Background Subtraction Based on Low-rank and Structured Sparse Decomposition

Xin Liu, Guoying Zhao, *Senior Member, IEEE,* Jiawen Yao, and Chun Qi, *Member, IEEE*

Abstract

Low rank and sparse representation based methods, which make few specific assumptions about the background, have recently attracted wide attention in background modeling. With these methods, moving objects in the scene are modeled as pixel-wised sparse outliers. However, in many practical scenarios, the distributions of these moving parts are not truly pixel-wised sparse but structurally sparse. Meanwhile a robust analysis mechanism is required to handle background regions or foreground movements with varying scales. Based on these two observations, we first introduce a class of structured sparsity-inducing norms to model moving objects in videos. In our approach, we regard the observed sequence as being constituted of two terms, a low-rank matrix (background) and a structured sparse outlier matrix (foreground). Next, in virtue of adaptive parameters for dynamic videos, we propose a saliency measurement to dynamically estimate the support of the foreground. Experiments on challenging well

This work was supported in part by the National Natural Science Foundation of China (Grant No.60972124), the National High-tech Research and Development Program of China (863 Program) (Grant No.2009AA01Z321), the Research Fund for the Doctoral Program of Higher Education of China (Grant No.20110201110012), and in part by the Academy of Finland and the strategic funds from University of Oulu.

Xin Liu and Chun Qi are with the School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, China, and Xin Liu is also with the Center for Machine Vision Research, Department of Computer Science and Engineering, University of Oulu, Oulu FI-90014, Finland. E-mail: linuxsino@gmail.com, qichun@mail.xjtu.edu.cn

Guoying Zhao is with the Center for Machine Vision Research, Department of Computer Science and Engineering, University of Oulu, Oulu FI-90014, Finland. E-mail: gyzhao@ee.oulu.fi

Jiawen Yao is with the Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76010 USA. E-mail: yjiaweneecs@gmail.com

Chun Qi is the corresponding author.

known data-sets demonstrate that the proposed approach outperforms the state-of-the-art methods and works effectively on a wide range of complex videos.

Index Terms

Background subtraction, Background modeling, Structured sparsity, Low-rank modeling, Foreground detection,

I. INTRODUCTION

Foreground object segmentation from a video stream is a fundamental and critical step for many high level computer vision tasks, such as traffic control, object-based video encoding, social signal processing [1] and human-machine interactions. The accuracy of segmentation can significantly affect the overall performance of the application employing it. Background subtraction is generally regarded as an effective method for extracting the foreground. However, the background in a complex environment may include distracting motions and hence makes precise segmentation challenging.

In the past decade, great progress in improving the performance of foreground detection has been reported. Background subtraction is a major technique used to detect moving parts by subtracting them from the established background. This means that video frames firstly are compared with a background model, and then changes are identified as the foreground. A very popular background subtraction approach is to model each pixel with a mixture of Gaussians [2], proposed by Stauffer and Grimson. Due to its effectiveness in sustaining background variations, a large amount of further developments [3] [4] [5] have been proposed. In [6], Elgammal $et\ al.$ proposed a non-parametric kernel density estimate (KDE) method for background modeling. In [7], the Principle Component Analysis (PCA) method for background modeling was proposed. In this method, the new frame was projected onto the subspace spanned by the trained principle components, and the residues indicate the presence of new foreground objects. In [8], Li $et\ al.$ utilize spatio-temporal features (color co-occurrence) to model complex backgrounds. The method in [9] made use of prior information of the neighborhood spatial context by a MAP-MRF framework. Heikkilä and Pietikäinen [10] developed an efficient texture-based method by using adaptive Local Binary Pattern (LBP) histograms to capture background statistics of each pixel. In the algorithms presented in [11] [12], each pixel is represented by a code-book. In [13] [14],

Maddalena $et\ al.$ proposed a self-organizing artificial neural network for background subtraction (SOBS). In the ViBe [15] and PBAS [16], background modeling is based on the collection and update pixel samples. In [17], foreground detection was cast as an outlier signal estimation problem in a linear regression model. A more detailed discussion of these conventional techniques can be found in recent surveys [18] [19]. However, because of the restrictive assumptions on background, and few spatial connections used in pixel-based modeling, these methods may fail in real scenarios, especially when handling a dynamic background.

Recently, low-rank and sparse decomposition methods have shown promising performance in foreground detection. The only assumption made on the background is that any variation in its appearance can be captured by the low rank matrix [20]. In this simple form, a matrix composed of the observed video frames can be decomposed into a low-rank matrix representing the background and a sparse matrix consisting of the foreground objects treated as the sparse foreground [21]. This is the well known Robust Principal Component Analysis (RPCA) and it has been widely studied in recent works [22] [23] [24] [25] [26] [27] [28] [21].

In earlier work [20] [29], Wright $et\ al.$ proposed using a ℓ_1-norm to constrain the sparse matrix and assumed that background images are linearly correlated with each other, forming a low-rank matrix L. Unlike conventional pixel-based modeling methods, the only assumption about the background is the low-rank property, and the foreground regions mean intensity changes which cannot be fitted into the low-rank model of background, and thus should be treated as outliers. In the low-rank representation, the foreground must be a sparse matrix with a small fraction of nonzero entries. However, this approach has two issues required to be discussed:

- The ℓ_1-norm treats each entry (pixel) independently. It did not consider the spatial connection of the foreground sparse pixel and it is know that in many practical scenarios these parts are distributed with structures.
- The static and global setting of λ in the formulation cannot handle complex scenes with highly dynamic background movements (λ controls the amount of outliers in the RPCA decomposition). An example of such a scene is shown in Fig. 1.

To solve the first problem, we consider possible relationships among subsets of the entries in the sparse matrix. In fact, the context of statistical signal processing tells us that when using the ℓ_1-norm to promote the sparsity in the sparse matrix, it assumes that each pixel is independently corrupted. However, in foreground detection, sparse outliers are typically spatially contiguous.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing

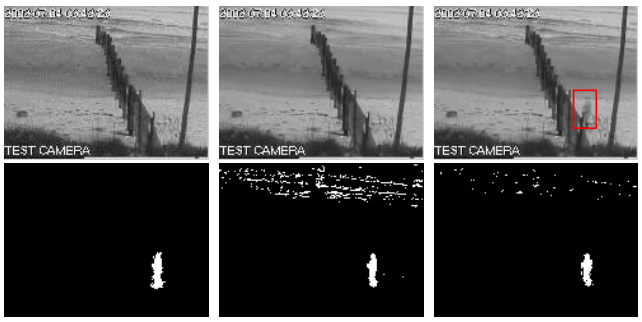


Fig. 1. Recovered background and foreground detection (top and bottom rows respectively). Left: our method. Middle: RPCA [20] with the smaller $\lambda = 1/\sqrt{\max(m,n)}$ (m and n are defined in Section II) produces a clean background recovery but detects more background movements. Right: the higher $\lambda = 2/\sqrt{\max(m,n)}$ achieves few false detection, but gives ghost appearance in background (see the red rectangle).

It is difficult to model these variations using ℓ_1-norm in a dynamic background. To solve this issue, in [30], the $l_{2,1}$-norm was adopted for detecting outliers with column-wise sparsity. Work is currently ongoing with the authors of [31] [23], and the low-rank and block-sparse matrix decomposition method (RPCA-LBD) has been proposed. This decomposition in the RPCA-LBD method enforces the low-rankness of the background and the block-sparsity aspect of the foreground. However, the block-sparsity property still has no structured information to model sparse outliers. In [27], a locally low-rank model was proposed. In this method, the Total Variation (TV) penalty on the sparse deviations was employed to better handle noisy data. The most recent work is incorporated with Markov Random Field (MRF) prior [28]. The smoothness effects of MRF can effectively eliminate noise and small background movements, but foreground regions tend to be over-smoothed due to the smoothness constraint.

Structural information does exist in real scenarios, and the recent methods mentioned above had few considerations on this issue. However, the structural information about nonzero patterns of variables has been developed and used in sparse signal recovery, and several approaches have been applied successfully, such as Lattice Matching Pursuit (LaMP) [32], Dynamic Group Sparsity (DGS) recovery [33], Bayesian Robust Matrix Factorization (BRMF) [26], and the Proximal Operator using Network Flow (ProxFlow) [34]. In [26], the group sparsity problem was casted into probabilistic models, which used a hierarchical view of Laplace distribution as the

residuals model, and the MCMC sampling for model inference was applied. The ProxFlow [34] used a structured regularizer to encode the prior that nonzero entries of sparse signal should be in a group structure. Jia $et\ al.$ [35] introduced this structured norm to the sparse representation based classification (SRC) framework, to model various corruptions in face images and obtained robust and practical face recognition results. Their work illustrated that the structured sparsity norm encourages outliers to distribute in patterns and thus can better model the real distribution of face variations than the ℓ_1-norm. In the original work in [34], given a training sequence not containing any foreground objects, ProxFlow aims to detect new objects in a new testing image. However, in some practical problems, a well-defined training sequence should not be easily obtained, and thus this approach may fail to handle the sudden changes in background such as illumination.

