IEEE Open Journal of Signal Processing

Received 6 July 2020; revised 5 November 2020; accepted 5 November 2020. Date of publication 19 November 2020; date of current version 15 December 2020. The review of this article was arranged by Associate Editor Prof. Pascal Frossard.

Digital Object Identifier 10.1109/OJSP.2020.3039325

Masked-RPCA: Moving Object Detection With an Overlaying Model

AMIRHOSSEIN KHALILIAN-GOURTANI [1] (Student Member, IEEE), SHERVIN MINAEE [2] (Member, IEEE), AND YAO WANG [1] (Fellow, IEEE)

[1]Electrical and Computer Engineering Department, New York University Tandon School of Engineering, Brooklyn, NY 11201 USA
[2]Snap, Inc., Machine Learning R&D, Seattle, WA 98121 USA

CORRESPONDING AUTHOR: AMIRHOSSEIN KHALILIAN-GOURTANI (e-mail: akg404@nyu.edu).

The work of Amirhossein Khalilian-Gourtani was supported by the New York University Ernst Weber Fellowship.

ABSTRACT Moving object detection in a given video sequence is a pivotal step in many computer vision applications such as video surveillance. Robust Principal Component Analysis (RPCA) performs low-rank and sparse decomposition to accomplish such a task when the background is stationary and the foreground is dynamic and relatively small. A fundamental issue with the RPCA is the assumption that the low-rank and sparse components are added at each pixel, whereas in reality, the moving foreground is overlaid on the background. We propose the masked decomposition (i.e. an overlaying model) where each element either belongs to the low-rank or the sparse component, decided by a mask. We introduce the Masked-RPCA (MRPCA) algorithm to recover the mask (hence the sparse object) and the low-rank components simultaneously, via a non-convex formulation. An adapted version of the Douglas–Rachford splitting algorithm is utilized to solve the proposed formulation. Our experiments using real-world video sequences show consistently better performance for both cases of static and dynamic background videos compared to RPCA and its variants based on the additive model. Additionally, we show that utilizing non-convex priors in our formulation leads to improved results without any added complexity compared to a relaxed formulation using convex surrogates and methods based on the additive model.

INDEX TERMS Moving object detection, foreground-background subtraction, nuclear-norm minimization, low-rank matrices, sparsity, ℓ_0-pseudo-norm minimization, video surveillance.

I. INTRODUCTION

A. MOTIVATION AND PRIOR WORK

Moving object detection is the initial step to many video analysis applications. Given a sequence of video frames, the goal is to separate the moving object (called the "foreground") from the static parts of each frame (called the "background"). For instance, in video surveillance applications, we may need to detect the activities that stand out from the background. Many different approaches have been proposed over the past two decades to tackle this problem. Initial approach was to model the static background using Principal Component Analysis (PCA) [1]. This method provides a model for the background and the foreground detection is consequently achieved by thresholding the difference between the original frame and the generated background.

Let us consider the matrix $M \in \mathbb{R}^{mn\times k}$ constructed by vectorizing the video frames of size $m \times n$ in to mn-vectors and stacking them as columns of the matrix M (see Fig. 1). The matrix M should approximately be low-rank (in an extreme case that the video is captured from an static scene, the columns of matrix M are nearly identical). The classical PCA approach seeks the best low-rank matrix L to approximate M

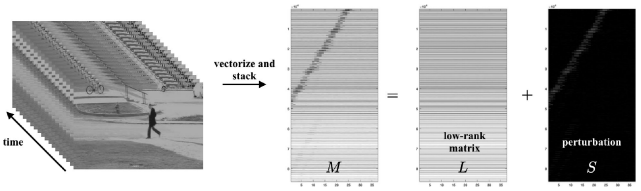


FIGURE 1. Frames of a video stacked as the columns of the matrix M. RPCA models the background as a low-rank matrix (L) and the moving object as addivive purturbations (S) to the background, i.e. $M = L + S$.

by solving

$$\underset{L}{\text{minimize}} \ \|M - L\|_2 \quad \text{subject to rank}(L) \leq k \qquad (1)$$

where $\|X\|_2$ denotes the 2-norm of a matrix X (largest singular value of X).

Although PCA is a very powerful technique for data analysis, its limitations with respect to outliers and noise put its validity for the moving object detection task in question. PCA models the cases where small, i.i.d Gaussian perturbations are applied to an underlying low-rank matrix. Considering the moving objects as the perturbations to the low-rank background, one can observe that the changes can often have large magnitude and do not necessarily follow a Gaussian distribution. Several PCA improvements have been proposed to address these limitations [2].

Another approach to improve over the PCA formulation is to decompose the data matrix M into two components, one low-rank and the other sparse [3], [4]. Robust Principal Component Analysis (RPCA) [3] attempts to efficiently find matrices L and S such that $M = L + S$ with L as the low-rank and S as the sparse component (as in Fig. 1). RPCA has no particular assumption on the magnitude of the perturbation. Additionally, there is no need to know the locations of the perturbations, the only requirement is the perturbations to be sparse. Although it might seem surprising from the classical statistics point of view, this separation was proved to be possible via a convex optimization program[3] (see Theorem 1.1). The notion of incoherence introduced in [5] for the matrix completion problem and briefly discussed in [3] provides a condition that makes the separation problem meaningful. Without going into details, the incoherence condition asserts that the underlying low-rank component is not sparse [3].

The foreground-background subtraction via RPCA model can be achieved by solving the convex optimization problem as

$$\underset{L,S}{\text{minimize}} \quad \|L\|_* + \lambda\|S\|_1 \quad \text{subject to} \quad M = L + S \qquad (2)$$

with variables L and S, where $\|X\|_*$ denotes the nuclear norm (sum of singular values), $\|X\|_1$ denotes the sum of absolute value of the entries (not the ℓ_1-norm in the matrix sense), and λ is a regularization parameter. This formulation is also known as principal component pursuit (PCP) method in the literature and can be solved by utilizing the augmented Lagrange multiplier (ALM) method [6], [7]. A sample result obtained by applying the PCP on 100 frames of the "Pedestrians" video [8] is shown in Fig. 2.

There has been several improvements over the PCP formulation in the past few years [9], [10]. One limitation of the formulation as in (2) is the fact that the presence of noise or dynamic changes in the background is not modeled. In real application, video sequence is also corrupted with entry-wise noise or there might be small dynamic changes in the background. Stable principal component pursuit (SPCP) [11] was proposed to solve the following optimization problem as an

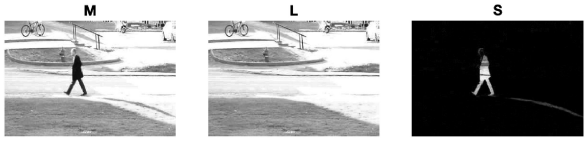


FIGURE 2. Sample result after applying the PCP algorithm reshaped into images. (left to right) original frame, recovered low-rank background, and foreground object are shown.

stable version of the PCP:

$$\underset{L,S}{\text{minimize}} \ \|L\|_* + \lambda\|S\|_1 \text{ subject to } \|M - L - S\|_F \leq \delta. \quad (3)$$

This formulation assumes that $M = L + S + E$ where in this case E represents the i.i.d. noise present on each entry of the matrix. In order to address the presence of quantization error, a similar inequality constraint approach is proposed in [12]. Furthermore, tensor based decomposition methods have been studied to improve the stability of the algorithms when noise is present [13], [14].

Another RPCA based method is DEtecting Contiguous Outliers in the LOw-rank Representation (DECOLOR) [15]. In this formulation, spatial continuity of the moving object is considered and graph-cut algorithm is used efficiently find the moving object mask [15]. In order to overcome the challenges arising from larger datasets, superpixel-based method that better regularizes the sparse term are utilized [16]. Additionally, the idea of total variation regularization of the sparse component is investigated in total variation regularized RPCA (TVRPCA) algorithm [17]. In addition, other regularization techniques regarding the completeness of the moving object especially in cases of a moving background have been explored in the literature [18]–[20]. Furthermore, in cases where the background illumination changes, decomposition into more than two components has shown promising results [14], [21], [22]. Other extensions such as considering camera motion with RPCA model have also been studied in the literature [23]. Further discussion and survey of other improvements are provided in [9], [10], [24]–[26].

B. CONTRIBUTION OF THIS STUDY

The RPCA model as presented in (2) assumes that the foreground object is *added* to the static background. Although this assumption might be true for many other applications of low-rank and sparse decomposition, additive perturbation is not a realistic assumption in video foreground-background decomposition. One immediate consequence of the additive model is that the recovered foreground object would not have the correct color, as clearly demonstrated in Fig. 2.

In this study, we forgo the additive assumption and instead propose a formulation of an overlaying model, which acknowledges that the foreground object is overlaid on top of the background and is occluding it (rather than simply being added). In our model each pixel either belongs to the foreground or to the background. We use a binary mask W to encode the foreground object location, and require the mask to

be sparse and the background to be low rank. Mathematically, our model can be described by $M = (1 - W) \circ L + W \circ S$ where $\circ$ denotes the elementwise matrix multiplication. In order to find the components introduced by the mask decomposition model, we formulate an optimization problem.

