International Journal of
Intelligent Systems and
Applications in Engineering
ISSN:2147-6799 www.atscience.org/IJISAE
Advanced Technology and Science
Original Research Paper

A new subspace based solution to background modelling and change detection

Kemal Özkan*[1], Şahin Işık[2], Ömer Nezih Gerek[3], Muzaffer Doğan[4]

Accepted 3rd September 2016

Abstract: For surveillance system, the background subtraction plays an important role for moving object detection with an algorithm embedded in the camera. Since the existence algorithms cannot satisfy the good accuracy on complex backgrounds including illumination change and dynamic objects, we have put forward the concept of Common Vector Approach (CVA) as a new idea for background modelling. Effectiveness of proposed method is presented through the experiments on popular Wallflower dataset. The obtained visual outputs are compared with well-known methods based on the subjective and objective criteria. From the overall evaluation, we can note the proposed method is not only exhibit successful foreground detection results, but also promises an effective and efficient system for background modelling.

Keywords: *Common Vector Approach, Background Modelling, Foreground Detection, Moving Object Detection, Change Detection.*

1. Introduction

Background subtraction, determining changes in the sequence of images, is an important and painful task in computer vision. One key problem in background detection is coping with dynamic backgrounds, which involve shadows, highlights, waving trees, camera jitter, camouflage, fountains and similar movements. The key idea is deriving a model that comprises the rich information about processed scene and taking difference between the model and current image in order to yield the foreground, which is usually called as change detection. Although utilizing this idea is convenient for static background, but for dynamic backgrounds, it is neither applicable and nor promising.

Until now, various methodologies are applied to alleviate problems encountered from dynamic backgrounds. The proposed methods can be grouped in two ways; pixel or block based approaches. While in pixel approaches, a model is constructed for each pixel by taking the history of them, in other side, in block based approaches, the contribution of neighbour pixels is taking into account in case of modelling the background.

A crowded set of algorithms have been utilized to demonstrate satisfactory results for non-stationary background. A survey is presented by Bouwmans to reveal the performance of subspace based background learning methods [1]. The impact of Principal Component Analysis (PCA) for background learning is firstly investigated in the work of Oliver et. Al [2]. They applied the concept of PCA on a model of the probability distribution function of the background. Since the PCA works based on the least square estimation as sensitive to outlier, an alternative approach was developed by Torre and Block. It is called as Robust Principal Component Analysis (RPCA) [3]. Comparing with PCA and RPCA, it should be noted that effects of outliers are suppressed in case of linear based optimization when compared with nonlinear based optimization as utilized in PCA. Inspiring from the theory of work performed by Torre and Block, some variants of RPCA [4] have been developed and utilized for subspace based background learning. With a different idea, the Independent Component Analysis has been attempted with a purpose of background modelling [5]. The aim is obtaining the background model $\mathbf{Y} = \mathbf{W}\mathbf{X}_\mathbf{T}$, where $\mathbf{W}$ and $\mathbf{X}_\mathbf{T}$ denote the de-mixing and mixing matrices, respectively. $\mathbf{X}_\mathbf{T}$ contains background and foreground images. The size of $\mathbf{X}_\mathbf{T}$ is $\mathbf{2xK}$ as $\mathbf{K}$ are stored in vector format. Yet another method is Gaussian model based background modelling, which was proposed by Wren et. al in order to tracking person body, named as Pfinder [6]. In referred work, a 2-D model based on the Maximum Posteriori Probability (MAP) was introduced for detecting and tracking human body. By focusing the changes in the region of interest, a blob model is proceeded to reveal the person body. Moreover, a comprehensive survey is available in study performed by Bouwman [7].

The capacity of each method is limited when utilized to overcome challenges caused from dynamic backgrounds. For this reason, we have proposed a new nonparametric and subspace based background modelling technique, which relies on the concept of common vector approach (CVA). The ability of Common Vector Approach [8] for background subtraction is firstly analysed in the present work. The proposed background subtraction system involves two stages; (i) the background modelling by using training images and (ii) detecting foreground objects in test image sequence. To evaluate the system performance, an experiment is conducted on well-known Microsoft's Wallflower dataset [9, 10]. The obtained good visual and statistical results implies that the CVA can be applied for background modelling and change detection.

[1] Eskisehir Osmangazi University, Computer Engineering Department
[2]Eskisehir Osmangazi University, Computer Engineering Department
[3]Anadolu University, Electrical & Electronics Engineering Department
[4]Anadolu University, Computer Engineering Department
** Corresponding Author: Email: kozkan@ogu.edu.tr*
Note: This paper has been presented at the 3rd International Conference on Advanced Technology & Sciences (ICAT'16) held in Konya (Turkey), September 01-03, 2016.

