TechRxiv
Powered by IEEE

# Robust low rank and sparse decomposition from blurred video frames

This paper was downloaded from TechRxiv (https://www.techrxiv.org).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

27-04-2020 / 19-08-2020

CITATION

bokaei, Mohammad (2020): Robust low rank and sparse decomposition from blurred video frames. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.12199094.v2

DOI

10.36227/techrxiv.12199094.v2

# Robust Low Rank and Sparse Decomposition from Blurred Video Frames

1st Mohammad Bokaei
*Electrical and Electronic Engineering*
*Sharif University of Technology*
Tehran, Iran
mohammad.bokaei@ee.sharif.edu

2nd Saeed RazaviKia
*Electrical and Electronic Engineering*
*Sharif University of Technology*
Tehran, Iran
saeedrazavikia@ee.sharif.edu

3rd Peyman Rostami
*Electrical and Electronic Engineering*
*Sharif University of Technology*
Tehran, Iran
peyman.rostami@ee.sharif.edu

***Abstract*—Recently, the low-rank and sparse decomposition problem has attracted attention in several applications, especially surveillance videos. Due to the physical limitations in acquisition systems, measured frames are blurred by a low-pass filter. In this article, we aim to decompose blurred videos' frames into low-rank and sparse components, in order to extract the background. Unlike conventional methods, we simultaneously take into account the blurring effect, as well as the missing data. Our simulation results confirmed the advantage of this approach in extracting low-rank components in surveillance videos.**

***Index Terms*—Low-rank recovery, background, modeling , video, video surveillance**

## I. Introduction

With the growth in population and public transportation, moving object detection and background extraction have become an essential problem in video surveillance systems [1]. Furthermore, the problem of separating the background and the foreground [2]–[5] has several applications such as safety and security, retail, and home automation [6]. One of the well-studied approaches to deal with this issue is the low-rank and sparse decomposition (LRSD), that are based upon reshaping the video into low-rank and sparse structures. Indeed, by vertically embedding the consecutive frames inside a matrix, we have a matrix composed of low-rank and sparse parts. The background in each frame can be seen as a low-rank structure, while the moving parts form a sparse one. Principle component analysis is the first attempt used for extracting the low-rank background of a video stream from Gaussian noise [7]. In [8], the authors considered the sparse part as noise and they exploited the robust principal component analysis (RPCA) method. This was done by conventional algorithms such as singular value thresholding and inexact augmented Lagrangian method. In another work [9], a stable version of the RPCA method was deployed to deal with noisy data.

In some cases, due to the perturbation of read or write processes during the recording of the frames; frame pixels might be either corrupted or missed. In [5], [10] and [11], algorithms were proposed to solve the LRSD problem with missing data. Also, in [12], [13], fast singular value decomposition (SVD) free methods were proposed. Indeed, the RPCA method in [14] could solve foreground and background extraction via gradient descent (RPCA-GD).

Due to physical limitations in imaging systems such as lens effect, lengthy shutter exposure, high-speed motions, etc. the measured images are blurred [15], [16]. Hence, this problem obliges us to consider the blurring effect in the LRSD problem model.

In this paper, we aim to solve the background extraction problem from blurred and sub-sampled frames (of surveillance videos), where the blurring kernel is assumed to be known. First, we assume each frame to be convolved with a blurring kernel (typically a low-pass filter); Next, the sub-samples of the blurred frames are measured. This model differs from [13] as we consider missing data. Unlike the methods in [5], [10], we take into account the blurring effect at each frame of the video. Furthermore, our simulation results demonstrate that our model can achieve better low-rank background estimation in surveillance videos than other conventional models in the LRSD problem.

Our paper is organized as follows: In Section II, we explain our problem model and propose the algorithm to solve it. Next, in Section III experimental results are presented. Finally, in Section IV, we conclude the paper.