Note that imposing constraints on the support of the sparse component (especially connectivity of the support as a region) has been previously studied [15], [18], [19]. As we show in Section II, we forgo solving for the sparse component and directly solve for the mask of the moving object. Assisting and restricting the search for the moving object via saliency map [27] or mask obtained via optical flow [28] have been also studied in the literature. Additionally, finding a binary mask by utilizing a first-order Markov random field and refining the background in cases where dynamic background changes are present has shown effective [29]. Rather than precomputing and utilizing a mask, we will try to directly find the mask of the moving object.

Note that the idea of an overlaying model or masked decomposition is first introduced in [30] in the context of separating foreground text or graphics from the background in images. There, the background is assumed to belong to a subspace specified by a known set of bases. The problem is then to find the sparse mask and the subspace coefficients for the background. For separating moving objects from stationary background in a video, the low rank assumption on the background is more appropriate. However, simultaneously imposing the low rank prior on the background, sparsity prior on the mask, and the overlaying model constraint leads to a challenging optimization problem. Recently, Jin *et al.* [31] proposed a masked model where both ℓ_1-norm of the mask and the sparse component are considered in the optimization formulation. Additionally, nuclear norm has been utilized as a surrogate for the rank. The formulation as presented in [31] involves combinatorial optimization and is not robust to dynamic changes or noise in the background.

In our formulation, we will use the ℓ_0-pseudo-norm instead of the ℓ_1-norm of the mask as the regularizer to get better sparsity properties and relax the problem in hand from an integer program. Additionally, rank$(\cdot)$ and nuclear-norm will be considered and compared when imposing the low-rank prior on the background component. Furthermore, robustness to dynamic changes in the background and noise has been considered in our proposed formulation. The formulated optimization problem is nonconvex and not differentiable. An adapted version of the Douglas–Rachford splitting algorithm for a class of nonconvex problems [32] is utilized to solve the formulated problem. All required sub-problems are studied and solved by efficient algorithms. Finally, the proposed method and algorithmic solution are tested on different real-world video sequences and promising performance is shown.

C. ORGANIZATION OF THE PAPER

The rest of this paper is organized as follows. In Section II, we will present the overlaying model and discuss how the problem is formulated as a nonconvex optimization problem. In Section II-B, we will review the Douglas-Rachford method for a class of nonconvex problems and we will utilize this approach in Section II-C. In Section III, we present results for different video sequences and discuss our findings. Finally, in Section IV we concludes the paper. Additionally, some proofs and numerical experiments that help us better understand the sub-problems required for our method are provided in the appendix section.

D. NOTATION

For a matrix $X \in \mathbb{R}^{m\times n}$, $\|X\|_*$ denotes the nuclear norm of X (sum of singular values), and $\|X\|_F$ denotes the the Frobenius norm of X. Additionally, $\|X\|_1 = \sum_{i,j} |X_{ij}|$ denotes the ℓ_1-norm of vectorized form of X. $\|X\|_0 = \#\{X_{ij} \neq 0\}$ denotes the ℓ_0-pseudo-norm (number of non-zero elements). We further introduce the following indicator functions for a set $\mathcal{A}$ and any given scalar x:

$$\mathbf{1}_{\mathcal{A}}(x) := \begin{cases} 1 & \text{if } x \in \mathcal{A}, \\ 0 & \text{if } x \notin \mathcal{A} \end{cases} \text{ and } \iota_{\mathcal{A}}(x) := \begin{cases} 0 & \text{if } x \in \mathcal{A}, \\ \infty & \text{if } x \notin \mathcal{A} \end{cases}.$$

For matrix X, we denote $\iota_{\mathcal{A}}(X) = \sum_{i,j} \iota_{\mathcal{A}}(X_{ij})$.

II. OVERLAYING MODEL, PROBLEM FORMULATION, AND SOLUTION

A. MASKED ROBUST PRINCIPAL COMPONENT ANALYSIS

Given a video sequence, let matrix $M \in \mathbb{R}^{mn\times k}$ denote the matrix obtained by vectorizing and concatenating the frames as columns of M. Additionally, let $W \in \{0, 1\}^{mn\times k}$ be a binary mask encoding the foreground (foreground support). As briefly discussed in Section I-B, a plausible decomposition (the overlaying model) is $M = (1 - W) \circ L + W \circ S$ where, $L \in \mathbb{R}^{mn\times k}$ has low-rank and $S \in \mathbb{R}^{mn\times k}$ encodes the moving object. Note that given a foreground mask W one can obtain the moving object as $S = M \circ W$. As a result, the variable S can be omitted from the model to remove redundancy and the sparsity prior could be directly applied to the mask.

Following our discussions, the goal is to find matrices $L \in \mathbb{R}^{mn\times k}$ (with low-rank structure) and $W \in \{0, 1\}^{mn\times k}$ (with sparse structure) such that the overlaying model is satisfied. A plausible formulation of such problem can be written as

$$\begin{aligned} \underset{L,W}{\text{minimize}} \quad & \Phi(L) + \lambda\Psi(W) \\ \text{subject to } & (1 - W) \circ (M - L) = 0,\ W \in \{0, 1\}^{mn\times k} \end{aligned} \tag{4}$$

where $\Phi(\cdot)$ and $\Psi(\cdot)$ encode our prior knowledge of L (e.g. low-rank assumption) and W (e.g. sparsity), respectively. Additionally, λ is a regularization hyper-parameter. The equality constraint in (4) is the key to the overlaying model. To reiterate, for pixels in which the mask is inactive (indicating background), the constraint is only satisfied when L matches the value of M in that entry. On the other hand, for pixels where the mask is active (indicating foreground) the constraint is satisfied and value of L can defer from M.

The problem as formulated in (4) is not tractable. The binary nature of the variable W requires solving a combinatorial

problem. To make the problem more tractable, the variable W can be relaxed to take values inside the interval of zero and one namely, $W \in [0,1]^{mn\times k}$. Consequently, we can write the relaxed problem as

$$\begin{aligned} &\underset{L,W}{\text{minimize}} \quad \Phi(L)+\lambda\Psi(W) \\ &\text{subject to } (1-W)\circ(M-L)=0,\ W\in[0,1]^{mn\times k}. \end{aligned} \tag{5}$$

The formulation in (5) is prone to similar problem as the original RPCA (2) in presence of additive Gaussian noise or dynamic background (such as water waves, tree leaves). In this paper, we use the term "noise" to refer to both types of small temporal variations of the background. Consider a case where such noises are present as part of the background. In the original RPCA formulation such changes would be modeled as dynamic perturbations to the static background and will be separated to the sparse component [11]. Similarly, the formulation in (5) is susceptible to such problem. Note that for small values of λ there is less weight on the sparsity of the W and many isolated pixels in the background (caused by noise) would be considered as in the foreground. On the other hand, for large values of λ, such perturbations and possibly parts of the moving object would be decomposed incorrectly into the low-rank term L. As a result, in the presence of background noise, formulation in (5) does not achieve desired results.

In order to make the solution robust to additive noise (or dynamics in the background), we model the noise as an extra error term. In other words, instead of equating the constraint to zero, we can consider $(1-W)\circ(M-L)=E$ where E is the error term associated with the noise. Now we can penalize the effect of the noise. Since we have the i.i.d Gaussian prior assumption about the noise, the summation of squared error terms ($\|E\|_F^2$) can be used as the penalty function.

In our formulation, we require the recovered mask to be sparse and we use a sparsity inducing function as the regularization term on W. Here, we choose the ℓ_0-pseudo-norm which is known to give sparse solutions [33]. In order to impose the low-rank structure of L, we can either consider rank$(\cdot)$ or nuclear-norm ($\|L\|_*$) [3], [34]. Consequently, we consider the following two formulations in Lagrangian form

$$\begin{aligned} &\underset{L,W}{\text{minimize}}\ \|L\|_*+\lambda\|W\|_0+\iota_{[0,1]}(W) \\ &\qquad\qquad +\frac{\rho}{2}\|(W-1)\circ(L-M)\|_F^2 \end{aligned} \tag{6}$$

(**MRPCA1**)

and

$$\begin{aligned} &\underset{L,W}{\text{minimize}}\ \text{rank}(L)+\lambda\|W\|_0+\iota_{[0,1]}(W) \\ &\qquad\qquad +\frac{\rho}{2}\|(W-1)\circ(L-M)\|_F^2 \end{aligned} \tag{7}$$

(**MRPCA2**)

where λ is a regularization parameter and ρ is a penalty parameter. The indicator function $\iota_{[0,1]}(W)$ ensures $W\in[0,1]^{mn\times k}$. These problems are nonconvex and parts of the objective functions are not differentiable so the algorithm to solve them is not immediately apparent. Recently, Li *et al.* adapted the Douglas–Rachford (DR) splitting for a class of nonconvex optimization problems [32]. In the following sections we will first review the DR splitting method and then derive an algorithmic solution for our problems (6) and (7).

Algorithm 1: Douglas-Rachford Splitting Method.

