A Unified Threshold Updating Strategy for Multivariate Gaussian Mixture Based Moving Object Detection

Akilan Thangarajah, Q. M. Jonathan Wu, and Jie Huo
Department of Electrical and Computer Engineering
University of Windsor
Windsor, Canada

Abstract— **Moving object detection is vitally used in video surveillance applications. Traditional Gaussian mixture model (GMM) based background subtraction (BGS) methods are usually performs well when background is stationary. However, they require parameter tuning to deal with dynamic backgrounds, whose background pixel values change over time. Particularly, the threshold which determines the pixels associated with moving objects from the resultant of BGS. To tackle this problem there is no ultimate solution. Considering that, this paper intents to present a novel idea to update the threshold of GMM based BGS with respect to color distortion, similarity and illumination measures in pixel level. Extensive experiments were carried out to demonstrate the effectiveness of the proposed method in comparison to some of the long-familiar GMM based BGS methods in literature. However, note that this paper is not attempted to provide a real-time technique, but rather to investigate the potential utilization of the aforementioned measures to set a threshold automatically to detect moving objects in video sequences.**

Keywords- **Background/ foreground classification; Moving object detection; Gaussian mixture model; Multivariate distribution**

I. INTRODUCTION

Moving object detection is a crucial process in computer vision applications mainly in video surveillance. This process ignores trivial information in a scene and raises attention to moving objects. In order to achieve this, over the past years, many algorithms have been proposed either based on a predictive or probabilistic mechanism [1] - [4]. For example, approaches that utilize filters like Wiener [5] come under the first category while the approaches which model the background (BG) based on probabilistic distributions like Gaussian [6] come under the second category. In general, all these approaches fall into three strategies namely pixel-, region-, and hybrid- based methods.

Among them the Gaussian distribution based approaches have received greater attraction since Stauffer and Grimson [6] proposed GMM for real-time tracking in a video surveillance. Due to its applicability, there have been several methods such as [7] - [11] proposed to improve the performance of [6]. Zhou et al. [7] introduces Markov random field (MRF) in an iterative process to refine the foreground (FG) with expectation maximization (EM). Mukherjee et al. [8] come up with a support weight mechanism (SWM) and histogram of gradients (HoGs) for better distance measurement and in [9] they use wavelet-based decomposition and a variable number of clustering technique to improve the GMM. Yang et al. [10] employ conditional random field (CRF) in spatial domain to support GMM based video segmentation. In [11], Lee attempts to incorporate incremental EM type of learning into a recursive filter such that parameter learning of each Gaussian follows a predefined schedule.

In all the standard GMM based FG detection algorithms the threshold which classifies the FG and BG is tuned exclusively for each video sequence [12]. This approach does not perform effectively, in times, even after the threshold is tuned for the particular video due to illumination changes or when objects appear with similar color as the BG. To address this issue this paper attempts to achieve a unified model which exploits pixel based color similarity and distortion measures along with illumination coefficient to update the threshold adaptively.

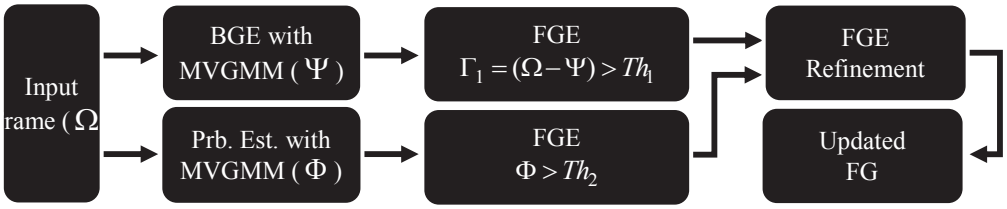
The rest of this paper is organized to provide details of the algorithm are described in Section II. The proposed method is applied for moving object detection on various datasets and the results are presented in Section III while conclusions are drawn in Section 4.

