Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 70 (2015) 55 – 61

Procedia Computer Science

4th International Conference on Eco-friendly Computing and Communication Systems, ICECCS 2015

Video foreground detection in non-static background using multi-dimensional color space

Akilan Thangarajah[a]*, Q.M. Jonathan Wu[a], A.K. Singh[b], B. Mandon[c], A.K. Chowdhury[c]

[a]*University of Windsor, Windsor, Canada*

[b]*National Institute of Technology, Kurukshsetra, India*

[c]*University College of Technology Sarawk, Malaysia*

Abstract

Detecting foreground (FG) and suppressing background (BG) is a vital task in video sequence analysis. This task is very challenging when the BG is non-static. Although there have been many algorithms proposed in literature, most of them are complex in terms of either mathematical modeling or computational requirements. In this paper, we experiment two simple algorithms for video FG detection using multi-dimensional color space when the BG is non-static. The algorithms utilize pixel level temporal intensity for FG and BG classification. The algorithms are tested on two sets of outdoor video sequences where the backgrounds are non-static. The experiment results show that the algorithms adequately perform well on the given environments.

Peer-review under responsibility of the Organizing Committee of ICECCS 2015

Keywords: Foreground detection, Non-static background, Background suppression/subtraction, Color models/spaces, Probability distribution, Euclidean distance, Video surveillance.

1. Introduction

Over the past decade utilizing video surveillance systems has largely increased since it is a way of eco-friendly approach in comparison to using multiple electromagnetic (EM) wave based sensors and sensor networks for surveillance purposes. As a consequence, researchers in computer vision related fields proposed many robust BG

* Corresponding author.
E-mail address: thangara@uwindsor.ca

Peer-review under responsibility of the Organizing Committee of ICECCS 2015
doi:10.1016/j.procs.2015.10.030

suppression/subtraction models for real-time FG detection[1-9]. For an automated video surveillance system BG suppression plays an inevitable role. BG subtraction fundamentally helps the system to ignore unwanted area of a scene being monitored and bring the attention to moving objects, for instance, a moving car or a walking man. Thus, important foreground information can be extracted for further processing such as traffic monitoring (vehicle detection, counting, and tracking), human activity recognition (run, walk, jump, squat, etc.), human-machine interaction or interface (HMI), moving object tracking (many live sports telecasting channels adopts this) and so forth.

A reliable BG subtraction algorithm should be robust and able to handle sudden or gradual illumination changes, high frequency moving objects, repetitive motion in the background (such as tree leaves, flags, waves, etc.) and long-term scene changes (a car is parked for a month, for instance). Thus, there were many algorithms proposed fundamentally based on Gaussian mixture models (GMM) by researches over the past two decades since the pioneer work reported by Stauffer and Grimson[10]. For instance, Effective GMM[11], GMM based conditional random filed[12,13], Variational clustered GMM[14], and Wavelet transformation based GMM[15]. However, such high complexity algorithms are not necessary for certain surveillance purposes such as monitoring an automatic teller machine (ATM) in a shopping complex or bank. Because, in such cases the surveillance camera is fixed at a place and the background environment is known prior to monitoring. In such conditions, it is recommended to employ simplistic models to detect foregrounds i.e. the moving objects in the given environment being monitored. This paper presents two simplistic algorithms: probabilistic based model with non-supervised thresholding and 3D color space model using distance vector for background suppression. The contribution of this paper is twofold; proposing simple techniques for non-static background suppression for a constrained environment and reporting their experiment results and limitations. The remainder of this paper is organized as follow: Section 2 describes the algorithms in detail; Section 3 presents the experiment results, and Section 4 concludes the paper with discussions and recommendations for future work.

Nomenclature

3D	3-dimension
BG	Background
FG	Foreground
FN	False negative
FoM	Figure of merit
FP	False positive
GMM	Gaussian mixture model
PMF	Probability mass function
TN	True negative
TP	True positive