Initialize: x^0, and **set:** $\gamma>0$;
while *not converged* **do**
  $y^{t+1}=\arg\min_y\ f(y)+\frac{1}{2\gamma}\|y-x^t\|^2$;
  $z^{t+1}=\arg\min_z\ g(z)+\frac{1}{2\gamma}\|2\,y^{t+1}-x^t-z\|^2$;
  $x^{t+1}=x^t+(z^{t+1}-y^{t+1})$;
Output: $y^*=z^*$

B. REVIEW OF DOUGLAS–RACHFORD SPLITTING FOR NONCONVEX OPTIMIZATION

In this section, we briefly review the Douglas-Rachford (DR) splitting method that was recently adapted for a class of nonconvex problems by Li *et al.* [32]. Further details can be found in [32]. Consider the optimization problem

$$\underset{u}{\text{minimize}} \quad f(u)+g(u)$$

where f is differentiable and is assumed to have Lipschitz continuous gradient and g is a proper closed function. Additionally, assume that for any given parameter $\gamma>0$, the proximal mapping of γg, is well defined and easy to compute. In other words, the minimizer of the problem

$$\text{prox}_{\gamma g}(x)=\arg\min_z\ \gamma\, g(z)+\frac{1}{2}\|z-x\|^2$$

exists and is easy to compute for different values of γ and x.

The adaptation of the Douglas–Rachford algorithm for this class of problems is presented in Algorithm 1. Motivated by the notion of DR envelope as in [35], the DR merit function for the nonconvex case [32] can be defined as

$$D_\gamma(y,z,x)=f(y)+g(z)+\frac{1}{2\gamma}\left(\|x-y\|^2-\|x-z\|^2\right) \tag{8}$$

where $\gamma>0$. Note that γ is the step-size in algorithm 1. Theorem 1 in [32] shows that for sufficiently small $\gamma>0$ and variables (y^t,z^t,x^t) generated by algorithm 1, the sequence $\{D_\gamma(y^t,z^t,x^t)\}_{t\geq1}$ is nonincreasing.

C. NONCONVEX DR SPLITTING APPLIED TO MRPCA

In this section, we adopt the DR splitting method, solve the required sub-problems, and provide the algorithmic solution. Considering the general form our formulation as

$$\text{minimize} \quad \underbrace{\Phi(L)+\lambda\Psi(W)}_{g(L,W)\text{–non differentiable}} \ +\ \underbrace{\frac{\rho}{2}\|(W-1)\circ(L-M)\|_F^2}_{f(L,W)\text{–differentiable}},$$

we can use the DR splitting algorithm by defining

$$f(L,W):=\frac{\rho}{2}\|(W-1)\circ(L-M)\|_F^2\,, \tag{9}$$

and in the case of solving MRPCA1 (6)

$$g(L,W) := \|L\|_* + \lambda\|W\|_0 + \iota_{[0,1]}(W), \tag{10}$$

and in the case of solving MRPCA2 (7)

$$g(L,W) := \operatorname{rank}(L) + \lambda\|W\|_0 + \iota_{[0,1]}(W). \tag{11}$$

Note that, our iterates in this case are the ordered pair $Y = (L_y, W_y)$ and $Z = (L_z, W_z)$. Also, $X = (X_l, X_w)$ are dual varibles of the DR algorithm. Let the superscript t denote the variable in iteration t, then the main steps of the DR algorithm are as follows

$$\begin{aligned}(L_y^{t+1}, W_y^{t+1}) &= \arg\min_{L,W}\ f(L,W) \\ &\quad + \frac{1}{2\gamma}\left\|L - X_l^t\right\|_F^2 + \frac{1}{2\gamma}\left\|W - X_w^t\right\|_F^2 \\ (L_z^{t+1}, W_z^{t+1}) &= \arg\min_{L,W}\ g(L,W) \\ &\quad + \frac{1}{2\gamma}\left[\left\|L - 2L_y^{t+1} + X_l^t\right\|_F^2 + \left\|W - 2W_y^{t+1} + X_w^t\right\|_F^2\right] \\ \left(X_l^{t+1}, X_w^{t+1}\right) &= \left(X_l^t, X_w^t\right) + \left[\left(L_z^{t+1}, W_z^{t+1}\right) - \left(L_y^{t+1}, W_y^{t+1}\right)\right].\end{aligned}$$

Here, we investigate each of the sub-problems and introduce efficient algorithm for solving them. The update for (L_y, W_y) involves solving the minimization problem

$$\begin{aligned}&\arg\min_{L,W}\ \frac{\rho}{2}\|(W-1)\circ(L-M)\|_F^2 \\ &\quad + \frac{1}{2\gamma}\left\|L - X_l^t\right\|_F^2 + \frac{1}{2\gamma}\left\|W - X_w^t\right\|_F^2.\end{aligned} \tag{12}$$

This problem is separable for entries of the matrices L and W so we can separately solve for the pairs (L_{ij}, W_{ij}) for all entries i, j. The reduced two variable problem can be written as

$$\begin{aligned}&\arg\min_{L_{ij},W_{ij}}\ \frac{\rho}{2}\left(\left(W_{ij}-1\right)\left(L_{ij}-M_{ij}\right)\right)^2 \\ &\quad + \frac{1}{2\gamma}\left(L_{ij} - X_{l_{ij}}^t\right)^2 + \frac{1}{2\gamma}\left(W_{ij} - X_{w_{ij}}^t\right)^2.\end{aligned} \tag{13}$$

A simple approach for solving this problem is to use the nonlinear Gauss–Seidel method [36] to find the solution. In this method, we fix one variable and solve for the other until convergence. The update rules are formulated by equating the derivative with respect to each variable to zero.

$$\begin{aligned}L_{ij}^+ &= \frac{X_{l_{ij}}^t + \gamma\rho\left(W_{ij}-1\right)^2 M_{ij}}{\gamma\rho\left(W_{ij}-1\right)^2 + 1}, \\ W_{ij}^+ &= \frac{X_{w_{ij}}^t + \gamma\rho\left(L_{ij}^+ - M_{ij}\right)^2}{\gamma\rho\left(L_{ij}^+ - M_{ij}\right)^2 + 1}.\end{aligned} \tag{14}$$

Note that these updates can be implemented as element-wise operations on matrices. Additionally, each update reduces the objective value in (13) and every limit point of the sequence is a critical point of the objective [36, Corollary 2]. Steps of the algorithm are summarized in Algorithm 2. Comparison of this method with some Newton based methods with numerical results for example cases of the function are provided in the appendix section. As our numerical comparisons on some small problems show, this easy method has good performance, is easy to implement for large-scale matrices and does not add extra parameters (like step-size, etc.) to the problem. Note that the resulting W_y found by Gauss-Seidel iteration does not necessarily satisfy the constraint $W_y \in [0,1]^{mn\times k}$. A projection onto the set can be utilized to satisfy the constraint.

Algorithm 2: Element-Wise Nonlinear Gauss-Seidel Iteration Applied to (12).

Input: M, X_l^t, X_w^t, and initial points L^0, and W^0;
Initialize: $L \leftarrow L^0$, and $W \leftarrow W^0$;
while *not converged* **do**
  //get indices which have not converged
  $\mathcal{A} =$ $\{(i,j) \mid$ associated problem for entry i,j has not converged$\}$;
  //for the indices in $\mathcal{A}$ update L and W using (14)
  $L_{\mathcal{A}} \leftarrow \frac{X_{l_{\mathcal{A}}}^t + \gamma\rho(W_{\mathcal{A}}-1)^2 M_{\mathcal{A}}}{\gamma\rho(W_{\mathcal{A}}-1)^2+1}$ //element-wise ops
  $W_{\mathcal{A}} \leftarrow \frac{X_{w_{\mathcal{A}}}^t + \gamma\rho\left(L_{\mathcal{A}}^+ - M_{\mathcal{A}}\right)^2}{\gamma\rho\left(L_{\mathcal{A}}^+ - M_{\mathcal{A}}\right)^2+1}$ //element-wise ops
Output: L, and W

The update rule for pair (L_z, W_z) involves solving the optimization problem

$$\begin{aligned}&\arg\min_{L,W}\ \|L\|_* + \lambda\|W\|_0 + \iota_{[0,1]}(W) \\ &\quad + \frac{1}{2\gamma}\left[\left\|L - 2L_y^{t+1} + X_l^t\right\|_F^2 + \left\|W - 2W_y^{t+1} + X_w^t\right\|_F^2\right]\end{aligned} \tag{15}$$

in the case of (6) and

$$\begin{aligned}&\arg\min_{L,W}\ \operatorname{rank}(L) + \lambda\|W\|_0 + \iota_{[0,1]}(W) \\ &\quad + \frac{1}{2\gamma}\left[\left\|L - 2L_y^{t+1} + X_l^t\right\|_F^2 + \left\|W - 2W_y^{t+1} + X_w^t\right\|_F^2\right]\end{aligned} \tag{16}$$

in the case of (7). These problems can be solved separately for variables L and W and close form solution in each case exists as a proximal operator. We first introduce these operators for nuclear norm, rank, and ℓ_0-pseudo-norm. Then, we can write the closed form solution for (15) and (16).