## II. Problem Model

Let $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_t \in \mathbb{R}^{n_1 \times n_2}$ be $t$ frames of the target video. A blurring kernel $\boldsymbol{\phi}$ is convolved with each frame $\mathbf{I}_i$ to model the lens effect. Next, the resulting images are uniformly sampled to reach measured matrices $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_t \in \mathbb{R}^m$ as follows:

$$\mathbf{B}_i = \mathcal{P}_\Omega(\boldsymbol{\phi} * \mathbf{I}_i) \quad \forall \quad i \in [t], \tag{1}$$

where $\Omega$ is the index set, uniformly chosen in $[n_1] \times [n_2]$ with size $|\Omega| = m$, and $\mathcal{P}_\Omega(\mathbf{Z}) : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$ is the sampling operator which measures the observed elements in $\Omega$. By concatenating each vectorized image frame $\mathrm{vec}(\mathbf{I}_i)$, we can build the frames matrix $\mathbf{F}_I \in \mathbb{R}^{n_1 n_2 \times t}$. Therefore, one can rewrite (1) in matrix format as:

$$\mathbf{M}_{\mathrm{o}} = \mathcal{P}_\Omega(\boldsymbol{\Phi}\mathbf{F}_I), \tag{2}$$

where $\boldsymbol{\Phi}$ is a block Hankel structure corresponding to the 2-D convolution of the blurring kernel $\boldsymbol{\phi}$ with each frame. Also, $\mathbf{M}_{\mathrm{o}}$ represents the matrix of measured frames.

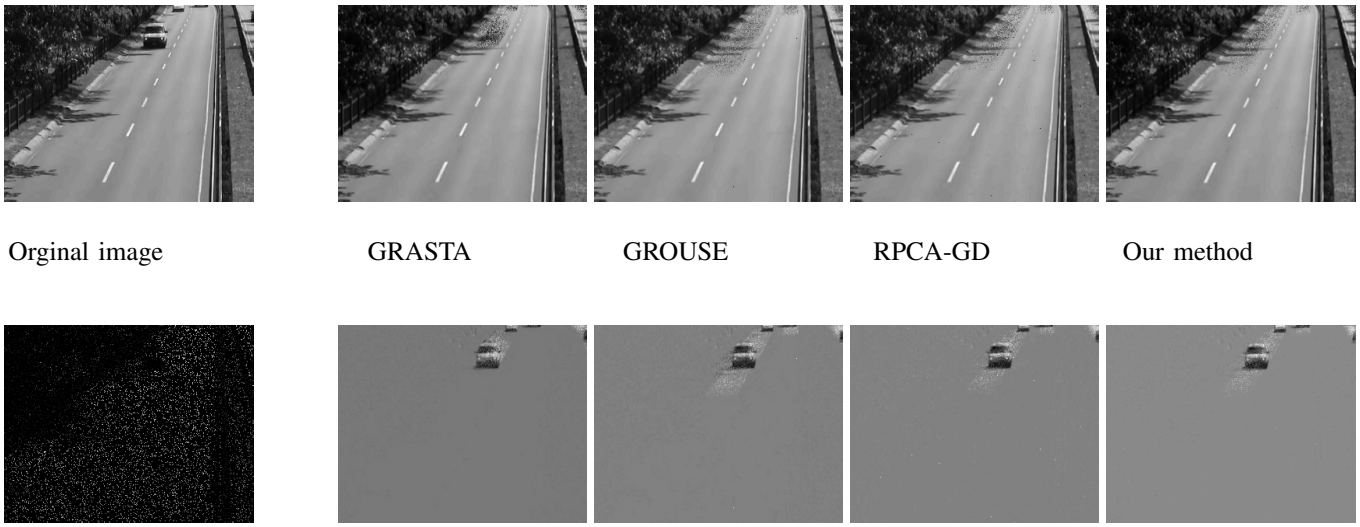


Fig. 1: Comparison result of the background and foreground extraction between our method and other state of the art methods on the Highway data set

In the surveillance videos, for the frames matrix, one can see that $\mathbf{F}_I = \mathbf{L}_I + \mathbf{S}_I$, where $\mathbf{L}_I, \mathbf{S}_I \in \mathbb{R}^{n_1 n_2 \times t}$ are low-rank and sparse matrices, respectively. The recovery problem here refers to extracting $\mathbf{L}_I$ and $\mathbf{S}_I$ from their mixture $\mathbf{F}_I$. Although the main goal is to achieve an estimate of the low-rank component $\mathbf{L}_I$, the sparse matrix $\mathbf{S}_I$ might also contain other information, such as the moving objects in the video surveillance example.