2. The Algorithm

2.1. Probabilistic based background suppression with non-supervised thresholding

For a known environment its BG can be modelled probabilistically based on some collected samples prior to actual monitoring. Every pixel in the scene will have its own probability mass function, $PMF_{N,c}(I_{u,v})$ with intensity ranging from 0 to 2^n where, N is number of samples collected to estimate the BG, the channel parameter $c \in \mathbb{R}^D$. In the case of RGB colour space, $c \in \mathbb{R}^3 \equiv \{Red, Green, Blue\}$, $I_{u,v}$ is the pixel intensity value at image coordinate (u, v) and n is number of bits used to represent the intensity values of each channel. Figure 1 describes this concept where, a scene is monitored over the time $t = 1$ to $t = N$.Then, PMF of the pixel at coordinate (u, v) is calculated for the first channel, for instance, red channel followed by for all other channels, for instance, blue and green channels. Similarly, the PMF for all other pixels in the scene respect to all the channels, also, can be calculated easily. Once the probabilistic based BG model of each pixel respect to concerned channels is calculated, then total likelihood probability of an incoming new pixel $\overline{I_{u,v}}$ to be the BG can be calculated as in (1).

$$P\left(\overline{I_{u,v}}\right) = \prod_{c=1}^{D} P\left(I_{u,v} | \overline{I_{u,v}}\right) \tag{1}$$

Where, the conditional probability of the pixel is $P\left(I_{u,v}|\overline{I_{u,v}}\right)$ while D is the dimension of the colour model and c is the channel index. If it is RGB colour model then, $D = 3$ and $C \equiv \{1 = Red, 2 = Green, 3 = Blue\}$. Now the new pixel $\overline{I_{u,v}}$ can be classified as either FG or BG based on the condition given in (2).

$$FG_{u,v} = P\left(\overline{I_{u,v}}\right) < Ts_{u,v} \tag{2}$$

Where, Ts the threshold specific to the pixel at (u, v) can be estimated in an unsupervised manner from the prior BG probabilities of the pixel respect to each channel as given in (3). The value, Ts determines the minimum prior probability of the pixel to be in the BG. Thus, if a new pixel has a likelihood probability, $P\left(\overline{I_{u,v}}\right) \geq Ts_{u,v}$ then $I_{u,v} \in BG$ else $I_{u,v} \in FG$.

$$Ts_{u,v} = argmin\left(P_R\left(I_{u,v}\right), P_G\left(I_{u,v}\right), P_B\left(I_{u,v}\right)\right) \tag{3}$$

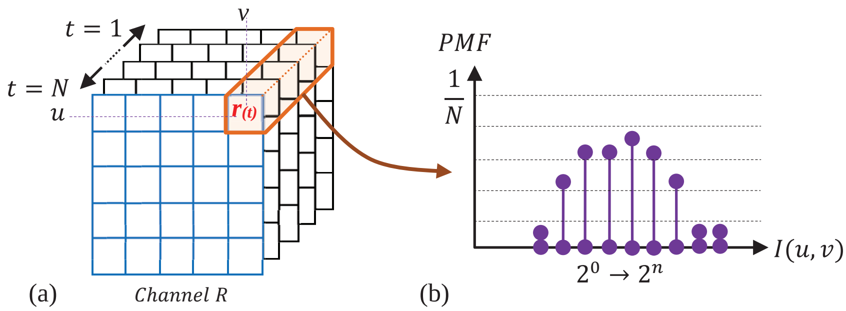

Fig. 1. (a) Extracted red channel of a scene over a period N and (b) PMF of a pixel at (u, v) observed over the same period N.

2.2. Euclidean distance based background suppression in 3D color spaces

A point in 3D space can be, uniquely, represented as shown in Fig. 2, where $P(x, y, z)$ and $P'(x', y', z')$ are two points located at the coordinates (x, y, z) and (x', y', z'), respectively.

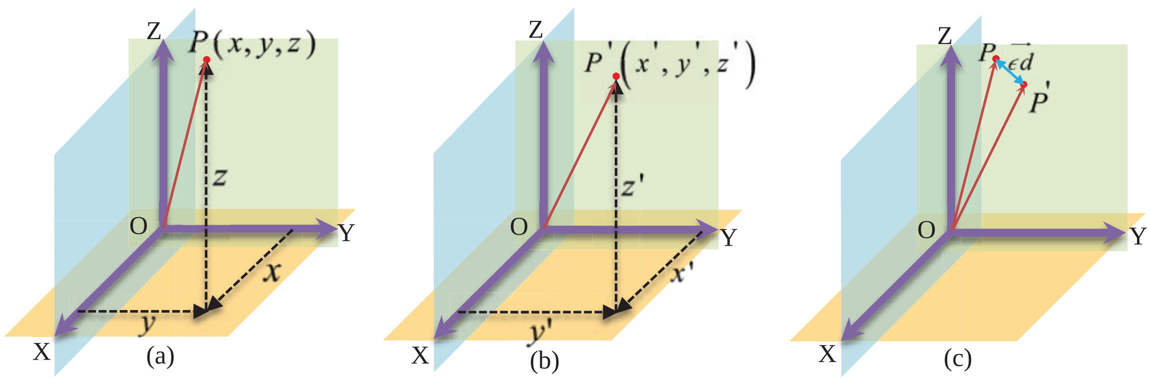

Fig. 2. Geometrical representation of (a). Point P, (b). Point P′, and (c) The distance between P and P′.

