RESEARCH ARTICLE OPEN ACCESS

Complex Background Subtraction Using Kalman Filter

Mr G.Sekar[1], M.Deepika[2]
Assistant Proffesor[1] Adhiparasakthi Engineering College Melmaruvathur-603319.
II ME CSE[2] Adhiparasakthi Engineering College Melmaruvathur-603319.

Abstract
Background subtraction from dynamic background, At any location of the scene, this system extract a sequence of regular video bricks, i.e., video volumes spanning over both spatial and temporal domain. The background modeling is thus posed as pursuing subspaces within the video bricks while adapting the scene variations. For each sequence of video bricks, it pursues the subspace by employing the auto regressive moving average model that jointly characterizes the appearance consistency and temporal coherence of the observations. During online processing, it use tracking algorithm kalman's filter for background/foreground classification and incrementally update the subspaces to cope with disturbances from foreground objects and scene changes.
Index Terms— Background modeling, visual surveillance, spatio-temporal representation.

I. INTRODUCTION

BACKGROUND subtraction (also referred as foreground extraction) has been extensively studied in decades,yet it still remains open in real surveillance applications due to the following challenges:

- Dynamic backgrounds. A scene environment is not always static but sometimes highly dynamic, *e.g.*, rippling water, heavy rain and camera jitter.
- Lighting and illumination variations, particularly with sudden changes.
- Indistinct foreground objects having similar appearances with surrounding backgrounds.

In this paper, we address the above mentioned difficulties by building the background models with the online pursuit of spatio-temporal models and kalman filter. Some results generated by system for the challenging scenarios are exhibited . Prior to unfolding the proposed approach, we first review the existing works in literature.

II. RELATED WORK

Due to their pervasiveness in various applications, there is no unique categorization on the existing works of background subtraction. Here we introduce the related methods basically according to their representations, to distinguish with our approach.

The **pixel-processing** approaches modeled observed scenes as a set of independent pixel processes, and they were widely applied in video surveillance applications [6], [7] . In these methods [1], [2], [8], [9], each pixel in the scene can be described by different parametric distributions (*e.g.* Gaussian Mixture Models) to temporally adapt to the environment changes. The parametric models, however, were not always compatible with real complex data, as they were defined based upon some underlying assumptions. To overcome this problem, some other non-parametric estimations [10]–[13] were proposed, and effectively improved the robustness. For example, Barnich et al. [13] presented a sample-based classification model that maintained a fixed number of samples for each pixel and classified a new observation as background when it matched with a predefined number of samples. Liao et al. [14] recently employed the kernel density estimation (KDE) technique to capture pixel-level variations. Some distinct scene variations, *i.e.* illumination changes and shadows, can be explicitly alleviated by introducing the extra estimations [15]. Guyon et al. [16] proposed to utilize the low rank matrix decomposition for background modeling, where the foreground objects constituted the correlated sparse outliers. Despite acknowledged successes, this category of approaches may have limitations on complex scenarios, as the pixel-wise representations overlooked the spatial correlations between pixels.

The **region-based** methods built background models by taking advantages of inter-pixel relations, demonstrating impressive results on handling dynamic scenes. A batch of diverse approaches were proposed to model spatial struc-tures of scenes, such as joint distributions of neighboring pixels [11], [17], block-wise classifiers [18], structured adja-cency graphs [19], auto-regression models [20], [21], random fields [22], and multi-layer models [23] etc. And a number of fast learning algorithms were discussed to maintain their models online, accounting for environment variations or any structural changes. For example, Monnet et al. [20] trained and updated the region-based model by the generative sub-space learning. Cheng et al. [19] employed the generalized 1-SVM algorithm for model learning and foreground

pre-diction. In general, methods in this category separated the spatial and temporal information, and their performances were somewhat limited in some highly dynamic scenarios, *e.g.* heavy rains or sudden illumination changes.