We consider the following convex optimization problem to extract the low-rank and the sparse parts:

$$\begin{aligned} \underset{\mathbf{L},\mathbf{S}\in\mathbb{R}^{n\times t},}{\text{minimize}} \quad & \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \\ \text{subject to} \quad & \mathcal{P}_\Omega(\mathbf{\Phi}\mathbf{L} + \mathbf{\Phi}\mathbf{S}) = \mathbf{M}_\text{o}, \end{aligned} \tag{3}$$

where $\mathbf{L}, \mathbf{S} \in \mathbb{R}^{n\times t}$ are the low-rank and the sparse desired matrices in which $n = n_1 n_2$, and $\mathbf{\Phi} \in \mathbb{R}^{n\times n}$ is a block Hankel structured matrix. $\|.\|_*$ denotes the nuclear norm which equals to the sum of the singular values and $\|.\|_1$ is the $\ell_1$ norm. Also, $\lambda > 0$ is a weighting factor, regularizing the sparsity and the rank. There are many suggested approaches to tackle the nuclear norm in optimization 3 such as smooth rank approximation [11], [17], [18].

### A. Algorithm

Regarding the upper bound of the nuclear norm of an arbitrary matrix $\mathbf{A}$ in [19], [20], we are using the following SVD free structure:

$$\|\mathbf{A}\|_* = \underset{\mathbf{U}\in\mathbb{C}^{r\times n},\mathbf{V}\in\mathbb{C}^{r\times t},\mathbf{A}=\mathbf{U}\mathbf{V}^\text{H}}{\text{minimize}} \frac{1}{2}(\|\mathbf{U}\|_\text{F}^2 + \|\mathbf{V}\|_\text{F}^2). \tag{4}$$

Therefore, problem (3) can be rewritten in the following setting:

$$\begin{aligned} \underset{\mathbf{S},\mathbf{U},\mathbf{V},\mathbf{L}=\mathbf{U}\mathbf{V}^\text{H}}{\text{minimize}} \quad & \|\mathbf{U}\|_\text{F}^2 + \|\mathbf{V}\|_\text{F}^2 + \lambda\|\mathbf{S}\|_1 \\ \text{subject to} \quad & \mathcal{P}_\Omega(\mathbf{\Phi}(\mathbf{L}+\mathbf{S})) = \mathbf{M}_\text{o}. \end{aligned} \tag{5}$$

We use the alternating direction method of multipliers (ADMM) technique for solving the problem (5). By combining the two constraints, we reach the following Augmented Lagrangian cost function:

$$\begin{aligned} \mathcal{L}(\mathbf{U},\mathbf{V},\mathbf{L},\mathbf{S},\mathbf{\Lambda}_1,\mathbf{\Lambda}_2) := \frac{1}{2}(\|\mathbf{U}\|_\text{F}^2 + \|\mathbf{V}\|_\text{F}^2 + 2\lambda\|\mathbf{S}\|_1) + \\ \frac{\mu_1}{2}\|\mathbf{L} - \mathbf{U}\mathbf{V}^\text{H} + \mathbf{\Lambda}_1\|_\text{F}^2 + \frac{\mu_2}{2}\|\mathcal{P}_\Omega(\mathbf{\Phi}(\mathbf{L}+\mathbf{S})) - \mathbf{M}_\text{o} + \mathbf{\Lambda}_2\|_\text{F}^2 \end{aligned}$$

Now, the update rules of $\mathbf{L}^{(k+1)}$, $\mathbf{U}^{(k+1)}$, $\mathbf{V}^{(k+1)}$, and $\mathbf{S}^{(k+1)}$ in $k-$th iteration can be obtained by sequentially