With the second issue, λ controls the amount of outliers in the RPCA. In the original method [29], a global and static small constant is commonly used for decomposition. Unfortunately, this setting leads to the system being more easily to be perturbed by noise and background motion. Setting λ to a large value would suppress the disturbance caused by background movement, while some foreground objects are also absorbed into the background. Setting a global value of the λ leaves no space for adjusting the scale. In [36], the authors discussed the parameter regularity issues of RPCA and employed the Minimum Description Length (MDL) principle to select an optimum low-rank approximation. To address the λ control problem, we introduce a feedback process, which can filter out most non-stationary background motions and at the same time maintain foreground motions. Other similar feedback mechanisms can be found in [24]. However, in [24], the final results rely heavily on the first-pass RPCA has and there is no discussion about the first issue, which the structural distribution of sparse outliers. In section V, we will investigate how RPCA achieves incomplete detections (or many false positives when using a small λ) in some cases, for which it will be difficult to estimate foreground candidates for the second process in [24]. Therefore, compared to the work in [24], the proposed method which incorporates structured sparsity will provide more complete foreground candidates than RPCA.

In this paper, we propose a novel algorithm for foreground detection, which falls into the category of low-rank based methods. We formulate the problem in a unified framework named Low-rank and Structured sparse Decomposition (LSD).

The main contributions are summarized as follows:

1. We propose a new formulation of foreground detection in the low-rank representation. It explicitly considers spatial information in sparse outliers rather than other related methods. The new structured sparsity-inducing norms can better model the foreground. The differing from previous works, such as [32] [33] [34], is that in the decomposition, the outlier support and the low-rank matrix are estimated simultaneously, and the clean background for training are not required.

2. We adopt a group-sparse RPCA to solve the regularizing parameter issue in RPCA. The static setting of the regularizing parameter is replaced by the adaptive settings for image regions with distinct properties for each frame. Hence, the proposed method is able to tolerate sudden background variations like the changing weather conditions or turn on/off lights, without losing sensitivity to detect real foreground objects. The proposed method achieves better accuracy in terms of both foreground detection and background estimation in comparison with state-of-the-art algorithms.

This paper is organized as follows: Section II reviews prior work and related methods. In Section III, the low rank and structured sparsity decomposition are proposed which induces the structured property of foreground. In Section IV details the framework of the proposed method. Experiments and discussions are presented in Section V. Finally, conclusions are drawn in Section VI.

The preliminary work has appeared in [37].

II. Related Methods

A. Sparse Signal Recovery

With the success of Compressive Sensing [38] [39], Sparse Signal Recovery provides a popular framework to deal with various problems in machine learning and signal processing. In sparse signal recovery for background subtraction, Dikmen and Huang [40] firstly impose the sparsity constraint on the residual (foreground) term, where a new frame $y \in \mathbb{R}^m$ can be modeled as a sparse linear combination of p previous images $X \in \mathbb{R}^{m \times p}$, plus a sparse error term (residual) $e \in \mathbb{R}^m$:

$$y = Xw + e \tag{1}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing

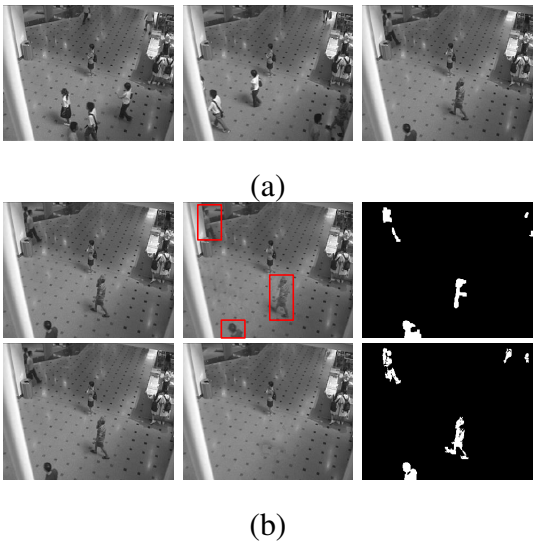


Fig. 2. An example illustrating the difference between Our method ((b) bottom) and ProxFlow [34] ((b) top). (a) The first, middle, and last frame of a sequence of 40 images. (b) The 40$^{\text{th}}$ frame ((b) left), the estimated background ((b) middle) and the detected foreground ((b) right). ProxFlow ((b) top) used the first 39 frames as the basis matrix X in Eqs. (1) and (3).

$w \in \mathbb{R}^p$ is the coefficient vector. The term Xw accounts for background parts, while the sparse error e corresponds to the foreground in y. The essential purpose of the method is to estimate the coefficients from a training dataset by minimizing a given objective function. The famous Lasso model [41] is the most commonly used objective:

$$\hat{w} = \arg\min_{w}(\|y - Xw\|_2^2 + \lambda\|w\|_1) \tag{2}$$

where $\|\cdot\|_2$ denotes the ℓ_2-norm. The estimated coefficient $\hat{w}$ can be obtained according to (2). Then the foreground is obtained as $e = y - X\hat{w}$. However, no prior knowledge on the spatial distribution of outliers e was considered. To improve the above Lasso issue, Huang $et\ al.$ proposed a Dynamic Group Sparsity (DGS) [42] [33] recovery method by making use of group clustering priors. However, in DGS, the sparsity degree number must be known in advance, otherwise the method has to set the lower bound of the sparsity range to zero, and running by incrementing the step size until certain halting conditions are satisfied, which results in a very low processing ability. Another limitation of DGS is that it requires a training sequences composed of clean background.

The method ProxFlow [34] in sparse signal recovery solves the following optimization to

recover w and e:

$$\min_{w,e} \frac{1}{2}\|y - Xw - e\|_2^2 + \lambda_1\|w\|_1 + \lambda_2\|e\|_{\ell_1/\ell_\infty} \tag{3}$$

where $\|\cdot\|_{\ell_1/\ell_\infty}$ is a norm to induce the structural sparsity. Mention that the structured sparsity inducing-norm was firstly developed in ProxFlow, but ProxFlow needs a training sequence not containing any foreground objects. However, in some actual problems, a training sequence X with only background may not be easily obtained. Fig. 2(a) gives such a sequence from an indoor surveillance video, where people are always in the scene. Fig. 2(b) shows the foreground detection result of the 40^{th} frame. Since the subspace spanned by previous frames also include foreground objects, ProxFlow cannot recover the background and thus gives incomplete segmentation. Different from [34], which employs a linear regression with fixed bases, the proposed method can estimate the foreground and background jointly by outlier detection during matrix learning without knowing X. As shown in the last row of Fig. 2, the results are obviously improved.

B. Robust Principle Component Analysis

Recently, some approaches considered foreground detection from a viewpoint of decomposition and optimization problem, which can be expressed as follows:

$$D = L + S \tag{4}$$

where $D \in \mathbb{R}^{m\times n}$ is the observed videos (n frames), L and S denote the background and foreground signals respectively. Many properties of L and S have been explored for the decomposition. Robust PCA via Principal Component Pursuit (PCP) [20] was firstly proposed to use a ℓ_1-norm to constrain the foreground matrix because these regions must be a sparse matrix with a small fraction of nonzero entries. It is also assumed that the background images are linearly correlated with each other, forming a low-rank matrix L. The decomposition could be solved by the following convex optimization:

$$\min_{L,S} \|L\|_* + \lambda\|S\|_1 \quad s.t. \quad D = L + S \tag{5}$$

In the PCP formulation, $\|L\|_*$ means the nuclear norm of matrix L, the sum of its singular values, and $\|S\|_1$ denotes the ℓ_1-norm of S. However, with this approach there remain two issues to be settled: Firstly, because the geometry of ℓ_1-norm is diamond shaped and this regularization treats each entry (pixel) independently, it does not take into account any specific structure or any

possible relations among subsets of entries [35]. While in real videos, outliers treated as moving parts usually have the structural properties of spatial contiguity and locality. Secondly, the scale issue exists in the PCP, which means that a single value of the regularizing parameter λ in the formulation cannot handle foreground or motion regions for all kinds of sizes (see Fig. 1).

In this work, to encode the prior structure of sparse outliers, we introduce the structured sparsity norm to a new formulation for foreground detection in the low-rank representation, in which the structured outlier support and the low-rank matrix are estimated simultaneously. Moreover, we adopt a new framework to solve the regularizing parameter issue. Since there is no particular value of the parameter λ to handle foreground objects of all sizes, the issue about tuning the parameter is a perennial challenge in complex background videos. We use a group-sparse RPCA with the parameter value set according to the motion saliency map. The setting involves a higher regularizing parameter λ in RPCA to lower motion salient regions (e.g. background motion), which will enable these parts to be absorbed into the background. For higher salient regions, we lower the parameter values to ensure that all the changes caused by the foreground are entirely captured.

