FOREGROUND DETECTION BASED ON LOW-RANK AND BLOCK-SPARSE MATRIX DECOMPOSITION

Charles Guyon, Thierry Bouwmans, El-Hadi Zahzah

Laboratoire MIA, Univ. La Rochelle, 17000 La Rochelle, France

ABSTRACT

Foreground detection is the first step in video surveillance system to detect moving objects. Principal Components Analysis (PCA) shows a nice framework to separate moving objects from the background but without a mechanism of robust analysis, the moving objects may be absorbed into the background model. This drawback can be solved by recent researches on Robust Principal Component Analysis (RPCA). The background sequence is then modeled by a low rank subspace that can gradually change over time, while the moving foreground objects constitute the correlated sparse outliers. In this paper, we propose to use a RPCA method based on low-rank and block-sparse matrix decomposition to achieve foreground detection. This decomposition enforces the low-rankness of the background and the block-sparsity aspect of the foreground. Experimental results on different datasets show the pertinence of the proposed approach.

***Index Terms*—** Foreground Detection, Robust Principal Component Analysis.

1. INTRODUCTION

The detection of moving objects is the basic low-level operations in video analysis. This detection is usually done using foreground detection. This basic operation consists of separating the moving objects called "foreground" from the static information called "background". Many foreground detection methods have been developed [1][2][3]. In 1999, Oliver et al. [4] are the first authors to model the background by Principal Component Analysis (PCA). Foreground detection is then achieved by thresholding the difference between the generated background image and the current image. PCA provides a robust model of the probability distribution function of the background, but not of the moving objects while they do not have a significant contribution to the model. However, this model presents several limitations as developped in [3]. The first limitation of this model is that the size of the foreground object must be small and don't appear in the same location during a long period in the training sequence. The second limitation is that the outliers or foreground objects may be absorbed into the background mode without a mechanism of robust analysis. Recent researches on robust PCA via Principal Component Pursuit (RPCA-PCP) [5][6] can be used to alleviate these limitations. However, these methods didn't address the spatial connexity of the foreground pixel. To take it into account, we propose to use a decomposition that enforces the low-rankness of the background and the block-sparsity aspect of the foreground. The rest of this paper is organized as follows. In Section 2, we present a short survey on recent robust PCA. Then, we present how the background and the foreground can be separated via the low-rank and block sparse matrix decomposition in Section 3. Finally, performance evaluation and comparison with three other algorithms are given in Section 4.

2. RELATED WORK

Many algorithms [3][7] have been proposed to address the limitation of classical PCA with respect to outlier (sparse) noise, yielding the field of robust PCA. However, these previous methods do not possess the strong performance guarantees provided by very recent works. Candes et al. [5] have proposed a convex optimization to address the robust PCA problem. The observation matrix is assumed represented as:

$$A = L + S \tag{1}$$

where L is a low-rank matrix and S must be sparse matrix with a small fraction of nonzero entries. The straightforward formulation is to use l_0 norm to minimize the energy function:

$$\min_{L,S} \; Rank(L) + \lambda ||S||_0 \quad \text{subj} \quad A = L + S \tag{2}$$

where $\lambda > 0$ is an arbitrary balanced parameter. But this problem is NP-hard, typical solution might involve a search with combinatorial complexity. This research seeks to solve for L with the following optimization problem:

$$\min_{L,S} \; ||L||_* + \lambda ||S||_1 \quad \text{subj} \quad A = L + S \tag{3}$$

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. [5] 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 background modeling are often corrupted by noise affecting every entry of the data matrix. Therefore, Zhou et al. [6] proposed a stable PCP that guarantee stable and accurate recovery in the presence of entry-wise noise.So, this measurement model assumes that:

$$A = L + S + Z \tag{4}$$

where Z is a noise term (say i.i.d. noise on each entry of the matrix) and $||Z||_F < \delta$ for some $\delta > 0$. To recover L and S, Zhou et al. [6] proposed to solve the following optimization problem, as a relaxed version to PCP:

$$\min_{L,S} \ ||L||_* + \lambda ||S||_1 \quad \text{subj} \quad ||A - L - S||_F < \delta \tag{5}$$

where $\lambda = \frac{1}{\sqrt{n}}$. However, these two methods 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. Recently, Tang and Nehorai [8] proposed a RPCA via a decomposition that enforces the low-rankness of one part and the block sparsity of the other part. This decomposition called RPCA-LBD allows to separate robustly the principal components from the outliers. In the following section, we present how the background and the foreground can be separated with this decomposition.