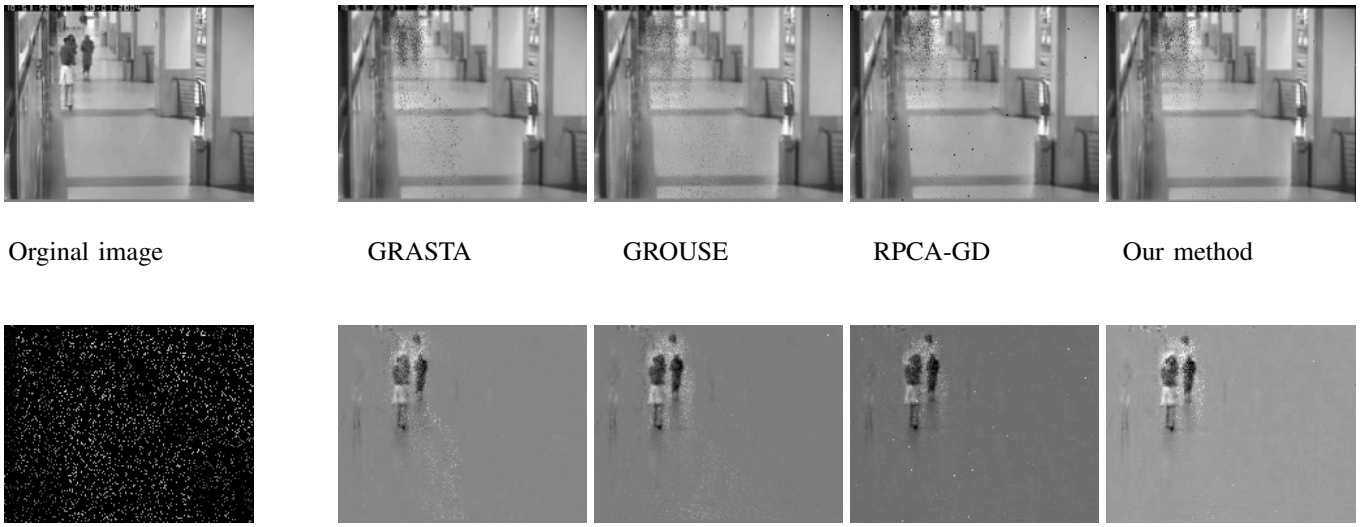


Fig. 2: Comparison result of the background and foreground extraction between our method and other state of the art methods on the Mall data set

solving the following alternative optimization problems:

$$\begin{aligned}\mathbf{L}^{(k+1)} =& \arg\min_{\mathbf{L}\in\mathbb{C}^{n\times t}} \frac{\mu_1}{2}\|\mathbf{L}-\mathbf{U}^{(k)}\mathbf{V}^{(k)\mathrm{H}}+\mathbf{\Lambda}_1^{(k)}\|_{\mathrm{F}}^2 \\ &+\frac{\mu_2}{2}\|\mathbf{M}_{\mathrm{o}}-\mathcal{P}_\Omega(\mathbf{\Phi}\mathbf{L})-\mathcal{P}_\Omega(\mathbf{\Phi}\mathbf{S}^{(k)})+\mathbf{\Lambda}_2^{(k)}\|_{\mathrm{F}}^2,\end{aligned} \tag{6}$$

$$\begin{aligned}\mathbf{U}^{(k+1)} =& \arg\min_{\mathbf{U}} \frac{1}{2}\|\mathbf{U}\|_{\mathrm{F}}^2 \\ &+\frac{\mu_1}{2}\|\mathbf{L}^{(k+1)}-\mathbf{U}\mathbf{V}^{(k)\mathrm{H}}+\mathbf{\Lambda}_1^{(k)}\|_{\mathrm{F}}^2,\end{aligned} \tag{7}$$

$$\begin{aligned}\mathbf{V}^{(k+1)} =& \arg\min_{\mathbf{V}} \frac{1}{2}\|\mathbf{V}\|_{\mathrm{F}}^2 \\ &+\frac{\mu_1}{2}\|\mathbf{L}^{(k+1)}-\mathbf{U}^{(k+1)}\mathbf{V}^{\mathrm{H}}+\mathbf{\Lambda}_1^{(k)}\|_{\mathrm{F}}^2,\end{aligned} \tag{8}$$

$$\begin{aligned}\mathbf{S}^{(k+1)} =& \arg\min_{\mathbf{S}\in\mathbb{C}^{n\times t}} \lambda\|\mathbf{S}\|_1 \\ &+\frac{\mu_2}{2}\|\mathbf{M}_{\mathrm{o}}-\mathcal{P}_\Omega(\mathbf{\Phi}\mathbf{L}^{(k+1)})-\mathcal{P}_\Omega(\mathbf{\Phi}\mathbf{S})+\mathbf{\Lambda}_2^{(k)}\|_{\mathrm{F}}^2.\end{aligned} \tag{9}$$