III. Low-rank and Structured sparsity decomposition (LSD) for foreground detection

A. Structured sparsity-inducing norms for modeling sparse outliers

Suppose we have an observed matrix $D \in \mathbb{R}^{m\times n}$ (a video sequence in our problem) and $D{=}\{I_1 \ldots, I_n\}$ where I_t is the frame at time t, and n is the total number of frames. To encode the prior knowledge, we consider the structured sparsity-inducing norm that involves overlapping groups of variables, inspired by recent advances in structured sparsity [34] [35]. First, the structured sparsity norm is defined as follows:

$$\Omega(S) = \sum_{j=1}^{n} \sum_{g\in\mathcal{G}} \|s_g^j\|_\infty \tag{6}$$

In Eq.(6), given $S \in \mathbb{R}^{m\times n}$, the j^{th} column $s^j \in \mathbb{R}^m$ in S has m variables with indices $\{1, \ldots, m\}$. These indices can be partitioned into overlapping groups, and each group $g \in \mathcal{G}$ contains a subset of these indices. In this paper, we define 3×3 overlapping-patch groups, to be the same as in [34], and then each group overlaps 6 pixels with its neighbors. $\|.\|_\infty$ denotes ℓ_∞-norm and it is the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing

(a) (b) (c) (d)

Fig. 3. (a)-(b) Two distributions of sparse entries in a 8×8 frame. (c)-(d) Several 3×3 overlapping groups on two cases.

maximum value of pixels in a group. The ℓ_∞-norm encourages the rest of the pixels to take arbitrary values, which we can expect that similar error regions have similar large magnitude [34].

The foreground is usually spatially contiguous and assumed to occupy a portion of the scene, so it will be highly appropriate to model the foreground using the structured sparsity norm, because it can reflect the spatial distribution of nonzero variables and thus promote the structural distribution of sparse outliers during the minimization. To illustrate this advantage, we assume two different distributions of a sparse foreground in an 8×8 ($m = 64$) sized image (see Fig. 3), in which white pixels correspond to outliers with high values and black pixels correspond to small values. Since the ℓ_1-norm sums up the absolute values of all pixels, it will have similar values in two cases (Fig. 3(a,b)). The structured sparsity norm in Eq. (6) only sums up the largest one in each pre-designed group (36 overlapping groups in 8×8 frame), and hence it will have distinct different values: the value in the first case (Fig. 3(c)) is much smaller than that in the latter case (Fig. 3(d)) because more groups containing large value pixels appear in the latter case (Fig. 3(d)). This example shows the ℓ_1-norm treats each pixel independently and the structured sparsity norm can take into account possible relations among subsets of the entries.

B. Optimization Method

To formalize structure prior on the outliers and also promote structured sparsity of sparse outliers, we introduce the structured sparsity norm and then propose Low-rank and Structured sparse Decomposition (LSD) method, as

$$\min_{L,S} \|L\|_* + \lambda\Omega(S) \quad s.t. \quad D = L + S \tag{7}$$

where $\|L\|_*$ means the nuclear norm of matrix L, the sum of its singular values. $\Omega(S)$ means the structured sparsity norm which has been defined in Eq. (6). Eq. (7) remains an

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing

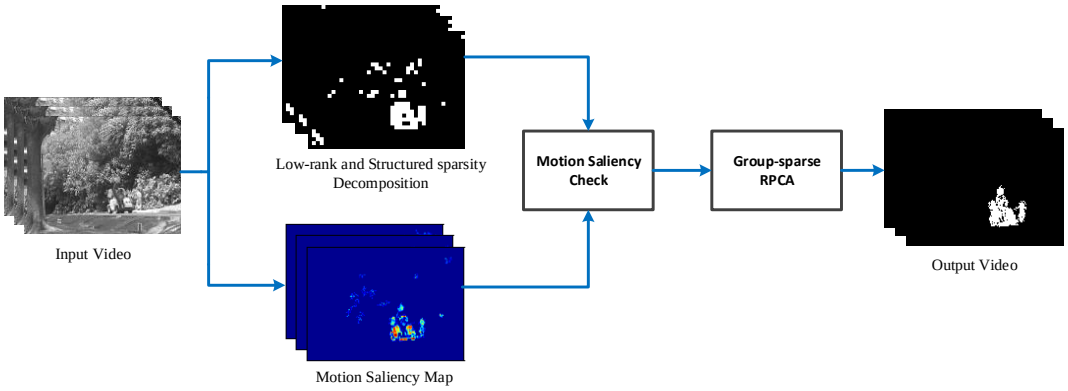


Fig. 4. Illustration of framework of the proposed method.

optimization problem and we could solve it based on Augmented Lagrange Multiplier (ALM) method [43], because ALM has good balance between efficiency and accuracy in solving related RPCA methods [22] [23] [24]. The augmented Lagrangian function is defined as:

$$\begin{aligned}\mathcal{L}_1(L,S,Y;\mu) = \|L\|_* + \lambda\Omega(S) \\ + \langle Y, D-L-S\rangle + \frac{\mu}{2}\|D-L-S\|_F^2\end{aligned} \tag{8}$$

where Y is a vector of Lagrange multipliers, μ is a positive scalar. ALM solves (8) by alternating between optimizing the primal variables L, S and updating the dual variable Y, which solves the following three sub-problems:

$$\begin{cases} L_{k+1} = \arg\min_L \mathcal{L}_1\left(L, S_k, Y_k; \mu\right) \\ S_{k+1} = \arg\min_S \mathcal{L}_1\left(L_{k+1}, S, Y_k; \mu\right) \\ Y_{k+1} = Y_k + \mu\left(D - L_{k+1} - S_{k+1}\right) \end{cases} \tag{9}$$

The first problem in (9) which solves for L with fixed S, Y can be explicitly expressed as the following form:

$$\min_L \|L\|_* + \frac{\mu}{2}\left\|\left(D - S_k + \mu^{-1}Y_k\right) - L\right\|_F^2 \tag{10}$$

In each iteration, the (10) can be rewritten as

$$L_{k+1} = \arg\min_L \left\{\|L\|_* + \frac{\mu}{2}\left\|G^L - L\right\|_F^2\right\} \tag{11}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing

where $G^L = D - S_k + \frac{1}{\mu}Y_k$. The subproblem (11) has the closed-form solution by matrix shrinkage operator. Suppose the singular value decomposition of G^L is $G^L = U\Sigma V^T$, then the matrix shrinkage operator of G^L is $U\mathcal{S}_{1/\mu}(\Sigma)V^T$ where the $\mathcal{S}$ is a soft-thresholding operator, defined as

$$\mathcal{S}_v(x) = max(0, x - v) \quad x \geq 0, v > 0 \tag{12}$$

With the same idea of developing (10), the second problem in (9) can be shown as the following equivalent formula:

$$\min_S \frac{\mu}{2}\|\left(D - L_{k+1} + \mu^{-1}Y_k\right) - S\|_F^2 + \lambda\Omega(S) \tag{13}$$

where $\Omega(S)$ is the structured sparsity norm defined in Eq. (6). The form in Eq. (13) turns out to be the proximal operator associated with a structured sparsity-inducing norm, which solutions can be obtained by solving a quadratic min-cost flow[1] problem [34].

Algorithm 1 Inexact Low-rank and Structured sparsity decomposition.

Input: Given Matrix $D \in R^{m\times n}$ and the parameter λ.

Output: Estimate of (L, S).

1: **Parameters initialization**:
2: **While** not converged **do**
3: //Line 4-5 solve $L_{k+1} = \arg\min_L \mathcal{L}_1(L, S_k, Y_k)$, as Eq. (9).
4: $G^L = D - S_k + \mu_k^{-1}Y_k$.
5: $L_{k+1} = U\mathcal{S}_{1/\mu_k}(\Sigma)V^T$, where $\mathcal{S}$ is a soft-thresholding operator
6: //Line 7-8 solve $S_{k+1} = \arg\min_S \mathcal{L}_1(L_{k+1}, S, Y_k)$.
7: $G^S = D - L_{k+1} + \mu_k^{-1}Y_k$.
8: $S_{k+1} = prox_g\left(G^S\right)$.
9: $Y_{k+1} = Y_k + \mu_k(D - L_{k+1} - S_{k+1})$.
10: $\mu_{k+1} = \rho\mu_k$; $k \leftarrow k + 1$.
11: **end while**.
12: $L \leftarrow L_k$, $S \leftarrow S_k$.

The whole algorithm is shown in Algorithm 1. In the algorithm, the error in outer loop is computed as $\|D - L_k - S_k\|_F / \|D\|_F$. The outer loop stops when it reaches the value lower than

[1]http://www.di.ens.fr/willow/SPAMS/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing

10^{-7} or the maximal iteration number 500 is reached. The ALM parameter is set $\rho = 1.1$. Please refer to [34] [31] for more details.

IV. Framework for robust foreground detection

In this section, we give details of the proposed method. Inspired by [24], for efficient computation, we operate on scaled-down low resolution video sequences (sub-sampled at a four to one ratio) by the proposed low rank and structured sparsity decomposition (LSD). After the decomposition, many foreground candidates will be detected and then a motion saliency map will be used to remove background motion and weigh the motion salient groups. Finally, a group-sparse RPCA will be introduced to obtain the final detections from groups whose size and locations have been detected (see Fig. 4).

A. Foreground group candidates via LSD

Firstly, we found that a structured sparsity based RPCA scheme can better estimate background, but it is also sensitive to some dynamic background motions. Hence, the obtained candidate groups denote both foreground objects and a few background motions. Note that pixels identified as outliers in the low resolution image correspond to a $4 * 4$ regions in the full resolution image.