The third category modeled scene backgrounds by exploiting both spatial and temporal information. Mahadevan et al. [24] proposed to separate foreground objects from surroundings by judging the distinguished video patches, which contained different motions and appearances compared with the majority of the whole scene. Zhao et al. [25] addressed the outdoor night background modeling by performing subspace learning within video patches. Spatio-temporal representations were also extensively discussed in other vision tasks such as action recognition [26] and trajectory parsing [27]. These methods motivated us to build models upon the spatio-temporal representations, *i.e.* video bricks.

In addition, several **saliency-based** approaches provided alternative ways based on spatio-temporal saliency estima-tions [24], [28], [29]. The moving objects can be extracted according to their salient appearances and/or motions against the scene backgrounds. For example, Wixson et al. [28] detected the salient objects according to their consistent mov-ing directions over time. Kim et al. [30] used a discrimi-nant center-surround hypothesis to extract foreground objects around their surroundings.

Along with the above mentioned background models, a number of reliable image features were utilized to better handle the background noise [31]. Exemplars included the Local Binary Pattern (LBP) features [32]–[34] and color tex-ture histograms [35]. The LBP operators described each pixel by the relative graylevels of its neighboring pixels, and their effectiveness has been demonstrated in several vision tasks such as face recognition and object detection [32], [36], [37]. The Center-Symmetric LBP was proposed in [34] to further improve the computational efficiency. Tan and Triggs [33] extended LBP to LTP (Local Ternary Pattern) by thresholding the graylevel differences with a small value, to enhance the effectiveness on flat image regions.

III. OVERVIEW

In this work, we propose to learn and maintain the dynamic models within spatio-temporal video patches (*i.e.* video bricks) and Kalman filter, accounting for real challenges in surveillance scenar-ios [7]. The algorithm can process 15 ∼ 20 frames per second in the resolution 352 × 288 (pixels) on average. We briefly overview the proposed framework of background modeling in the following aspects.

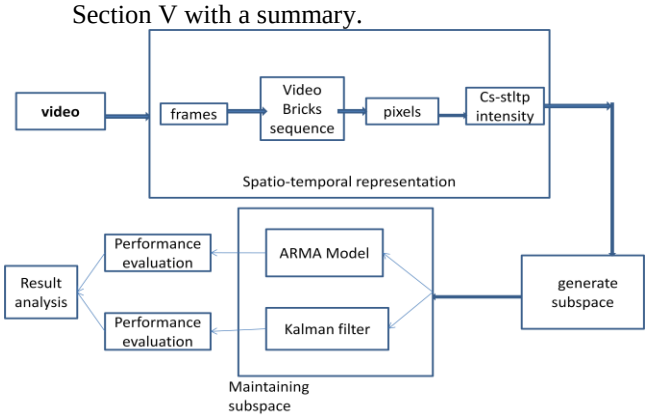
1) Spatio-Temporal Representations: We represent the observed scene by video bricks, *i.e.* video volumes spanning over both spatial and temporal domain, in order to jointly model spatial and temporal information. Specifically, at every location of the scene, a sequence of video bricks are extracted as the observations, within which we can learn and update the background models. Moreover, to compactly encode the video bricks against illumination variations, we design a brick-based descriptor, namely Center Symmetric Spatio-Temporal Local Ternary Pattern (CS-STLTP), which is inspired by the 2D scale invariant local pattern operator proposed in [14]. Its effectiveness is also validated in the experiments.

2) Pursuing Dynamic Subspaces: We treat each sequence of video bricks at a certain location as a consecutive signal, and generate the subspace within these video bricks. The linear dynamic system (*i.e.* Auto Regressive Moving Average, ARMA model [38]) is adopted to characterize the spatio-temporal statistics of the subspace. Specifically, given the observed video bricks, we express them by a data matrix, in which each column contains the feature of a video brick. The basis vectors (*i.e.* eigenvectors) of the matrix can be then estimated analytically, representing the appearance parameters of the subspace, and the parameters of dynamical variations are further computed based on the fixed appearance parameters. It is worth mentioning that our background model jointly captures the information of appearance and motion as the data (*i.e.* features of the video bricks) are extracted over both spatial and temporal domains.

