International Journal of
Intelligent Systems and Applications in Engineering

ISSN:2147-6799 www.ijisae.org Original Research Paper

A Note on Background Subtraction by Utilizing a New Tensor Approach

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

Accepted 3rd September 2016

Abstract: This study deals with determining the foreground region by background subtraction based on a new tensor decomposition method. With this aim, the concept of Common Matrix Approach (CMA) is utilized with a purpose of background modelling. The performance of proposed method is validated by making experiments on real videos provided by Wallflower dataset. The obtained results are compared with well-known methods based on subjective on objective evaluation measures. The obtained good results indicate that using the CMA algorithm for background modelling is a simple and effective technique in terms computational cost and implementation. As an eventual result, we have observed that the superior results are determined on complex backgrounds including dynamic objects and illumination variation in image sets.

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

1. Introduction

Foreground detection is the principal interest topic of computer vision based applications such as intelligent visual surveillance, intelligent visual observation of animals and insects, optical motion capture, human-machine interaction, content based video coding, etc. The most extensively utilized areas can be given as road surveillance, airplane surveillance, maritime surveillance, boats and store surveillance systems, in where "people" is the main point of interest [1].

Major challenges associated with background subtraction can be noted as shadow, waving trees, foundations, intensity changes and camera jitter, which are called as dynamic backgrounds. Although a perfect solution has not been proposed to cope with these problems, but an affirmed method should be capable to alleviate all dynamic problems. The general idea is actuating a mathematical model to represent all image sequences of the processed background scene with a rich information one. Once the background model obtained, the difference between the test frame and model is considered as foreground in terms of traditional background modelling.

Numerous algorithms are proposed for background subtraction with a statistical or mathematical theory. By taking the handling strategy of images, the categorization of them can be grouped in two ways as 2-D based methods or tensor based methods. Technically, in the concept of 2-D based methods, each **MxN** frame is converted into vector format and a 2-D matrix is constructed with **(M.N)xK** dimension as **K** denotes number of frames in training set. Conversely, in tensor based one (3-D), a set of 2-D frames are combined and background is modelled through the tensor without converting frame into vector format. The 2-D based methods have disadvantages when compared with the tensor one. Specifically, in vector based methods the spatial information behind the neighbourhood pixels are neglected as all columns in a frame are connected as back to back in case of converting frame into vector format.

Various tensor decomposition based methods have been illustrated in research area of background subtraction. The Diffusion Bases (DB) [2] methodology has been adopted by decomposing 3-D data into 2-D plane, which denotes the found out background model. The capability of incremental tensor based background modelling [3] has been investigated with application for foreground segmentation and tracking. Another alternative method versus Principal Component Analysis (PCA) has been utilized by applying the concept of Locality Preserving Projections (LPP) [4], which is called as LoPP. An optimal rank-(R1, R2, …, Rn) tensor decomposition [5] model has been proposed in order to the high-dimensional tensor to low dimensional as sparse irregular patterns. Also, the Tensor Singular Value Decomposition on Fourier Domain has been analyzed for multilinear data completion and denoising, which is named as t-SVD [6].

Because of different challenges in the concept of background dataset, proposed methods do not meet all expectations. With this aim, a new tensor based background learning and change detection algorithm is presented in order to successful discrimination of foreground and background. Specifically, the theory of Common Matrix Approach (CMA) is applied to decompose 3D dimensional data (tensor) [7]. In case of orthogonal decomposition, the motivation of Gram-Schmidt orthogonalization is adopted. After projection stage, a common matrix that refers to obtained background model is determined. To report the statistical and visual results, the test stage is conducted on Wallflower dataset [8,9]. By comparing the statistical results with some of other tensor based approaches, one

[1] *Eskisehir Osmangazi University, Computer Engineering Department*
[2] *Eskisehir Osmangazi University, Computer Engineering Department*
[3] *Anadolu University, Computer Engineering Department*
[4] *Anadolu University, Electrical & Electronics Engineering Department*
* *Corresponding Author: Email: sahini@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.

can observe that the proposed method provides impressive and dominant results.

The remain part of paper is designed as follows. In section 2, the CMA and its application to foreground extraction is presented. In section 3, the obtained objective and subjective results are compared with other tensor based approaches. Finally, a conclusion is touched.

2. Principle of CMA and Its Application to Background Subtraction

The CMA algorithm is an extended form of Common Vector Approach, which is a subspace based method and utilized for face recognition [10], spam classification [11], image denoising [12] and edge detection [13] tasks. However, the ability of CMA for background modelling has not been realized in literature of computer vision. In case of CVA the data is handled in vector format as a 2-D matrix is constituted from training set and matrix decomposition strategy is applied on constructed 2D data, whereas for CMA, a tensor is generated from 2-D frames.