B. Motion saliency check

Secondly, the likelihood of a group containing foreground should be checked out and those with little motions should be suppressed. Since background motion is usually smaller and more regular than foreground object motion, the foreground object will form a distinct trajectory from the background in a temporal slice on the X-T and Y-T planes [44]. The analysis of temporal slices will detect such distinct trajectories of foreground objects and generate a motion saliency map. In the motion saliency map, larger values typically correspond to the more motion salient pixels (e.g. $\overline{G}_A$ in Fig. 5). Since we have foreground group candidates from LSD, we can calculate each group's average saliency $\overline{G}$ by counting the pixels of each group using the motion saliency map [22] (See Fig. 5).

After calculating all group candidates' motion saliency values, we employ an adaptive threshold selection step to remove the motion non-salient groups. Similar to [44], we assume that the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing

distribution of the average values of salient groups to satisfy the Gaussian distribution (μ, σ). To remove groups with lower average values, we adopt $Th = \mu + \sigma$ as the global threshold. Besides, the relatively small groups are also rejected based on the observation that these small motion salient groups are likely to be falsely positive. Therefore, we also employ a group size threshold $T_{size} = (m \times n)/1500$ to reject the small groups.

C. Group-sparse RPCA

After the motion saliency check, many non-stationary background motions are filtered out. We can lower the regularizing parameter λ to ensure that all the changes caused by foreground motion will be entirely captured to the outlier matrix. In this step, inspired by [24], for all group i with considerable motion saliency, we set the corresponding λ_i:

$$\lambda_i = \frac{1}{\delta\sqrt{\max(m,n)}} \frac{\overline{G}_{\min}}{\overline{G}_i} \tag{14}$$

Where the last factor normalizes the computed saliency measurement $\overline{G}_i$ for each group with respect to the minimum $\overline{G}$ detected among all groups with high motion saliency values. For group with lower motion saliency value, the λ_i is set to a large value to reduce the false detection. According to [24] and our experimental results, we set $\delta = 10$.

In fact, after the two earlier steps in our proposed method, we estimated the location and size of the likely groups and also weigh each group with a different saliency measure. Then, a group-sparse RPCA is used to carry out the final foreground detection from those motion saliency groups.

$$\min_{L,S} \|L\|_* + \sum_i \lambda_i \|M_i(S)\|_F \quad s.t. \quad D = L + S \tag{15}$$

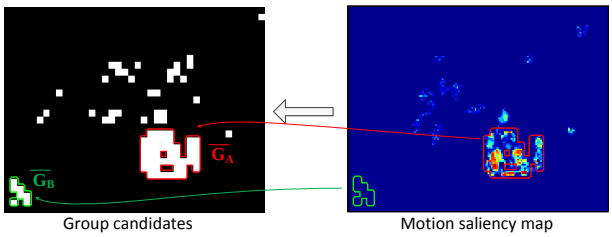


Fig. 5. Group candidates with different average motion saliency values by motion saliency map.

where $\|\cdot\|_F$ is the Frobenius norm of a matrix, and M_i represents the matrix from group i in each column of S. Eq. (15) is the group-sparse version of the RPCA, which remains a convex optimization problem and can be again solved via the inexact ALM method. The procedure of the ALM method is the same as that detailed in section III:

$$\begin{aligned} f_\mu(L,S,Y) =& \|L\|_* + \sum_i \lambda_i \|M_i(S)\|_F \\ &+ \langle Y, D-L-S\rangle + \frac{\mu}{2}\|D-L-S\|_F \end{aligned} \tag{16}$$

A similar optimization solution described in [43] [24] is adopted, and the parameter settings and conditions recommended in [43] [24] are also used, please see it for details.

V. Experimental Results

A. *Effects of separate steps*

In order to better understand the performance of the proposed algorithm, we analyzed the effects of separate steps one by one. First, to evaluate the effectiveness of the low-rank and structured sparse decomposition (LSD) method, we used LSD (excluding using motion saliency check and group-sparse RPCA) to detect foreground and compared its performance with the following approaches: the original PCP [20] ($\lambda = b/\sqrt{\max(m,n)}, b = 1$) uses ℓ_1-norm, LBD [31] [23] employs the $\ell_{2,1}$-norm for block-sparse constraint, and DECOLOR (DEC) [28] uses the MRF smoothness constraint. Due to the page limitation, in Fig. 6, we provide a short qualitative analysis on two videos from dataset I2R [8]. In fact, all sequences in I2R were tested in this paper, and we present a full quantitative (*F-measure*) evaluation in Fig. 7.

Fig. 6 shows the results of the recovery of background and detected foreground from four methods. The "Lobby" is a video with the light switch on/off, and the "Water Surface" contains a flickering water surface, those typical dynamic background changes should be quickly updated into background models, and the system should not lose its sensitivity to detect real foreground objects. From Fig. 6, we see that the algorithm LBD using the $\ell_{2,1}$-norm outperforms the PCP, the reason being that LBD enforces the low-rankness of the background part and the block-sparsity of the foreground part [31] [23]. However, the block-sparsity property still has no structured information to model sparse outliers. Hence, we also observe that the problem of foreground

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing

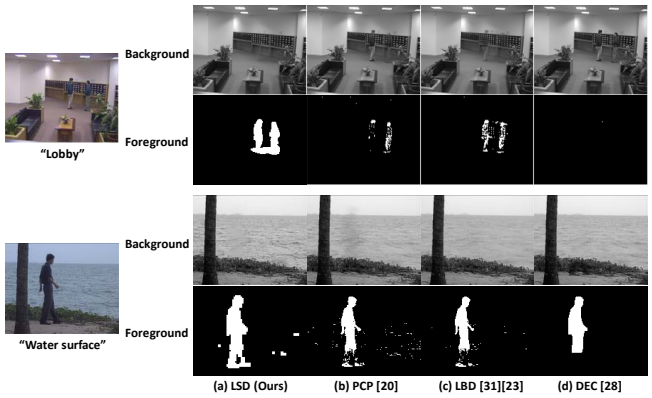
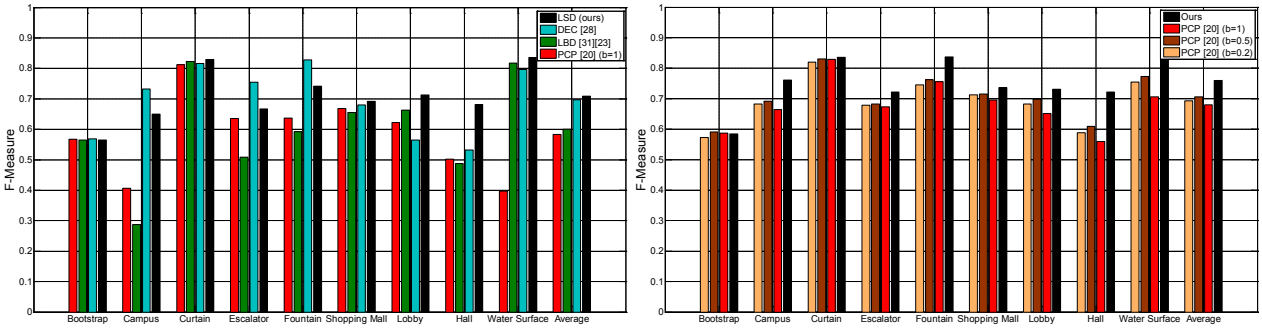


Fig. 6. Comparisons of low-rank and structured sparse decomposition (LSD) with other recent state-of-art methods.



(a) (b)

Fig. 7. *F-measure* comparisons of sequences from I2R [8] for demonstrating the effects of separate steps. (a) Comparisons of LSD with other methods. (b) Comparisons of the proposed method with PCP [20] (different λ) using motion saliency check and our group-sparse RPCA.

false negatives cannot be solved by LBD. DECOLOR failed to detect the people in "Lobby", and lost a lot of foreground in the "Water surface". It is noticed that the proposed LSD method provides much cleaner background than other methods since the structured sparsity-inducing norm encourages the structured sparsity of the foreground, and better captures the outliers than block-sparse norms or the MRF smoothness constraint. According to the average F-measure shown in Fig. 7(a), the proposed LSD can yield superior performance over the other state-of-art methods.

Following that, we focused on the regularizing parameter control and group-sparse RPCA. We consider scenarios with a dynamic background and the sequences used here are "Water

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing

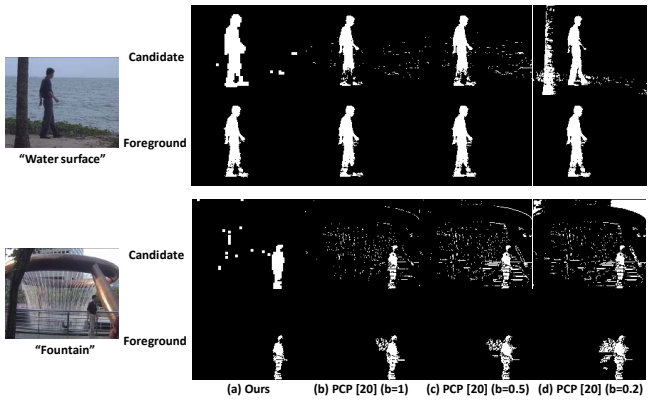


Fig. 8. Comparisons of our method with PCP [20]. The first row of each test video is the detection results of the proposed method and PCP with different λ , as the foreground group candidate. The second row is the final foreground results after motion saliency check and our group-sparse RPCA.