II. THE ALGORITHEM

A. Multivariate Gaussian Mixture Model

The Gaussian mixture model used in this work differs from the standard GMM introduced by Stauffer and Grimson [6] in the following ways: (i) It assumes that the each channel is independent with unequal variance value while [6] assumes that each channel is independent with same variance value, (ii) It uses a new distance measure based on Bhattacharyya and Mahalanobis distances in choosing appropriate Gaussian components which model the BG while [6] uses Mahalanobis distance measure for the same purpose, (iii) In the standard GMM based BGS, FG classification is based on a fixed threshold which has to be tuned for better performance exclusive to each video sequence. This is the same technique used in some of the improved versions such as in Yang et al. [10], Lee [11], and Mukherjee et al. [9]. Contrastingly, in this work the

978-1-5090-2088-1/16/$31.00 ©2016 IEEE



BGE – Background estimation, FGE – Foreground estimation, Prb. – Probability, Est. – Estimation

Figure 2. Processflow of the propossed algorithm.

threshold is automatically updated based on color distortion, color similarity, and illumination coefficient measures, and (iv) This work also process a FG refinement based on FG estimation directly from pixel wise posterior probability.

Block diagram shown in Fig.1 depicts the process flow of the proposed algorithm. In this work, Multivariate Gaussian Mixture Model (MVGMM) is utilized for estimating the BG and probability of each pixel to belong to the BG. The probability calculation of a multivariate Gaussian distribution is expressed as in (1).

$$p(x;\mu,\Sigma)=\frac{1}{(2\pi)^{n/2}}\frac{1}{|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}\Psi^{T}\Sigma^{-1}\Psi\right], \quad (1)$$

where x is a vector-valued random variable $X=[\mathrm{X}_1 \ldots X_n]^T$, mean vector $\mu \in R^n$, covariance matrix $\Sigma \in S_{++}^n$, and n is the dimension of the multivariate variable. The S_{++}^n is the space of symmetric positive definite $n \times n$ matrices and the coefficient $1/\left[(2\pi)^{n/2}\times|\Sigma|^{0.5}\right]$ is a normalizing factor, and $\Psi=(x-\mu)$. The multivariate variable in this case is pixel intensities correspond to each pixel values in RGB color space. As noted earlier channels of the color space are assumed to be independent and have unequal variance level. Thus, the covariance matrix of the color space become a diagonal matrix as in (2), which results an efficient determinant and inverse matrix computation of the covariance matrix.

$$\Sigma_{j,t}=\sigma_{j,t}^2.I, \quad (2)$$

where I is the identity matrix and j is the pixel index of rolled input frame at time t in a video sequence. Then, to determine the matching Gaussian distributions of each new pixel a new distance measure is used by utilizing parts of Bhattacharyya measure and Mahalanobis distance. It is because, among various distance measures such as Chi-squared, Mahalanobis, Euclidian, and Matusita, the Bhattacharyya measure is stable, unbiased, self-consistent, and applicable to any distribution [13]. In Bhattacharyya measure, similarity between two multivariate distributions is calculated as;

$$B=\frac{1}{8}(M_2-M_1)^{T}\left[\frac{\Sigma_1-\Sigma_2}{2}\right]^{-1}(M_2-M_1)+\frac{1}{2}\ln\frac{\Gamma}{\Upsilon}, \quad (3)$$

where M_i is the mean vector Σ_i is the covariance matrix of i^{th} distribution $\Gamma=\left|(\Sigma_1+\Sigma_2)/2\right|$ and $\Upsilon=\sqrt{|\Sigma_1||\Sigma_2|}$. Meanwhile, the Mahalanobis distance (D_M) of multivariate distribution is given by;

$$D_M^2=\sum_{i=1}^{n}\left[(\mathrm{x}_i-\mu_i)/\sigma_i\right]^2, \quad (4)$$

which is already part of probability calculation of multivariate GMM as in (1). Thus, to alleviate computational burden the Bhattacharyya measure in (3) is modified as in (5).

$$newB=\frac{1}{8}\sum_{i=1}^{n}\left[(\mathrm{x}_i-\mu_i)/\sigma_i\right]^2+\frac{1}{f}\ln\left(\sqrt{|\sigma_i|^2}\right), \quad (5)$$

where f is the number features involve for example, if a 3-D color space is used then $f=3$ and $\Sigma=\prod_{c=1}^{f}\sigma_c^2$. Introducing the logarithmic term of variance values provides better stability in the distance measure. Hence, accuracy in the BGS and FG detection are enhanced. Then, parameters of the multivariate Gaussian distributions are updated if the following condition is met.