Remark 1: [37, Theorem 2.1] Let $\mathcal{S}_\tau : \mathbb{R} \to \mathbb{R}$ be the soft-thresholding operator $\mathcal{S}_\tau(x) = \operatorname{sgn}(x)\max\{|x| - \tau, 0\}$ and similarly for matrices applied element-wise. Let $\mathcal{D}_\tau(X)$ denote the singular value thresholding operator of matrix X as $\mathcal{D}_\tau(X) = U\mathcal{S}_\tau(\Sigma)V^*$, where $X = U\Sigma V^*$ is the singular value decomposition. Then,

$$\mathcal{D}_\tau(Y) = \arg\min_X \frac{1}{2}\|X - Y\|_F^2 + \tau\|X\|_*.$$

Remark 2: [33] Let $\mathcal{H}_{\sqrt{2\tau}}(x) = \mathbf{1}_{|x|\geq\sqrt{2\tau}}x$ be the hard-thresholding operator and similarly extend for matrices by applying element-wise. Then,

$$\mathcal{H}_{\sqrt{2\tau}}(Y) = \arg\min_X \frac{1}{2}\|X-Y\|_F^2 + \tau\|X\|_0.$$

Remark 3: [38]–[40] Let $\mathcal{F}_\tau(X) = U\mathcal{H}_\tau(\Sigma)V^*$ be the hard-thresholding on the singular values where $X = U\Sigma V^*$ is the singular value decomposition, then

$$\mathcal{F}_{\sqrt{2\tau}}(Y) = \arg\min_X \operatorname{rank}(X) + \frac{1}{2\tau}\|X-Y\|_F^2.$$

Remark 4: Let $\mathcal{H}_\tau^{[0,1]}(y)$ be defined as

$$\mathcal{H}_\tau^{[0,1]}(y) = \begin{cases} y & \text{if} \quad 0\leq y\leq 1 \quad \text{and} \quad y\geq\sqrt{2\tau} \\ 0 & \text{if} \quad 0\leq y\leq 1 \quad \text{and} \quad y\leq\sqrt{2\tau} \\ 1 & \text{if} \quad y\geq 1 \quad \text{and} \quad y\geq\frac{1}{2}+\tau \\ 0 & \text{if} \quad y\geq 1 \quad \text{and} \quad y\leq\frac{1}{2}+\tau \\ 0 & \text{if} \quad y\leq 0 \end{cases}$$

and extended similarly for matrices by applying element-wise. Then,

$$\mathcal{H}_\tau^{[0,1]}(Y) = \arg\min_X \frac{1}{2}\|X-Y\|_F^2 + \iota_{[0,1]}(X) + \tau\|X\|_0.$$

Using these results, we can write the close form for the update of (L_z, W_z) when solving for (6) as

$$\begin{aligned} L_z^{t+1} &= \arg\min_L \|L\|_* + \frac{1}{2\gamma}\left\|L - 2L_y^{t+1} + X_l^t\right\|_F^2 \\ &= \mathcal{D}_\gamma\left(2L_y^{t+1} - X_l^t\right), \text{ and} \end{aligned} \tag{17}$$

$$\begin{aligned} W_z^{t+1} &= \arg\min_W \lambda\|W\|_0 + \iota_{[0,1]}(W) \\ &+ \frac{1}{2\gamma}\left\|W - 2W_y^{t+1} + X_w^t\right\|_F^2 = \mathcal{H}_{2\gamma\lambda}^{[0,1]}\left(2W_y^{t+1} - X_w^t\right). \end{aligned} \tag{18}$$

When solving for (7) the update for W_z is similar and the update for L_z can be written as

$$\begin{aligned} L_z^{t+1} &= \arg\min_L \operatorname{rank}(L) + \frac{1}{2\gamma}\left\|L - 2L_y^{t+1} + X_l^t\right\|_F^2 \\ &= \mathcal{F}_{\sqrt{2\gamma}}\left(2L_y^{t+1} - X_l^t\right). \end{aligned} \tag{19}$$

The parameter γ plays the role of the step-size in our algorithm. We empirically initialize γ to be one and as suggested in [32], decrease the value by a constant ratio at each iteration. Parameters λ and ρ are tunable hyper-parameters controlling the effect of each regularization term. We provide some heuristic values for these parameters in Section III. The DR splitting algorithm applied to the MRPCA problem (6) and (7) is summarized in Algorithm 3.

Algorithm 3: DR Splitting Method Applied to MRPCA (6) and (7).

Input: M, λ, and ρ **set:** γ and $\bar{\gamma}$;
Initialize: X_l, and X_w;
while *not converged* **do**
  //Update L_y, W_y using Algorithm 2 initialized with most recent values of L_y and W_y.
  $L_y, W_y \leftarrow \arg\min_{L,W} f(L,W) + \frac{1}{2\gamma}\left(\|L - X_l\|_F^2 + \|W - X_w\|_F^2\right)$;
  //Project W_y to the interval $[0,1]$.
  $W_y \leftarrow \mathcal{P}_{[0,1]}(W_y)$ //element-wise clipping
  //Update L_z using closed form updates
  if *Solving for MRPCA1* (6) **then**
    $L_z \leftarrow \mathcal{D}_\gamma(2L_y - X_l)$ //using (17)
  else if *Solving for MRPCA2* (7) **then**
    $L_z \leftarrow \mathcal{F}_{\sqrt{2\gamma}}(2L_y - X_l)$ //using (19)
  //Update W_z using closed form update
  $W_z \leftarrow \mathcal{H}_{2\gamma\lambda}^{[0,1]}\left(2W_y^{t+1} - X_w^t\right)$; //using (18)
  //Update X_l, X_w using linear equations
  $X_l \leftarrow X_l + (L_z - L_y)$;
  $X_w \leftarrow X_w + (W_z - W_y)$;
  //Decrease γ
  $\gamma \leftarrow \max\{\bar{\gamma}, 0.99\,\gamma\}$
Output: L_y^*, L_z^*, W_y^*, W_z^*

D. COMPUTATIONAL COMPLEXITY

Considering k frames of size $m \times n$ of a video, the data matrix M is of size $mn \times k$. The update for variables L_y and W_y is iterative but only involves element-wise operations. The updates for variables W_z has a simple closed form involving element-wise operation. The closed form update for L_z involves finding the singular value decomposition (SVD) of a matrix of size $mn \times k$ which is the most computationally demanding step of the algorithm. Assuming there are less frames in the video than the number of pixels in each frame ($mn > k$), the complexity of computing the SVD is $\mathcal{O}(mnk^2)$ [41]. Otherwise, in general, the computational complexity of the SVD is $\mathcal{O}(\min\{mnk^2, m^2n^2k\})$ [41].

Note that the computational complexity of most RPCA-based methods is also similar to or more than the proposed algorithm. This is due to the fact that all these methods also need to calculate the SVD of a similarly sized matrix at every iteration. Additionally, more advanced techniques for computing the SVD such as the algorithms proposed in [42]–[45] can be adopted in our proposed framework to reduce the computation cost of SVD.

III. EXPERIMENTAL RESULTS

In this section, we provide results of the proposed algorithm on real-world videos. The videos used are from the change detection (CDnet) dataset [8] and I2R dataset [46]. We will use both visual and statistical measures to evaluate the recovery results. To quantify the goodness of the recovered mask we use the F-measure. Considering the number of true positive

TABLE 1. Heuristic values for λ and ρ given a data matrix $M \in \mathbb{R}^{mn \times k}$

Video Type	Static		Dynamic	
Parameter	λ	ρ	λ	ρ
MRPCA1	$\frac{10}{\sqrt{\max\{mn,k\}}}$	$\frac{10}{mnk}$	$\frac{20}{\sqrt{\max\{mn,k\}}}$	$\frac{1}{mnk}$
MRPCA2	$\frac{10}{\max\{mn,k\}}$	$\frac{1}{mnk}$	$\frac{20}{\max\{mn,k\}}$	$\frac{0.1}{mnk}$

(tp), true negative (tn), false positive (fp), and false negative (fn), these statistical measures are defined as Recall $= \frac{\text{tp}}{\text{tp+fn}}$, Precision $= \frac{\text{tp}}{\text{tp+fp}}$, and

$$\text{F–measure} = 2\,\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

We provide results for both proposed formulation. Here, **MRPCA1** and **MRPCA2** refer to the formulations in (6) and (7), respectively. In order to show the effect of non-convex priors in these formulations we provide comparisons with a relaxation of the model formulated as

$$\begin{aligned} &\underset{L,W}{\text{minimize}}\ \|L\|_* + \lambda\|W\|_1 + \iota_{[0,1]}(W) \\ &\qquad + \frac{\rho}{2}\,\|(W-1)\circ(L-M)\|_F^2\,. \end{aligned} \tag{20}$$

(MRPCA3)

Note that nuclear norm and ℓ_1-norm are used in this formulation which are convex surrogates for rank$(\cdot)$ and ℓ_0-pseudo-norm, respectively. Note that MRPCA3 can be also solved via similar DR framework as in algorithm 3 or other techniques such as linearizing the augmented Lagrangian [47], [48]. In our experiments we use the DR framework for MRPCA3.