3. FOREGROUND DETECTION VIA THE LOW-RANK AND BLOCK SPARSE MATRIX DECOMPOSITION

Denote the training video sequences $D = \{I_1, ...I_N\}$ where I_t is the frame at time t and N is the number of training frames. Let each pixel (x,y) be characterized by its intensity in the grey scale. The decomposition involves the following model:

$$D = L + S \tag{6}$$

where L and S are the low-rank component and block-sparse component of D, respectively. The low-rank matrix L contains the background and the block-sparse matrix S contains mostly zero columns, with several non-zero ones corresponding to the foreground. In order to eliminate ambiguity, the columns of the low-rank matrix L corresponding to the outlier columns are assumed to be zeros. The low-rank matrix L and the block-sparsity matrix S can be recovered by the following convex program [8]:

$$\min_{L,S} \ ||L||_* + \kappa(1-\lambda)||L||_{2,1} + \kappa\lambda||S||_{2,1} \quad \text{subj} \quad D = L + S \tag{7}$$

where $||.||_*$ and $||.||_{2,1}$ are the nuclear norm and l_1 norm of the vector formed by taking the l_2 norms of the columns of the underlying matrix, respectively. The term $\kappa(1-\lambda)||L||_{2,1}$ ensures that the recovered matrix L has exact zero columns corresponding to the outliers. The decomposition can be solved by a convex program based on the augmented Lagrange Multiplier (ALM) method. The ALM method is a well-known method to successfully perform low-rank and sparse matrix decomposition for large data problems. For this work, we used the general augmented Lagrange multiplier method [8] to obtain the decomposition. The proposed algorithm to separate the background and the foreground via RPCA-LBD is as follows.

Algorithm for Foreground Detection via RPCA-LBD

Input data: Training sequence with N images that is contained in D and the parameters κ, λ.
Output data: Background in L and foreground in S.
Parameters initialization: $L_0 \leftarrow D$; $S_0 \leftarrow O$; $\mu_0 = 30/||sign(D)||_2$; $\rho > 0$, $\beta > 0$, $\alpha \in (0,1)$, $k \leftarrow 0$.
Background and Foreground Generation:
while not converged **do**
  $G^L = D - S_k + \frac{1}{\mu_k}Y_k$; $j \leftarrow 0$, $L^0_{k+1} = G^L$.
  while not converged **do**
  $L^{(j+1/2)}_{k+1} = US_\beta(\Sigma)V^T$ where $L^{(j)}_{k+1} = U\Sigma V^T$ is the SVD of $L^{(j)}_{k+1}$.
  $L^{(j+1)}_{k+1} = L^{(j)}_{k+1} + \alpha(\tau_{\frac{\beta\kappa(1-\lambda)}{(1+\beta\mu_k)}}(\frac{2L^{(j+1/2)}_{k+1} - L^j_{k+1} + \beta\mu_k G^L}{1+\beta\mu_k}) - L^{(j+1/2)}_{k+1})$.
  $j \leftarrow j+1$.
  end while
  $L_{k+1} = L^{(j+1/2)}_{k+1}$.
  $G^S = D - L_{k+1} + \frac{1}{\mu_k}Y_k$.
  $S_{k+1} = \tau_{\frac{\kappa\lambda}{\mu_k}}(G^S)$.
  $Y_{k+1} = Y_k + \mu_k * (D - L_{k+1} - S_{k+1})$.
  $\mu_{k+1} = \rho\mu_k$; $k \leftarrow k+1$.
end while
$L \leftarrow L_k$, $S \leftarrow S_k$.
Foreground Detection: Threshold the matrix S to obtain the foreground mask.

