Moving Object Detection via Robust Low Rank Matrix Decomposition with IRLS scheme

Charles Guyon, Thierry Bouwmans, El-Hadi Zahzah

Laboratoire MIA (Mathematiques, Image et Applications)
University of La Rochelle
thierry.bouwmans@univ-lr.fr

Abstract. Moving object detection is a key step in video surveillance system. Recently, Robust Principal Components Analysis (RPCA) shows a nice framework to separate moving objects from the background when the camera is fixed. The background sequence is then modeled by a low rank subspace that can gradually change over time, while the moving objects constitute the correlated sparse outliers. In this paper, we propose to use a low-rank matrix factorization with IRLS (Iteratively Reweighted Least Squares) scheme for RPCA decomposition and to address in the minimization process the spatial connexity of the pixels. Experimental results on different datasets show the pertinence of the proposed method.

1   Introduction

The detection of moving objects is the basic low-level operations in video analysis. The basic operation consists of separating the moving objects called "foreground" from the static information called "background" [1][2][3]. Recent reseach on robust PCA shows qualitative visual results with the background variations appromatively lying in a low dimension subspace, and the sparse part being the moving objects. First, Candes et al. [4] proposed a convex optimization to address the robust PCA problem. The observation matrix is assumed represented as $A = L + S$ where L is a low-rank matrix and S must be sparse matrix with a small fraction of nonzero entries. This research seeks to solve for L with the following optimization problem:

$$\min_{L,S} \ ||L||_* + \lambda ||S||_1 \quad \text{subj} \quad A = L + S \tag{1}$$

where $||.||_*$ and $||.||_1$ are the nuclear norm (which is the L_1 norm of singular value) and l_1 norm, respectively, and $\lambda > 0$ is an arbitrary balanced parameter. Under these minimal assumptions, this approach called Principal Component Pursuit (PCP) solution perfectly recovers the low-rank and the sparse matrices. Candes et al. [4] showed results on face images and background modeling that demonstrated encouraging performance. However, PCP is limited to the low-rank component being exactly low-rank and the sparse component being exactly sparse but the observations in real applications are often corrupted by noise affecting every entry of the data matrix. Therefore, Zhou et al. [5] proposed a

stable PCP (SPCP) that guarantee stable and accurate recovery in the presence of entry-wise noise. However, PCP and SPCP present the following limitation: When few columns of the data matrix are generated by mechanims different from the rest of the columns, the existence of these outlying columns tends to destroy the low-rank structure of the data matrix.

In this paper, we propose a novel algorithm for moving object detection based on a robust matrix factorization. For a data matrix A, we assume that it is approximatively low-rank, and a small part of this matrix is corrupted by the outliers. The aim of the proposed method is to alleviate the limitation of PCP and SPCP by addressing the spatial connexity of the pixel to obtain a robust estimation of the true low-rank and the sparse structure of the matrices L and S. Our contributions can be summarized as follows: 1) Addition of spatial constraint to minimization process, 2) IRLS alternating scheme for weighted the 2-parameters $||.||_{\alpha,\beta}$ for matrix low rank decomposition. The rest of this paper is organized as follows. In Section 2 and 3, we present the proposed method based on a robust low-rank matrix factorization which allows us to detect moving objects in dynamic backgrounds. Then, in Section 4, we present comparison and evaluation versus the state-of-the-art methods. Finally, the conclusion is provided in Section 5.

2   1-D Case: L_1 Minimization with spatial constraint

In most applications, video surveillance data is assumed to be compose of background and moving objects components. We improved the decomposition (1) by adding a third component which models the noise. We believe the regression task as a crucial part of the proposed decomposition algorithm and illustrate the main keys through a simple 1-D example shown in Fig. 1. Consider the following minimization problem, where A is a dictionary matrix (row order) and b is a row vector,

$$\operatorname*{argmin}_{x}\ ||Ax - b||_{\alpha} + \mu||\nabla_s \phi(E)||_1, \quad E = |b - Ax| \qquad (2)$$