Surface" and "Fountain". To demonstrate the effectiveness of the proposed λ setting strategy and group-sparse RPCA, we also employed PCP to detect the foreground group candidate, then check motion saliency and use the proposed framework to obtain the final foreground results. In Fig. 8, qualitative comparisons among four methods are presented (quantitative results are illustrated in Fig. 7 (b)). It can be seen that the proposed group-sparse RPCA can promote the detection results of the different methods. By checking motion saliency and applying a regularizing parameter control, many foreground false positives can be filtered out. The original PCP ($\lambda = b/\sqrt{\max(m,n)}, b = 1$) achieved incomplete detections in "Water Surface", which make the final foreground result lost many true positives in Fig. 8(b). In particular, we chose the results obtained by PCP with a smaller λ ($b = 0.2\ and\ 0.5$), as it can provide more foreground group candidates than the original PCP. However, as shown in Fig. 8(c), PCP with a smaller λ produced more false positives than original PCP method in the "Fountain", and degraded the performance of group-RPCA on the final foreground result. It is noticeable that the proposed method can yield superior performance over PCP with different λ, since the LSD can provide more accurate foreground candidates than others by imposing the structural properties of outliers. Visually, the proposed method obtains the best foreground mask among all methods. This is confirmed by the performance of the *F-measure* in Fig. 7(b).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing

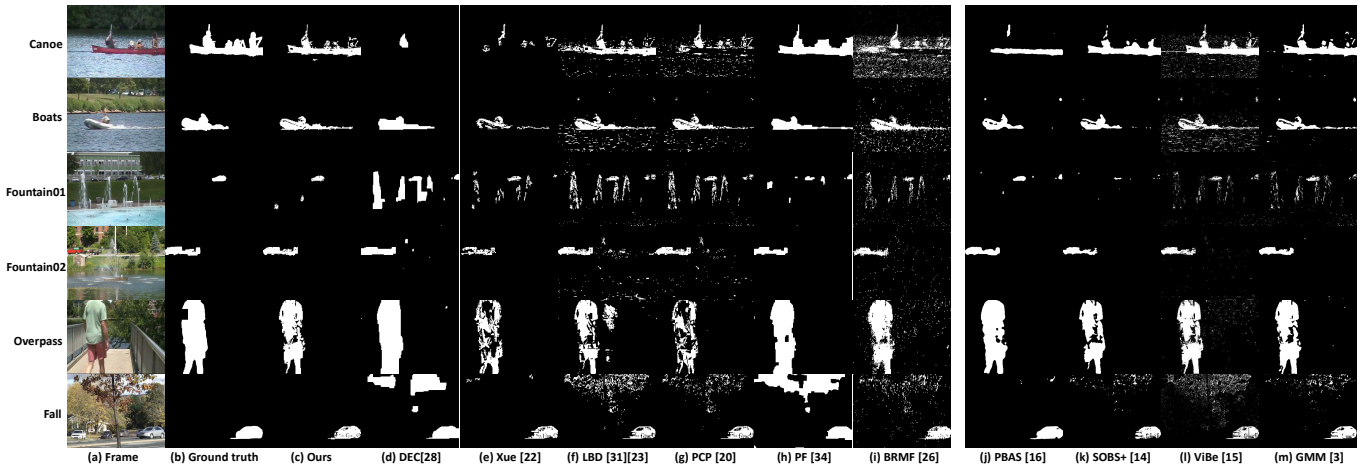


Fig. 9. Detected foreground results of six dynamic background videos from CDnet [45]. (c)-(i) are RPCA and sparse-related algorithms, and (j)-(m) are pixel-based methods.

B. Data and Qualitative Results

The experiments are conducted qualitatively on real sequences from the CDnet[1] dataset [45], and I2R[2] dataset [8]. We compare the proposed algorithm with the six RPCA and sparse-related algorithms, namely DECOLOR (DEC) [28], LBD [31] [23], PCP [20], Xue's method (Xue) of [22], ProxFlow (PF) [34], and BRMF [26], and four state-of-the-art pixel-based background modeling algorithms: PBAS [16], SC-SOBS (SOBS+) [14], ViBe [15] and the improved Gaussian Mixture Model by Zivkovic (GMM) [3]. For the RPCA-related method, a threshold criterion is required to get the final foreground mask and we adopt the same threshold strategy as in [24]. For other approaches, we use the default parameters for experiments. In the experiments, the proposed and PRCA-related algorithms were implemented in batch mode with the same input frames and no post-processing was applied to any of methods for fair comparisons (i.e., any morphological operations were not conducted).

Fig. 9 shows the detected foreground masks on videos from the CDnet dataset [45]. There are six videos in this category depicting outdoor scenes exhibiting dynamic background motion. The objective of the experiments is to illustrate the ability of our proposed method for dealing with complex dynamic scenes. The first two rows depict boats on shimmering water, but

[1]www.changedetection.net

[2]perception.i2r.a-star.edu.sg

only our method, ProxFlow [34], and SC-SOBS [14] preformed relatively well. GMM [3] and ViBe [15] can detect most of foreground regions, but they also produces plenty of false positives. BRMF [26] and DECOLOR [28] failed to correctly detect the canoe. It is noticed that LBD did not manage to judge water motion. By adding constraint for background in our method, we ensure that foregrounds are not be absorbed into background, and our method returned a small motion saliency value for water motion and did not regard it as a foreground motion. We adopted Xue's method [22] as potential saliency map in motion saliency measurement but we evaluated average motion saliency for each group to have different λ values. Therefore, in the last group-sparse RPCA step, we can refine the results but Xue's method is unable to do so. The third and fourth rows of Fig. 9 represent cars passing next to a fountain, it is not difficult to find that PBAS [16] and the proposed are the best two methods. The proposed method can handle dynamic background and remove background motion as well as preserve foreground. It can be seen that BRMF cannot always remove the dynamic background motions. DECOLOR typically can detect the most foreground pixels, but it will produce more false alarms due to the smoothness constraint imposed on the foreground shapes. SC-SOBS cannot get the complete foreground results of the moving car in "Fountain01". As discussed earlier, RPCA-based methods have problems in setting a correct regularizing parameter λ that cannot handle regions or motions of varying scales. The last two rows show pedestrians, cars and trucks passing in front of a tree shaken by the wind. With the exceptions of PBAS and our method, none of others can eliminate background motion in the scene "Fall". Yet, we would like to point out a weakness in the proposed and RPCA-related methods. As can be seen in the fifth row, when a large object moves into the scenario, the sparse constraint of the outlier may lead to incomplete detection of the foreground. In most of the sequences, the proposed method and PBAS can tolerate such background motions, while PBAS may lose sensitivity to detect real foreground completely, as in "Canoe". After all, in RPCA and sparse-related algorithms, the proposed method obtained the best performance in most videos.

To provide a better evaluation, we tested the proposed method on another widely used I2R dataset [8], which includes videos with typical dynamic background, indoor scenarios and sudden illumination changes. The results are shown in Fig. 10. In the first row, this scenario contained significant motion of the curtain, as well as the background changes caused by automatic gain adjustment. A person is seen wearing a bright color cloth which is similar to the color of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing

