Proceedings of 2022 IEEE
International Conference on Mechatronics and Automation
August 7 - 10, Guilin China

Robust Dual-Graph Regularized Moving Object Detection

Jing Qin
Department of Mathematics
University of Kentucky
Lexington, KY 40506, USA
jing.qin@uky.edu

Ruilong Shen, Ruihan Zhu and Biyun Xie
Department of Electrical and Computer Engineering
University of Kentucky
Lexington, KY 40506, USA
{*Ruilong.Shen, Ruihan.Zhu, Biyun.Xie*}*@uky.edu*

2022 IEEE International Conference on Mechatronics and Automation (ICMA) | 978-1-6654-0853-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICMA54519.2022.9856248

***Abstract*— Moving object detection and its associated background-foreground separation have been widely used in a lot of applications, including computer vision, transportation and surveillance. Due to the presence of the static background, a video can be naturally decomposed into a low-rank background and a sparse foreground. Many regularization techniques, such as matrix nuclear norm, can therefore be imposed on the background. In the meanwhile, sparsity or smoothness based regularizations, such as total variation and ℓ_1, can be imposed on the foreground. Moreover, graph Laplacians are further used to capture the complicated geometry of background images. Recently, weighted regularization techniques including the weighted nuclear norm regularization have been proposed in the image processing community to promote adaptive sparsity while achieving efficient performance. In this paper, we propose a robust dual-graph regularized moving object detection model based on a new weighted nuclear norm regularization and spatiotemporal graph Laplacians, which is solved by the alternating direction method of multipliers (ADMM). Numerical experiments on realistic body movement data sets have demonstrated the effectiveness of this method in separating moving objects from background, and the great potential in robotic applications.**

***Index Terms*— Moving object detection, sparsity, graph Laplacian, weighted nuclear norm, alternating direction method of multipliers**

I. Introduction

The development of advanced robotic technologies has released traditional robots isolated by fences or other protective barriers to environments with human beings [1]. Such kinds of robots that are safe and intelligent enough to work alongside or directly interact with humans are called collaborative robots, including lightweight industrial robots, social robots, and service robots [2]. Human motion detection plays a significant role in the motion planning and control of collaborative robots to improve the safety and efficiency of human-robot interaction. On the one hand, the detected human motion will be used as the input information of various real-time motion planning algorithms to prevent the potential collision between a robot and a human subject and guarantee the safety of human-robot interaction [3]. On the other hand, the detected human motion can be further used for human motion analysis and prediction to enable robots to comprehend human intention and enhance the efficiency of human-robot interaction [4]. In this paper, we aim to develop an effective human motion detection algorithm with excellent accuracy and efficiency.

Detection of moving objects in a video with static background is usually done by separating foreground from background, and the moving objects are typically considered as the foreground. Background modeling is crucial in designing a moving object detection algorithm. Many subspace learning methods such as principal component analysis (PCA) have been developed to model background [5] by reducing the dimensionality and learning the intrinsic low-dimensional subspaces. In practice, a background matrix can be generated by concatenating the vectorized versions of background images of a video, which naturally possesses the low-rank structure. Thus sparsity of singular values is expected for a background matrix. In one of the most popular methods - robust PCA (RPCA) [6], nuclear-norm regularization is used to enforce the matrix low-rankness as a convex relaxation of the matrix rank. Numerous variants of RPCA have been proposed [7], [8] and a comprehensive review can be found [9]. Recently, adaptive regularization techniques have been developed to promote sparsity and achieve fast convergence of the regularized algorithms. For example, the weighted nuclear norm (WNN) regularization has shown effectiveness in various image and data processing applications [10], which can be considered as a natural extension of reweighted L1 [11] and a more general error function based regularization (ERF) [12].