By this, we impose the error E should be a connexe shape, through the TV (Total Variation) of the error must be small, where ∇_s is the spatial gradient. Note that the problem is convexe for $\alpha > 1$ and the standard IRLS (Iteratively reweighted least squares) scheme is given by,

$$\operatorname*{argmin}_{x}\ ||Ax - b||_{\alpha} \quad \left| \quad \begin{array}{l} D^{(i)} = \operatorname{diag}((\varepsilon + |b - Ax^{(i)}|)^{\alpha-2}) \\ x^{(i+1)} = (A^t D^{(i)} A)^{-1} A^t D^{(i)} b \end{array} \right. \qquad (3)$$

As note in [6] and [7], it was proven that a suitable IRLS method is convergent for $1 < \alpha < 3$. Some modifications allow to get more numericaly stability and let us to choose freely $\alpha \in [1, \infty[$ with an adapted λ_{opt}.

$$\begin{array}{l} r^{(i)} = b - Ax^{(i)} \\ D = \operatorname{diag}((\epsilon + |r^{(i)}|)^{\alpha-2}) \\ y^{(i)} = (A'DA)^{-1} A'Dr^{(i)} \\ x^{(i+1)} = x^{(i)} + (1 + \lambda_{opt})y^{(i)} \end{array} \qquad (4)$$
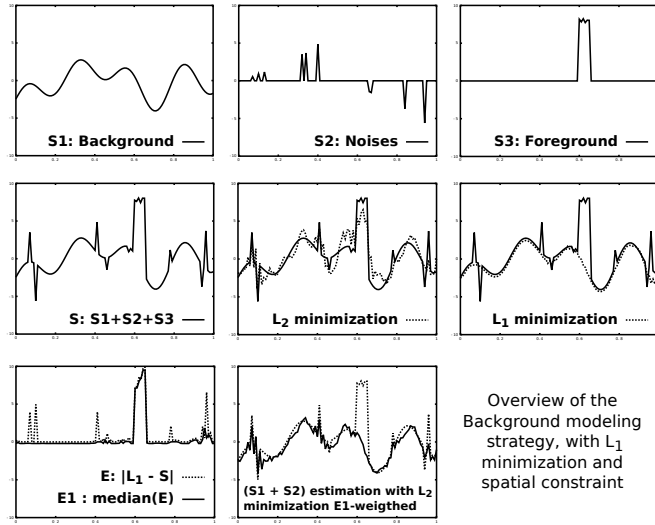
Fig. 1. Schematic example for the 1-D case: We illustrate different fitting strategies with random dictionary basis (cosine function, wavelets, ...) and a composite signal S.

The algorithm is twice iterative, where we try to get an optimal x and an optimal λ at each step.

$$
\begin{aligned}
&c^{(i)} = Ay^{(i)} \\
&d^{(i)} = b - A(x^{(i)} + y^{(i)}) \\
&\lambda_{opt} = \underset{\lambda}{\operatorname{argmin}} \ ||c^{(i)}\lambda - d^{(i)}||_\alpha
\end{aligned}
\qquad
\begin{aligned}
&\lambda^{(0)} = \Lambda(\alpha) \\
&s^{(k)} = d - c\lambda^{(k)} \\
&E = \operatorname{diag}((\epsilon + |s^{(k)}|)^{\alpha-2}) \\
&\lambda^{(k+1)} = \lambda^{(k)} + (1 + \Lambda(\alpha))\frac{c^t E s^{(k)}}{c^t E c}
\end{aligned}
\tag{5}
$$

With a fixed λ_{opt}, we should choose $\lambda_{opt} = \Lambda(\alpha)$ to allow $1 < \alpha < \infty$ and distinguish three cases:

$$
\Lambda(\alpha) = \begin{cases} 5/6 & \text{if} \quad \alpha \leq 1 \\ -\frac{2}{3}\alpha + \frac{3}{2} & \text{if} \quad 1 < \alpha < 1 + \frac{3}{4} \\ \frac{1}{\alpha-1} - 1 & \text{if} \quad \alpha \geq 1 + \frac{3}{4} \end{cases}
\tag{6}
$$

Observe for case $\alpha > 2$, convergence is achieved when $0 < 1 + \lambda < \frac{2}{\alpha-1}$. More generally, we can solve matrix regression problem with two parameters norm (α, β) and a weighted matrix (W).