In Fig. 2 (a), $\overrightarrow{OP}$ is the distance vector of the point, P from the origin, O and its magnitude is measured as in (4).

$$|\overrightarrow{OP}| = \sqrt{x^2 + y^2 + z^2} \tag{4}$$

Similarly, magnitude of the distance vector, $\overrightarrow{OP'}$ is calculated as $|\overrightarrow{OP'}| = \sqrt{x'^2 + y'^2 + z'^2}$ by using (4). Then if P and P' refer the same point in the space, the Euclidean distance (ϵd) between them will be a null vector, $\vec{0}$. The Euclidean distance between any two points in a 3D space is measured as in (5).

$$|\overrightarrow{\epsilon d}| = \sqrt{(\mathrm{x} - x')^2 + (\mathrm{y} - y')^2 + (\mathrm{z} - \mathrm{z}')^2} = \delta \tag{5}$$

Where, $\delta \to 0$ as P and P' come closer. Thus, to check if two points refer the same location in the space (5) can be used with a preferred precision (against a threshold value). This simple mathematical model can be used to classify background and foreground pixels in a scene if the scene is represented in a 3D-color space, for instance, RGB, YCbCr, YUV, or YIQ. The reader may refer Douglas[16], Gonzalez[17], and Nixon[18] for detail of various colour models.

Generally, in color spaces, the type of lights can be distinguished wholly in terms of human perception of colors; and each color has two technical aspects: (i) Luminance, the indication of brightness of light, (ii) Chromaticity, the property that distinguishes red from blue and red from pink[16]. Thus, color spaces facilitate the specification of colors in some widely accepted standards. Technically, a color space is a specification of a coordinate system and subspace within that system, where each color is represented by a single point[17]. There are numerous color models in use today due to the fact that color science is a broad field that encompasses many areas of application. In this paper, some of the interesting ones, namely, RGB, YCbCr, YIQ, and YUV are used. In the case of RGB color space, to eliminate illumination variance in the frames the model in (6)[19] is employed. The illumination variance generally occur due to exposure of camera rapidly changes while scene capturing process. It will cause misclassification of significant number of image pixels.

$$\left[\hat{R}, \hat{G}, \hat{B}\right]^T = \left[\left((R-G)\big/\sqrt{2}\right), \left((R+G-2B)\big/\sqrt{6}\right), \left((R+G+B)\big/\sqrt{3}\right)\right]^T \tag{6}$$

In this algorithm, from a set of collected frames prior to actual monitoring task a generalized BG model is formed for each channel by using median filtering at pixel level as in (7).

$$I_{C,\psi}(u,v) = \arg median\left(I_C(u,v,k)\right) \tag{7}$$

Where, $k = 1, 2, \ldots N$, the sequence of samples of a pixel at image coordinate (u, v) of the channel c and ψ denotes the generalized background model of the channel. If $M_{c,\psi} = argmedian(I_{c,k})$, then the overall generalized background model will be $BG = [M_{R,\psi}, M_{G,\psi}, M_{B,\psi}]$. Then an incoming new pixel at image coordinate (u, v) can be classified as foreground if it satisfies the condition in (8) which is similar to (5).

$$\left|\overrightarrow{\epsilon d}_{(u,v)}\right| = \sqrt{\left[M_{R,\psi,(u,v)} - \mathrm{I}_{R,(u,v)}\right]^2 + \left[M_{G,\psi,(u,v)} - \mathrm{I}_{G,(u,v)}\right]^2 + \left[M_{B,\psi,(u,v)} - \mathrm{I}_{B,(u,v)}\right]^2} < \delta_{u,v} \tag{8}$$

Where, the pixel level threshold $\delta_{u,v}$ can be determined as in (9) from Euclidian distances of each pixel between the first frame and the subsequence frames till the last one collected in modelling the generalized BG in (7).