In our experiments, we will provide examples for both static and dynamic background situations. In the static case, the background is not changing and only foreground is moving and Gaussian noise might also be present. In the dynamic case, there are changes in the background which are not part of the moving object. For example, waves on water, moving leaves, and motion of water fountains. In the dynamic case, ideally we want to detect the moving object but not the motion related to the dynamics of background. We will provide results for both scenarios. The code for the proposed algorithms are available *here*[49].

For both cases of static and dynamic background, we provide visual and statistical comparisons between the proposed models and PCP [3], SPCP [11], and DeColor [15] (as implemented in [50]). The hyper-parameters of each model are tuned by a grid search. Note that theoretical value for parameter λ in the PCP formulation is discussed in [3], [51], [52]. The threshold for getting the mask for PCP and SPCP are calculated by grid search around the threshold given by the Otsu algorithm [53] to maximize the F-measure. DeColor provides the mask by utilizing the graph-cut algorithm [54] and the associated hyper-parameter is tuned for the best performing F-measure. Heuristic rules for setting hyper-parameters for MRPCA1 (6) and MRPCA2 (7) are given in Table I. We use a grid search around these heuristic values to find the best performing models. Additionally, by investigating the F-measure around these heuristic values (similar to the experiment in [14]) we observe that the results are stable around the proposed heuristics.

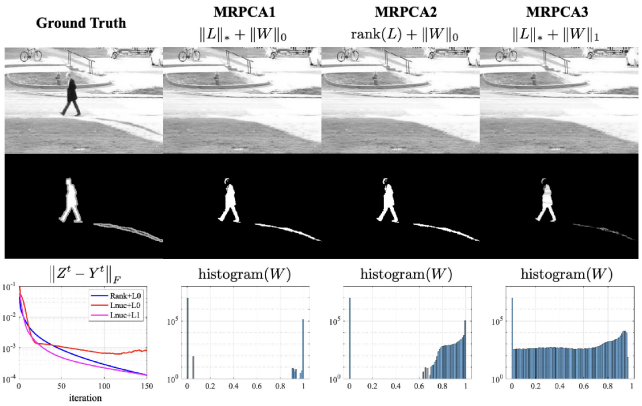


FIGURE 3. Comparison of the results for MRPCA1 (6), MRPCA2 (7), and MRPCA3 (20) for static background.

Note that the proposed formulations (6) and (7) are not convex and can benefit from a good initialization point. In our experiments, we use the median filter over all video frames to get a robust estimate of the background for initialization [55]. The mask is then initialized by subtracting this estimated background from each frame.

A. STATIC BACKGROUND

In this section, we consider the static background videos from the CDnet dataset[8] and provide results for the proposed model as well as some RPCA based models for comparison. A frame of the resulting background and mask for MRPCA1 (6), MRPCA2 (7), and MRPCA3 (20) are shown in Fig. 3. Additionally, last row of Fig. 3 shows the duality gap between the split variables Z^t and Y^t per iteration. The value of $\|Z^t - Y^t\|_F$ is an indicator for the convergence of the algorithm showing that all proposed methods converge within few tens of iterations.

Histogram of the recovered mask for all different methods is shown in Fig. 3. As we can observe, the non-convex priors lead to a separation of the $\{0, 1\}$ values. This effect is not observed when solving the relaxed problem as MRPCA3 (20). As a result, non-convex priors in the model lead to a solution closer to the desired binary mask without any added complexity. Note that in most of the RPCA-based methods the mask is obtained by thresholding the recovered sparse component. As shown in Fig. 3, the result of the proposed method (MRPCA1 or MRPCA2) is not sensitive to such threshold values and we set the value of 0.5 in our experiments whereas the optimal threshold needs to be found for most of the RPCA-based methods.

Visual and statistical comparison of the proposed models with RPCA-based methods are shown in Fig. 4 and Table II for static background videos. The proposed overlaying model performs better than other methods both in terms of visual

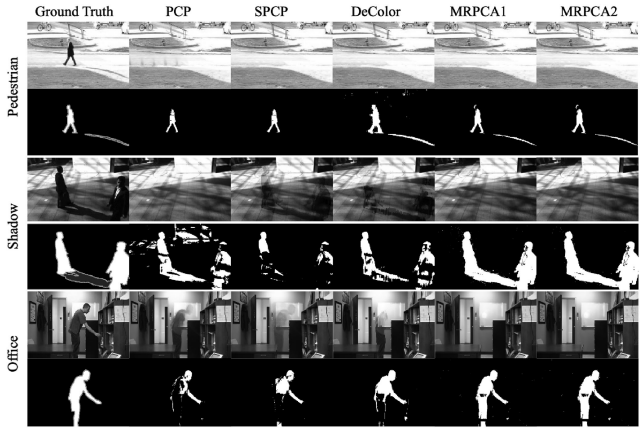


FIGURE 4. Sample results comparing PCP [3], SPCP [11], and DeColor [15] with MRPCA1 (6) and MRPCA2 (7). Recovered background and detected moving object mask are shown for example frames of several static background videos.

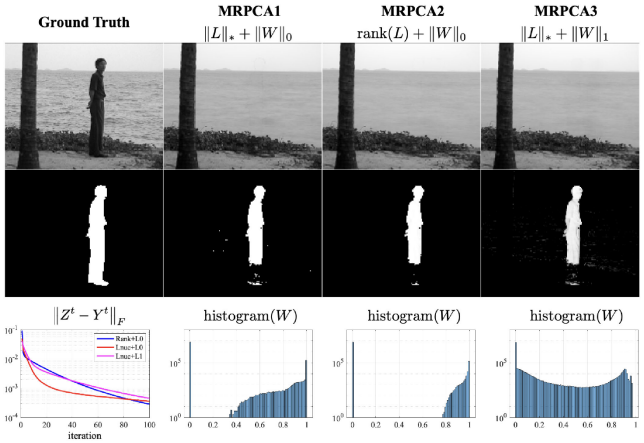


FIGURE 5. Comparison of the results for MRPCA1 (6), MRPCA2 (7), and MRPCA3 (20) for dynamic background.

TABLE 2. F-measure for static background video sequences

Video	PCP	SPCP	DeColor	MRPCA1	MRPCA2
Pedestrian	0.90	0.93	0.86	**0.96**	**0.96**
Shade	0.71	0.60	0.81	0.90	**0.92**
Office	0.62	0.66	0.71	0.83	**0.85**
highway	0.64	0.64	0.72	**0.79**	0.78
winter	0.42	0.53	0.69	**0.72**	**0.72**

TABLE 3. F-measure for dynamic background video sequences

Video	PCP	SPCP	DeColor	MRPCA1	MRPCA2
Water Surface	0.41	0.52	0.79	0.81	**0.85**
Trees	0.43	0.71	0.81	0.82	**0.84**
Fountain	0.43	0.62	0.69	0.71	**0.74**
Overpass	0.46	0.71	0.70	0.74	**0.79**
Fall	0.35	0.74	0.81	0.79	**0.83**
Canoe	0.32	0.45	0.64	**0.68**	**0.68**

and F-measure results. As the results for the "Office" sequence illustrate, the proposed model provides a more complete mask for the case where the person is standing relatively still over some video frames. The background recovered by the overlaying model is visually more appealing compared to the other models and better separation of the moving object is observed.

B. DYNAMIC BACKGROUND

In this section, we consider dynamic background videos from the CDnet [8] and I2R datasets [46]. We compare the proposed models with the RPCA-based approaches. As for the static background case, the comparison between MRPCA1 (6), MRPCA2 (7), and the relaxation as in MRPCA3 (20) are shown in Fig. 5. Similar convergence behavior and separation of the mask into $\{0, 1\}$ values is observed in the dynamic case as well.

Comparison of the proposed formulations with other methods is shown in Table III, Fig. 6, and Fig. 7. The proposed model outperforms the other models visually and numerically. In the case of the "WaterSurface" sequence, Fig. 6, the motion

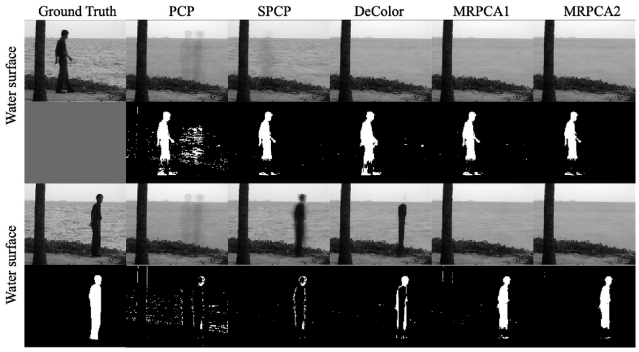


FIGURE 6. Sample results comparing PCP [3], SPCP [11], and DeColor [15] with MRPCA1 (6) and MRPCA2 (7). Recovered background and detected moving object mask are shown for two frames of the "WaterSurface" video with dynamic background (ground-truth is only provided for a selection of frames with a moving object in the dataset).

