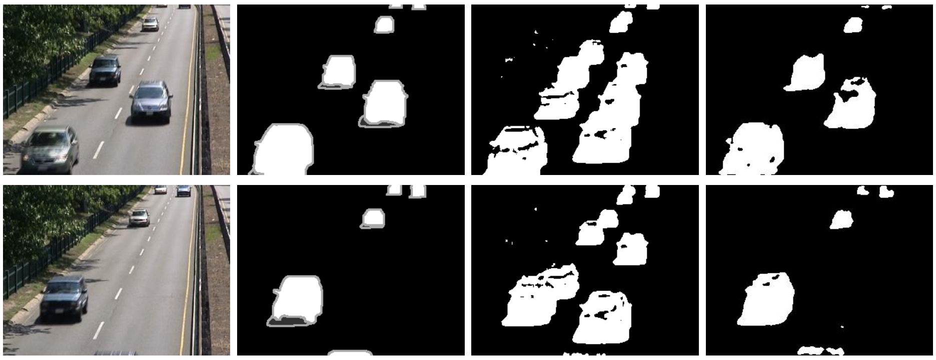

Figure 3. Robustness of the proposed algorithm to ghost sutuation. 1^{st} column shows the original,2^{nd}: The ground truth,3^{rd}: AGMM[10],and 4^{th}: The Proposed.

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

[3] S.-C. S. Cheung and C. Kamath, “Robust techniques for background subtraction in urban traffic video,” *In Proceedings of SPIE*, vol. 5308, pp. 881–892, 2004.

[4] Lacassagne. L. et al., “High performance motion detection: some trends toward new embedded architectures for vision systems,” *Journal of Real-Time Image Processing*, vol. 4-2, pp. 127–146, 2009.

[5] M. Van Droogenbroeck and O. Paquot, “Background subtraction: Experiments and improvements for vibe,” *In Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 32–37, 2012.

[6] M. Li, J. Fan, Y. Zhang, R. Zhang, X. Weijing, and H. Dingding, “A moving object detection algorithm based on multiple judgments,” *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 11, no. 10, pp. 5539–5544, 2013.

[7] Shisong et al., “Moving vehicle detection and tracking algorithm in traffic video,” *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 11, no. 06, pp. 3053–3059, 2013.

[8] B. Y. et al., “Comparative study of background subtraction algorithms,” *Journal of Electronic Imaging*, vol. 19-3, 2010.

[9] Goyette, N. et al., “Changedetection.net: A new change detection benchmark dataset,” *In-Computer Vision and Pattern Recognition Workshops (CVPRW)*, vol. -, pp. 1–8, 2012.

[10] Zivkovic, Z. “Improved adaptive gaussian mixture model for background subtraction,” *In Pattern Recognition, 2004. ICPR*, vol. 2, pp. 28–31, 2004.