The main idea behind the CMA is combining background information from different frames and obtaining a single frame, which summarizes cues about background locations. Assuming that we have given *n* sample frames $(S_1, S_2, \ldots, S_n)$ and each frame in 2-D form. In the context of CMA, a frame can be represented with common and difference frames as shown in Eq. (1).

$$S_k = S_{com} + S_{k,diff} \tag{1}$$

(1) Where the S_{com} and $S_{k,diff}$ refers to common and difference frames, respectively. In order to calculate the Common frame, a tensor with 3-D size is constructed and the concept of Gram Schmidt is applied to derive orthogonal and orthonormal basis. First of all, difference matrices are calculated by a taking a first frame as reference. Instead of first frame, a different frame can be chosen among others as reference.

$$\begin{aligned} D_1 &= S_2 - S_1 \\ D_2 &= S_3 - S_1 \\ &\ldots. \\ D_{n-1} &= S_n - S_1 \end{aligned} \tag{2}$$

(2) Once a tensor $T = \{D_1, D_2, \ldots, D_{(n-1)}\}$ is obtained, the Gram-Schmidt procedure is activated on elements of T, which is shown in Eq. (3) and Eq. (4).

$$V_1 = D_1 \; and \; U_1 = \frac{V_1}{|V_1|} \tag{3}$$

$$V_i = D_i - \sum_{j=1}^{i-1} \langle D_i, U_j \rangle U_j \quad and \quad U_i = \frac{V_i}{|V_i|} \qquad i = 1, \ldots, n-1 \tag{4}$$

Where, $\langle D_i, U_j \rangle$ indicates dot product of two vectors and $|V_i|$ denotes the Frobenious norm of each vector Each of the orthogonal matrices V_i is divided by their Frobenious norm to make them normalized. After Gram-Schmidt orthogonalization procedure the orthogonal $(V_1, V_2, \ldots, V_{(n-1)})$ and $(U_1, U_2, \ldots, U_{(n-1)})$ orthonormal sets are extracted to compute difference matrix.

(3) The next stage of CMA based background modelling algorithm is computing the difference and common matrices based upon orthonormal basis after computed with below formula.

$$S_{k,diff} = \langle S_k, U_1 \rangle U_1 + \langle S_k, U_2 \rangle U_2 + \ldots + \langle S_k, U_{(n-1)} \rangle U_{(n-1)} \tag{5}$$

(4) Finally, subtracting the $S_{k,diff}$ from S_k gives the common matrix of the processed class, where $k = 1$ and S_{com} refers to common matrix of class.

$$S_{com} = S_k - S_{k,diff} \tag{6}$$

With this way, the training of set background can be represented by a unique 2-D frame, which is named as, common matrix. In other side, all details including noises and outliers of training set are stored in difference matrix $S_{k,diff}$.

When the rank of data becomes smaller than 2 in case of highly correlated data, then the CMA procedure concluded with a not meaningful common matrix that is undistinguishable with human eye. To overcome this problem, a low noise value between 0-1 is injected to each difference subspace in Eq. 2 in terms of reducing the correlation ratio among the processed images.

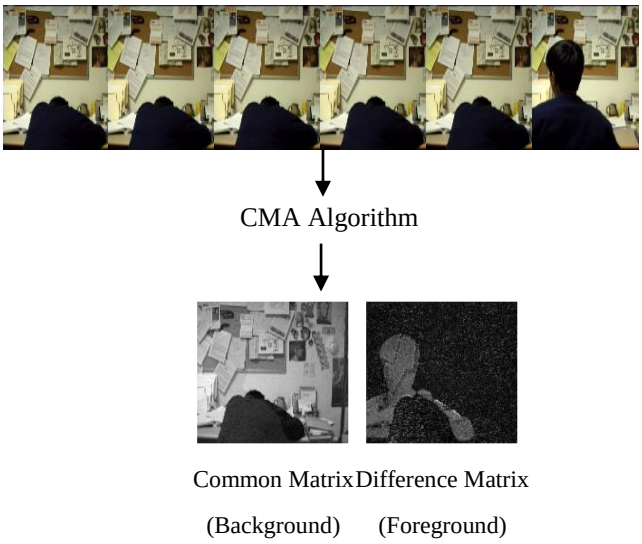


Figure 1. Demonstration of proposed method

From the Fig. 1, we can observe that the decomposed tensor generates two components:

(1) first component reserves the common matrix of training set, which denotes the acquired background model.
(2) the other component involves the difference matrix that refers to detail features of training set.