$$\delta_{(u,v)} = \arg mean\left[|\overrightarrow{\epsilon d}_{(u,v,1,2)}|, |\overrightarrow{\epsilon d}_{(u,v,2,3)}|, \ldots, |\overrightarrow{\epsilon d}_{(u,v,k-1,k)}|\right] \tag{9}$$

3. Experiments and Results

This section provides experimental results of the presented algorithms for two different video sequences in four different colour spaces: RGB, YCbCr, YIQ, and YUV. The *WavingTree* video sequence has 287 frames taken from Wallflower[19]. In this sequence, there is a tree swaying continuously which causes a non-static background while a man entering the scene at frame no. 242 from right and leaving to left at frame no. 260. The *WaterSurface* video sequence is taken from perceptual computing datasets[21]. It has 560 frames of a waving sea at the background while a man entering the scene at frame no. 483 and staying at the middle of the scene for a while. In both the cases, the background is non-static. Figures 3 and 4 show visual outputs of the algorithms: Algorithm I – simple probabilistic and Algorithm II – Euclidian distance based with respect to hand segmented ground truths of notable few frames. The results also provide a comparison with two GMM based background subtraction models; the standard GMM model introduced by Stauffer and Grimson[10] and the GMM model with wavelet transformation proposed by Mukherjee et al.[15]. These two models are denoted as oriGMM and wavGMM in short here after. Note that, wavGMM perfoms on grayscale input data only.

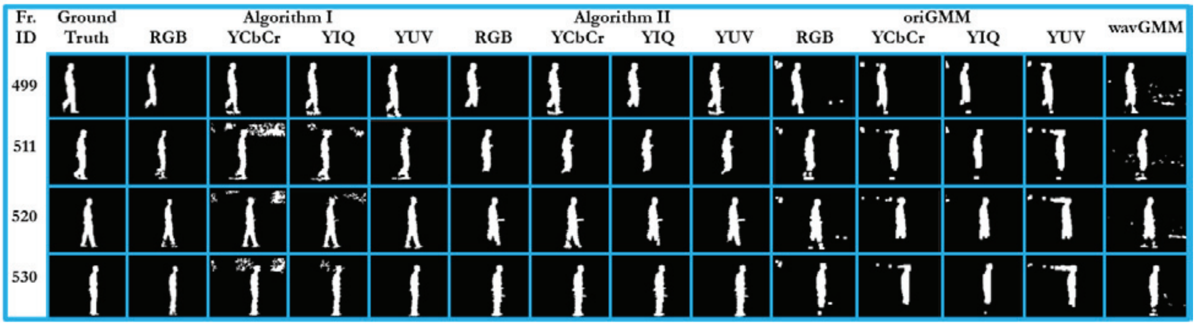

Fig. 3. Results (detected FG region) of the algorithms for WaterSurface dataset.

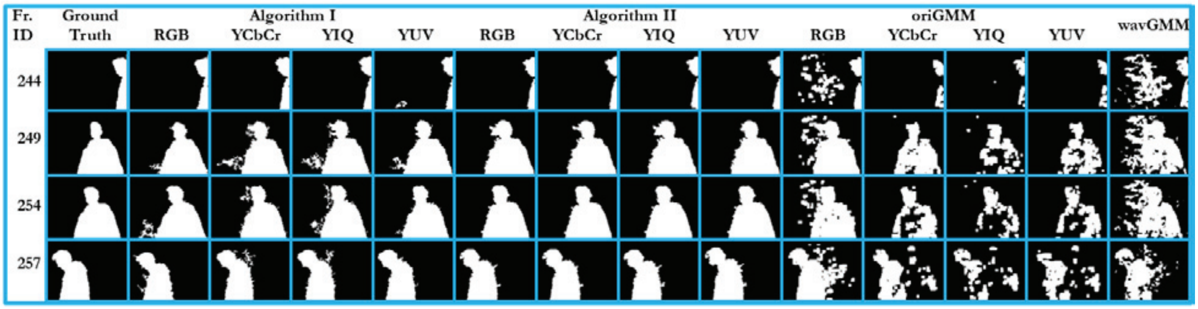

Fig. 4. Results (detected FG region) of the algorithms for WavingTree dataset.

To analyse the performance of the proposed algorithms and to compare it with other algorithms stated earlier four different measures are taken, namely Accuracy, Recall, Precision, and Figure of merit (FoM). All these measures are taken as average over the number of available ground truths for each dataset. To understand the aforementioned measures confusion matrix in Table 1 is used, where TP, TN, FP, and FN denote true positive, true negative, false positive and false negative, respectively.