Then, the update rules of Lagrangian multipliers is given by

$$\begin{aligned}\mathbf{\Lambda}_1^{(k+1)} =& \mathbf{\Lambda}_1^{(k)}+\mathbf{L}^{(k+1)}+\mathbf{S}^{(k+1)}-\mathbf{U}^{(k+1)}\mathbf{V}^{(k+1)\mathrm{H}}, \\ \mathbf{\Lambda}_2^{(k+1)} =& \mathbf{\Lambda}_2^{(k)}+\mathbf{M}_{\mathrm{o}}-\mathcal{P}_\Omega(\mathbf{\Phi}\mathbf{L}^{(k+1)}+\mathbf{\Phi}\mathbf{S}^{(k+1)}).\end{aligned} \tag{10}$$

Next, the sub-problems on $\mathbf{L}$, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{S}$ can be solved by taking the gradient concerning each variable and setting the gradients to zero. Hence,

$$\begin{aligned}\mathbf{L}^{(k+1)} = \mathcal{J}^{-1}\Big(&\mu_1(\mathbf{U}^{(k)}\mathbf{V}^{(k)\mathrm{H}}-\mathbf{\Lambda}_1^{(k)})+ \\ &\mu_2\mathbf{\Phi}^{\mathrm{T}}\mathcal{P}_\Omega(\mathbf{M}_{\mathrm{o}}+\mathbf{\Lambda}_2^{(k)}-\mathbf{\Phi}\mathbf{S}^{(k)})\Big),\end{aligned} \tag{11}$$

where the linear operator $\mathcal{J}:\mathbb{C}^{n\times t}\mapsto\mathbb{C}^{n\times t}$ is defined as:

$$\mathcal{J}(\mathbf{Z}) = \mu_1\mathbf{Z}+\mu_2\mathbf{\Phi}^{\mathrm{T}}\mathcal{P}_\Omega(\mathbf{\Phi}\mathbf{Z}). \tag{12}$$

Afterward, we update $\mathbf{S}$ by

$$\mathbf{S}^{(k+1)} = \mathcal{S}_{\lambda/\mu_2}(\mathbf{M}_{\mathrm{o}}+\mathbf{\Lambda}_2^{(k)}-\mathcal{P}_\Omega(\mathbf{\Phi}\mathbf{L}^{(k+1)})) \tag{13}$$

where $\mathcal{S}_{\lambda/\mu_2}:\mathbb{R}^{n\times t}\mapsto\mathbb{R}^{n\times t}$ is the soft-threshold function which is defined as:

$$[\mathcal{S}_\tau(\mathbf{Z})]_{i,j} = \begin{cases}\mathrm{sgn}(z_{i,j})(|z_{i,j}|-\tau) & |z_{i,j}|\geq\tau \\ 0 & \text{o.w.}\end{cases}$$

where $z_{i,j}$ is the $(i,j)$-th element of matrix $\mathbf{Z}$. Next, for $\mathbf{U}$ and $\mathbf{V}$ we have:

$$\begin{aligned}\mathbf{U}^{(k+1)} =& \mu_1\big(\mathbf{L}^{(k+1)}+\mathbf{\Lambda}_1^{(k)}\big)\mathbf{V}^{(k)}\cdot \\ &\big(\mathbf{I}_{t\times t}+\mu_1\mathbf{V}^{(k)\mathrm{H}}\mathbf{V}^{(k)}\big)^{-1},\end{aligned} \tag{14}$$

$$\begin{aligned}\mathbf{V}^{(k+1)} =& \mu_1\big(\mathbf{L}^{(k+1)}+\mathbf{\Lambda}_1^{(k)}\big)\mathbf{U}^{(k+1)}\cdot \\ &\big(\mathbf{I}_{n\times n}+\mu_1\mathbf{U}^{(k+1)\mathrm{H}}\mathbf{U}^{(k+1)}\big)^{-1},\end{aligned} \tag{15}$$

