OBJECTS TRACKING FROM VIDEO IN URBAN ENVIRONMENT BY LOW RANK RECOVERY

Ivo R. Draganov, Rumen P. Mironov

Radiocommunications and Videotechnologies Dept.
Faculty of Telecommunications, Technical University of Sofia
8 Kliment Ohridski Blvd., 1756 Sofia, Bulgaria

e-mail: {idraganov, rmironov}@tu-sofia.bg

Abstract

In this paper a novel scheme is presented for objects tracking in urban environment from a video, represented as three-way tensor. The low rank recovery (LRR) method is applied in 4 of its variations – the Robust Orthogonal Subspace Learning (ROSL), the Fast Linearized Alternating Direction Method with Adaptive Penalty (FastLADMAP), the Inexact Augmented Lagrange Multiplier (IALM), and the Linearized ADM with Adaptive Penalty (LADMAP). Two fusion functions are proposed for combining binary segmented frames from the top two performing decompositions – LRR FastLADMAP and ROSL, to get higher detection rate for foreground objects. The proposed approach is considered applicable in fields like video surveillance, production automation, public entertainment, etc.

1. INTRODUCTION

Moving objects detection on a frame-by-frame basis in video plays important role in numerous applications, such as video surveillance, vehicle traffic control, industrial production, logistics and many others. Detecting the boundaries of foreground objects over a complex background as accurately as possible over time is the first step towards further efficient analysis of the scene. Many methods rely on representing the input video as three-way tensor, which is being further decomposed under various schemes to sparse and low-rank components, which further analysis lead, in some of the applications, to detection of the moving objects – a hard task, when the background is complex.

In [1] Shijila B. et al. use low rank approximation as a mean for denoising and in parallel to it, moving objects detection in videos. They apply the l_1-Total Variation (TV) regularization approach and try to consolidate it with the nuclear norm, and the l_2-norm in a single framework. F1-measure of the detection process has been reported to vary between 0.98 and 0.99 over videos with Gaussian noise. The same authors had proposed another technique, using TVRPCA with a convex optimization, where the convergence of the resulting algorithm is being tried [2]. Reduced computational complexity is reported, compared to TVRPCA, DECOLOR and other algorithms, while the F1-measure changes between 0.52 and 0.90 among 6 test videos, comprising of various elements, including camera jitter, shadows, dynamic background, etc.

Yang et al. [3], following the basic idea of approximation based on rank functions and sparse conditions, try to find different mean than the nuclear norm. The authors propose nonconvex function, using Generalized Singular Value Thresholding (GSVT) and Alternating Direction Method of Multipliers (ADMM). The proposed approach appears effective in noise suppression, leading to Peak Signal to Noise Ratio (PSNR) of filtered frames between 35.2 and 39.4 dB. The F-measure of the object detection process changes between 0.3969 and 0.9153 for a set of 7 test videos, which proves its general applicability.

Matrix recovery with a low rank and using weights through a spectral graph is the approach, developed by Chen et al. [4], to detect salient objects. Both the low-rank and sparse matrices are employed in the process. The average F-measure deviates between 0.5678 and 0.8437 from testing with 4 datasets.

Sobral, in his thesis [5], looks towards both the matrix and tensor decomposition to low-rank and sparse components for object detection in video and other applications. Double-constrained version of the RPCA algorithm is developed for enhanced foreground detection. Spatial saliency maps ease the process in the case of dynamic scenes. Additionally, two decomposition algorithms, based on tensor representation and incrementally organized help to separate better the foreground from the background in multidimensional data. The F-measure reaches 85.96% for a RGB test sequence and 95.17% for a multispectral one for one of the implementations, following this approach.