By using the CMA, we can see that foreground and changes are observed in difference matrix. Therefore, the strategy behind CMA provides a new way to detect moving and stable objects in a given dataset.

In order to reveal the foreground objects, the common matrix of test frame *(F)* is determined from the projection of incoming test frame onto the orthonormal basis returned by Gram-Schmit

procedure [14]. First of all, the difference matrix related to the test frame is calculated as shown in below equation.

$$F_{diff} = \langle F,U_1\rangle U_1 + \langle F,U_2\rangle U_2 +, \ldots, + \langle F,U_{(n-1)}\rangle U_{(n-1)} \quad (7)$$

Again, the common matrix corresponding to the test frame is computed by subtracting test frame from the difference matrix.

$$F_{com} = F - F_{diff} \quad (8)$$

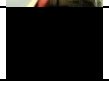
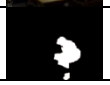
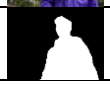
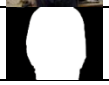
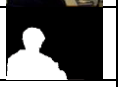
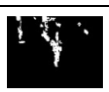
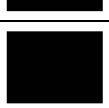
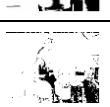
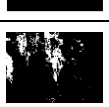
In case of revealing the foreground objects the difference between the common matrix of processed video and common matrix of processed frame is taken into account.

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

As shown in equation above, the difference of two common matrix presents foreground objects. In case of Moved Object and Camouflage videos, the difference of two common matrices are considered to find the foreground regions for other ones the absolute difference taken into account. The threshold value for each video are determined as follows; 0.1 for Camouflage, Bootstrap and Waving Trees, 0.2 for Foreground Aperture, Light Switch and Moved Object, and 0.3 for Time of Day video, respectively.

To obtain the pleasing visual results, some fixed morphological operations are applied on the foreground mask. Firstly, 5x5 median filter are utilized on the binary image. The connected components with the size of less than 20, are considered as ghost are removed by area open morphological operator. Then, the morphological closing procedure is utilized with disk structural element having size of 5 and binary holes are filled with morphological filling operator. Finally, 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.

Sequence	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.*							
CMA *Proposed*							

3. Performance Evaluation

3.1. Dataset

The experimental stages are conducted on well-known Wallflower Dataset. Numerous methods have been experimented on this dataset in order to objective and subjective performance comparison. Wallflower dataset [9] includes real-world background datasets as associated with dynamic events including Moved Object, Time of Day, Light Switch, Waving Trees, Camouflage, Bootstrapping and Foreground Aperture. In case of background modelling (obtaining Common Matrix), we have utilized predetermined train images, which are specified by the authors of dataset [8]. For each video, the first 199 images are taken to learn the background frame in case of training stage.

3.2. Subjective Results

In the present work, a simple thresholding methodology is realized in case of revealing the binary skeleton of objects. Since the difference of two common matrix gives changes, a fixed thresholding is carried over the absolute difference. The obtained visual results are demonstrated on Table 1.

To subjectively judge performance of both methods, the obtained visual results are compared with state of the art subspace and other methods, which are given as Single Gaussian (SG) [15], Mixture of Gaussian (MOG) [16], Kernel Density Estimation (KDE) [17], Subspace Learning PCA (SL-PCA) [18], Subspace Learning ICA (SL-ICA) [19], Subspace Learning via Incremental Non Negative Matrix Factorization (SL-INMF) [20] and Subspace Learning via Incremental Rank-(R1, R2, R3) Tensor (SL-IRT) [21]. For this purpose, the visual results determined in the work of Bouwman [22] are taken as ground on in case of performance comparison.

In Table 1, the first column denotes method's name, the other columns show video's name, respectively. Again, the first row and second row exhibit test image and related ground truth, and other rows demonstrates visual results returned from each method. From the exhibited results, we can observe that each method presents similar foreground objects in the meaning of obtained foreground skeleton.

By analysing results, one can note that results of MOG and KDE are closest to each other and more dominant than the SG method. The performance of SG, MOG and KDE are weakness to illumination changes due to stochastic characteristic of them as working based on the historical probability of pixels. To continue, we can see that subspace based method are more robust to light changes.

By comparing the PCA, ICA, INMF and IRT, we can emphasize that the result of IRT is the worst one in terms of preserving foreground skeleton. While the INMF shows good results in case of bootstrap video, but the same performance has not sustained in case of camouflage video.