where $\mathbf{I}_{n\times n}$ and $\mathbf{I}_{t\times t}$ represent identity matrices with size $n\times n$ and $t\times t$, respectively. The initial values of $\mathbf{L}^{(0)}$ and $\mathbf{S}^{(0)}$ come from the least square problem solution, i.e.,

$$\mathbf{F}^{(0)} = \operatorname*{argmin}_{\mathbf{X}\in\mathbb{R}^{n\times t}}\|\mathbf{\Phi}(\mathbf{M}_{\mathrm{o}}-\mathbf{X})\|_{\mathrm{F}}^2, \tag{16}$$

where $\mathbf{F}^{(0)} = \mathbf{L}^{(0)}+\mathbf{S}^{(0)}$. Then, $\mathbf{U}^{(0)}$ and $\mathbf{V}^{(0)}$ can be computed by the polar decomposition of $\mathbf{L}^{(0)}$. Also, we set $\mathbf{\Lambda}_1^{(0)} = \emptyset$ and $\mathbf{\Lambda}_2^{(0)} = \emptyset$. The convergence analyses of the

ADMM algorithm for minimizing the sum of multi non-smooth convex separable functions have been studied in [21], [22], in which it is shown that if the penalty parameters are chosen to be sufficiently large, the classical ADMM converges to the set of stationary solutions.

## III. Simulation Results

In this section, we apply our model to the foreground and background of real video data. We compare methods such as Grassmannian Robust Adaptive Subspace Tracking Algorithm (GRASTA) [23], Grassmannian Rank-One Update Subspace Estimation(GROUSE) [24], and Robust PCA via Non-convex Gradient Descent (RPCA-GD) [14] with each other. These algorithms and ours are implemented in MATLAB. We used three types of real surveillance videos in our simulations (i.e. traffic data, shopping center data, and escalator data). In addition, Gaussian filters are used to simulate blurring effect in datasets. The videos in these datasets have different frame sizes, but we used 100 frames of each one. Moreover, the input data is generated by randomly setting 90% of the pixels of each frame as missing data.

Figure 1 demonstrates the foreground and background separation for the Highway dataset by applying the mentioned methods. The original image is in the first row, and other rows show the outputs of the other algorithms. Figures 2, 3 show the same results for Escalator, and Mall data set, respectively. The results show that the accuracy of our method is higher than those of the other tested methods.

## IV. Conclusion

In this paper, we investigated the background and foreground separation problem on blurred surveillance video datasets. Moreover, we proposed a problem considering the blurring kernel effects. Next, we developed an ADMM algorithm to solve the sparse and low-rank matrix decomposition problem. At last, we compared our method with state-of-the-art methods. Numerical simulation results that came from real-world datasets demonstrates that our method works better than the other simulated approaches.

## References


[1] S. Brutzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *CVPR 2011*, pp. 1937–1944, IEEE, 2011.

[2] M. Piccardi, "Background subtraction techniques: a review," in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, vol. 4, pp. 3099–3104, IEEE, 2004.

[3] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, 2014.

[4] O. Barnich and M. Van Droogenbroeck, "Vibe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image processing*, vol. 20, no. 6, pp. 1709–1724, 2010.

[5] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *Computer Science Review*, vol. 23, pp. 1–71, 2017.

[6] B. S. I. Association *et al.*, "An introduction to video content analysis," 2009.

[7] I. Jolliffe, *Principal component analysis*. Springer, 2011.

[8] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.

[9] T. Zhou and D. Tao, "Godec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, 2011.

[10] R. Otazo, E. Candes, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magnetic Resonance in Medicine*, vol. 73, no. 3, pp. 1125–1136, 2015.

[11] M. Azghani, A. Esmaeili, K. Behdin, and F. Marvasti, "Missing low-rank and sparse decomposition based on smoothed nuclear norm," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[12] F. Shang, Y. Liu, J. Cheng, and H. Cheng, "Robust principal component analysis with missing data," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pp. 1149–1158, ACM, 2014.

[13] X. Yuan, "Nuclear-norm-free variational models for background extraction from surveillance video," in *Cross-straits Optimization Workshop, COW 2013*, 2013.

[14] X. Yi, D. Park, Y. Chen, and C. Caramanis, "Fast algorithms for robust PCA via gradient descent," in *Advances in neural information processing systems*, pp. 4152–4160, 2016.