In this paper, we use the ERF-weighted nuclear norm regularization (ERF-WNN) imposed on the matrix singular values to enforce the adaptive low-rankness. In addition, a video usually has complicated geometry and varying smoothness in either the spatial domain or the temporal domain. To preserve those geometrical structures in the background, we create a spatial graph and a temporal graph, which are then embedded in the graph regularizations of the background matrix. Generation of the spatial graph is implemented by comparing the patchwise similarity to exploit the nonlocal similarity. To reduce the computational cost, we only consider the k-nearest neighboring pixels in terms of similarity when calculating the pairwise similarity. On the other hand, the ℓ_1-regularization is imposed on the foreground due to its sparsity. Thus far, we propose a spatiotemporal dual-graph regularized moving object detection model, which is solved by the alternating direction method of multipliers (ADMM). After introducing a few auxiliary variables and splitting regularizers, we obtain a sequence of subproblems.

978-1-6654-0853-0/22/$31.00 ©2022 IEEE

One quadratic subproblem is solved by gradient descent, and the other subproblems all have closed-form solutions which can be implemented efficiently. Furthermore, we test our algorithm on the two real RGB videos containing a whole-body motion and an arm motion under a static background, respectively. Performance is compared with other related methods in terms of background recovery and foreground detection accuracy.

The rest of this paper is organized as follows. In Section II, we provide a brief introduction of moving object detection and low-rank based models. In Section III, we propose a novel spatiotemporal dual graph regularized moving object detection method based on the ERF-WNN regularization. Numerical experiments on two realistic videos with moving objects and the results are reported in Section IV. Finally, conclusions of this research and future work are presented in Section V.

II. Low-Rank Models

Throughout the paper, we use boldface lowercase letters to denote vectors and uppercase letters to denote matrices. For $p \geq 1$, the ℓ_p-norm of a vector $\mathbf{x} \in \mathbb{R}^n$ is given by $\|\mathbf{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. The entry-wise ℓ_1-norm of a matrix $X \in \mathbb{R}^{n\times m}$ is defined as $\|X\|_1 = \sum_{i,j} |x_{ij}|$ where x_{ij} is the (i,j)-th entry of X. The Frobenious norm of X, denoted by $\|X\|_F$, is defined as $\sqrt{\sum_{i,j} |x_{ij}|^2}$. The nuclear norm of X, denoted by $\|X\|_*$, is defined as the sum of all singular values of X. We use the symbol $\mathrm{diag}(\mathbf{x})$ to denote a diagonal matrix whose diagonal entries form the vector $\mathbf{x}$, and I_n as the n-by-n identity matrix.

Consider a video with a static background consisting of m frames of gray-scale images with size $n_1 \times n_2$. By reshaping each image as a vector, we convert a video to a matrix D of size $n \times m$ where $n = n_1 n_2$ is the number of total spatial pixels. Assume that D can be decomposed into the background component L and the foreground component S, where $L, S \in \mathbb{R}^{n\times m}$. Here we let S correspond to the moving object. That is, we have $D = L + S$ in the noise-free case. In order to retrieve L and S from D simultaneously, we apply regularization techniques on both variables. Since the background is static, the matrix L typically has low-rank structures. In the meanwhile, the object occupies a small portion of each frame and thereby S is sparse. Thus we consider the problem

$$\min_{L,S} \mathrm{rank}(L) + \lambda \|S\|_1 \quad \text{s.t.} \quad D = L + S.$$

Here $\lambda > 0$ is a regularization parameter and $\mathrm{rank}(L)$ equals the number of nonzero singular values of L. Since this problem is NP-hard, matrix rank is replaced by the nuclear norm which leads to the RPCA model [6]

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

In some RPCA variants [13], [7], the matrix max-norm based regularizer has been used to replace the nuclear norm

$$\min_{L,S} \|L\|_{\max} + \lambda \|S\|_1 \quad \text{s.t.} \quad D = L + S.$$

Here the max-norm of L is given by $\|L\|_{\max} = \min_{L=UV'} \|U\|_{2\to\infty} \|V\|_{2\to\infty}$ where V' is the transpose of V and $\|U\|_{2\to\infty} = \max_{\|\mathbf{x}\|_2=1} \|U\mathbf{x}\|_\infty$. See [14] for the connections between the matrix nuclear norm and the max-norm. They both are convex and can be used to describe the low-rankness of the background matrix. Recently, weighted nuclear norm minimization (WNNM) has been proposed and shown outstanding performance in a lot of image processing applications [10]. Specifically, weighted nuclear norm (WNN) is defined as