3) Maintaining Dynamic Subspaces Online: Given the newly appearing video bricks with our model, moving fore-ground objects are segmented by estimating the residuals within the related subspaces of the scene, while the back-ground models are maintained simultaneously to account for scene changes. The raising problem is to update parame-ters of the subspaces incrementally against disturbance from foreground objects and background noise. The new obser-vation may include noise pixels (*i.e.* outliers), resulting in degeneration of model updating [20], [25]. Furthermore, one video brick could be partially occluded by foreground objects in our representation, *i.e.* only some of pixels in the brick are true positives. To overcome this problem, we present a novel approach to compensate observations (*i.e.* the observed video bricks) by generating data from the current models. Specifically, we replace the pixels labeled as non-background by the generated pixels to synthesize the new observations. The algorithm for online model updating includes two steps:

4) (i) update appearance parameters using the incremental sub-space learning technique, and (ii) update dynamical variation parameters by analytically solving the linear reconstruction. The experiments show that the proposed method effectively improves the robustness during the online processing.

5) The remainder of this paper is arranged as follows. We first present the model representation in Section II, and then discuss the initial learning, foreground segmentation and online updat-ing mechanism in Section III, respectively. The experiments and comparisons are demonstrated in Section IV and finally comes the conclusion in Section V with a summary.



IV. DYNAMIC SPATIO-TEMPORAL MODEL

In this section, we introduce the background of our model, and then discuss the video brick representation and our model definition, respectively.

A. Background

In general, a complex surveillance background may include diverse appearances that sometimes move and change dynami-cally and randomly over time flying [39]. There is a branch of works on time-varying texture modeling [40]–[42] in computer vision. They often treated the scene as a whole, and pursued a global subspace by utilizing the linear dynamic system (LDS). These models worked well on some natural scenes mostly including a few homogeneous textures, as the LDS characterizes the subspace with a set of linearly combined components. However, under real surveillance challenges, it could be intractable to pursue the global subspace. In this work, we represent the observed scene by an array of small and independent subspaces, each of which is defined by the linear system, so that our model is able to handle better challenging scene variations. Our background model can be viewed as a mixed compositional model consisting of the linear subspaces. In particular, we conduct the background subtraction with our model based on the following observations.

Assumption 1: The local scene variants (*i.e.* appearance and motion changing over time) can be captured by the low-dimensional subspace.

Assumption 2: It is feasible to separate foreground moving objects from the scene background by fully exploiting spatio-temporal statistics.

B. Spatio-Temporal Video Brick

Given the surveillance video of one scene, we first decom-pose it with a batch of small brick-like volumes. We consider the video brick of small size (*e.g.*, 4 × 4 × 5 pixels) includes relative simple content, which can be thus generated by few bases (components). And the brick volume integrates both spatial and temporal information, that we can better capture complex appearance and motion variations compared with the traditional image patch representations.

We divide each frame I_i , *(*i = 1, 2, . . . , n*)* into a set of image patches with the width w and height h. A number t of patches at the same location across the frames are combined together to form a brick. In this way, we extract a sequence of video bricks $V = \{v_1, v_2, \ldots, v_n\}$ at every location for the scene.

V. KALMAN FILTER

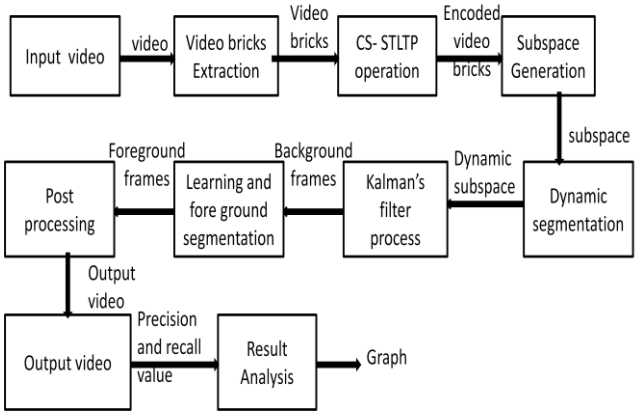
The Kalman ¯filter [1,5] is an optimal estimator of the state of processes which satisfies: (a) they can be modeled by a linear system, (b) the measurement and the process noise are white, and have zero mean gaussian distributions. Under these conditions, knowing the input (external controls u_t) and the output (measurements z_t) of the system, the Kalman ¯filter provides an optimal estimate of the state of the process (x_t), by minimizing the variance of the estimation error and constraining the average of the estimated outputs and the average of the measures to be the same. It is characterized by two main equations: the state equation and the measurement equation