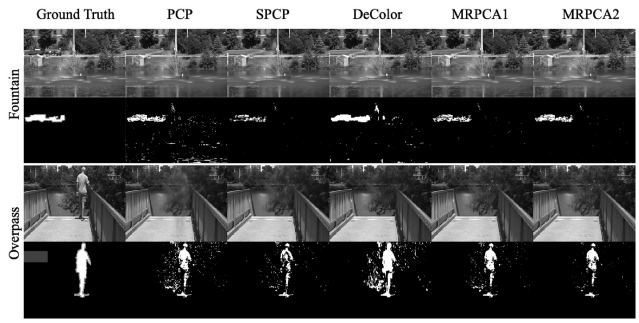


FIGURE 7. Sample results comparing PCP [3], SPCP [11], and DeColor [15] with MRPCA1 (6) and MRPCA2 (7). Recovered background and detected moving object mask are shown for a frame of the dynamic background videos.

on the water follows the Gaussian assumption and is removed from the mask effectively. Additionally, for the frames where the moving person stops, the proposed method provides a more consistent mask. Less artifact from the dynamics of the

background are present on the mask detected by MRPCA2 compared to MRPCA1.

Examples of the cases where the dynamics of the background does not follow the Gaussian assumption are shown in Fig. 7 (motion of top of the fountain or motion of the clusters of leaves). In such cases, all models suffer from artifacts in the recovered masks. As shown in Fig. 7, MRPCA2 leads to a more connected mask with less false positives compared to MRPCA1 and other compared methods and it is less affected by the background dynamics.

C. DISCUSSION

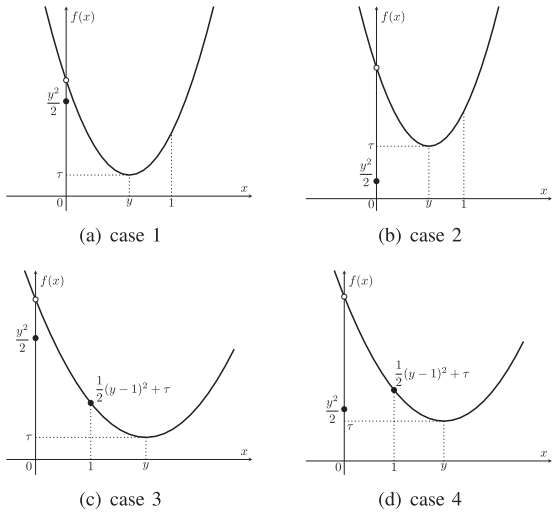
From all the experiment results presented, we observe that the MRPCA1 (6) and MRPCA2 (7) perform closely for the static background cases. For the videos with dynamic background, there is a slight advantage for the MRPCA2 (7) and the resulting masks are less affected by the dynamics of the background. Furthermore, for both static and dynamic cases, the recovered mask via the non-convex priors (e.g. MRPCA1 and MRPCA2) is closer to the desired binary solution without any added computational complexity compared to the relaxed formulation as MRPCA3 (20). In terms of computational complexity the proposed algorithms are similar to or better than the RPCA-based methods. Note that the recovered sparse component from the additive models need to be converted to the mask via thresholding and an optimal threshold needs to be found. Whereas in the overlaying model, we directly solve for the mask and the non-convex formulation leads to a close to binary solution. As a result, searching for the optimal threshold value is not required in this case.

In comparison to the results of the RPCA-based methods, the proposed overlaying model outperforms in both static and dynamic background scenarios. As visually shown, the proposed model provides a better recovery of the background and better separation of the moving object from the background. Additionally, non-convexities in the formulation help us to better mask the moving object while keeping the estimated background low-rank. This leads to a more complete mask compared to the other additive models in challenging cases. For instance, when the moving object stays still over a period of time we observe a more complete mask. Furthermore, in the presence of dynamic changes, MRPCA can recover the mask with less artifacts from the background. Whereas in RPCA-based methods, the false positives due to the dynamics of the background (even when total variation of the support of sparse component is regularized) are present over the recovered results.

IV. CONCLUSION

In this study, we considered the moving object detection problem using low-rank and sparse decomposition. We formulated the problem as an overlaying model where we simultaneously recover the low-rank background and the mask of the moving object. The proposed formulation is non-convex and we provided an algorithmic solution via an adaptation of the Douglas-Rachford method. The experimental results showed effective recovery in both dynamic and static background scenarios. We empirically showed that solving for the non-convex priors leads to a closer to binary mask without any added computational complexity compared to solving for the relaxed priors with convex surrogates. The proposed algorithm outperforms the models based on the additive assumption with similar or lower complexity.

In future work, we aim to investigate the effect of utilizing other regularization terms in the formulation to tackle more complex scenarios. For example, videos with camera motion (either shift of view or jitter) can be interesting for investigation. Additionally, proposed framework can be adopted for other applications where the overlaying model is valid such as waterfall artifacts in electron microscopy images.



(a) case 1 (b) case 2 (c) case 3 (d) case 4

FIGURE 8. Different cases possible for the objective function in (21) for $y > 0$.

APPENDIX

A. PROOF FOR DOMAIN CONSTRAINED HARD-THRESHOLDING

In this section we provide the proof for *Remark 4*. Note that the problem

$$\arg\min_X \frac{1}{2}\|X - Y\|_F^2 + \iota_{[0,1]}(X) + \tau\|X\|_0$$

is separable for each element of the matrix X as

$$x^* = \arg\min_x \frac{1}{2}(x-y)^2 + \iota_{[0,1]}(x) + \tau\|x\|_0. \qquad (21)$$

Note that $\|x\|_0 = 1$ for any $x \neq 0$ and is only zero for $x = 0$. Additionally, the indicator function $\iota_{[0,1]}(x)$ ensures that the minimizer is in the interval [0,1]. Since the only other term is a quadratic function, the solution to the problem can be found by considering all possible values of y and τ and their relative relation. These possible cases are shown in Fig. 8. The resulting minimizer can be written as the following function

based on different values of y and τ.

$$x^* = \begin{cases} y \text{ if } 0 \le y \le 1 \text{ and } y \ge \sqrt{2\tau} & \text{(case 1)} \\ 0 \text{ if } 0 \le y \le 1 \text{ and } y \le \sqrt{2\tau} & \text{(case 2)} \\ 1 \text{ if } \quad y \ge 1 \quad \text{and } y \ge \frac{1}{2} + \tau & \text{(case 3)} \\ 0 \text{ if } \quad y \ge 1 \quad \text{and } y \le \frac{1}{2} + \tau & \text{(case 4)} \\ 0 \text{ if } \quad y \le 0 & \text{(case 5)} \end{cases}$$

B. INVESTIGATING THE 2 DIMENSIONAL PROBLEM

In this section, we consider the problem in (12) and investigate the behavior of the function and different minimization methods for the two dimensional reduced problem. With some basic change of variables the general problem can be rewritten as

$$\text{minimize } \frac{\rho}{2}\|W \circ L\|_F^2 + \frac{1}{2\gamma}\|L - X_l\|_F^2 + \frac{1}{2\gamma}\|W - X_w\|_F^2.$$

Since this problem is separable for each entry of L and W, we can consider the two dimensional problem as

$$\text{minimize } \frac{\rho}{2}w^2l^2 + \frac{1}{2\gamma}(l - x_l)^2 + \frac{1}{2\gamma}(w - x_w)^2,$$

or equivalently,

$$\text{minimize } \frac{1}{2}w^2l^2 + \frac{c}{2}(l - x_l)^2 + \frac{c}{2}(w - x_w)^2 \tag{22}$$

where $c = \dfrac{1}{\gamma\rho}$ (note that we use lower-case letters to emphasize that the variables are scalar valued in this case). Considering the objective function

$$h(l, w) = \frac{1}{2}w^2l^2 + \frac{c}{2}\left[(l - x_l)^2 + (w - x_w)^2\right],$$

we have the following expressions for the gradient and the Hessain

$$\nabla h(l, w) = \begin{pmatrix} w^2 l + c\,(l - x_l) \\ l^2 w + c\,(w - x_w) \end{pmatrix} \tag{23}$$

$$\nabla^2 h(l, w) = \begin{pmatrix} w^2 + c & 2wl \\ 2wl & l^2 + c \end{pmatrix}. \tag{24}$$

Note that the determinant of the Hessian is

$$\det(\nabla^2 h) = \left(l^2 + c\right)\left(w^2 + c\right) - 4\,w^2\,l^2 \tag{25}$$

and we have

$$\nabla^2 h \succ 0 \Leftrightarrow c > -\frac{1}{2}\left(w^2 + l^2\right) + \sqrt{3\,w^2\,l^2 + \frac{1}{4}\left(w^2 + l^2\right)^2}.$$

Eigenvalues of the Hessian are

$$\mu_{1,2} = \frac{1}{2}\left(w^2 + l^2 + 2c\right) \pm \sqrt{3\,w^2\,l^2 + \frac{1}{4}\left(w^2 + l^2\right)^2}.$$

As a result, the Newton method with Hessian modification can be used to solve the problem. Since we want to have a method which can be extended to many problems solved using element-wise operations on matrices, we will use simple modification techniques which involve modifying the eigenvalues of the Hessain. These methods are simple enough and can be extended easily for our general case. In the next section, we introduce some examples of the methods that can be used and compare them with the presented nonlinear Gauss-Seidel method [56].