Wang and Huang [6] use the $l_{2,1}$-norm minimization, while performing the low-rank approximation. The approach is applied over a set of multi-scale features, incorporating color, shape, texture, and other descriptors. All three parameters – Precision, Recall and F-measure reach values close to 0.9 during experimentation with MSRA10K dataset. More recently, Yang et al. [7] propose matrix recovery, based on spatiotemporal representation, in which approach the scalability is one of the imposed properties. The movement of objects in the video is detected by the optical flow algorithm and further the background is eliminated by low-rank regularization with consequent affine transform for suppressing its slight variation (motion) over time. PSNR in background recovery reaches 40 dB and SSIM – 0.9939 in some instances. The average F-measure in detecting foreground objects is 0.87, higher than numerous other decomposition methods. Low-rank tensor representation of videos is employed in their decomposition along with fused-sparse representation of salient type [8]. Three-way tensors assure preservation of the spatial and time relations among objects on a continuous basis over time. Three-dimensional local adaptive regression kernel (3D-LARK) redounds for finding the movement saliency on both the space and time independent coordinates in relation to the foreground. Salient objects are also being detected with a weighted matrix recovery process, where the table arrangement of data has a low rank [9]. Color, mutual location, and contour connectivity are some of the properties embedded in that representation to find the regions, classified as part of the background. Positive results are reported after comparing the algorithm efficiency with other 24 implementations in the practice. Structured sparse outliers [10] are another view over the moving objects in video, which is employed in the low-rank and sparse decompositions. Shkeri and Zhang propose with addition to that view the prior map, which assists in the reduction of the negative effect of significant illumination changes in the scene. F-measure of the background subtraction from testing with popular datasets reaches 0.8033 in some instances.

In this study the main aim is to evaluate the Low Rank Recovery (LRR) method in 4 of its base implementations - the Robust Orthogonal Subspace Learning (ROSL), the Fast Linearized Alternating Direction Method with Adaptive Penalty (FastLADMAP), the Inexact Augmented Lagrange Multiplier (IALM), and the Linearized ADM with Adaptive Penalty (LADMAP) over a popular video test set. Based on the obtained results two new fusion schemes are proposed and tested, which yield higher Detection Rate in one instance and higher Precision in the other. In Section 2 of the paper, the description of the algorithms is given, followed by experimental results in Section 3 and discussion in Section 4. Section 5 contains the conclusion.

2. ALGORITHMS DESCRIPTION

2.1. LRR IALM

In the base of the LRR IALM algorithm lays the matrix completion task. It has been proven that for a matrix A of rank r (typically a low one), having some missing elements, the following optimization procedure may lead to its restoration [11]:

$$\min_A \|A\|_*, given\ A_{ij} = D_{ij}, \forall (i,j) \epsilon \Omega, \quad (1)$$

where Ω is the entity of samples' indices, D – an input matrix of real noisy elements, mxn in number. The exact Augmented Lagrange Multiplier (ALM) method could be applied in this case, following [11]:

$$\min_A \|A\|_*, \quad given\ A + E = D, \quad \pi_\Omega(E) = 0, \quad (2)$$

where π_Ω is a transformation that puts all elements, falling outside Ω to be 0; E – matrix of the additive errors. The partial augmented Lagrangian function with the multiplier Y in it, then, could be found from [11]:

$$L(A,E,Y,\mu) = \|A\|_* + \langle Y, D - A - E \rangle + \frac{\mu}{2}\|D - A - E\|_F^2, \quad (3)$$

where $\mu > 0$ is a scalar, and F denotes the Frobenius norm. Depending on the constraints imposed over the values of $\pi_\Omega(E)$, the IALM algorithm takes its form.