The rest of paper is arranged as follows. In section 2, the CVA and its application to background modelling is presented. In section 3, the experimental results and performance comparison with well-known methods is carried out.

2. CVA with Application to Background Modelling

CVA is a popular subspace based classification algorithm as applied for face recognition [11], spam classification [12], image denoising [13] and edge detection [14] tasks. The motivation of CVA is inspired from theory behind the PCA. While in PCA, the data is recovered by using eigenvectors corresponding to largest eigenvalues, but it has been emphasized that using null space of data gives more impressive accuracy in case of classification [8]. Depending on this fact, CVA algorithm has been put forward by authors of study in [8, 15, 16]. Specifically, by using CVA algorithm, a frame is represented with two components, which are common and difference as shown in Eq. (1). There are two cases in CVA algorithm as sufficient and insufficient data cases. If the number of vectors is less than dimension vectors, then it is called as insufficient data case, otherwise, it is sufficient data case. In case of insufficient data case, common and difference frames can be calculated by using the Gram Schmidt procedure.

In this study, the motivation under the CVA algorithm is adopted for background modelling. The key point of algorithm is encapsulating background information of different frames in order to obtain a single and meaningful background frame. Similar to PCA, each frame is converted in to vector format in case of background modelling.

Assuming that we have given n samples $(\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_n)$ and each frame in 1-D. With CVA algorithm, it is accepted that a given frame a_i can be separated into two parts as common and difference frame, which is denoted in Eq. (1).

$$a_k = a_{com} + a_{k,diff} \tag{1}$$

(1) Where the a_{com} and $a_{k,diff}$ refers to common and difference frames, respectively. In order to obtain orthogonal and orthonormal basis, the concept of Gram Schmidt is carried out on given vector set $(\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_n)$. As a first stage, the selected reference frame is subtracted from remain vectors as shown in Eq. (2). In this study, the first frame $(k=1)$ is considered as reference frame for the sake of simplicity.

$$\begin{aligned} d_1 &= a_2 - a_1 \\ d_2 &= a_3 - a_1 \\ &\;\;.... \\ d_{n-1} &= a_n - a_1 \end{aligned} \tag{2}$$

(2) From the combination of difference vectors, a matrix $M = \{d_1, d_2, \ldots, d_{(n\text{-}1)}\}$ is obtained. The next stage is computing the orthonormal and orthogonal vectors with the idea of Gram-Schmidt procedure which is shown in Eq. (3) and Eq. (4).

$$v_1 = d_1 \;\; and \;\; u_1 = \frac{v_1}{|v_1|} \tag{3}$$

$$v_i = d_i - \sum_{j=1}^{i\text{-}1} \langle d_i, u_j \rangle u_j \quad and \quad u_i = \frac{v_i}{|v_i|} \qquad i = 1, \ldots, n\text{-}1 \tag{4}$$

Where, $\langle d_i, u_j \rangle$ refers to dot product of two vectors and $|v_i|$ denotes the l_2 norm of each vector. Each vector is normalized by dividing with their l_2 norm. At the end of Gram-Schmidt orthogonalization procedure the $(u_1, u_2, \ldots, u_{(n-1)})$ orthonormal and orthogonal $(v_1, v_2, \ldots, v_{(n\text{-}1)})$ sets are obtained to yield difference frame.

(3) Once the orthonormal sets are obtained, the difference frame is determined as given in the below formula. Specifically, the selected reference frame is projected on orthonormal vectors and summed up to obtain the difference frame. In this study, the first frame is taken as reference, and $k=1$.

$$a_{k,\,diff} = \langle a_k, u_1 \rangle u_1 + \langle a_k, u_2 \rangle u_2 +, \;\ldots\; ,+ \langle a_k, u_{(n\text{-}1)} \rangle u_{(n\text{-}1)} \tag{5}$$

(4) As a result, the common vector a_{com} is derived by subtracting the $a_{k,\,diff}$ from a_k.