$$
\min_X \ ||AX - B||_{\underset{W}{\alpha,\beta}} \quad \text{where} \quad ||M_{ij}||_{\underset{W}{\alpha,\beta}} = \Big(\sum_{i=1}^n \big(\sum_{j=1}^m W_{ij}\,|M_{ij}|^\beta\big)^{\frac{\alpha}{\beta}}\Big)^{\frac{1}{\alpha}}
\tag{7}
$$

This is solved by the following algorithm:

$$
\begin{aligned}
&\text{Until X is stable, repeat on each } k\text{-columns} \\
&\left|
\begin{aligned}
R \ &\leftarrow B - AX \\
S \ &\leftarrow \varepsilon + |R| \\
D \ &\leftarrow \operatorname{diag}(S_{ik}^{\beta-2} \circ (\textstyle\sum_j (S_{ij}^\beta \circ W_{ij}))^{\frac{\alpha}{\beta}-1} \circ W_{ik}) \\
X_{ik} &\leftarrow X_{ik} + (1 + \Lambda(\max(\alpha,\beta)))(A^t D A)^{-1} A^t D R_{ik}
\end{aligned}
\right.
\end{aligned}
\tag{8}
$$

3   Moving Object Detection via Robust Low-Rank Matrix Factorization

Let us denote the training video sequences $A = \{I_1, ...I_m\}, A \in \mathbb{R}^{n \times m}$ where I_j is a vectorized frame at time j and m is the number of training frames. Let each pixel (x,y) be characterized by its intensity in the gray scale. The decomposition involves the following model:

$$A = L + S = BC + S \tag{9}$$

where B is a low-rank matrix corresponding to the background model plus noise, and C allows us to reconstruct L by linear combination. S is a matrix which corresponds to the moving objects. The model involves the error reconstruction determined by the following constraints:

$$\min_{B \in \mathbb{R}^{n \times p}, C \in \mathbb{R}^{p \times m}} ||(A - BC) \circ W||_{\alpha,\beta} + \mu ||BC||_* \tag{10}$$

where $||.||_*$ denote the nuclear norm. The decomposition is split into two parts. firstly, we track 1-Rank decomposition since the first eigen-vector is strongly dominant in video surveillance.

$$
\begin{aligned}
R_1 &= A - B_1 C_1 \\
R &= A - B_1 C_1 - B_r C_r
\end{aligned}
\left|
\begin{aligned}
&(1) \ \min_{B_1, C_1} ||R_1||_{1,1} \\
&(2) \ \min_{B_r, C_r} ||R \circ \phi(R_1)||_{2,1 \to 0}
\end{aligned}
\right.
\tag{11}
$$

We use $||.||_{2,1 \to 0}$ instead of usual $||.||_{1,1}$ because it forces spatial homogeneous fitting. $\beta = (1 \to 0)$ means the β parameter decreases during iteration. First, we search a solution of the convex problem $||.||_{2,1}$, then use the solution as an initial guess for non-convex problem $||.||_{2,(1-\epsilon)}$. Finally, we find a local minimum of $||.||_{2,0}$ and hope that is near of the global minimum. In the case where $\alpha = \beta = 2$, the decomposition is usually solved by a SVD (Singular Value Decomposition). Thus, our SVD algorithm can be seen as an iterative regression. The proposed scheme determines alternatively the optimal coefficients, it means searching C for B fixed and searching B for C fixed.

$$
\begin{aligned}
C^{(k+1)} &= (A^t A)^{-1} A^t B^{(k)} \\
\bar{C}^{(k+1)} &= C^{(k+1)} \sqrt{C^{t(k+1)} C^{(k+1)}}^{-1} \\
B^{(k+1)} &= (A^t A)^{-1} A^t \bar{C}^{(k+1)}
\end{aligned}
\tag{12}
$$

Additionnaly this alternating regression framework allows to associate a weigthed matrix W which is entrywise multiplied to the error ($\circ$ denotes the Hadamard product).

$$\min_{B,C} ||(A - BC) \circ W||_{\alpha,\beta} \tag{13}$$

We define a function ϕ that have two goals, smooth the error (like spatial median filtering) and transform the error for obtain a suitable mask for regression.

$$W = \exp(-\sigma \, \phi(|A - BC|)) \tag{14}$$

By including local penalty as a constraint in a RPCA decomposition, this explicitly increases local coherence of the error (therefore foreground/moving object).