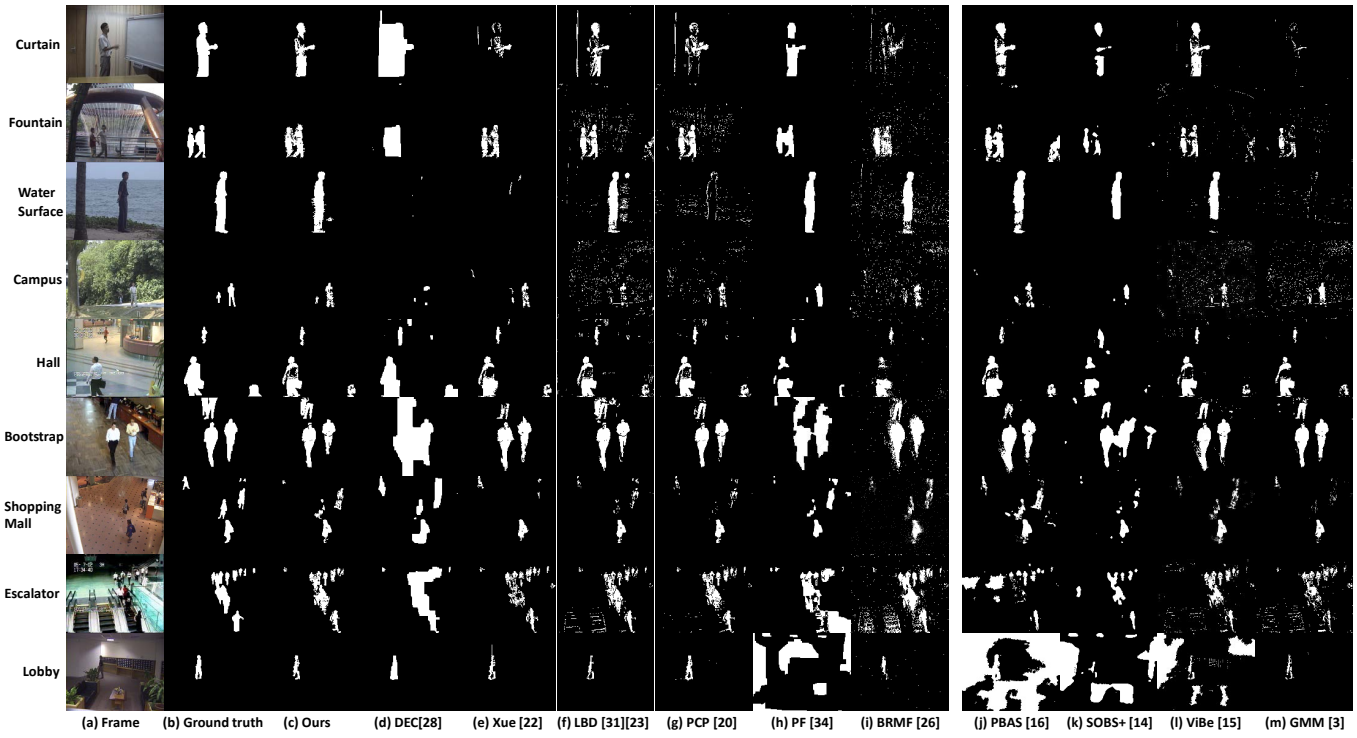


Fig. 10. Detected foreground results of nine videos from I2R dataset [8]. (c)-(i) are RPCA and sparse-related algorithms, and (j)-(m) are pixel-based methods.

the curtain. The results shown that the proposed method has detected the person quite well in such an environment. The next two rows show the sequences about the water, with a non-stationary background of the fountain and water waves. In the "Fountain" scene, all first three methods and ProxFlow can handle water movement, but among them, DECOLOR and Xue failed to detect the target in some cases. LBD and BRMF can detect the complete foreground objects, but these methods produce plenty of false positives. As shown in the "Water surface", ProxFlow, PBAS and the proposed method yield superior performance over others. "Campus" shows a scene with a typical dynamic background caused by motion of tree branches. The results have shown that our method can provide better performance in handling such non-stationary background than others. In the next three rows, these three test sequences provide the indoor environment. As discussed in Section II, a key distinction between the proposed and ProxFlow is the assumption about the availability of training sequences with/without a foreground object. For ProxFlow, background modeling using sparse signal recovery requires a set of background

frames without foreground, which is not always available, especially for surveillance of crowded scenes. “Shopping Mall” gives such an environment where people walking are always in the scene. ProxFlow cannot recover the background and gives inaccurate segmentation. Instead, RPCA-related methods can estimate a clean background from occluded data, while DECOLOR tended to detect bulb shapes with large false alarm. On the other hand, without a structured sparsity constraint, Xue, LBD, and PCP fail to obtain complete foreground. For other methods, SC-SOBS detected many wrong regions where no foreground actually existed, and lost much of the real foreground. The proposed method produced similar visual results to BRMF, PBAS and ViBe, and are found to be better than the others. In the sequence “Escalator”, the motion of escalators would make the background motion regions difficult to be removed. Further, the background model is hard to establish if there is a steady stream of human flow in the scenes. LBD and PCP failed to eliminate motion of escalator. PBAS performed poorly with more false positives and missing pixels. The proposed method lost fewer foreground pixels than SC-SOBS, and had fewer false positives than DECOLOR. In the last row, the scene comprises of a light that is switch on/off. What is noteworthy is that ProxFlow failed to handle such light variation. This is not surprising because the training sequences composed of background variation are required for ProxFlow. Thus, except PABS, SC-SOBS and ViBe, other methods can avoid the problem of sudden changes in lighting.

Visually, the results of the proposed method look better and are the closest to ground-truth references. A quantitative evaluation provides more solid conclusions on the performance of the proposed method.

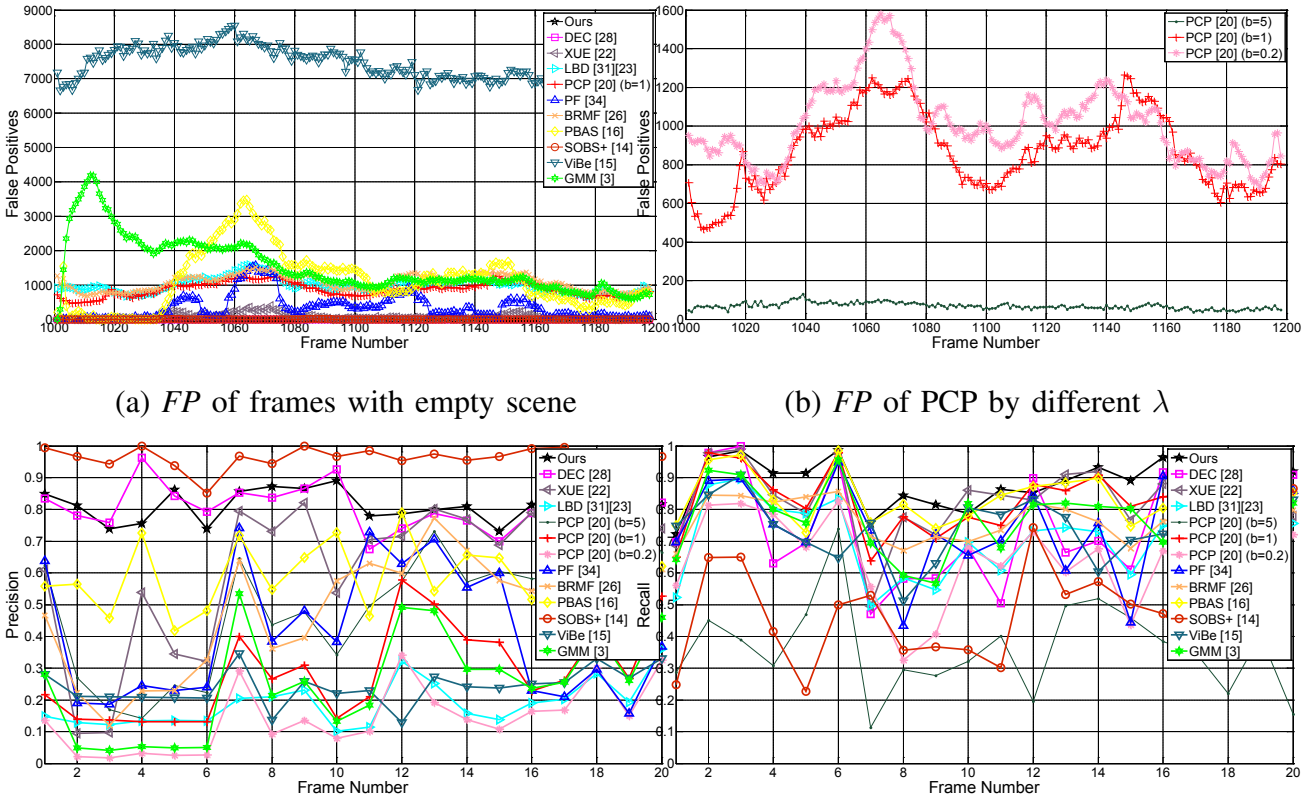
C. Quantitative evaluations

To get an accurate evaluation of the proposed method, the criteria of *recall* and *precision* [19] are employed:

$$recall = \frac{TP}{TP + FN}, \quad precision = \frac{TP}{TP + FP} \tag{17}$$

In the *recall* calculation, ***TP*** is the total number of correctly classified foreground (true positives), and ***FN*** is the total number of false negatives, which accounts for the incorrect number of foreground pixels classified as background. In *precision* calculation, ***FP*** is the total number of false positives, which means the pixels are incorrectly classified as foreground.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing



(a) *FP* of frames with empty scene

(b) *FP* of PCP by different λ

(c) *precision* of 20 frames with moving object

(d) *recall* of 20 frames with moving object

Fig. 11. False positives(*FP*), *precision* and *recall* comparisons of sequence "Campus".