$$a_{com} = a_k - a_{k,diff} \tag{6}$$

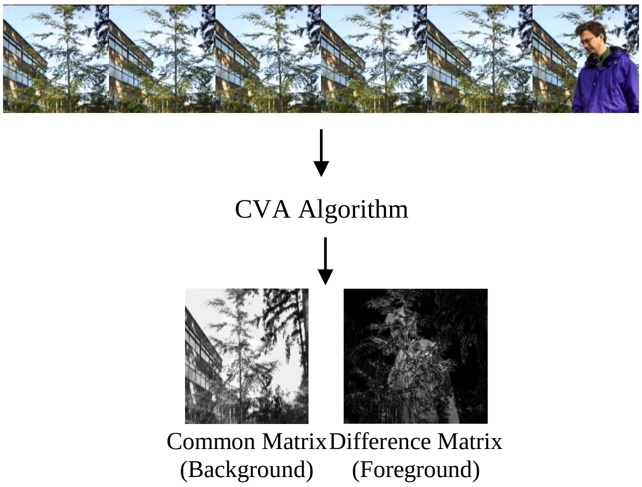

Figure 1. Demonstration of proposed method

As an improvement on CVA, a low noise value between 0-1 is inserted to each difference subspace in Eq. 2 in terms of making high correlated data as low correlated form. The reason of making data low correlated is explained with idea that if the data is highly correlated then the rank becomes smaller than 2. As a result of small rank value, the obtained common vector does not become meaningful to eye. With this way, a background model with training data set is constructed as common frame refers to background, whereas the difference frame indicates foreground.

The motivation behind the CVA based background modelling is exhibited in Fig. 1. As we can observe from the Fig. 1, there are two components of a frame as:

(1) first component provides the common frame of training set, which refers to obtained background model.
(2) other component denotes the difference frame that exhibits details including moving objects and changes of training set.

From the Fig. 1, the ability of CVA for change detection can be observed clearly. Inspired from this fact, we have utilized the CVA algorithm for background modelling and change detection.

In case of foreground extraction, the common vector of processed test frame (t) is computed as projecting the test frame onto the orthonormal basis generated by Gram-Schmit procedure [15]. As a first stage, the difference vectors corresponding to the test frame is obtained with Eq. 7.

$$t_{diff} = \langle t,u_1 \rangle u_1 + \langle t,u_2 \rangle u_2 +, \dots ,+ \langle t,u_{(n-1)} \rangle u_{(n-1)} \tag{7}$$

Once the difference vector is subtracted from the test vector, the common vector of processed test frame is determined as shown in Eq. 8.

$$t_{com} = t - t_{diff} \tag{8}$$

The difference between the two common vectors is considered in terms of observing the foreground regions.

$$\forall(i,j), I(i,j) = \begin{cases} 1 & abs(t_{com} - a_{com}) > threshold \\ 0 & otherwise \end{cases} \tag{9}$$

As indicated in Eq. 9, for each pixel location $\forall(i,j)$, if the absolute difference is greater than a fixed threshold value, then foreground mask is marked as 1, otherwise marked as 0. However, taking the absolute difference for Moved Object, Light Switch, Camouflage videos, produces a lot of erroneous pixels in foreground mask. To overcome this challenge, only difference of two common vectors is put into the thresholding procedure. The utilized threshold value for each video are predetermined as follows; 0.1 for Camouflage, Bootstrap, Light Switch, Waving Trees, 0.2 for Foreground Aperture and 0.3 for Time of Day and Moved Object video, respectively.

After thresholding procedure, it has been observed that some morphological procedure is greatly required to obtain best results. For this purpose, firstly, a 5x5 median filter is applied on the binary foreground mask. Then, the connected components having size of less than 20, are considered as ghosts and ignored by applying the area open morphological operator.

To close the holes in binary region, the morphological closing procedure is performed with disk structural element having size of 5 and binary holes are filled with morphological filling operator. As a last step, morphological opening with disk structural element having size of 5 is performed to mitigate the effect of closing operator.

Table 1: Subjective results on Wallflower dataset.

Method	Moved Objects	Time of Day	Light Switch	Waving Trees	Camou -flage	Boot -strap	Foreg. Aperture
Test image							
Ground truth							
SG Wren *et al.*							
MOG Stauffer *et al.*							
KDE Elgammal *et al.*							
SL-PCA Oliver *et al.*							
SL-ICA Tsai and Lai							
SL-INMF Bucak *et al.*							
SL-IRT Li *et al.*							
CVA *Proposed*							

3. Performance Evaluation

3.1. Dataset

To comment the performance proposed method, some experimental are conducted on popular Wallflower Dataset. Technically Wallflower dataset [8] provides different classes of about dynamic backgrounds which are Moved Object, Time of Day, Light Switch, Waving Trees, Camouflage, Bootstrapping and Foreground Aperture. Until now, various methods have been made experimental on this dataset. The priorly specified training and test images with their ground truth [10] are utilized to obtain subjective and objective results.