$$\|L\|_{W,*} := \sum_i w_i \sigma_i(L), \tag{1}$$

where $\sigma_i(L)$ is the i-th singular value of L in the decreasing order and $w_i \geq 0$ is the i-th weight. The selection of weights is related to adaptive sparsity regularizers such as iteratively reweighted L1 (IRL1) [11]. More recently, ERF generalizes IRL1 with improved sparsity and convergence speed [12]. Both can be naturally extended to the singular values in the WNN framework to promote the low-rankness. In this paper, we adopt a novel ERF-WNN as the regularizer that will be detailed in the next section.

III. Proposed Method

Moving object detection (MOD) is one fundamental task in robotic applications. The problem can be cast as the foreground and the background separation. In addition to the low-rankness assumption of the background matrix, we can use spatial and temporal graph regularizations to preserve the sophisticated geometry. To split multiple regularization terms in the proposed MOD model, we apply ADMM to derive an efficient algorithm.

A. Spatial and Temporal Graph Laplacians

In what follows, we will describe the generation of spatial and temporal graph Laplacians and their corresponding graph regularizers on the background.

For a reshaped video $D \in \mathbb{R}^{n\times m}$, rows and columns of D correspond to the spatial and the temporal samples, respectively. Consider a weighted temporal graph $G_t = (V_t, E_t, A_t)$ where $V_t = \{\mathbf{v}_i^t\}_{i=1}^m$ is a set of temporal samples, E_t is an edge set and $A_t \in \mathbb{R}^{m\times m}$ is the adjacency matrix which defines the weights. First, we generate an adjacency matrix A_t whose (i,j)-th entry is given by

$$(A_t)_{i,j} = \exp\left(-\frac{\|\mathbf{v}_i^t - \mathbf{v}_j^t\|_2^2}{h_t^2}\right), \quad i,j \in \{1,\ldots,m\}$$

where $h_t > 0$ is a temporal filtering parameter. Let W_t be the degree matrix of G_t where $(W_t)_{i,i} = \sum_{j=1}^m (A_t)_{i,j}$. Next we define a symmetrically normalized temporal graph Laplacian $\Phi_t \in \mathbb{R}^{m\times m}$ given by

$$\Phi_t = I_m - W_t^{-1/2} A_t W_t^{-1/2}.$$

Note that $W_t^{-1/2}$ is a diagonal matrix whose i-th diagonal entry is $(W_t)_{i,i}^{-1/2}$.

Likewise, we consider a weighted spatial graph $G_s = (V_s, E_s, A_s)$ where $V_s = \{\mathbf{v}_i^s\}_{i=1}^n$ is a set of spatial samples,

E_s is the edge set and $A_s \in \mathbb{R}^{n\times n}$ is a spatial adjacency matrix. Slightly different from the construction of A_t, we consider the patchwise similarity in the spatial domain for A_s. Specifically, the (i,j)-th entry of A_s is given by

$$(A_s)_{i,j} = \exp\left(-\frac{\|\mathcal{N}(\mathbf{v}_i^s) - \mathcal{N}(\mathbf{v}_j^s)\|_F^2}{h_s^2}\right),\ i,j \in \{1,\dots,n\}$$

where $\mathcal{N}(\mathbf{v}_i^s) \in \mathbb{R}^{p^2\times m}$ is a reshaped version of the video patch centered at the i-th pixel and $h_s > 0$ is the spatial filtering parameter. To reduce the computational cost, we consider the k-nearest neighbors in terms of location for calculating A_s. Specifically, we use the four-nearest neighboring spatial pixels to compute the patch-based similarity for generating the spatial adjacency matrix A_s. Likewise we use the four-nearest neighboring temporal pixels to compute A_t. Moreover, it is worth noting that Gaussian smoothing could be embedded to the calculation of patchwise similarity in the presence of noise. Now we define the symmetrically normalized graph Laplacian in the spatial domain as

$$\Phi_s = I_n - W_s^{-1/2} A_s W_s^{-1/2}.$$

Similar to W_t, W_s is the degree matrix corresponding to G_s which can be obtained using A_s. Furthermore, we save all graph Laplacians as sparse matrices to circumvent the out-of-memory issue.