4   Experimental Results

We have compared the proposed approach with recent RPCA approaches: LBD [8], LRR [9], SADAL [10] and GRASTA [11] algorithms. The experiments were conducted qualitatively and quantitatively on the Wallflower[1] dataset [12], I2R[2] dataset [13] and CDW[3] dataset [14]. The algorithms are implemented in Matlab.

4.1   Wallflower dataset

This dataset provided by Toyama et al. [12] consists of seven video sequences, with each sequence presenting one of the difficulties a practical task is likely to encounter. The images are 160×120 pixels. For each sequence, the ground truth is provided for one image when the algorithm has to show its robustness to a specific change in the scene. Thus, the performance is evaluated against hand-segmented ground truth. The Fig. 2 and Fig. 3 show the qualitative results. For the quantitative evaluation, we used metrics based on the true negative (TN), true positive (TP), false negative (FN), false positive (FP) detections. Then, we computed the detection rate, the precision and the F-measure [15]. A good performance is then obtained when the F-measure is closed to 1. Table 1 shows the results obtained for each algorithms. The F-measure value of "Moved Object" sequence can't be computed due to the absence of true positives in its ground-truth. We have highlighted when one algorithm outperforms the other ones. As these encouraging results are obtained by using one ground-truth image, we have evaluated the proposed method on a dataset with more ground-truth images in the following sub-section.

Dataset	Sequence	Frame	LBD	LRR	SADAL	GRASTA	Our method
	Bootstrap	00299	**70.18**	69.40	67.02	55.37	58.78
	Foreground Aperture	00489	60.69	50.10	71.55	75.14	**75.19**
	Light Switch	01865	57.74	36.76	**69.33**	28.35	58.49
Wallflower	Moved Objects	00985	-	-	-	-	-
	Time of Day	01850	71.43	54.41	**80.84**	79.80	80.43
	Waving Trees	00247	62.65	50.74	81.67	84.16	**84.70**
	Camouflage	00251	70.58	70.49	76.00	70.34	**82.27**

Table 1. F-measure in percentage for LBD [8], LRR [9], SADAL [10], GRASTA [11], and our method (direct one-to-one correspondence with Fig. 3).

4.2   I2R dataset

This dataset provided by [13] consists of nine video sequences, which each sequence presenting dynamic backgrounds or illumination changes. The size of the images is 176×144 pixels. For each sequence, the ground truth is provided for 20 images. Among this dataset, we have chosen to show results on six sequences due the limitation of pages (see Fig. 4 and Fig 5). Tab. 2 shows the corresponding qualitative results.

---

[1] http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/testimages.htm

[2] http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html

[3] http://www.changedetection.net/

Fig. 2. Wallflower dataset. From left to right (split in two columns): The original image, the background model computed by our method and their difference with enhanced variance.
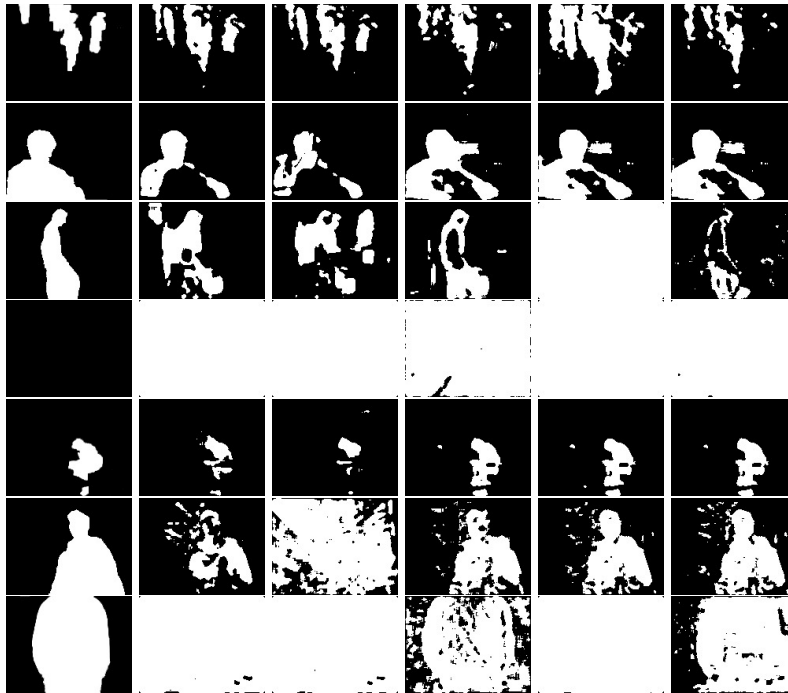