3.2. Subjective Results

In order to comment the obtained results, we have compared the produced results with other ones. For this purpose, the subjective outputs are presented on Table 1. Specifically, the visual results that are presented in the study of Bouwman [1] are considered as reference in case of performance comparison. For a benchmark comparison, the obtained visual results are compared with popular subspace and other methods, which are given as Single Gaussian (SG) [6], Mixture of Gaussian (MOG) [17], Kernel Density Estimation (KDE) [18], Subspace Learning PCA (SL-PCA) [19], Subspace Learning ICA (SL-PCA), Subspace Learning ICA (SL-PCA) [20], Subspace Learning via Incremental Non Negative Matrix Factorization (SL-INMF) [21] and Subspace Learning via Incremental Rank-(R1, R2, R3) Tensor (SL-IRT) [22].

The all of visual results are exhibited in Table 1. The first column indicates the method's name and the rest of columns show the performance of each aforementioned method. Also, the first row denotes the processed image, second row indicates the ground truth related to given image and other rows show visual results generated by each method.

Table 2: Objective results on Wallflower dataset.

Method	Error	Problem Type: Moved Object	Time of Day	Light Switch	Waving Trees	Camou-flage	Bootstrap	Foreground Aperture	Total Errors	TE without LS	TE without C
SG	FN	0	949	1857	3110	4101	2215	3464			
Wren *et al.*	FP	0	535	15123	357	2040	92	1290	35133	18153	28992
MOG	FN	0	1008	1633	1323	398	1874	2442			
Stauffer *et al.*	FP	0	20	14169	341	3098	217	530	27053	11251	23557
KDE	FN	0	1298	760	170	238	1755	2413			
Elgammal *et al.*	FP	0	125	14153	589	3392	933	624	26450	11537	22175
SL-PCA	FN	0	879	962	1027	350	304	2441			
Oliver *et al.*	FP	1065	16	362	2057	1548	6129	537	17677	16353	15779
SL-ICA	FN	0	1199	1557	3372	3054	2560	2721			
Tsai and Lai	FP	0	0	210	148	43	16	428	15308	13541	12211
SL-INMF	FN	0	724	1593	3317	6626	1401	3412			
Bucak *et al.*	FP	0	481	303	652	234	190	165	19098	17202	12238
SL-IRT	FN	0	1282	2822	4525	1491	1734	2438			
Li *et al.*	FP	0	159	389	7	114	2080	12	17053	13842	15448
CVA	FN	0	1012	946	766	708	982	2537			
Proposed.	FP	0	0	320	20	8	130	482	7891	6625	7175

At a first glance, we can observe that similar outputs are obtained from each method. Upon inspecting results, one can emphasize that probabilistic based methods including MOG and KDE produce similar results in terms of foreground region detection. The results of KDE and MOG are superior than SG, since background modelling with single Gaussian is a short-side in term of complex background. Again, we can emphasize that SG, MOG and KDE are sensitive illumination changes because of considering the historical probability of each pixel.

On the other side, the subspace based methods are more robust to illumination and complex background changes. By examining results of PCA, ICA, INMF and IRT, it can be seen that the visuals result of IRT are not converged to ground truth as some objects are disappeared in foreground mask. Moreover, although the PCA method exhibits good results in case of Time of Day, Light Switch, Waving Trees, Camouflage, Foreground Aperture, but some erroneous pixels are obtained for Moved Objects and Bootstrap videos. Furthermore, visual outputs of ICA and INMF are similar to each other, however, the performance of ICA is more dominant for Camouflage and Bootstrap videos.

Finally, we can observe that CVA and PCA generate closest results, however, the PCA method fails in case of indoor crowded scene (bootstrap). Also, one can note that the proposed method can perfectly model the clean background in case of illumination changes as well as crowded scenes and other complex backgrounds. As a result, good foreground masks are determined for all videos.

3.3. Objective Results

In addition to subjective evaluation, the objective results for each method is determined with respect to statistical metrics, called false positive (FP) and false negative (FN). While the FP indicates the pixel marked as foreground in processed image but it is background in ground truth image, conversely the FN refers to the pixel marked as background in processed image but it is foreground in ground truth image. If a pixel is marked as 1 in processed image, but it is 0 in ground truth image, then the count of FP is incremented by 1. Similarly, if a pixel is marked as 0 in processed image, but it is 1 in ground truth image, then the count of FN is incremented by 1. By combining these error values, the Total Error (TE) metric is computed as a sum of FP and FN. The lower value of error value denotes the best performance in the concept of foreground segmentation. Also, the Total Errors without light switch (TE without LS) and Total Errors without Camouflage switch (TE without Camouflage) are presented on the last columns of Table 2.