TABLE 4. Different parameters for the first experiment

parameter	x_l	x_c	γ	ρ	l_0	w_0
value	10	10	4	4	-4	5

1) NONLINEAR GAUSS-SEIDEL METHOD COMPARED TO NEWTON BASED METHODS

A very simple methods for modifying the Hessian is to replace it by a suitable positive definite matrix $H + \tau I$ if the Hessian H is not positive definite [57]. One simple approach is to set $\tau = \mu_{\min} + \delta$ for some small value of δ. As another approach, [57, Algorithm 3.3] we can check if the Hessian is positive definite and if not keep adding τI to the Hessian for some τ until the modified matrix becomes positive definite (this was proposed for the case where you do not know the minimum eigenvalue and cannot easily be extended to the matrix case but we will include this method in our comparisons as well). Another method is to replace the negative eigenvalues by their absolute value.

We can compare these methods for some prototype problems (different values of x_l, x_w, γ, and ρ). The methods compared here are

- **nonlinear Gauss-Seidel–** the updates of this method (as presented in the main text) are obtained by setting gradient w.r.t. each variable separately and updating the variables in an alternating way.
- **Diagonal loading–** in this method, we use the Newton step with backtracking line-search. In this case, if the Hessian at a point is not positive definite, we keep adding τI (starting with $\tau = 10^{-3}$) to the Hessian until the resulting matrix is positive definite[57, Algorithm 3.3].
- **Adding the minimum eigenvalue–** in this method, we use the Newton step with backtracking line-search. If the Hessian is not positive definite, we add $(|\mu_{\min} + \delta|I)$ to the Hessian (we choose $\delta = 0.01$).
- **Replace negative eigenvalues with their absolute value**in this method, we use the Newton step with backtracking line-search. If the Hessian is not positive definite, we replace the negative eigenvalue with its absolute value.

2) EXAMPLES

Here, we present some examples of the function minimized by different methods. For all the methods we choose to terminate the iteration if norm of the gradient is less than a tolerance (here we set to `1e-9`) or a maximum iteration number is reached (here we choose 100). For example, we set the values according to the Table IV, V, and VI and results for different methods are shown in Figures 9, 10, and 11. As we can

TABLE 5. Different parameters for the third experiment

parameter	x_l	x_c	γ	ρ	l_0	w_0
value	0.1	0.1	2	2	2	-4

TABLE 6. Different parameters for the fourth experiment

parameter	x_l	x_c	γ	ρ	l_0	w_0
value	1	1	2	2	2	-4

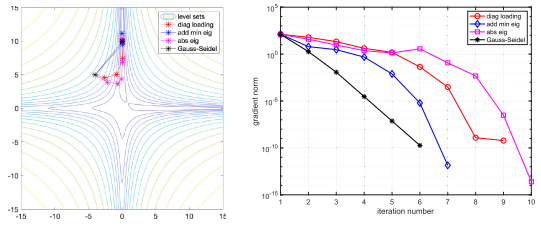


FIGURE 9. Results for Example 1: (left) iterates for different methods plotted over the level-sets of the function (right) norm of the gradient per iteration for different methods.

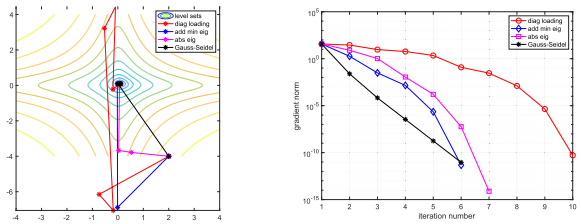


FIGURE 10. Results for Example 2: (left) iterates for different methods plotted over the level-sets of the function (right) norm of the gradient per iteration for different methods.

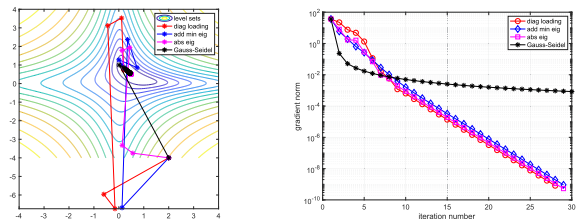


FIGURE 11. Results for Example 3: (left) iterates for different methods plotted over the level-sets of the function (right) norm of the gradient per iteration for different methods.

observe in Figs. 9 and 10, the Gauss-Seidel approach achieves the tolerance in less number of iterations. Here, we note that the Hessian of the function for first few iterations, is not positive definite and the step obtained from different simple modification methods are slowly moving toward the solution. Quadratic convergence can be observed near the optimum point for Newton based methods.

In Fig. 11, we observe that the Gauss-Seidel method moves to a close neighborhood of the solution in the first few iterations but then it has a very slow convergence. This slow convergence behavior is due to the fact that the function is very flat near optimum. The Newton based methods do not move close to the solution at first but show better overall convergence property. In this case, quadratic convergence is not observed for Newton based methods.

All in all, we can observe that each iteration of the Gauss-Seidel method is less computationally demanding than the Newton based methods and does not require a step-size parameter to be tuned. This method moves to a close neighborhood of the solution in a few iterations and obtains reasonable accuracy in few iterations. In some cases the method can be slow near the optimal point like our last example.

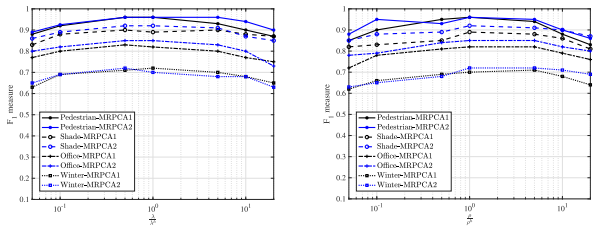


FIGURE 12. Stability wrt hyper-parameters for static background: (left) stability for λ when ρ is set to heuristic value ρ^h. (right) stability for ρ when λ is set to heuristic value λ^h.

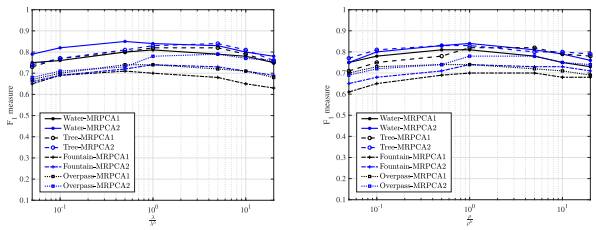


FIGURE 13. Stability wrt hyper-parameters for dynamic background: (left) stability for λ when ρ is set to heuristic value, ρ^h. (right) stability for ρ when λ is set to heuristic value, λ^h.

C. STABILITY WITH RESPECT TO HYPER-PARAMETERS

In this section, we show the stability of MRPCA1 (6), and MRPCA2 (7) w.r.t. the hyper-parameters λ and ρ. Following the setup in [14] and denoting the heuristic values in Table I as λ^h, and ρ^h, we show the F-measure for different video sequences around the heuristic values. As shown in figures 12 and 13, the F-measure is stable around the proposed heuristic values in both cases of static and dynamic background.

REFERENCES

[1] 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, Aug. 2000.

[2] F. De la Torre and M. J. Black, "Robust principal component analysis for computer vision," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, 2001, vol. 1, pp. 362–369.

[3] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM (JACM)*, New York, NY, USA: ACM, vol. 58, no. 3, pp. 1–37, 2011.

[4] 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.

[5] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations Comput. Math.*, vol. 9, no. 6, p. 717, 2009.

[6] S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[7] Z. Lin, M. Chen, and Y. Ma, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 612–620.

[8] Y. Wang, P. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "CDnet 2014: An expanded change detection benchmark dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 393–400.

[9] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust pca, robust subspace tracking, and robust subspace recovery," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 32–55, Jul. 2018.

[10] A. Sobral, T. Bouwmans, and E.-H. Zahzah, *Handbook of Robust Low Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*. Boca Raton, FL, USA: CRC Press, 2016.

[11] Z. Zhou, X. Li, J. Wright, E. Candès, and Y. Ma, "Stable principal component pursuit," in *Proc. IEEE Int. Symp. Inf. Theory.*, 2010, pp. 1518–1522.

[12] S. Becker, "Templates for convex cone problems with applications to sparse signal recovery," *Math. Program. Comput.*, vol. 3, no. 3, p. 165, 2011.

[13] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5249–5257.

[14] M. Shakeri and H. Zhang, "Moving object detection under discontinuous change in illumination using tensor low-rank and invariant sparse decomposition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 7221–7230.

[15] 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, Mar. 2013.

[16] S. Javed, S. H. Oh, A. Sobral, T. Bouwmans, and S. K. Jung, "Background subtraction via superpixel-based online matrix decomposition with structured foreground constraints," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2015, pp. 90–98.