B. Robust Dual-Graph Regularized Method

Let $D \in \mathbb{R}^{n\times m}$ be the reshaped video with n spatial pixels and m temporal frames. Assume that $\Phi_s \in \mathbb{R}^{n\times n}$ and $\Phi_t \in \mathbb{R}^{m\times m}$ are the respective spatial and temporal graph Laplacians, which are obtained from Section III-A. We propose a robust foreground-background separation model of the form

$$\min_{L,S\in\mathbb{R}^{n\times m}} \|D-L-S\|_1 + \lambda_1\|L\|_{W,*} + \lambda_2\|S\|_1 + \frac{\gamma_1}{2}\operatorname{tr}(L^T\Phi_s L) + \frac{\gamma_2}{2}\operatorname{tr}(L\Phi_t L^T).$$

Here we adopt the L_1-norm in the first data fidelity term to enforce the robustness of the method and suppress the outliers for recovering the low-rank component, and $\|\cdot\|_{W,*}$ is the WNN defined in (1) with weights generated by ERF. The last two graph regularization terms are used to enforce the spatiotemporal smoothness for the background. By introducing an auxiliary variables U and V, we rewrite the above problem

$$\min_{L,S,U,V} \|V\|_1 + \lambda_1\|U\|_{W,*} + \lambda_2\|S\|_1 + \frac{\gamma_1}{2}\operatorname{tr}(L^T\Phi_s L) + \frac{\gamma_2}{2}\operatorname{tr}(L\Phi_t L^T), \quad \text{s.t.} \quad U = L,\ D-L-S = V.$$

Define the augmented Lagrangian

$$\mathcal{L} = \|V\|_1 + \lambda_1\|U\|_{W,*} + \lambda_2\|S\|_1 + \frac{\gamma_1}{2}\operatorname{tr}(L^T\Phi_s L) + \frac{\gamma_2}{2}\operatorname{tr}(L\Phi_t L^T) + \frac{\rho_1}{2}\|U-L+\widetilde{U}\|_F^2 + \frac{\rho_2}{2}\|D-L-S+V+\widetilde{V}\|_F^2.$$

Based on the ADMM framework, we obtain the algorithm

$$\begin{cases} L \leftarrow \operatorname*{argmin}_L \frac{\gamma_1}{2}\operatorname{tr}(L^T\Phi_s L) + \frac{\gamma_2}{2}\operatorname{tr}(L\Phi_t L^T) \\ \quad + \frac{\rho_1}{2}\|U-L+\widetilde{U}\|_F^2 + \frac{\rho_2}{2}\|D-L-S+V+\widetilde{V}\|_F^2 \\ S \leftarrow \operatorname*{argmin}_S \lambda_2\|S\|_1 + \frac{1}{2}\|D-L-S\|_F^2 \\ U \leftarrow \operatorname*{argmin}_U \lambda_1\|U\|_{W,*} + \frac{\rho_1}{2}\|U-L+\widetilde{U}\|_F^2 \\ \quad = \operatorname*{argmin}_U \frac{\lambda_1}{\rho_1}\|U\|_{W,*} + \frac{1}{2}\|U-L+\widetilde{U}\|_F^2 \\ V \leftarrow \operatorname*{argmin}_V \|V\|_1 + \frac{\rho_2}{2}\|D-L-S+V+\widetilde{V}\|_F^2 \\ \widetilde{U} \leftarrow \widetilde{U} + (U-L) \\ \widetilde{V} \leftarrow \widetilde{V} + (D-L-S+V) \end{cases}$$