Fig. 3. Moving object detection masks on the Wallflower dataset. From left to right: Ground Truth, LBD, LRR, SADAL, GRASTA, our method. For *Moved Object* and *Camouflage* sequences, two fails occurs due of breaking *a priori*: Area of background must be bigger than foreground and inherently of evaluation procedure (*i.e.* F-measure), ground truth must not be void.

Dataset	Sequence	Frame	LBD	LRR	SADAL	GRASTA	Our method
I2R	Campus	01650	59.93	59.99	68.14	61.84	**75.07**
	Curtain	22772	**91.08**	88.34	91.01	89.88	90.73
	Escalator	02424	65.28	63.41	59.81	**68.31**	63.68
	Hall	02926	73.58	69.75	77.57	78.25	**79.26**
	Shopping Mall	01862	80.15	77.88	**82.80**	79.54	81.94
	Water Surface	01499	90.37	83.57	**92.01**	90.97	91.02

Table 2. F-measure in percentage for LBD [8], LRR [9], SADAL [10], GRASTA [11], and our method (direct one-to-one correspondence with Fig. 5).

Fig. 4. I2R dataset. From left to right (split in two columns): The original image,the background model and their difference.
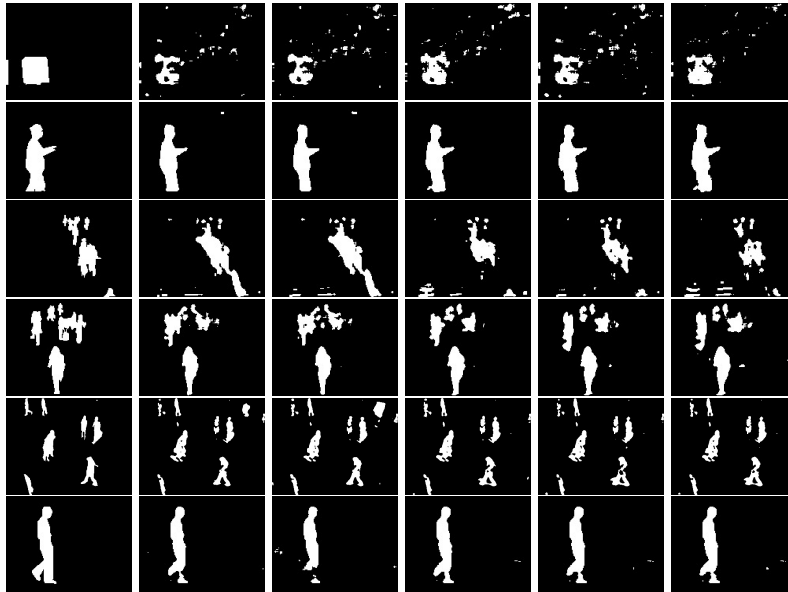

Fig. 5. Moving object detection masks on the I2R dataset.From left to right: Ground Truth, LBD, LRR, SADAL, GRASTA, our method.

4.3   Change Detection Workshop dataset

This dataset provided by [13] consists of 31 video sequences. We have chosen to show results of ten videos, corresponding of the subsets *baseline* and *dynamic backgrounds* categories. The size of the images is variable. For each sequence, the ground truth is provided for each frame. Among this dataset, the results (see Fig. 5 and Fig. 6). Tab. 3 shows the corresponding qualitative results.

Dataset	Sequence	Frame	LBD	LRR	SADAL	GRASTA	Our method
Baseline	PETS2006	00500	56.50	55.25	74.09	69.20	**78.12**
	highway	00800	87.48	83.99	93.82	92.29	**94.44**
	office	01886	76.01	58.42	75.67	78.36	**80.71**
	pedestrians	00471	**82.74**	79.02	81.79	81.24	81.54
Dynamic Back.	boats	02000	84.01	82.29	83.36	**84.02**	83.39
	canoe	00966	78.82	75.47	58.69	**79.73**	68.19
	fall	01492	62.22	59.46	78.49	65.27	**88.39**
	fountain01	00718	35.85	32.82	38.14	37.25	**54.46**
	fountain02	00740	85.21	83.93	87.24	**87.25**	86.23
	overpass	02529	72.01	72.55	**78.83**	78.65	70.33

Table 3. F-measure results of LBD [8], LRR [9], SADAL [10], GRASTA [11], and our method apply on partial CDW dataset (direct one-to-one correspondence with Fig. 6).