2.2. LRR LADMAP

The linearized ADM could be expressed as [12]:

$$\boldsymbol{x}_{k+1} = arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \frac{\beta}{2}\|\boldsymbol{\mathcal{A}}(\boldsymbol{x}) + \boldsymbol{\mathcal{B}}(\boldsymbol{y}_k) - \boldsymbol{c} + \lambda_k/\beta\|^2, \quad (4)$$

where $\boldsymbol{x}$ (typically, the noisy input), $\boldsymbol{y}$ and $\boldsymbol{c}$ are matrices (could be also vectors, which is not of interest to this study, since processing is done over video frames), f – a convex function, λ – Lagrange multiplier, β – penalty parameter, k – the number of the current iteration, and $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ – linear transformations. Equation (4) could be approximated as [12]:

$$\boldsymbol{x}_{k+1} = \underset{\boldsymbol{x}}{\operatorname{argmin}} f(\boldsymbol{x}) + \frac{\beta\eta_A}{2}\left\|\boldsymbol{x} - \boldsymbol{x}_k + \boldsymbol{\mathcal{A}}^*\left(\lambda_k + \beta(\boldsymbol{\mathcal{A}}(\boldsymbol{x}_k) + \boldsymbol{\mathcal{B}}(\boldsymbol{y}_k) - \boldsymbol{c})\right)/(\beta\eta_A)\right\|^2, \quad (5)$$

where $\boldsymbol{\mathcal{A}}^*$ is the adjoint of $\boldsymbol{\mathcal{A}}$, η_A – scaling parameter with positive value. The adaptive penalty could be found by updating process, according to [12]:

$$\beta_{k+1} = \min(\beta_{max}, \rho\beta_k) \quad (6)$$

with ρ – a scalar, greater or equal to 1, found separately.

2.3. LRR FastLADMAP

The LRR solution to the problem, posed in Section 2.2, could be found approximately following the criterion [12]:

$$\min_{\boldsymbol{Z},\boldsymbol{E}} \beta\left(\|\boldsymbol{Z}\|_* + \mu\|\boldsymbol{E}\|_{2,1}\right) + \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{R} - \boldsymbol{E}\|^2, \quad (7)$$

where β is relaxation parameter with a non-negative value, $\boldsymbol{Z}$ – coefficient matrix, $\boldsymbol{R}$ – an additional matrix. Gradual decrease of β could be achieved by [12]:

$$\beta_{k+1} = \max(\beta_{min}, \theta\beta_k), \quad (8)$$

where θ is a constant.

2.4. LRR ROSL

Following the general case of RPCA recovering of a low-rank matrix A, given input matrix X with noise [13]:

$$\min_{A,E}\|A\|_* + \lambda\|E\|_1, given\ A + E = X, \quad (9)$$

where $||.||_*$ is the nuclear norm and $||.||_1$ – the l_1-norm. It has been proven that [13]:

$$\|A\|_* = \min_{D,\alpha}\frac{1}{2}(\|D\|_F^2 + \|\alpha\|_F^2), given\ A = D\alpha, \quad (10)$$

$$\|A\|_* = \|\alpha\|_{row-1}, given\ A = D\alpha, D^TD = I_k, \quad (11)$$

where D is a spanning matrix of the ordinary orthonormal space the data is being present, α – vector of coefficients, revealing the influence degree of the components of D, and I – the identity matrix. From (10) and (11) it has been shown that the low-rank recovery by the ROSL algorithm could be accomplished by solving [13]:

$$\min_{E,D,\alpha}\|\alpha\|_{row-1} + \lambda\|E\|_1, given\ D\alpha + E = X, D^TD = I_k, \forall i. \quad (12)$$

2.5. Fusion OR and AND

In order to get higher Detection Rate over the pixels of moving objects the logical operation OR is performed between the binary frames, obtained from the LRR decomposition algorithms. The result is another binary frame. For the purposes of getting higher Precision, the logical operation AND is performed in analogous way. It is expected to have, in this second case, finer detection of the boundaries of moving objects, which are more contrast to the background with less False Positives.

3. EXPERIMENTAL RESULTS

The hardware platform, used for testing, consists of 64-bit Intel Core i5 processor with 4 cores, working on a base frequency of 3.1 GHz, along with 12 GB of RAM and 2 TB 7200 rpm HDD. It is controlled by Ubuntu 14.04 LTS operating system, and all tests are implemented in the Matlab R2016a simulation environment. All decompositions are based on implementations from the LRS Library v. 1.0.10 [14]. The video dataset contains 6 24-bit color videos, captured in urban environment (in- and outdoor with at least 1 person or a vehicle moving over complex background), some of which with changing illumination conditions (Table 1). They are derived from the LASIESTA database [15], in which every video has its groundtruth correspondent (binary) video.