The first L-subproblem can be solved by gradient descent. Specifically, the gradient of the objective function is

$$\nabla f(L) = \gamma_1\Phi_s L + \gamma_2 L\Phi_t + \rho_1(L-U-\widetilde{U}) + \rho_2(L-D+S-V-\widetilde{V}).$$

Note that $\frac{d}{dX}\operatorname{tr}(X^TAX) = (A+A^T)X = 2AX$ if A is a symmetric matrix.Then at each step, we update L with fixed $S, U, \widetilde{U}, V, \widetilde{V}$ via

$$L \leftarrow L - dt\cdot\nabla f(L), \tag{2}$$

where $dt > 0$ is a step size. It can be empirically shown that only a few steps of gradient descent are sufficient. Next, the S-subproblem has the closed-form solution

$$S \leftarrow \operatorname{shrink}(D-L, \lambda_2). \tag{3}$$

Here the shrinkage operator is defined as $(\operatorname{shrink}(A,\mu))_{ij} = \operatorname{sign}(a_{ij})\cdot\max\{|a_{ij}|-\mu, 0\}$ where a_{ij} is the (i,j)-th entry of A. One can show that the U-subproblem has the closed-form solution via a weighted version of the singular value thresholding operator (SVT)

$$U \leftarrow A\widetilde{\Sigma}B, \quad \widetilde{\Sigma} = \operatorname{diag}(\operatorname{shrink}(\sigma(\widehat{L}), w_i\lambda_1/\rho_1)) \tag{4}$$

where $A\Sigma B$ is the singular value decomposition (SVD) form of the matrix $\widehat{L} := (L-\widetilde{U})$, $\sigma(\widehat{L})$ is the vector containing all the singular values of $\widehat{L}$ with $\sigma_i(\widehat{L})$ as its i-th component. Here the weights are constructed iteratively based on the singular values of the matrix L from the previous iteration based on ERF [12]:

$$w_i = \exp(-\sigma_i^2(\widehat{L})/\sigma^2). \tag{5}$$

Finally, the V-subproblem is similar to the S-subproblem with the closed-form solution and thereby V is updated via

$$V \leftarrow \operatorname{shrink}(L+S-D-\widehat{V}, \rho_2). \tag{6}$$

As one crucial preprocessing step, we remove motionless frames in the data set if two consecutive frames have small overall changes, i.e., the ℓ_1-norm of the difference vector of the two adjacent columns of D is below a threshold. The stopping criteria are based on the relative changes in

L and S, i.e., $\frac{\|L^{i+1}-L^i\|_F}{\|L^i\|_F} < tol$ and $\frac{\|S^{i+1}-S^i\|_F}{\|S^i\|_F} < tol$ where L^i and S^i are the obtained background and foreground matrices at the i-th iteration and tol is tolerance. Notice that the parameters - λ_2, ρ_2 - in the shrinkage operator can be adaptively updated. The entire algorithm is summarized in Algorithm 1, which can be extended to handle RGB data sets channelwise. In this work, we focus on gray scale videos by converting all RGB data to gray scale ones.

Algorithm 1 Robust Dual-Graph Regularized Moving Object Detection

Inputs: reshaped test video $D \in \mathbb{R}^{n\times m}$, graph filtering parameters $h_s, h_t > 0$, parameters $\lambda_1, \lambda_2, \gamma_1, \gamma_2, \rho_1, \rho_2 > 0$, maximum outer loops T_{out}, maximum inner loops T_{in}, tolerance tol
Outputs: background L and foreground S
Generate graph Laplacians Φ_t and Φ_s
Initialize L and S
for $i = 1, 2, \ldots, T_{out}$ **do**
  for $j = 1, 2, \ldots, T_{in}$ **do**
    Update L via (2)
  end for
  Update S via (3)
  Update U via (4) and singular values $\sigma_i(\widehat{L})$
  Update W via (5)
  Update V via (6)
  $\widetilde{U} \leftarrow \widetilde{U} + (U - L)$
  $\widetilde{V} \leftarrow \widetilde{V} + (D - L - S + V)$
  Exit the loop if the stopping criteria are met.
end for