Table 1: Confusion Matrix.

		Predicted Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

Recall or detection rate given by (10) measures the percentage of predicted true positives as compared to the total number of actual positives in the ground truth.

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

Accuracy is the measure of correct prediction as compare to the total number predictions defined by (11).

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \tag{11}$$

Recall or accuracy alone do not suffice for a valid comparison of different methods, thus it is useful to consider the precision (known as positive prediction), which measures the percentage of correct detection as compared to the total number of detection as positives defined by (12).

$$Precison = TP / (TP + FP) \tag{12}$$

All the above measures are good enough to compare different methods. However, a weighted harmonic mean measure with recall and precision called FoM as in (13) provides a better comparison.

$$FoM = (2 \times Recall \times Precision) / (Recall + Precision) \tag{13}$$

Following tables tabulate the performance of the proposed and other algorithms based on the above measures.

Table 2: Performance comparison of the algorithms for WaterSurface dataset.

	Algorithm I				**Algorithm II**				**OriGMM**				**wavGMM**
	RGB	YCbCr	YIQ	YUV	RGB	YCbCr	YIQ	YUV	RGB	YCbCr	YIQ	YUV	Gray
Accuracy	0.9815	0.9527	0.9743	**0.9870**	0.9836	0.9849	0.9831	0.9842	0.9710	0.9744	0.9766	0.9653	0.9781
Recall	0.7584	0.3774	0.6652	0.8852	0.9233	0.9526	0.9147	0.9377	0.8890	0.8045	0.7768	0.7989	0.8711
Precision	0.9887	0.9736	0.9829	0.8952	0.8641	0.8579	0.8651	0.8612	0.7644	0.8367	0.8880	0.7463	0.8387
FoM	0.8584	0.5439	0.7934	0.8902	0.8927	**0.9028**	0.8892	0.8978	0.8220	0.8204	0.8288	0.7717	0.8546

Table 3: Performance comparison of the algorithms for WavingTree dataset.

	Algorithm I				**Algorithm II**				**OriGMM**				**wavGMM**
	RGB	YCbCr	YIQ	YUV	RGB	YCbCr	YIQ	YUV	RGB	YCbCr	YIQ	YUV	Gray
Accuracy	0.9810	0.8097	0.9366	0.9830	0.9857	0.9850	0.9832	**0.9858**	0.8970	0.9185	0.8440	0.8844	0.8330
Recall	0.9466	0.2754	0.6428	0.9701	0.9801	0.9838	0.9707	0.9782	0.9691	0.7781	0.4934	0.6412	0.8004
Precision	0.9545	0.9774	0.9804	0.9186	0.9615	0.9565	0.9614	0.9628	0.7150	0.9002	0.8705	0.8920	0.6497
FoM	0.9505	0.4297	0.7765	0.9436	**0.9707**	0.9699	0.9660	0.9704	0.8230	0.8347	0.6301	0.746	0.7172

4. Discussion and Conclusion

The results show that the simple algorithms perform well on the tested two non-static BG video sequences in comparison to the two GMM based algorithms. Although the algorithms find required thresholds which determine the foregrounds without any fixed parameters i.e. in a non-supervised manner, they are not robust enough for various scenarios. For instance, video streams which do not have set of initialization frames prior to actual monitoring task or the video inputs have congested moving objects in the scene. It is because performance of the algorithms largely depends on the collected samples which estimate each pixel value in the BG either by probability or Euclidian distance map. These models also do not have the ability to update themselves on-line to adopt changes in the background scene. In contrast, the GMM based models has the ability to absorb the scene progress since they model each pixel values as a mixture of k number of Gaussian distributions. Then, based on persistence and variance of each distribution $m|m<k$ distributions are chosen to represent the BG. However, in such models there are application depended parameters such as learning rates, threshold level which decides the amount of scene to be fall into BG have to be tuned for the best

performance. Therefore, the simplicity of proposed algorithms is useful for indoor surveillance systems where the cameras are stably fixed and used to monitor a predetermined zone. On the other hand, effect of various colour spaces is found to be useful in these BG subtraction techniques, but there is no significant improvement found in comparison to the results from RGB colour space.