Table 1. Video dataset for testing

Video	Width, px	Height, px	FPS	Frames
I_IL_01	352	288	10	300
O_CL_01	352	288	10	250
I_OC_02	352	288	10	300
I_SI_01	352	288	10	220
O_RA_02	352	288	10	370
O_SU_02	352	288	10	400

The following parameters are measured on a pixel basis along all frames from a video:

$$DR = TP/(TP+FN), \quad (13)$$

$$Precision = TP/(TP+FP), \quad (14)$$

$$F = 2DR.Precision/(DR+Precision), \quad (15)$$

where TP is the True Positive value, expressing the count of pixels correctly discovered as part of a moving object, FN – False Negatives – the count of pixels, belonging to moving objects, but marked as part of the background, and FP – False Positives – all pixels, labeled as part of moving objects, but being

part of the background. Processing time (*PT*) of just applying the decomposition and Full time (*FT*), including input-output operations, are also measured. They are shown in Fig. 1 and the average *DT*, *Precision* and *F* measure – in Fig. 2.

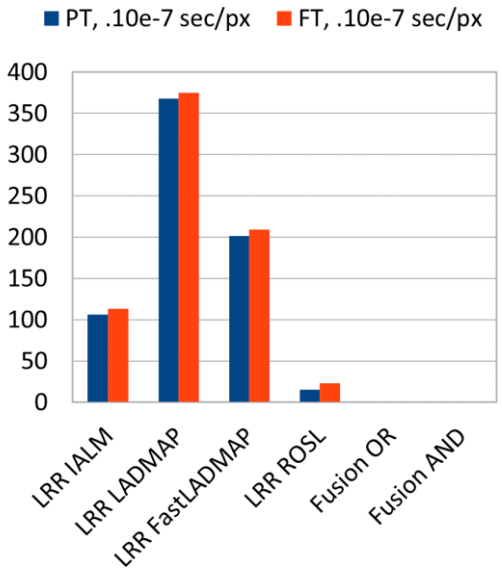

Fig. 1. Processing and Full times of the decomposition schemes

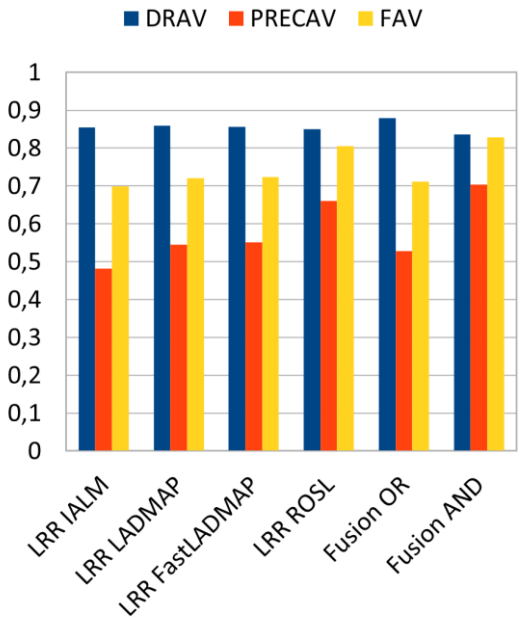

Fig. 2. Detection Rate, Precision and F measure of tested algorithms

The deviations of the same, accuracy defining parameters, are visible in Fig. 3.

Sample processed frames from the two most accurate decompositions and the newly proposed two fusion algorithms are included in Fig. 4.