[15] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless, "Single image deblurring using motion density functions," in *European Conference on Computer Vision*, pp. 171–184, Springer, 2010.

[16] S. Baker and S. K. Nayar, "A theory of catadioptric image formation," in *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*, pp. 35–42, IEEE, 1998.

[17] A. Esmaeili, K. Behdin, M. A. Fakharian, and F. Marvasti, "Transductive multi-label learning from missing data using smoothed rank function," *Pattern Analysis and Applications*, pp. 1–9, 2020.

[18] A. Esmaeili and F. Marvasti, "A novel approach to quantized matrix completion using huber loss measure," *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 337–341, 2019.

[19] N. Srebro, "Learning with matrix factorizations," 2004.

[20] M. Signoretto, V. Cevher, and J. Suykens, "An SVD-free approach to a class of structured low rank matrix optimization problems with application to system identification," in *IEEE Conference on Decision and Control (CDC)*, 2013.

[21] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *Mathematical Programming*, vol. 162, no. 1-2, pp. 165–199, 2017.

[22] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337–364, 2016.

[23] J. He, L. Balzano, and A. Szlam, "Incremental gradient on the grassmannian for online foreground and background separation in subsampled video," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1568–1575, IEEE, 2012.

[24] L. Balzano, R. Nowak, and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," in *2010 48th Annual allerton conference on communication, control, and computing (Allerton)*, pp. 704–711, IEEE, 2010.

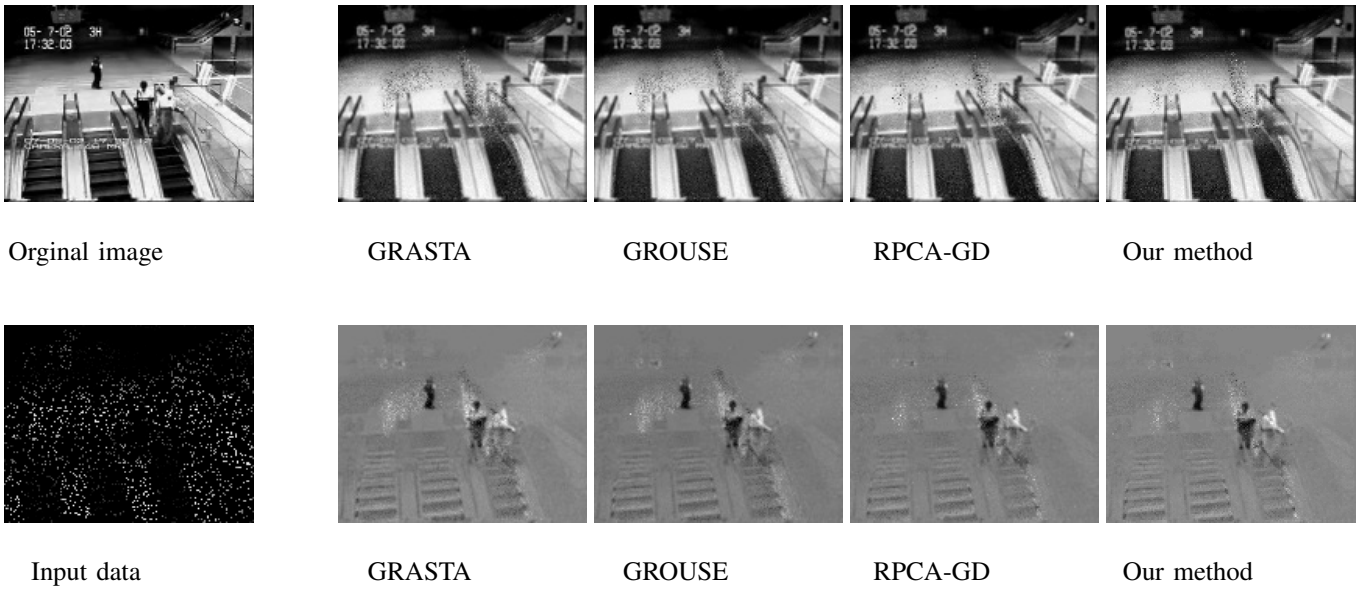


Fig. 3: Comparison result of the background and foreground extraction between our method and other state of the art methods on the escalator data set