$$if\ \sum_{f=1}^{d} newB_{j,t}<\gamma\sum_{f=1}^{d}\sigma_{j,t}, \quad (6)$$

where γ is a control value set to in the range of 2.0–2.5 and j is the pixel index of rolled input video frame at time t. Consequently, rest of the parameters are updated as suggested by Stauffer and Grimson [6];

$$\mu_t=(1-\beta)\mu_{t-1}+\beta X_t, \quad (7)$$

$$\sigma_t^2=(1-\beta)\sigma_{t-1}^2+\beta(X_t-\mu_t)^{T}(X_t-\mu_t), \quad (8)$$

$$\omega_t=(1-\alpha)\omega_{t-1}+\alpha, \quad (9)$$

where α and β are application dependent learning rates set to be in the range [0,1). For unmatched distributions means and variances remain unchanged while weight will be updated as $\omega_t = (1-\alpha)\omega_{t-1}$ i.e. reduced by the factor of $(1-\alpha)$. If a pixel does not match with any of the distributions, then the least weighted distribution is updated with: $\omega_t = \omega_{initial}$, σ_t^2 to a highest value, and $\mu_t = X_t$. Considering that changes in the BG is sparse, then the BG can be represented by the distributions associated with higher weights. Thus, once the distributions are ranked in descending order of the weight matrix, and the first L distributions satisfying the following precedent are selected to represent the background;

$$BG = \arg\min_{L} \left(\sum_{c=1}^{C} \omega_{t,c} > \mathrm{Th}_1 \right), \tag{10}$$

where Th_1 controls the minimum amount of data that form the BG. In the standard GMM based BGS algorithms, Th_1 is set to a fixed value throughout the whole sequence. This approach causes misclassification of pixels when the BG and FG pixels are similar in color. For instance, when the standard GMM model given in [6] is tuned for best performance for Watersurface video sequence[1], it fails to classify part of pixels inside the red rectangular region shown in Fig.2 as FG since BG pixels in that region are also quite similar to the FG color. To address this issue we consider there is no ultimate solution for a problem and explore a method which can utilize color distortion (∂), color similarity (ϖ), and illumination (ϑ) measures to set unique threshold for every pixel in the current frame. Thus, Th_1 is updated at pixel-level on the run as described in the following subsection.

B. *Updating Threshold*

Setting a fixed threshold to classify FG and BG causes misclassification due to color similarity, distortion, and illumination changes. In order to overcome this situation the following hypothesizes are taken.

i. The fixed threshold has to be lowered if a FG pixel in the current frame is very similar to BG pixel at the same coordinate, so that the pixel will be correctly classified as FG.
ii. Similarly, when a pixel experiences higher color distortion or illumination variance the threshold to classify the pixel to be FG has to be updated to account the changes.

Considering the aforementioned hypotheses, the fixed threshold Th_1 is modified as

$$\mathrm{Th}_1 = \mathrm{Th}_1 \times \aleph, \tag{11}$$

where $\aleph$ is a control variable which adjust the fixed threshold Th_1 which has an initial value of 60 to adopt the changes and it is calculated by (12). We achieve the mathematical derivations through empirical method to support the stated hypotheses.

$$\aleph = \frac{abs(\varpi - (\partial + \vartheta))}{(\partial + \vartheta)}. \tag{12}$$

The color similarity measure (ϖ) is calculated by (13) based on CIEDE2000 Color-Difference aka *ciede*00 formula as described in Sharma et al. [14]. The color similarity is calculated between the current frame and the previous frame. We would like to give credit to Sharma et al. for the source code of the color-difference calculation made publically available for researchers. Hence, the color distortion (∂) is measured by (14) as described in [15].