$$xt = Axt¡1 + But¡1 + wt¡1$$

$$z_t = Cx_t + v_t$$

A is the state transition matrix, B is the external control transition matrix, w represents the process noise, C is the transition matrix that maps the process state to the measurement, and v represents the measurement noise. The Kalman ¯lter works in two steps: prediction and correction steps. The former uses the state of the system and the external control at time t ¡ 1 to predict the current state ($\hat{x}^i_t$), the latter uses the current measure z_t to correct the state estimation ($\hat{x}_t$). The working schema of the Kalman ¯lter is illustrated in Figure 1. The factor K_t , the gain of the ¯lter, is chosen in order to minimize the variance of the estimate error (P_t). The di®erence

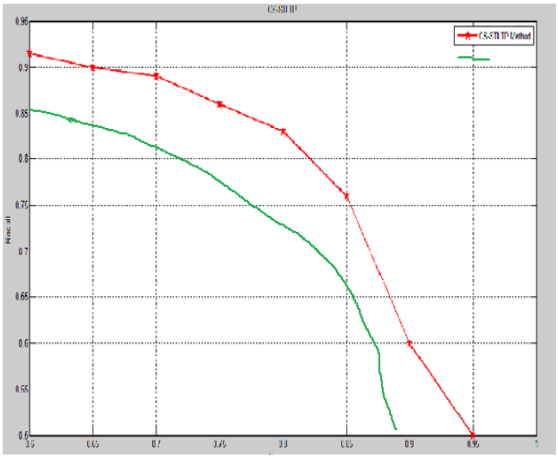
between the measure and the state predicted value (z_t ¡ $Cx^{\wedge i}_t$) is called innovation.



VI. RESULT ANALYSIS

This measure Intensity, precision and recall rate for background subtraction method and plot a precision versus recall graph.

- precision - fraction of retrieved instances that are relevant
- recall - fraction of relevant instances that are retrieved



VII. CONCLUSION

This paper studies an effective method for background subtraction, addressing the all challenges in real surveillance scenarios. In the method, we learn and maintain the dynamic texture models within spatio-temporal video patches (*i.e.* video bricks). Sufficient experiments as well as empirical analysis are presented to validate the advantages of our method.

REFERENCES

[1] C. Stauffer and W. Grimson, "*Adaptive background mixture models for real-time tracking*," in *Proc. IEEE Conf. CVPR*, Jun. 1999.

[2] T. Bouwmans, F. E. Baf, and B. Vachon, "*Background modeling using mixture of Gaussians for foreground detection-a survey*," *Recent Patents Comput. Sci.*, vol. 1, no. 3, pp. 219–237, 2008.

[3] L. Maddalena and A. Petrosino, "*A self-organizing approach to background subtraction for visual surveillance applications*," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1168–1177, Jul. 2008.

[4] D.-M. Tsai and S.-C. Lai, "*Independent component analysis-based background subtraction for indoor surveillance*," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 158–167, Jan. 2009.

[5] H. Chang, H. Jeong, and J. Choi, "*Active attentional sampling for speedup of background subtraction*," in *Proc. IEEE Conf. CVPR*, Jun. 2012, pp. 2088–2095.

[6] X. Liu, L. Lin, S. Yan, H. Jin, and W. Tao, "*Integrating spatio-temporal context with multiview representation for object recognition in visual surveillance*," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 393–407, Apr. 2011.