The first sequence tested here is the "Campus" [8], with two sources of dynamic motion: tree waving caused by wind, and their shadows on the ground surface. The "Campus" is a commonly used foreground segmentation test sequence, especially for evaluating the effectiveness of dynamic background modeling techniques. As shown in Fig. 11(a), the scene is empty from frames 1000 to 1200, and False Positives (FP) caused by background movement were observed. For other RPCA-based methods, such as LBD [31] [23] and PCP [20] produced a lot of false positives. By imposing structural properties of the outlier, the proposed method can handle dynamic background immediately and a negligible number of FP are detected. In order to better understand why the RPCA-based methods have problems in setting regularizing parameter, for PCP, we experimented with different λ. In Fig. 11(b), we observed that PCP with a relatively small regularizing value ($\lambda = b/\sqrt{\max(m,n)},\ b = 0.2\ or\ 1$) produced plenty of FP. In contrast,

setting a higher λ ($b = 5$) will decrease FP, but these positives cannot be entirely eliminated. On the other hand, in that case, a lot of foreground will be falsely treated as background, making the rate of *recall* very low. As shown in Fig. 11(c) and (d), *recall* and *precision* are obtained for the frames which contain the moving objects (e.g. person, car). It is noteworthy that PCP with smaller λ ($b = 0.2$) did not perform better than the default setting ($b = 1$) [20] in *precision*. This is not surprising because the small setting of λ makes the system very sensitive to background motion or noises. It also can be seen that no matter how the regularizing value is chosen, PCP cannot get a satisfactory result. And the proposed method with parameter control can yield superior performance with respect to the PCP. It is also noticeable that Xue [22] achieved a relatively low FP result by using the motion saliency map; however, the results of *recall* and *precision* demonstrate an unsteady performance. For other methods, PBAS [16], ViBe [15], and GMM [3] failed to handle such dynamic background according to the three criteria in Fig. 11. Because of the lack of clean frames to train the background model, ProxFlow [34] also has difficulty to get satisfactory results. BRMF [26] can detect the complete foreground objects, but it also produces plenty of background motions. DECOLOR [28] did not have any false positives and by enforcing the MRF constraint on the foreground region, DECOLOR can usually get a high *recall* rate. However, DECOLOR achieved a rather low *recall* in "Campus". It is noticeable that DECOLOR has a similar performance in "Canoe", as shown in Fig. 9. These results seem to indicate that DECOLOR may fail to carry out foreground detection in such a complex background. Another cause for concern is that SC-SOBS [14] is able to remove the dynamic background movement very well, but this method tend to produce an incomplete foreground object, as well as missing a small moving object. It also can be seen that SC-SOBS is the best performing algorithm in *precision* but also is one of the worst methods in *recall*. This is because a large part of the real foreground objects are wrongly classified as background (see Fig. 10). Compared to the above methods, the proposed method not only provided the highest rate in *recall* but also provides a balanced performance with respect to *precision*. To take both the *precision* and *recall* into account, the criteria of *F-measure* is used.

$$F = 2\frac{recall \cdot precision}{recall + precision} \tag{18}$$

In Table I, we compared the proposed method with other methods. It is noticeable that PBAS

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2419084, IEEE Transactions on Image Processing

TABLE I

PERFORMANCE OF *F-measure*(%) ON DATASETS I2R [8] AND CDNET [45](BEST: BOLD, SECOND BEST: UNDERLINE)

Video	Ours	DEC	Xue	LBD	PCP	PF	BRMF	PBAS	SOBS+	ViBe	GMM
Boats	79.71	49.48	40.18	44.69	34.15	74.96	52.88	36.11	**89.57**	63.30	74.74
Bootstrap	58.42	56.86	54.62	56.47	56.78	58.71	57.15	55.38	47.47	63.30	**66.36**
Campus	**76.13**	73.29	66.75	28.75	40.69	42.42	64.93	69.50	63.79	36.17	34.77
Canoe	83.51	6.67	36.47	58.14	37.09	80.22	59.17	71.96	**95.24**	68.44	81.51
Curtain	**83.57**	81.58	30.25	82.22	81.28	83.06	72.48	82.84	60.28	82.49	30.52
Escalator	72.14	**75.46**	66.53	50.90	63.54	56.69	70.52	31.13	57.52	57.04	50.44
Fall	74.27	54.20	60.95	26.80	34.90	21.25	45.83	**87.14**	27.75	32.77	42.38
Fountain	**83.71**	82.76	63.31	59.25	63.74	80.73	72.41	62.04	48.59	55.80	50.06
Fountain01	33.15	2.71	6.66	3.90	4.31	23.59	21.49	**41.73**	11.58	6.05	8.15
Fountain02	85.36	75.22	82.50	34.98	45.92	79.91	59.35	**93.55**	81.77	63.38	79.17
Hall	**72.22**	53.20	49.87	48.75	50.25	69.87	69.37	61.92	51.12	61.16	42.17
Lobby	**73.13**	56.52	46.88	66.26	62.18	48.66	53.44	56.74	30.50	26.41	42.62
Overpass	74.39	87.72	63.59	52.32	47.92	84.49	63.74	79.25	**88.37**	66.93	86.72
Shopping Mall	**73.62**	68.06	73.34	65.49	66.78	54.68	65.77	71.77	64.39	68.54	67.25
Water Surface	90.50	79.65	27.74	81.75	39.76	**93.11**	80.87	89.02	84.80	86.02	32.88
Average	**74.26**	60.23	51.31	50.71	48.62	63.49	60.63	66.01	60.18	55.85	52.65

Fig. 12. Results of the proposed method on nine real videos from BMC dataset [46].

TABLE II

PERFORMANCE OF *F-measure*(%) ON DATASET BMC [46] (BEST: BOLD, SECOND BEST: UNDERLINE)

Video	Ours	IALM	ADM	LADMAP	LSADM	LADM	BLWS	NSA	TFOCS	ALM	VBRPCA	SGoDec
001	79.20	78.70	**79.30**	**79.30**	74.10	67.80	62.50	74.10	78.50	62.80	71.20	75.20
002	**80.70**	75.40	74.90	75.60	76.80	74.50	80.50	**80.70**	75.10	**80.70**	76.60	73.30
003	94.10	92.80	91.50	91.50	**94.20**	93.10	89.90	**94.20**	92.30	91.60	**94.20**	93.80
004	**88.90**	83.30	82.50	82.40	87.50	85.20	82.30	87.60	83.30	84.20	87.40	83.70
005	**73.20**	70.10	69.90	69.70	69.90	65.80	59.80	59.60	70.40	59.60	68.70	70.40
006	80.50	79.20	78.60	78.80	80.50	78.60	75.70	80.50	79.30	76.30	**81.30**	78.70
007	**81.90**	68.60	68.00	68.10	70.00	67.90	81.80	70.00	68.70	81.70	70.50	67.50
008	**84.60**	77.40	76.90	77.20	77.60	76.60	65.50	77.60	77.90	69.80	76.80	76.00
009	**92.10**	90.40	90.60	90.60	90.10	85.90	81.20	84.30	90.40	84.30	89.60	89.40
Average	**83.90**	79.50	79.10	79.20	80.10	77.30	75.50	78.70	79.50	76.70	79.50	78.70

is one of the best methods according to the evaluation results on the CDnet. The proposed method obtained the best average *F-measure* against all the other methods, and for most parts of the sequences, our method ranked amongst the top two of all methods. For "Bootstrap" and "Overpass", results are still acceptable. In the "Overpass", DECOLOR reached the highest *F-measure* value in six RPCA and sparse related methods. This is because for a large sized object, such a false positives by the MRF constraint seems not to be seriously considered. But it is for that smoothness constraint, DECOLOR may produce many false alarms or lose foreground in a highly dynamic background e.g. "Fall", "Fountain01". For the "Fountain01", all methods failed to remove the moving background, therefore, the lasting fountain detection may seriously reduce the *precision* and also the *F-measure*. Though there were no considerable values, the proposed method achieved the second highest *F-measure*. After all, the proposed method can effectively remove background motions and the average *F-measure* shows significant improvement.

Moreover, for the sake of comparison, we tested the proposed method on the newest dataset, the Background Models Challenge (BMC)[1] dataset [46], which includes nine real videos. This dataset has been built in order to test the algorithms reliability during time and in difficult situations, such as outdoor scenes. So, really long videos (Video 005 and 009) are available that can present

[1]bmc.univ-bpclermont.fr

long time change in luminosity. This dataset allows us to test the influence of some difficulties (different ground type, presence of vegetation, casted shadows, presence of a continuous car flow near to the surveillance zone, etc.) encountered during the object extraction phase [46]. Similar to the recent survey paper [21], we compared the proposed method with six algorithms for solving RPCA-PCP : IALM [43], ADM [47], LADMAP [48], LSADM [49], LADM [50], BLWS [51], and RPCA-SPCP solved via NSA [52], RPCA-QPCP solved via TFOCS [53], RPCA-BPCP (LBD) solved via ALM [31] [23], Bayesian RPCA algorithm VBRPCA [54], and approximated RPCA algorithm SemiSoft GoDec (SGoDec) [55]. We visualize the results of the proposed method in Fig. 12. The *F-measure* scores[2] of all algorithms are compared in the Table II. Clearly, the proposed method separated the background and foreground satisfactorily. The results in Table II have shown that the proposed method outperforms other algorithms for the videos 002, 004, 005, 007, 008 and 009, and achieved the highest average *F-measure* on nine test videos.

Our method has been implemented in MATLAB. All methods were tested on a PC with a 3.0 GHz Intel Core Duo CPU and 4GB RAM. Finally, since the proposed and PRCA-related algorithms were executed in batch mode, we report the average processing time on the sequence of 200 frames to complete our analysis. Dealing with a resolution of 320×240, the proposed method needs 1062s, DECOLOR costs 869s, Xue uses 821s, LBD requires 854s, PCP needs 776s, and ProxFlow runs for nearly 950s. For resolution 160×120, the CPU time are 259, 204, 178, 199, 165, and 239 seconds, respectively. In the proposed method, the dominant cost comes from the computation of SVD in each iteration.