(a). Current Frame (b). Detected FG

Figure 2. Unclassifed FG Region.

$$\varpi_{j,t} = \frac{abs(ciede00_{j,t} - \mathrm{argMed}(ciede00)}{\mathrm{argStd}(ciede00)}, \tag{13}$$

where argMed() and argStd() are the operations to extract median and standard deviation values from the calculated *ciede*00. Note that lower value of *ciede*00 indicates that the two pixels are very similar in terms of appearance. So for that pixel considering the hypothesis I the threshold value has to be lowered so that chances of this pixel to be classified as FG increases. Keep in mind that (13) is derived empirically.

$$\partial_{j,t} = \sqrt{\left(\sum_{c=1,j,t}^{f} I_{c,j,t}\right)^2 - \left[\frac{\sum_{c=1,j,t}^{f} \left(I_{c,j,t} \times \mu_{c,j,t}\right)^2}{\sum_{c=1,j,t}^{f} \left(\mu_{c,j,t}\right)^2}\right]}, \tag{14}$$

where $I_{c,j,t}$ and $\mu_{c,j,t}$ are the intensity value and running mean respectively of a pixel j respect to time t and channel c. The illumination coefficient (ϑ) is defined by (15).

$$\vartheta_{j,t} = L^*_{j,t} / a^*_{j,t} + L^*_{j,t} / b^*_{j,t}, \tag{15}$$

where $L^*_{j,t}$, $a^*_{j,t}$, and $b^*_{j,t}$ are values of each channel in Lab color space of a pixel j at time t.

[1] http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html

C. Foreground Refinement

The detected FG is refined based on the following rules. First of all two different FG estimations are performed by (16) and (17).

$$\mathrm{FGE}_1 = abs(\mathrm{I}_t - \mathrm{B}_t) > \mathrm{Th}_1. \quad (16)$$

where I_t and B_t represent pixel values in the current frame and reconstructed BG by (10) respectively at time t and Th_1 is the threshold value defined by (11). At the same time, another FG estimation is carried as in (17) by thresholding the probability p calculated in (1).

$$\mathrm{FGE}_2 = p > ciede00 \times |\ln(\partial + 1/\vartheta)|, \quad (17)$$

where $ciede00$, ∂ and ϑ are color similarity, color distortion and illumination coefficient values, respectively. Then, to refine the FG outlier pixels (ζ) when compared FGE_1 and FGE_2 are taken to another FG validation process with lower threshold as;

$$\zeta = \mathrm{FGE}_1 \oplus \mathrm{FGE}_2, \quad (18)$$

where $\oplus$ represents a binary XOR operation.

$$\mathrm{FGE}_3 = abs(\mathrm{I}_{\zeta,t} - \mathrm{B}_{\zeta,t}) > \mathrm{Th}_{1,\zeta} \times \left(\frac{\aleph_\zeta}{ciede00_\zeta} \right). \quad (19)$$

Finally the validated FG is achieved by merging the FGs determined by (16) and (19) as

$$\mathrm{FG} = \mathrm{FGE}_1 \cup \mathrm{FGE}_3. \quad (20)$$

III. Experiments and Results

A. Nature of the Experiments

This section, demonstrates the performance of the proposed method on various datasets as a comparison to other algorithms: GMM [6], crfGMM [10], and Effective GMM (EGMM) [11]. A short description of the datasets used is given in Table I. The experiments were carried out with number of mixture components $k = 5$ while other parameters were tuned for best results. Processing time per frame (PTPF) is recorded based on MATLAB R2015a on Windows 8 64-bit, with i7-4770 CPU at 3.40 GHz PC. The proposed method utilizes a standard RGB to Lab color space conversion since the color similarity measure is based on Lab color space.

B. Visual Results

Visual results for key frames of the datasets are shown in Fig. 3 and 4, where columns from left to right show input frame in RGB, ground truth, and results from the proposed method, GMM, crfGMM, and EGMM respectively.

TABLE I. Description of the Datasets

Datasets (DS)	Sequence	Frame Size	Notes
(a). Wallflower	Waving Trees	120×160	Sawing tree in the BG as a person enters the scene.
(b). Complex Background	Water surface	128×160	Person walking in front of a water surface.
(c). Carnegie Mellon [16]	Camera motion	240×360	Man walks across as car enters the scene.
(d). Change Detection [17]	Pedestrian	240×360	Pedestrians and cyclist are crossing across.
(e). Change Detection [17]	Highway	240×320	Trees branches sway as vehicles move on.

a. http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/testimages.htm

b. http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html

C. Quantitative Analysis

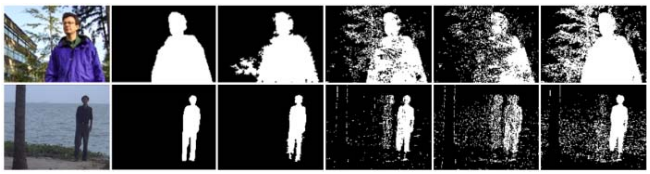


Figure 3. Visual comparisons of the results for frames no. 250 - 1st row and no. 559 - 2nd row form datasets (a) and (b), respectively.