4.4   Results interpretation

The proposed approach outperforms the other algorithms for three sequences for Wallower dataset, five sequences for CDW dataset and obtain similar performance with SADAL for I2R dataset. Globally, our method provide better result for ten sequences, followed by SADAL and for the rest results are still acceptable. For this comparison, we used default parameters setting provided by the authors for all the algorithms and computed low rank decomposition on time window of 200 consecutive frames. Evaluation of time computing is difficult because thoses algorithms are implemented in non-optimized matlab codes, but we can give a line of thought (in timing of algorithm by timing of SVD): LBD 50, LRR 10, SADAL 70, GRASTA 240, and our method 110 (in Time/(SVD Time)).

4.5   Note on the evaluation procedure

Only grayscale images are computed, otherwise green channel is selected. Some high resolution video are subsampled, reasonably for memory allocating since for example, a 200 frames video of 640×480 cost 492 Mb. Additionally, a 5×5 median filter is postprocessed in order to suppress peak noise but this does not offer much advantage over a noise sensitive algorithm. For each sequence, the thresholding value is automatically choose for maximize the F-measure. That is an optimistic evaluation and may differ from real case, but algorithms are compared on the same manner and are not favored by a particular threshold selection strategy.
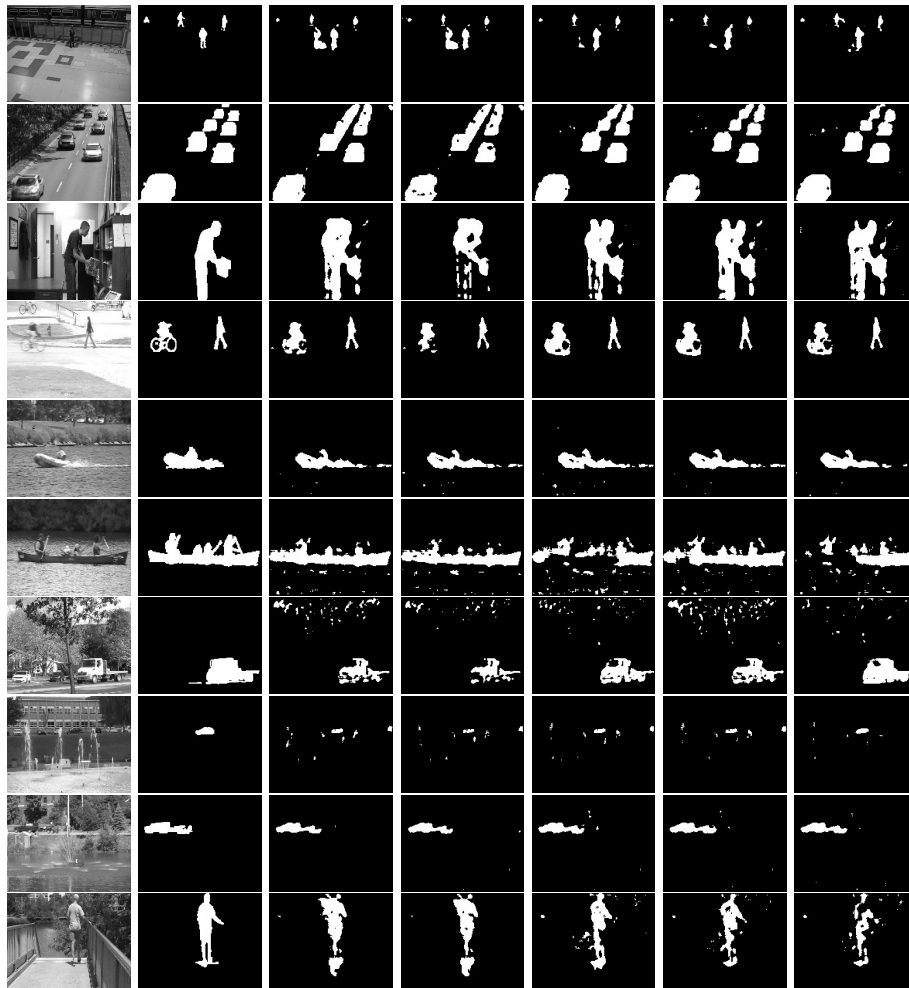
Fig. 6. Moving object detection masks on the CDW dataset. From left to right: Original, Ground Truth, LBD, LRR, SADAL, GRASTA, our method.
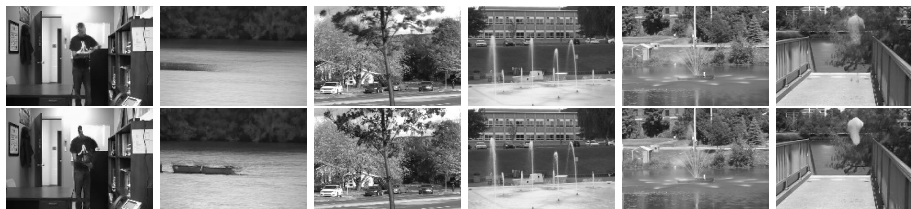