As for future improvement; (i). a weight parameter can be introduced to control the prior probabilities of each intensity level at each pixel. For instance, if an intensity value is classified as FG in the current frame it will have less probable to be in BG at the same pixel coordinate, so its weightage in the BG prior probability can be set to lower than its initial value. By doing so the model will have the ability to adopt a new BG intensity level and to remove an old intensity value which becomes least probable as the scene evolve, (ii). The Euclidean distance based algorithm can be improved by taking variance information of each channel into the distance calculation like in Mahalanobis distance so that it will be able to adopt rapid scene changes.

Acknowledgements

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

References

1. Y. Min-Hsiang, H. Chun-Rong, L. Wan-Chen, L. Shu-Zhe and C. Kun-Ta, "Binary descriptor based nonparametric background modeling for foreground extraction by using detection theory," *IEEE Trans. on Circu. Syst. Video Technol.,* vol. 25, no. 4, pp. 595 - 608, 2015.
2. J. L.-R. Francisco and L.-R. Ezequiel, "Features for stochastic approximation based foreground detection," in *Compu. Vis. Image Understanding,* vol. 133, pp. 30-50, 2015.
3. Y. Jian and O. Jean-Marc, "Multi-Layer background subtraction based on color and texture," in *Compu. Vis. Pattern Recongnit.-Visual Surveill.,* Minneapolis, 2007.
4. C. Shengyong, Z. Y. L. Jianhua and Z. Jianwei, "A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction," *IEEE Trans. on Industr. Informat.,* vol. 8, no. 1, pp. 118 - 127, 2012.
5. H. Bohyung and S. D. Larry, "Density-Based multifeature background subtraction with support vector machine," *IEEE Trans. on Pattern Anal. Mach. Intell.,* vol. 34, no. 5, pp. 1017-1023, 2012.
6. K. Wonjun and K. Changick, "Background subtraction for dynamic texture scenes using fuzzy color histograms," in *IEEE Sig. Process. Letters,* pp. 127-130, 02 January 2012.
7. T. Yonghong, W. Yaowei, H. Zhipeng and H. Tiejun, "Selective eigenbackground for background modeling and subtraction in crowded scenes," *IEEE Trans. on Circu. Syst. Video Technol.,* vol. 23, no. 11, pp. 1849-1864, 2013.
8. M. Haque, Gippsland and M. Murshed, "Perception-Inspired background subtraction," *IEEE Trans. on Circu. Syst. Video Technol.,* vol. 23, no. 12, pp. 2127 - 2140, 2013.
9. A. R.-Q. Juan and I. C.-M. Mario, "Self-adaptive SOM-CNN neural system for dynamic object detection in normal and complex scenarios," in *Pattern Recognit.,* vol. 48, no. 4, pp. 1133-1145, 2015.
10. C. Stauffer and W. Grimson, "Adaptive background mixture models for realtime tracking," in *Proc. IEEE Conf. on Comput. Vis. Pattern Recognit.,* vol. 2, pp. 246-252, 1999.
11. D.-S Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 27, no. 5, p. 827–832, May 2005.
12. Y. Wang, K. Loe and J. Wu, "“A dynamic conditional random field model for foreground and shadow segmentation," *IEEE Trans. on Pattern Anal. Mach. Intell.,* vol. 28, no. 2, p. 279–289, 2006.
13. Y. Wang, "Foreground and shadow detection based on conditional random field," in *Proc. Int. Conf. on Comput. Anal. Images Pattern,* p. 85–92, 2007.
14. Z. Chen and T. Ellis, "Self-adaptive Gaussian mixture model for urban traffic monitoring system," in *Proc. IEEE Int. Conf. on Comput. Vis. Workshops,* p. 1769–1776, 2011.
15. D. Mukherjee, Q. Wu and T. Nguyen, "Multiresolution based Gaussian mixture model for background suppression," *IEEE Trans. on Image Process.,* vol. 22, no. 12, pp. 5022-5035, October 2013.
16. A. K. Douglas, "The CIE XYZ and xyY color spaces," no. 1, March 2010.
17. R. C. Gonzalez and R. E. Woods, Digital Image Processing, New Jersey: Pearson, 2008.
18. M. Nixon and A. Aguado, Feature Extraction & Image Processing for Computer Vision, London: Elsevier, 2012.
19. E. A. V. Koen, G. Theo and G. S. Cees, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582-1596, 2010.
20. T. Kentaro, K. John, B. Barry and M. Brian, "Wallflower: Principles and practice of background maintenance," in *Seventh Inter. Conf. on Comput. Vis.,* Kerkyra, 1999.
21. "Statistical modeling of complex background for foreground object detection," [Online]. Available: http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html. [Accessed 22 May 2015].