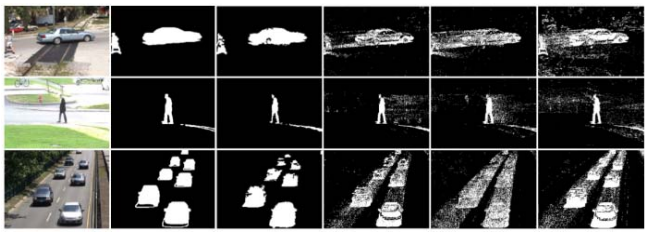


Figure 4. Visual comparisons of the results for frames 435 - 1st row, 626 - 2nd row, and 827 – 3rd row from datasets (c) - (e) respectively.

The quantitative analysis exploit evaluation matrices described in [17] and [18]. If TP, FP, and FN denote true positive, false positive and false negative respectively in the detected FG and BG. Then, the performance matrices can be expressed by Recall, Precision, and Figure of Merit (FoM). The recall given by (21) measures the percentage of predicted TP as compared to the total number of actual positives in the ground truth. The precision, which measures the percentage of correct detection as compared to the total number of detection as positives defined by (22). Precision and recall are totally different perspective measures of the performance. Thus, a weighted harmonic mean measure jointly with recall and precision called FoM (23) provides a better performance evaluation.

$$Recall = TP/(TP + FN). \quad (21)$$

$$Precision = TP/(TP + FP). \quad (22)$$

$$FoM = (2 \times Recall \times Precision)/(Recall + Precision). \quad (23)$$

Table II tabulate the performance evaluation of the proposed and other algorithms, where the bold face values represent the best performances. According to the visual and quantitative comparison the proposed algorithm produces competitive results. It is because the standard GMM and its improved versions such as EGMM and crfGMM do not focus on automatically setting a threshold based on pixel level variations presently occur in terms of color and illumination. On the hand, crfGMM takes greater processing time due to repeated neighborhood computations. Similarly, the EGMM also considerable computational cost since it employs an incremental EM based learning recursively to update parameter of each Gaussian. In our case, the computational cost higher than the GMM and EGMM due to validation process.

TABLE II. PERFORMANCE COMPARSION ON VARIOUS DATASETS

DS	Algorithm	Recall	Pre.	FoM	PTPF (ms)
(a)	Proposed	0.9252	0.9245	**0.9248**	114.56
	GMM	0.8042	0.3884	0.5238	**91.00**
	EGMM	0.9647	0.3033	0.4615	106.20
	crfGMM	0.6713	0.3906	0.4939	597.00
(b)	Proposed	0.9159	0.9114	**0.9137**	111.80
	GMM	0.7611	0.7683	0.7647	**65.60**
	EGMM	0.8294	0.5002	0.6241	129.10
	crfGMM	0.6798	0.4699	0.5557	782.60
(c)	Proposed	0.9133	0.9264	**0.9198**	429.90
	GMM	0.6795	0.5215	0.5901	**258.60**
	EGMM	0.8233	0.2812	0.4192	362.20
	crfGMM	0.7419	0.4649	0.5717	1399.40
(d)	Proposed	0.7237	0.9182	**0.8094**	431.10
	GMM	0.6420	0.5042	0.5648	**298.40**
	EGMM	0.7190	0.2804	0.4034	368.4
	crfGMM	0.6919	0.3742	0.4857	1530.8
(e)	Proposed	0.5890	0.7829	**0.6723**	403.90
	GMM	0.6531	0.5549	0.6000	**267.16**
	EGMM	0.7928	0.4428	0.5682	332.42
	crfGMM	0.6143	0.4524	0.5211	1409.92

CONCLUSION

The proposed algorithm exploits color distortion, color similarity, and illumination variation measures to derive a mathematical expression to update pixel based threshold per frame which detects moving objects through multivariate GMM based BG subtraction. The results proved that the taken hypothesizes and the empirical method used to drive the threshold automatically are valid for the datasets used in the experiments. Hence, the performance evaluation shows that the proposed algorithm more robust than the other compared algorithms in terms of FoM. However, the standard GMM takes lesser processing time per frame compared to the proposed algorithm.