VI. Conclusion

RPCA is a widely-used technology, while in the field of background subtraction, there are two issues that need to be discussed. To refine RPCA-based foreground detection, firstly, a novel low-rank and structured-sparse matrix decomposition method was proposed to take into account the spatial connection of the foreground regions. Then, we assign individual regularizing parameter for each pixel group with respect to space and time and updated them over time. Finally, a group-based PRCA was proposed to ensure that the system is able to tolerate dynamic background

[2]Experimental results under comparison are come from paper [21]

variations, without losing the sensitivity to detect real foreground objects. The proposed approach was tested in most of the challenging situations. Qualitative and quantitative evaluations have shown that an improved performance for foreground detection has been achieved especially for a dynamic background. In the future, further research is needed to investigate the meeting of real-time requirements.

REFERENCES

[1] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image Vison Comput.*, vol. 27, no. 12, pp. 1743–1759, 2009.

[2] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 1999.

[3] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Proc. IAPR Int. Conf. Pattern Recognit.*, vol. 2, 2004, pp. 28–31.

[4] H. Lin, J. Chuang, and T. Liu, "Regularized background adaptation: a novel learning rate control scheme for gaussian mixture modeling," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 822–836, 2011.

[5] X. Liu and C. Qi, "Future-data driven modeling of complex backgrounds using mixture of gaussians," *Neurocomputing*, vol. 119, pp. 439–453, 2013.

[6] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2000, pp. 751–767.

[7] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 831–843, 2000.

[8] L. Li, W. Huang, I. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1459–1472, 2004.

[9] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1778–1792, 2005.

[10] M. Heikkilä and M. Pietikäinen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657–662, 2006.

[11] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground–background segmentation using codebook model," *Real-time imaging*, vol. 11, no. 3, pp. 172–185, 2005.

[12] J. Guo, Y. Liu, C. Hsia, M. Shih, and C. Hsu, "Hierarchical method for foreground detection using codebook model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 6, pp. 804–815, 2011.

[13] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1168–1177, 2008.

[14] ——, "The SOBS algorithm: what are the limits?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2012, pp. 21–26.

[15] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, 2011.

[16] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2012, pp. 38–43.

[17] G. Xue, L. Song, and J. Sun, "Foreground estimation based on linear regression model with fused sparsity on outliers," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 8, pp. 1346–1357, 2013.

[18] A. Bugeau and P. Pérez, "Detection and segmentation of moving objects in complex scenes," *Comput. Vis. Image Underst.*, vol. 113, no. 4, pp. 459–476, 2009.

[19] S. Brutzer, B. Hoferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 1937–1944.

[20] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, 2011.

[21] T. Bouwmans and E. H. Zahzah, "Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance," *Comput. Vis. Image Underst.*, vol. 122, pp. 22–34, 2014.

[22] Y. Xue, X. Guo, and X. Cao, "Motion saliency detection using low rank and sparse decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 1485–1488.

[23] C. Guyon, T. Bouwmans, and E. Zahzah, "Foreground detection based on low-rank and block-sparse matrix decomposition," in *Proc. IEEE Int. Conf. Image Process.*, 2012, pp. 1225–1228.

[24] Z. Gao, L. Cheong, and M. Shan, "Block-sparse RPCA for consistent foreground detection," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2012, pp. 690–703.

[25] N. Wang, T. Yao, J. Wang, and D. Yeung, "A probabilistic approach to robust matrix factorization," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2012, pp. 126–139.

[26] N. Wang and D. Yeung, "Bayesian robust matrix factorization for image and video processing," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 1785–1792.

[27] B. Wohlberg, R. Chartrand, and J. Theiler, "Local principal component pursuit for nonlinear datasets," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 3925–3928.

[28] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597–610, 2013.

[29] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 2080–2088.

[30] H. Xu, C. Caramanis, and S. Sanghavi, "Robust PCA via outlier pursuit," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 2496–2504.

[31] G. Tang and A. Nehorai, "Robust principal component analysis based on low-rank and block-sparse matrix decomposition," in *Annual Conference on Information Sciences and Systems*. IEEE, 2011, pp. 1–5.

[32] V. Cevher, M. F. Duarte, C. Hegde, and R. Baraniuk, "Sparse signal recovery using markov random fields," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 257–264.

[33] J. Huang, X. Huang, and D. Metaxas, "Learning with dynamic group sparsity," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 64–71.

[34] J. Mairal, R. Jenatton, F. R. Bach, and G. R. Obozinski, "Network flow algorithms for structured sparsity," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1558–1566.

[35] K. Jia, T. Chan, and Y. Ma, "Robust and practical face recognition via structured sparsity," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2012, pp. 331–344.

[36] I. Ramírez and G. Sapiro, "Low-rank data modeling via the minimum description length principle," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 2165–2168.

[37] J. Yao, X. Liu, and C. Qi, "Foreground detection using low rank and structured sparsity," in *Proc. IEEE Int. Conf. Multimed. Expo.*, 2014, pp. 1–6.

[38] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[39] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[40] M. Dikmen and T. S. Huang, "Robust estimation of foreground in surveillance videos by sparse error estimation," in *Proc. IAPR Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.

[41] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[42] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *J. Mach. Learn. Res.*, vol. 12, pp. 3371–3412, 2011.

[43] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[44] X. Cui, Q. Liu, and D. Metaxas, "Temporal spectral residual: fast motion saliency detection," in *Proc. ACM Int. Conf. Multimed.* ACM, 2009, pp. 617–620.

[45] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection. net: A new change detection benchmark dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2012, pp. 1–8.

[46] A. Vacavant, T. Chateau, A. Wilhelm, and L. Lequièvre, "A benchmark dataset for outdoor foreground/background extraction," in *ACCV Workshops*. Springer, 2012, pp. 291–300.

[47] X. Yuan and J. Yang, "Sparse and low-rank matrix decomposition via alternating direction methods," *Optimization Online*, 2009.

[48] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 612–620.

[49] D. Goldfarb, S. Ma, and K. Scheinberg, "Fast alternating linearization methods for minimizing the sum of two convex functions," *Mathematical Programming*, vol. 141, no. 1-2, pp. 349–382, 2013.

[50] Y. Shen, Z. Wen, and Y. Zhang, "Augmented lagrangian alternating direction method for matrix separation based on low-rank factorization," *Optimization Methods and Software*, vol. 29, no. 2, pp. 239–263, 2014.

[51] Z. Lin and S. Wei, "A block lanczos with warm start technique for accelerating nuclear norm minimization algorithms," *arXiv preprint arXiv:1012.0365*, 2010.

[52] N. S. Aybat, D. Goldfarb, and G. Iyengar, "Fast first-order methods for stable principal component pursuit," *arXiv preprint arXiv:1105.2126*, 2011.

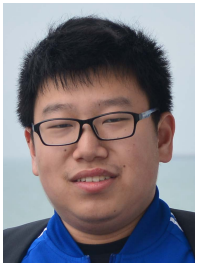
[53] S. Becker, E. J. Candès, and M. Grant, "Tfocs: Flexible first-order methods for rank minimization," in *Low-rank Matrix Optimization Symposium, SIAM Conf. on Optimization*, 2011.

[54] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse Bayesian methods for low-rank matrix estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, 2012.

[55] T. Zhou and D. Tao, "Godec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 33–40.

Xin Liu received the B.S. degree from the Changchun University of Science and Technology, Changchun, China, and the M.S. degree from Kunming University of Science and Technology, Kunming, China, in 2003 and 2007, respectively, both in computer science. He is a Ph.D. candidate at the School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, China. He is currently a Researcher with the Center of Machine Vision Research, Department of Computer Science and Engineering, University of Oulu, Oulu, Finland. His current research interests include object detection, social signal processing, gesture and action understanding.

Guoying Zhao (SM' 12) received the Ph.D. degree in computer science from the Chinese Academy of Sciences, Beijing, China, in 2005. She is currently an Associate Professor with the Center for Machine Vision Research, University of Oulu, Finland, where she has been a Researcher since 2005. Since 2011, she has held an Academy Research Fellow position of Academy of Finland. She has authored and edited three books and two special issues, and has authored and co-authored more than 100 papers in journals and conferences. She is a Co-Organizer of ECCV 2014 Workshop on Spontaneous Facial Behavior Analysis: Long Term Continuous Analysis of Facial Expressions and Microexpressions and ACCV 2014 Workshop on RoLoD: Robust local descriptors for computer vision. She is a Program Committee Member for many conferences. Her current research interests include gait analysis, dynamic-texture recognition, facial-expression recognition, human motion analysis, and person identification.



Jiawen Yao received the M.S. degree in information and communication engineering from Xi'an Jiaotong University, Xi'an, China, in 2014, and the B.S. degree in information engineering from Xi'an Jiaotong University in 2011. He is currently a Ph.D. student with the Department of Computer Science and Engineering, University of Texas, Arlington, USA. His current research interests include video surveillance and image processing.

Chun Qi received the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 2000. He is currently a Professor and Ph.D. supervisor at School of Electronics and Information Engineering, Xi'an Jiaotong University. His current research interests mainly include image processing, pattern recognition and signal processing.