[7] L. Lin, Y. Lu, Y. Pan, and X. Chen, "*Integrating graph partitioning and matching for trajectory analysis in video surveillance*," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4844–4857, Apr. 2012.

[8] Z. Zivkovic, "*Improved adaptive Gaussian mixture model for background subtraction*," in *Proc. 17th IEEE ICPR*, Aug. 2004, pp. 28–31.

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

[10] A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis, "*Background and foreground modeling using nonparametric kernel density estimation for visual surveillance*," *Proc. IEEE*, vol. 90, no. 7, pp. 1151–1163, Jun. 2002.

[11] Y. Sheikh and M. Shah, "*Bayesian modeling of dynamic scenes for object detection*," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1778–1792, Nov. 2005.

[12] C. Benedek and T. Sziranyi, "*Bayesian foreground and shadow detection in uncertain frame rate surveillance videos,*" *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 608–621, Apr. 2008.

[13] O. Barnich and M. Droogenbroeck, "*Vibe: A universal background subtraction algorithm for video sequences*," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[14] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikainen, and S. Li, "*Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes*," in *Proc. IEEE Int. Conf. CVPR*, Jun. 2010, pp. 1301–1306.

[15] J. Pilet, C. Strecha, and P. Fua, "*Making background subtraction robust to sudden illumination changes*," in *Proc. ECCV*, 2008, pp. 567–580.

[16] C. Guyon, T. Bouwmans, and E.-H. Zahzah, "*Foreground detection via robust low rank matrix decomposition including spatiotemporal constraint*," in *Proc. Comput. Vis. ACCV Workshops*, 2013, pp. 315–320.

[17] M. Wu and X. Peng, "*Spatio-temporal context for codebook-based dynamic background subtraction*," *AEU Int. J. Electron. Commun.*, vol. 64, no. 8, pp. 739–747, 2010.

[18] H.-H. Lin, T.-L. Liu, and J.-H. Chuang, "*Learning a scene background model via classification*," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1641–1654, May 2009.

[19] L. Cheng and M. Gong, "*Realtime background subtraction from dynamic scenes*," in *Proc. 12th IEEE ICCV*, Sep./Oct. 2009, pp. 2066–2073.

[20] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, "*Background modeling and subtraction of dynamic scenes*," in *Proc. 9th IEEE ICCV*, Oct. 2003, pp. 1305–1312.

[21] J. Zhong and S. Sclaroff, "*Segmenting foreground objects from a dynamic textured background via a robust Kalman filter*," in *Proc. 9th IEEE ICCV*, Oct. 2003, pp. 44–50.

[22] Y. Wang, K.-F. Loe, and J.-K. Wu, "*A dynamic conditional random field model for foreground and shadow segmentation*," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 279–289, Feb. 2007.

[23] K. A. Patwardhan, G. Sapiro, and V. Morellas, "*Robust foreground detection in video using pixel layers*," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 746–751, Apr. 2008.

[24] V. Mahadevan and N. Vasconcelos, "*Spatiotemporal saliency in dynamic scenes*," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 171–177, Jan. 2010.

[25] Y. Zhao, H. Gong, L. Lin, and Y. Jia, "*Spatio-temporal patches for night background modeling by subspace learning*," in *Proc. IEEE ICPR*, Dec. 2008, pp. 1–4.

[26] X. Liang, L. Lin, and L. Cao, "*Learning latent spatio-temporal compositional model for human action recognition,"* in *Proc. ACM Int. Conf. Multimedia*, Oct. 2013, pp. 263–272.

[27] X. Liu, L. Lin, and H. Jin, "*Contextualized trajectory parsing with spatio-temporal graph*," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 3010–3024, Dec. 2013.

[28] L. Wixson, "*Detecting salient motion by accumulating directionallyconsistent flow*," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 774–780, Aug. 2000.

[29] D. Gutchess, M. Trajkovics, E. Cohen-Solal, D. Lyons, and A. K. Jain, "A background model initialization algorithm for video surveillance," in *Proc. IEEE ICCV*, Jul. 2001, pp. 733–740.