For the initialization, the low-rank matrix L, the block-sparse matrix S, and the Lagrange multiplier Y are set respectively to the matrix D, O and O. The error in the outer loop is computed as $||D - L_k - S_k||_F/||D||_F$. The outer loop stops when this error is $< 10^{-7}$ or when the maximal iteration number is reached. The error in the inner loop is the Frobenius norm of the difference between successive matrices L^j_k. The inner loop has a tolerence error equal to 10^{-6} and a maximal iteration equal to 20. The computational cost in the RPCA-LBD algorithm is the singular value decomposition (SVD) in the DR iteration. It can be reduced significantly by using a partial SVD because only the first largest few singular values are needed. Practically, we used the implementation available in PROPACK[1]. The tuning parameters κ and λ are set exper-

Fig. 1. Original image (309), low-rank matrix L (background), block-sparse matrix S (foreground), foreground mask, ground truth.

imentally to 1.1 and 0.61. For the DR iteration, $\beta = 0.2$ and $\alpha = 1$. The ALM parameter ρ is set to 1.1. Fig. 1 shows the original frame 309 of the sequence from [9] and its decomposition into the low-rank matrix L and block-sparse matrix S. We can see that L corresponds to the background whereas S corresponds to the foreground. The fourth image shows the foreground mask obtained by thresholding the matrix S and the fifth image is the ground truth image.

4. EXPERIMENTAL RESULTS

We compared the proposed approach with the PCA[4], RSL [7], RPCA via Principal Component Pursuit (RPCA-PCP) [5]. The experiments were conducted qualitatively and quantitatively on the Wallflower dataset [10] and I2R dataset [11]. The algorithms were implemented in batch mode with matlab.

4.1. Wallflower dataset

We have chosen this particular dataset provided by Toyama et al. [10] because of how frequent it is used in this field. This frequency is due to its faithful representation of real-life situations typical of scenes susceptible to video surveillance. Moreover, it consists of seven video sequences, with each sequence presenting one of the difficulties a practical task is likely to encounter: Moved Object (MO), Time of Day (TD), Light Switch (LS), Waving Trees (WT), Camouflage (C), Bootstrapping (B) and Foreground Aperture (F). 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 figures 2 and 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. The detection rate is given as follows:

$$DR = \frac{TP}{TP + FN} \tag{8}$$

The precision is computed as follows:

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

[1] http://soi.stanford.edu/rmunk/PROPACK/

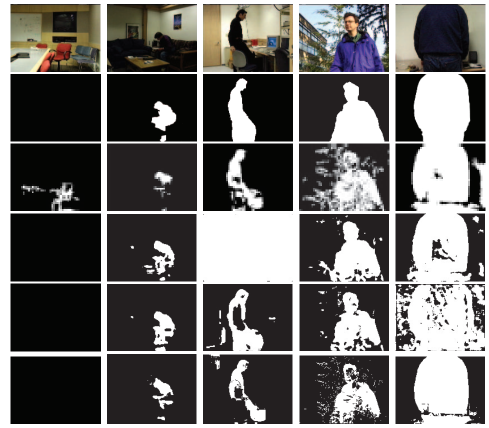

Fig. 2. From top to bottom: original image, ground truth, PCA, RSL, RPCA-PCP, RPCA-LBD. From left to right: MO (985), TD (1850), LS (1865), WT (247), C (251).