The Table 2 summarizes all of the objective results for aforementioned background modelling methods. As we can see that the performance MOG and KDE are closest to each other and show better performance than SG method. The performance of MOG and KDE are better when the light switch video excluded, but worse in case of TE metric. Comparing the PCA, ICA, INMF and IRT, one can observe that the performance of ICA is dominant in case of all metrics. On the other side, we can find that the CVA method combining with the basic post processing procedure show favourable results in terms of all metrics.

4. Conclusion

In the demonstrated work, a new idea is introduced for background modelling and foreground detection in a given video. Through experiments on real and complex videos, we have observed that the proposed method can efficiently detect the changes in a given set of images. The performance of proposed are compared with state of algorithms including SG, MOG, KDE, PCA, ICA, IRT and INMF and commented with respect to some objective and subjective measures. The obtained superior results indicate that it is appropriate to use the CVA method for background modelling. Also, we can emphasize that an intelligent post processing procedure is vitally needed in order to accurate foreground detection and segmentation.

References

[1] T. Bouwmans, Subspace learning for background modeling: A survey, Recent Patents on Computer Science, 2 (2009) 223-234.

[2] N.M. Oliver, B. Rosario, A.P. Pentland, A bayesian computer vision system for modeling human interactions, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22 (2000) 831-843.

[3] F. De La Torre, M.J. Black, A framework for robust subspace learning, International Journal of Computer Vision, 54 (2003) 117-142.

[4] C. Guyon, E.-h. Zahzah, T. Bouwmans, Robust principal component analysis for background subtraction: systematic evaluation and comparative analysis, Citeseer2012.

[5] D.-M. Tsai, S.-C. Lai, Independent component analysis-based background subtraction for indoor surveillance, image Processing, IEEE Transactions on, 18 (2009) 158-167.

[6] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, P nder: Real-time tracking of the human body. Media Lab 353, MIT1995.

[7] T. Bouwmans, Traditional and recent approaches in background modeling for foreground detection: An overview, Computer Science Review, 11 (2014) 31-66.

[8] M.B. Gülmezoğlu, V. Dzhafarov, A. Barkana, The common vector approach and its relation to principal component analysis, Speech and Audio Processing, IEEE Transactions on, 9 (2001) 655-662.

[9] Wallflower Dataset, http://research.microsoft.com/en-us/um/people/jckrumm/WallFlower/TestImages.htm.

[10] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, Wallflower: Principles and practice of background maintenance, Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, IEEE1999, pp. 255-261.

[11] S. Ergin, M.B. Gulmezoglu, A novel framework for partition-based face recognition, International Journal of Innovative Computing Information and Control, 9 (2013) 1819-1834.

[12] S. Gunal, S. Ergin, M.B. Gülmezoğlu, Ö.N. Gerek, On feature extraction for spam e-mail detection, Multimedia content representation, classification and security, Springer2006, pp. 635-642.

[13] K. Ozkan, E. Seke, Image denoising using common vector approach, Image Processing, IET, 9 (2015) 709-715.

[14] K. Ozkan, S. Isik, A novel multi-scale and multi-expert edge detector based on common vector approach, AEU-International Journal of Electronics and Communications, 69 (2015) 1272-1281.

[15] H. Cevikalp, M. Neamtu, M. Wilkes, A. Barkana, Discriminative common vectors for face recognition, IEEE Transactions on pattern analysis and machine intelligence, 27 (2005) 4-13.

[16] M. Koc, A. Barkana, O.N. Gerek, A fast method for the implementation of common vector approach, Information Sciences, 180 (2010) 4084-4098.

[17] C. Stauffer, W.E.L. Grimson, Adaptive background mixture models for real-time tracking, Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., IEEE1999.

[18] A. Elgammal, D. Harwood, L. Davis, Non-parametric model for background subtraction, European conference on computer vision, Springer2000, pp. 751-767.

[19] N.M. Oliver, B. Rosario, A.P. Pentland, A bayesian computer vision system for modeling human interactions, IEEE transactions on pattern analysis and machine intelligence, 22 (2000) 831-843.

[20] D.-M. Tsai, S.-C. Lai, Independent component analysis-based background subtraction for indoor surveillance, IEEE Transactions on Image Processing, 18 (2009) 158-167.

[21] S.S. Bucak, B. Günsel, O. Gursoy, Incremental Non-negative Matrix Factorization for Dynamic Background Modelling, PRIS2007, pp. 107-116.

[22] X. Li, W. Hu, Z. Zhang, X. Zhang, Robust foreground segmentation based on two effective background models, Proceedings of the 1st ACM international conference on Multimedia information retrieval, ACM2008, pp. 223-228.