Moreover, the results of PCA are similar to CMA method, however, the PCA method fails in case of indoor crowded scene (bootstrap). Furthermore, the proposed method not only robust to dynamic structures but also resistance to illumination change in case of foreground detection.

Table 2: Objective results on the Wallflower dataset.

Method	Error	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
CMA	FN	0	1017	882	26	172	929	2534			
Proposed.	FP	0	0	320	1106	616	157	485	8218	7016	7430

3.3. Objective Results

In addition to subjective results, the statistical results are obtained by considering the false positive (FP) and false negative (FN) pixels. With this aim, the ground truth images and foreground region are compared to find the number of erroneous pixels by counting the number of FP and FN. If a pixel marked as foreground in processed image, but marked as background in ground truth, then it is considered as FP. For opposite case, if a pixel marked as foreground by ground truth, but marked as background in processed image, then it is considered as FN. The sum of FP and FN denotes the error measure in terms of comparing the objective results. Specifically, the Total Errors, Total Errors without light switch (TE without LS) and Total Errors without Camouflage switch (TE without Camouflage) are demonstrated on the last columns of Table 2. The less error value indicates the best performance in terms of foreground segmentation.

The obtained statistical results are presented in Table 2. From the Table 2, one can derive that a superior performance is obtained by the proposed method, called CMA. In conjunction with visual results, the performance SG, MOG and KDE similar to each other. However, when the light switch video is excluded in case of performance evaluation, we can observe that the MOG and KDE generate better results than almost of all algorithms except CMA. These results are attributed to characteristic of probability

based foreground and change detection property. Moreover, when the camouflage video is excluded, the worst performance is produced by probabilistic based background subtraction methods. Also, comparing the subspace based methods including SL-PCA, SL-ICA and SL-INMF, one can note that the performance of SL-ICA is favourable against SL-PCA and SL-INMF. The performance of SL-PCA and SL-IRT are closest to each other, but difference bears in case of removing the light switch.

4. Conclusion

In this study, the impact of CMA is investigated for background modelling based foreground detection. The performance of the proposed method is compared with other well-known methods for dynamic backgrounds including Moved Objects, Time of Day, Light Switch Waving Trees, Camouflage, Bootstrap, Foreground Aperture. From the objective and subjective evaluation, it has observed that the proposed method exhibit eye pleasing results. The obtained experimental results present significant performance difference between PCA, ICA, INMF and probabilistic based methods (SG, MOG and KDE) in terms of accuracy and robustness to dynamic changes among the images for a given video. From the overall evaluation, one can emphasize that a smart post processing procedure is greatly needed to both accurately reveal the region of foreground object meanwhile eliminating the noisy pixels caused by uncontrolled changes, which are waving trees and illumination changes. As a future work, a comprehensive and universal background subtraction method is aimed to develop by using the concept of CMA.

References

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

[2] D. Dushnik, A. Schclar, A. Averbuch, Video segmentation via diffusion bases, arXiv preprint arXiv:1305.0218, (2013).

[3] W. Hu, X. Li, X. Zhang, X. Shi, S. Maybank, Z. Zhang, Incremental tensor subspace learning and its applications to foreground segmentation and tracking, International Journal of Computer Vision, 91 (2011) 303-327.

[4] M.G. Krishna, V.M. Aradhya, M. Ravishankar, D.R. Babu, LoPP: locality preserving projections for moving object detection, Procedia Technology, 4 (2012) 624-628.

[5] Y. Li, J. Yan, Y. Zhou, J. Yang, Optimum subspace learning and error correction for tensors, Computer Vision–ECCV 2010, Springer2010, pp. 790-803.

[6] Z. Zhang, G. Ely, S. Aeron, N. Hao, M. Kilmer, Novel methods for multilinear data completion and de-noising based on tensor-SVD, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition2014, pp. 3842-3849.

[7] S. Ergin, S. Çakir, Ö.N. Gerek, M.B. Gülmezoğlu, A new implementation of common matrix approach using third-order tensors for face recognition, Expert Systems with Applications, 38 (2011) 3246-3251.

[8] 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.

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

[10] 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.

[11] S. Günal, 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.

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

[13] K. Özkan, Ş. Işık, 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.

[14] 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.

[15] C.R. Wren, A. Azarbayejani, T. Darrell, A.P. Pentland, Pfinder: Real-time tracking of the human body, IEEE Transactions on pattern analysis and machine intelligence, 19 (1997) 780-785.

[16] 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.

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

[18] 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.

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

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

[21] 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.

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