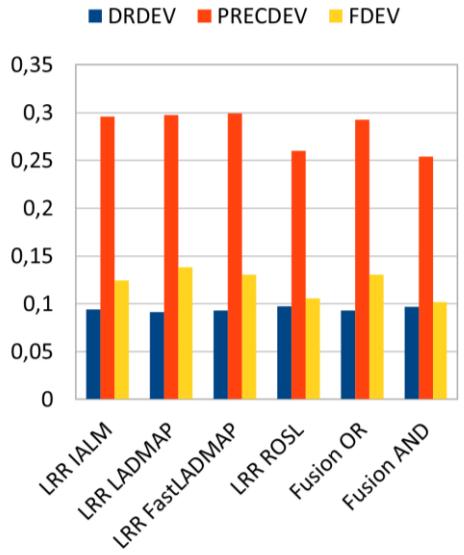

Fig. 3. Deviations of the Detection Rate, Precision and F measure

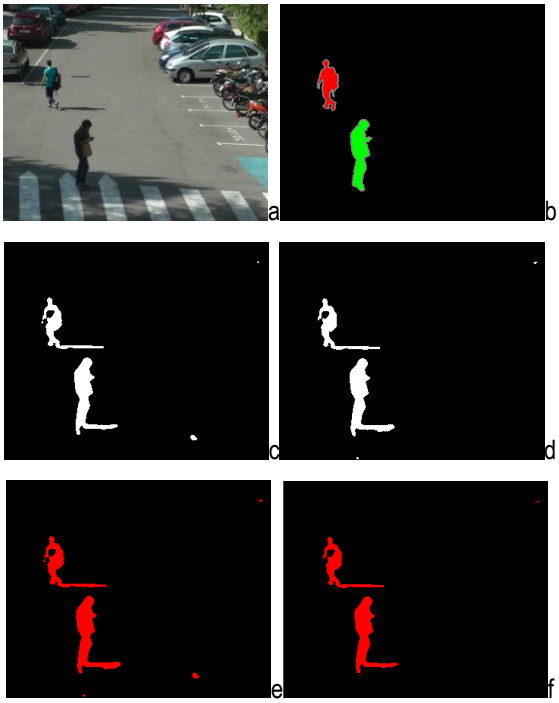

Fig. 4. Frame 257 from the O_SU_02 video: a – original, b – groundtruth, c – FastLADMAP, d – ROSL, e – Fusion OR, f – Fusion AND

4. DISCUSSION

The fastest algorithm is LRR ROSL with *PT* = $15.46.10^{-7}$ sec/px, almost 24 times faster than the slowest LRR LADMAP (Fig. 1). Fusion OR and Fusion AND take $0.58.10^{-7}$ and $0.42.10^{-7}$ sec/px, respectively. These times, although much smaller than the *PT* of all decompositions, are not negligible and should be taken into consideration.

Based on the achieved *F* measures, the most accurate single decomposition implementation is the LRR

ROSL (0.8050), followed by the LRR FastLADMAP with 0.7244 (Fig. 2). There is no significant difference in deviation of accuracy parameters among all tested implementations (Fig. 3).

Fusion OR over the ROSL and FastLADMAP leads to higher *DR* = 0.8791 than any of the four basic decompositions (Fig. 2). Fusion AND on the other hand yield higher *Precision* = 0,7040 – 1.07 times higher than the ROSL, turning out to be the most precise single algorithm (Fig. 2).

At visual inspection of the processed videos, it could be observed that the LRR basic decomposition algorithms produce in some of the cases spots within the binary frames, which are not part of a moving object, but are appearing artefact from the video compression or present noise. In some cases, such spots are result of slight movement of an object, which casts shadow over the visible portion in the frame. The higher the *DR* is, the more of these spots may appear. Examples for such False Positives are seen in Fig. 4 c and d in the lower and top-right area of the frame. Applying the Fusion OR do not remove them (Fig. 4 e), but the Fusion AND does – Fig. 4 f. In general, there is a reduction in the *DR* for the Fusion AND algorithm but getting higher *Precision* for the detection process. Another inconsistency is the appearance of holes in moving objects – at uneven illumination, for example due to reflected light (Fig. 4 a, c-f).