Fig. 7. Background modeling of selected sequences. By row: GRASTA and our method. Edges of shape are less fuzzy on the second row due to the add of the local constraint.

5   Conclusion

In this paper, we have presented a moving object detection method based on robust matrix factorization. This IRLS alternating scheme is conceptually simple, easy to implement and efficient for matrix low rank decomposition.
Furthermore, experiments on video surveillance datasets show that this approach is more robust than RPCA in the presence of dynamic backgrounds and illumination changes. Further research consists in developping an incremental version to update the model at every frame and to achieve real-time requirements. IRLS seems an good alternative to Lagrangian-based approach and is generally well suited to integrate an add-on process of acceleration of convergence.

References

1. Piccardi, M.: Background subtraction techniques: a review. (IEEE International Conference on Systems, Man and Cybernetics)
2. Bouwmans, T., Baf, F.E., Vachon, B.: Background modeling using mixture of gaussians for foreground detection - a survey. RPCS **1** (2008) 219–237
3. Bouwmans, T.: Recent advanced statistical background modeling for foreground detection: A systematic survey. RPCS **4** (2011) 147–176
4. Candes, E., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? International Journal of ACM **58** (2011)
5. Wright, J., Peng, Y., Ma, Y., Ganesh, A., Rao, S.: Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization. Neural Information Processing Systems, NIPS 2009 (2009)
6. Daubechies, I., Devore, R., Fornasier, M., Gntrk, C.S.: Iteratively reweighted least squares minimization for sparse recovery. Comm. Pure Appl. Math (2008)
7. Osborne, M.R.: Finite algorithms in optimization and data analysis. John Wiley & Sons (1985)
8. Tang, G., Nehorai, A.: Robust principal component analysis based on low-rank and block-sparse matrix decomposition. Annual Conference on Information Sciences and Systems (CISS2011) (2011)
9. Lin, Z., Liu, R., Su, Z.: Linearized alternating direction method with adaptive penalty for low-rank representation. NIPS 2011 (2011)
10. Ma, S.: Algorithms for sparse and low-rank optimization: Convergence, complexity and applications. Thesis (2011)
11. He, J., Balzano, L., Szlam, A.: Incremental gradient on the grassmannian for online foreground and background separation in subsampled video. International Conference on Computer Vision and Pattern Recognition (CVPR) (June 2012)
12. Toyama, K., Krumm, J., Brumitt, B., Meyers, B.: Wallflower: Principles and practice of background maintenance. ICCV 1999 (1999) 255–261
13. Li, L., Huang, W., Gu, I., Tian, Q.: Statistical modeling of complex backgrounds for foreground object detection. IEEE Transaction on Image Processing (2004) 1459–1472
14. Jodoin, P., Porikli, F., Konrad, J., Ishwar, P.: Change detection benchmark web site. IEEE Workshop on Change Detection (June 2012)
15. Maddalena, L., Petrosino, A.: A fuzzy spatial coherence-based approach to background foreground separation for moving object detection. Neural Computing and Applications (2010) 1–8