[30] W. Kim, C. Jung, and C. Kim, "*Spatiotemporal saliency detection and its applications in static and dynamic scenes*," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 446–456, Apr. 2011.

[31] L. Lin, X. Liu, and S.-C. Zhu, "*Layered graph matching with composite cluster sampling*," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1426–1442, Aug. 2010.

[32] T. Ojala, M. Pietikainen, and T. Maenpaa, "*Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[33] X. Tan and B. Triggs, "*Enhanced local texture feature sets for face recognition under difficult lighting conditions*," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 168–182, Oct. 2007.

[34] M. Heikkila, M. Pietikainen, and C. Schmid, "*Description of interest regions with local binary patterns*," *Pattern Recognit.*, vol. 42, no. 3, pp. 425–436, Mar. 2009.

[35] J. Yao and J. Odobez, "*Multi-layer background subtraction based on color and texture*," in *Proc. IEEE CVPR*, Jun. 2007, pp. 1–8.

[36] Z. Guo, L. Zhang, and D. Zhang, "*Rotation invariant texture classification using LBP variance (LBPV) with global matching*," *Pattern Recognit.*, vol. 43, no. 3, pp. 706–719, Mar. 2010.

[37] L. Lin, P. Luo, X. Chen, and K. Zeng, "*Representing and recognizing objects with massive local image patches*," *Pattern Recognit.*, vol. 45, no. 1, pp. 231–240, Jan. 2012.

[38] E. Hannan and M. Deistler, *Statistical Theory Of Linear Systems* (Probability and Mathematical Statistics). New York, NY, USA: Wiley, 1988.
[39] L. Lin, T. Wu, J. Porway, and Z. Xu, "*A stochastic graph grammar for compositional object representation and recognition*," *Pattern Recognit.*, vol. 42, no. 7, pp. 1297–1307, Jul. 2009.
[40] S. Soatto, G. Doretto, and Y. Wu, "*Dynamic textures*," *Int. J. Comput. Vis.*, vol. 52, no. 2, pp. 91–109, 2003.
[41] P. Saisan, G. Doretto, Y. Wu, and S. Soatto, "*Dynamic texture recognition*," in *Proc. IEEE CVPR*, Jun. 2001, pp. 58–63.
[42] G. Doretto, D. Cremers, P. Favaro, and S. Soatto, "*Dynamic texture segmentation*," in *Proc. 9th IEEE ICCV*, Oct. 2003, pp. 1236–1242.
[43] D. Skocaj and A. Leonardis, "*Weighted and robust incremental method for subspace learning*," in *Proc. 9th IEEE ICCV*, Oct. 2003, pp. 1494–1501.
[44] M. Artac, M. Jogan, and A. Leonardis, "*Incremental PCA for on-line visual learning and recognition,"* in *Proc. 16th ICPR*, 2002, pp. 781–784.
[45] A. Levy and M. Lindenbaum, "*Sequential karhunen-loeve basis extraction and its application to images*," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1371–1374, Aug. 2000.
[46] L. Wang, L. Wang, M. Wen, Q. Zhuo, and W. Wang, "*Background subtraction using incremental subspace learning*," in *Proc. IEEE ICIP*, Sep./Oct. 2007, pp. V-45–V-48.
[47] E. Candès, X. Li, Y. Ma, and J. Wright, "*Robust principal component analysis*?" *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.
[48] X. Ding, L. He, and L. Carin, "*Bayesian robust principal component analysis*," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3419–3430, Dec. 2011.
[49] D. Farcas, C. Marghes, and T. Bouwmans, "*Background subtraction via incremental maximum margin criterion: A discriminative subspace approach*," *Mach. Vis. Appl.*, vol. 23, no. 6, pp. 1083–1101, Nov. 2012.
[50] Y. Li, "*On incremental and robust subspsubspace learning*," *Pattern Recognit.*, vol. 37, no. 7, pp. 1509–1518, 2004