5. CONCLUSION

In this paper, 4 single LRR decomposition algorithms are tested over videos with varying conditions for detection of moving objects – the IALM, LADMAP, FastLADMAP, and the ROSL. Resulting frames from two of the most accurate – the FastLADMAP and ROSL are being passed to a fusion process – once with the logical OR operator and once – with the AND. Higher detection rate is the result in the first case with lower precision and the opposite in the second case. Both fusion schemes are considered applicable in numerous systems, employing video analysis, e.g., video surveillance, vehicle traffic control, automated production, etc.

ACKNOWLEDGEMENT

This work was supported by the National Science Fund at the Ministry of Education and Science, Republic of Bulgaria, within the project KP-06-H27/16 „Development of efficient methods and algorithms for tensor-based processing and analysis of multidimensional images with application in interdisciplinary areas “.

References

[1] Shijila, B., Tom, A. J., & George, S. N. (2019). Simultaneous denoising and moving object detection using low rank approximation. Future Generation Computer Systems, 90, 198-210.

[2] Shijila, B., Tom, A. J., & George, S. N. (2018). Moving object detection by low rank approximation and l1-TV regularization on RPCA framework. Journal of Visual Communication and Image Representation, 56, 188-200.

[3] Yang, Z., Fan, L., Yang, Y., Yang, Z., & Gui, G. (2019). Generalized singular value thresholding operator based nonconvex low-rank and sparse decomposition for moving object detection. Journal of the Franklin Institute, 356(16), 10138-10154.

[4] Chen, J., Chen, J., Ling, H., Cao, H., Sun, W., Fan, Y., & Wu, W. (2018). Salient object detection via spectral graph weighted low rank matrix recovery. Journal of Visual Communication and Image Representation, 50, 270-279.

[5] Sobral, A. C. (2017). Robust low-rank and sparse decomposition for moving object detection: from matrices to tensors (Doctoral dissertation, Université de La Rochelle).

[6] Wang, S., & Huang, A. (2017). Salient object detection with low-rank approximation and ℓ2, 1-norm minimization. Image and Vision Computing, 57, 67-77.

[7] Yang, J., Shi, W., Yue, H., Li, K., Ma, J., & Hou, C. (2020). Spatiotemporally scalable matrix recovery for background modeling and moving object detection. Signal Processing, 168, 107362.

[8] Hu, W., Yang, Y., Zhang, W., & Xie, Y. (2016). Moving object detection using tensor-based low-rank and saliently fused-sparse decomposition. IEEE Transactions on Image Processing, 26(2), 724-737.

[9] Tang, C., Wang, P., Zhang, C., & Li, W. (2016). Salient object detection via weighted low rank matrix recovery. IEEE Signal Processing Letters, 24(4), 490-494.

[10] Shakeri, M., & Zhang, H. (2017). Moving object detection in time-lapse or motion trigger image sequences using low-rank and invariant sparse decomposition. In Proceedings of the IEEE International Conference on Computer Vision (pp. 5123-5131).

[11] Lin, Z., Chen, M., & Ma, Y. (2010). The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint, arXiv:1009.5055.

[12] Lin, Z., Liu, R., & Su, Z. (2011). Linearized alternating direction method with adaptive penalty for low-rank representation. arXiv preprint, arXiv:1109.0367.

[13] Shu, X., Porikli, F., & Ahuja, N. (2014). Robust orthonormal subspace learning: Efficient recovery of corrupted low-rank matrices. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3874-3881).

[14] Sobral, A., Bouwmans, T., & Zahzah, E. H. (2016). Lrslibrary: Low-rank and sparse tools for background modeling and subtraction in videos. In: Bouwmans, T., Aybat, N., Zahzah, E.-H. (eds.) Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing, CRC Press.

[15] Cuevas, C., Yáñez, E. M., & García, N. (2016). Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA. Computer Vision and Image Understanding, 152, 103-117.