Future work can be dedicated to further exploring robust methods to utilize color distortion and similarity measures along illumination measures in region and global level to facilitate accurate foreground extraction.

ACKNOWLEDGMENT

The authors would like to thank to the reviewers for their helpful comments and suggestions on earlier version of the manuscript.

REFERENCES

[1] A. Pal, G. Schaefer and M. Celebi, "Robust codebook-based video background subtraction," *IEEE Int. Conf. on Acoust. Spee. Sig. Process.*, pp. 1146 - 1149, March 2010.

[2] S.-C. Wang, T.-F. Su and S.-H. Lai, "Detecting moving objects from dynamic background with shadow removal," *IEEE Int. Conf. on Acoust. Spee. Sig. Process.*, pp. 925 - 928, May 2011.

[3] H. Mansour and A. Vetro, "Video background subtraction using semi-supervised robust matrix completion," *IEEE Int. Conf. on Acoust. Spee. Sig. Process.*, pp. 6528 - 6532, May 2014.

[4] G. Warnell, D. Reddy and R. Chellappa, "Adaptive rate compressive sensing for background subtraction," *IEEE Int. Conf. on Acoust. Spee. Sig. Process.*, pp. 1477 - 1480, March 2012.

[5] K. Toyama, J. Krumm, B. Brumitt and B. Meyers, "Wallflower: Principles and practice of background maintenance," *Inter. Conf. on Computer Vision*, pp. 255–261, 1999.

[6] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," *IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, pp. 246-252, 1999.

[7] L. Zhou, H. Kaiqi and T. Tieniu, "Foreground object detection using top-down information based on EM framework," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4204-4217, September 2012.

[8] D. Mukherjee, Q. Wu and M.-N. Thanh, "Gaussian mixture model with advanced distance measure based on support weights and histogram of gradients for background suppression," *IEEE Trans. Industrial Info.*, vol. 10, no. 2, pp. 1086-1096, May 2014.

[9] D. Mukherjee, Q. Wu and T. Nguyen, "Multiresolution based gaussian mixture model for background suppression," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5022-5035, October 2013.

[10] W. Yang, L. Kia-Fock and W. Jian-Kang, "A dynamic conditional random field model for foreground and shadow segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 279-289, December 2006.

[11] D.-S. Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.* , vol. 27, no. 5, pp. 827 - 832, May 2005.

[12] T. Akilan, Q. M. Jonathan Wu, A.K. Singh, B. Mandon and A.K. Chowdhury, "Video foreground detection in non-static background using multidimensional color space," *in Proc. of the 4th Inter. Conf. on Eco-friendly Comput. Communi. Sys.*, vol. 70, pp. 55–61, 2015.

[13] N. Thacker, F. Aherne and P. Rockett, "The bhattacharyya metric as an absolute similarity measure for frequency coded data," *Kybernetika*, vol. 34, no. 4, pp. 9 - 11, June 1997.

[14] G. Sharma, W. Wu and E. Dalal, "The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations," *Color Research and Application*, vol. 30, no. 1, pp. 21-30, February 2005.

[15] A.H.Doli, S.P.Anton and A.Azizi, "Codebook model for real time robot soccer recognition: a comparative study," *FIRA RoboWorld Congress*, Taiwan, 2011.

[16] S. Yaser-Ajmal and M. Shah, "Bayesian modelling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 27, no. 11, pp. 1778 - 1792, November 2005.

[17] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth and P. Ishwar, "CDnet 2014: an expanded change detection benchmark dataset," *in Proc. IEEE Workshop on Change Detection*, pp. 387-394, 2014.

[18] S. Brutzer, B. Hoferlin and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," *IEEE Conference on Comput. Vis. Patt. Recognit.*, pp. 1937 - 1944, June 2011.