[17] X. Cao, L. Yang, and X. Guo, "Total variation regularized rpca for irregularly moving object detection under dynamic background," *IEEE Trans. Cybern.*, vol. 46, no. 4, pp. 1014–1027, Apr. 2016.

[18] B. Xin, Y. Tian, Y. Wang, and W. Gao, "Background subtraction via generalized fused lasso foreground modeling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 4676–4684.

[19] M. Shakeri and H. Zhang, "COROLA: A sequential solution to moving object detection using low-rank approximation," *Comput. Vis. Image Understanding*, vol. 146, pp. 27–39, 2016.

[20] X. Liu, G. Zhao, J. Yao, and C. Qi, "Background subtraction based on low-rank and structured sparse decomposition," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2502–2514, Aug. 2015.

[21] M. Shakeri and H. Zhang, "Moving object detection in time-lapse or motion trigger image sequences using low-rank and invariant sparse decomposition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5123–5131.

[22] M. Li *et al.*, "Video rain streak removal by multiscale convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6644–6653.

[23] B. E. Moore, C. Gao, and R. R. Nadakuditi, "Panoramic robust pca for foreground–background separation on noisy, free-motion camera video," *IEEE Trans. Comput. Imag.*, vol. 5, no. 2, pp. 195–211, Jun. 2019.

[24] C. Guyon, T. Bouwmans, and E.-H. Zahzah, "Robust principal component analysis for background subtraction: Systematic evaluation and comparative analysis," *Principal Compon. Anal.*, vol. 10, pp. 223–238, 2012.

[25] B. Garcia-Garcia, T. Bouwmans, and A. J. R. Silva, "Background subtraction in real applications: Challenges, current models and future directions," *Comput. Sci. Rev.*, vol. 35, 2020, Art. no. 100204.

[26] 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," *Comput. Sci. Rev.*, vol. 23, pp. 1–71, 2017.

[27] A. Sobral, T. Bouwmans, and E.-H. ZahZah, "Double-constrained RPCA based on saliency maps for foreground detection in automated maritime surveillance," in *Proc. 12th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, 2015, pp. 1–6.

[28] J. Yang, X. Sun, X. Ye, and K. Li, "Background extraction from video sequences via motion-assisted matrix completion," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 2437–2441.

[29] H. Ahn and M. Kang, "Dynamic background subtraction with masked rpca," *Signal, Image Video Process*, [Online]. Available: https://link.springer.com/article/10.1007%2Fs11760-020-01766-5#citeas

[30] S. Minaee and Y. Wang, "An admm approach to masked signal decomposition using subspace representation," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3192–3204, Jul. 2019.

[31] M. Jin and Y. Chen, "Robust image recovery via mask matrix," in *Proc. Int. Conf. Intell. Sci. Big Data Eng.*, 2019, pp. 349–361.

[32] G. Li and T. K. Pong, "Douglas–Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems," *Math. Program.*, vol. 159, no. 1/2, pp. 371–401, 2016.

[33] F. Bach *et al.*, "Optimization with sparsity-inducing penalties," *Foundations Trends Mach. Learn.*, vol. 4, no. 1, pp. 1–106, 2012.

[34] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[35] P. Patrinos, L. Stella, and A. Bemporad, "Douglas-rachford splitting: Complexity estimates and accelerated variants," in *Proc. 53rd IEEE Conf. Decis. Control.*, 2014, pp. 4234–4239.

[36] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear gauss–seidel method under convex constraints," *Operations Res. Lett.*, vol. 26, no. 3, pp. 127–136, 2000.

[37] J. Cai, E. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[38] N. Parikh *et al.*, "Proximal algorithms," *Foundations Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.

[39] H. H. Bauschke *et al.*, *Convex Anal. and Monotone Operator Theory in Hilbert Spaces*. Berlin, Germany: Springer, 2011, vol. 408.

[40] M. Fazel, H. Hindi, and S. Boyd, "Rank minimization and applications in system theory," in *Proc. Amer. Control Conf.*, 2004, vol. 4, pp. 3273–3278.

[41] A. K. Cline and I. S. Dhillon, "Computation of the singular value decomposition," in *Handbook Linear Algebra*. CRC press, 2006, ch. 45, pp. 1–11.

[42] Z. Allen-Zhu and Y. Li, "LazySVD: Even faster svd decomposition yet without agonizing pain," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 974–982.

[43] C. Musco and C. Musco, "Randomized block krylov methods for stronger and faster approximate singular value decomposition," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1396–1404.

[44] X. Feng, W. Yu, and Y. Li, "Faster matrix completion using randomized svd," in *Proc. IEEE 30th Int. Conf. Tools Artif. Intell.*, 2018, pp. 608–615.

[45] J.-M. Low, I. K. Tan, and C.-Y. Ting, "Recent developments in recommender systems," in *Proc. Int. Conf. Multi-Disciplinary Trends Artif. Intell.*, 2019, pp. 38–51.

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

[47] J. Zhang, R. Chen, C. Deng, and S. Wang, "Fast linearized augmented lagrangian method for euler's elastica model," *Numer. Math.: Theory, Methods Appl.*, vol. 10, no. 1, pp. 98–115, 2017.

[48] A. Khalilian-Gourtani, S. Minaee, and Y. Wang, "Masked-RPCA: Sparse and low-rank decomposition under overlaying model and application to moving object detection," 2019, *arXiv:1909.08049*.

[49] A. Khalilian-Gourtani, *Masked-RPCA Repository*, 2020. [Online]. Available: https://github.com/amirhkhalilian/Masked-RPCA

[50] A. Sobral, T. Bouwmans, and E.-H. Zahzah, "Lrslibrary: Low-rank and sparse tools for background modeling and subtraction in videos," in *Robust Low-Rank and Sparse Matrix Decomposition: Application in Image and Video Processing*. Boca Raton, FL, USA: CRC Press, Taylor and Francis Group.

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

[52] 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.
[53] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *J. Electron. Imag.*, vol. 13, no. 1, pp. 146–166, 2004.
[54] S. Vicente, V. Kolmogorov, and C. Rother, "Graph cut based image segmentation with connectivity priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
[55] M.-H. Hung, J.-S. Pan, and C.-H. Hsieh, "Speed up temporal median filter for background subtraction," in *Proc. 1st Int. Conf. Pervasive Comput., Signal Process. Appl.*, 2010, pp. 297–300.
[56] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear gauss–seidel method under convex constraints," *Operations Res. Lett.*, vol. 26, no. 3, pp. 127–136, 2000.
[57] J. Nocedal and S. Wright, *Numerical Optimization*. Berlin, Germany: Springer Science & Business Media, 2006.

AMIRHOSSEIN KHALILIAN-GOURTANI (Student Member, IEEE) received the B.Sc. degree in electrical engineering and communications from the Isfahan University of Technology, the M.Sc. degree in electrical engineering from New York University (NYU) in 2018. He is working toward the Ph.D. degree with New York University, Tandon School of Engineering. His main research interests include signal processing, machine learning, convex and non-smooth optimization, medical signal processing, and numerical analysis. He was the recipient of the Myron M. Rosenthal Award for Best MSc Academic Achievement in Electrical and Computer Engineering from New York University in 2018.

SHERVIN MINAEE (Member, IEEE) received the Ph.D. degree in electrical engineering and computer science from NYU, in 2018. He is a Machine Learning Lead with Snapchat Computer Vision Team. He has authored or coauthored more than 40 papers and patents during his PhD. He has previously worked as a Research Scientist with Samsung Research, AT&T Labs, Huawei Labs, and as a Data Scientist with Expedia group. His research interest includes computer vision, image segmentation, biometrics recognition, and applied deep learning. He is a Reviewer for more than 20 computer vision related journals from IEEE, ACM, and Elsevier.

YAO WANG (Fellow, IEEE) is a Professor with the New York University Tandon School of Engineering (formerly Polytechnic University, Brooklyn, NY), with joint appointment in Departments of Electrical and Computer Engineering and Biomedical Engineering. She is also an Associate Dean with Faculty Affairs for NYU Tandon since June 2019. She is the leading author of a textbook titled Video Processing and Communications, and has authored or coauthored more than 250 papers in journals and conference proceedings. Her research interests include video coding and streaming, multimedia signal processing, computer vision, and medical imaging. She was the recipient of the New York City Mayor's Award for Excellence in Science and Technology in the Young Investigator Category in year 2000. She was elected Fellow of the IEEE in 2004 for contributions to video processing and communications. She received the IEEE Communications Society Leonard G. Abraham Prize Paper Award in the Field of Communications Systems in 2004, and the IEEE Communications Society Multimedia Communication Technical Committee Best Paper Award in 2011. She was a keynote speaker at the 2010 International Packet Video Workshop, INFOCOM Workshop on Contemporary Video in 2014, the 2018 Picture Coding Symposium, and the 2020 ACM Multimedia Systems Conference (MMSys20). She received the NYU Tandon Distinguished Teacher Award in 2016.