Table 1. F-measure on the Wallflower dataset

	RSL	RPCA-PCP	RPCA-LBD
TD	75.73	81.18	71.43
LS	28.36	70.86	**71.64**
WT	89.69	86.40	86.16
C	91.78	75.43	**96.78**
B	69.38	74.4	**74.9**
FA	74.37	72.07	**88.46**
Average	71.55	76.73	**79.90**

A good performance is obtained when the detection rate is high without altering the precision. This can be measured by the F-measure [12] as follows:

$$F = \frac{2 \times DR \times Precision}{DR + Precision} \tag{10}$$

A good performance is then obtained when the F-measure is closed to 1. Table 1 shows the F-measure for each sequence and the average of the F-measure on the dataset. The F-measure value of MO sequence can't be computed due to the absence of true positives in its ground-truth. RCPA-LBD outperforms globally RSL and RPCA-PCP. For the four sequences (LS, C, B, FA), the proposed method gives better results than RSL and RPCA-PCP. For the sequence WT and TD, results are still acceptable. 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.

4.2. I2R dataset

This dataset provided by [11] consists of nine video sequences, which each sequence presenting dynamic back-

Fig. 3. From top to bottom: original image, ground truth, RSL, RPCA-PCP, RPCA-LBD. From left to right:**Wallflower dataset:** B (2832), FA (449). **I2R dataset:** shopping mall (1980), water surface (1594), curtain (22772).

Table 2. F-measure on the I2R dataset

	RSL	RPCA-PCP	RPCA-LBD
Shopping mall	58.64	76.07	**77**
Water surface	34.42	45	**87.69**
Curtain	91.04	90.96	**91.09**

grounds 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 three sequences that are the following ones: shopping mall, water surface and curtain. Fig. 3 shows the qualitative results. For example, we can see that the block-sparsity contraints allows to detect the complete silhouette in the sequence called "water surface". Table 2 shows the average F-measure in percentage that is obtained on 20 ground truth images for each sequence. We can see that RCPA-LBD outperforms RSL and RPCA-PCP for each sequence.

5. CONCLUSION

In this paper, we have presented a foreground detection method based on low-rank and block-sparse matrix decomposition. This decomposition allows to take into account the spatial connexity of the foreground pixels. Furthermore, experiments on video surveillance datasets show that this approach is more robust than RSL and RPCA-PCP in the presence of dynamic backgrounds and illumination changes. Further research consists in developping an incremental RPCA-LBD to update the model at every frame and to achieve real-time requirements.

6. ACKNOWLEDGEMENTS

The authors would like to thank Congguo Tang (Dept. of Electical System Engineering, Washington Univ., USA) who has kindly provided the batch algorithm of RPCA-LBD.

7. REFERENCES

[1] T. Bouwmans, "Recent advanced statistical background modeling for foreground detection: A systematic survey," *RPCS*, vol. 4, no. 3, pp. 147–176, Nov. 2011.

[2] T. Bouwmans, F. El Baf, and B. Vachon, "Background modeling using mixture of gaussians for foreground detection - a survey," *RPCS*, vol. 1, no. 3, pp. 219–237, Nov. 2008.

[3] T. Bouwmans, "Subspace learning for background modeling: A survey," *RPCS*, vol. 2, no. 3, pp. 223–234, Nov. 2009.

[4] N. Oliver, B. Rosario, and A. Pentland, "A bayesian computer vision system for modeling human interactions," *ICVS 1999*, Jan. 1999.

[5] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *International Journal of ACM*, vol. 58, no. 3, May 2011.

[6] Z. Zhou, X. Li, J. Wright, E. Candes, and Y. Ma, "Stable principal component pursuit," *IEEE ISIT Proceedings*, pp. 1518–1522, Jun. 2010.

[7] F. De La Torre and M. Black, "A framework for robust subspace learning," *International Journal on Computer Vision*, pp. 117–142, 2003.

[8] G. Tang and A. Nehorai, "Robust principal component analysis based on low-rank and block-sparse matrix decomposition," *CISS 2011*, 2011.

[9] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE T-PAMI*, vol. 27, pp. 1778–1792, 2005.

[10] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," *ICCV 1999*, pp. 255–261, Sept. 1999.

[11] L. Li, W. Huang, I. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE T-IP*, pp. 1459–1472, 2004.

[12] L. Maddalena and A. Petrosino, "A fuzzy spatial coherence-based approach to background foreground separation for moving object detection," *Neural Computing and Applications*, pp. 1–8, 2010.