IV. Numerical Experiments

In this section, we will test the proposed Algorithm 1 on two simulated moving object images. For comparison, we include three closely related algorithms based on the fast robust principal component analysis (RPCA) [6]: (1) Largangian optimization method for unconstrained RPCA (LAGO) (2) stable principal component pursuit (SPCP) [15] and (3) SPGL1 [16] for solving the problem $\min_{L,S} \max\{\|L\|_*, \lambda\|S\|_1\}$ subject to $\|D - L - S\|_F \le \varepsilon$. Their source codes can be found in fastRPCA `https://github.com/stephenbeckr/fastRPCA` [17]. There are two groups of metrics for comparing the performance, i.e., comparing the foreground and the background. First, the static background image is extracted from the low-rank component of the given video. We take the mean column of the low-rank matrix L and then reshape it as a matrix. We use the following metrics to evaluate the background recovery quality:

- relative error (RE): $\mathrm{RE}(\widehat{L}, L) = \frac{\|L-\widehat{L}\|_F}{\|L\|_F}$;
- peak signal-to-noise ratio (PSNR): $\mathrm{PSNR}(\widehat{L}, L) = 20\log(I_{\max}/\sqrt{\|\widehat{L} - L\|_F^2/(n_1 n_2)})$.

Here $\widehat{L}$ is the estimate of the ground truth $L \in \mathbb{R}^{n_1\times n_2}$, and $I_{\max}$ is the maximum image intensity set as 1. In our experiments, all of the videos to be processed are scaled to the range $[0, 1]$.

For the foreground assessment, we apply the hard thresholding to extract the foreground masks and then compute the following metrics. Here ground truth foreground masks are manually made. Let TP be the true positive counting the foreground pixels correctly labeled as foreground, FP be the false positive counting the background pixels incorrectly labeled as foreground, and FN be the false negative counting the foreground pixels incorrectly labeled as background. The three metrics are defined as follows.

- Precision (Pr): $\mathrm{Pr} = TP/(TP + FP)$
- Recall (Re): $\mathrm{Re} = TP(TP + FN)$
- F-measure (Fm): $\mathrm{Fm} = 2\mathrm{Re}/\mathrm{Pr}$

All the three metrics are between 0 and 1. The higher the value is, the more accurate the result is. We also find that various hard thresholding strategies may cause different one or two metrics high while the remaining ones are low.

A Microsoft Azure Kinect Sensor was used to record human motion, including one 1-MP depth sensor, one 7-microphone array, one 12-MP RGB video camera, and one accelerometer and gyroscope (IMU) sensor. Designed to pull together multiple AI sensors in a single device, Azure Kinect sensors have been employed for various applications, such as building telerehabilitation solutions, democratizing home fitness, etc. In this study, only the RGB video camera was used to record human motion and test the proposed algorithm. All numerical experiments were run in Matlab R2021a on a desktop computer with Intel CPU i9-9960X RAM 64GB and GPU Dual Nvidia Quadro RTX5000 with Windows 10 Pro.

A. Experiment 1: Whole Body Movement Video

For the first experiment, we consider a video capturing whole-body movement, which was recorded when one student volunteer was walking naturally at an average speed in a lab room. The video of interest consists of 60 frames where each frame has 150×200 pixels. Due to the limited lighting conditions, there are inevitable shadows of the person and brightness variations in the foreground. In Fig. 1, we compare the recovered background from the various methods. For each method, we take the mean column of the recovered L followed by reshaping it as a matrix, i.e., we use the mean of the obtained backgrounds over 60 frames. There are some white spots in the blackboard mistakenly recognized as foreground in the LAGO result and quite a few still exist in the SPCP result. Both SPGL1 and our results can recover the background well except the light shadow on the ground. In Fig. 2, we show the recovered foregrounds at the first and the last frame. The LAGO result has blurry edges for the human body, and both SPGL1 and our results can detect the shadow motion. The quantitative comparison of the recovered foreground and background for all methods is reported in Table I. Our method performs best in terms of all the comparison metrics for this video data.

Fig. 1. Recovered backgrounds of the walking video via various methods.

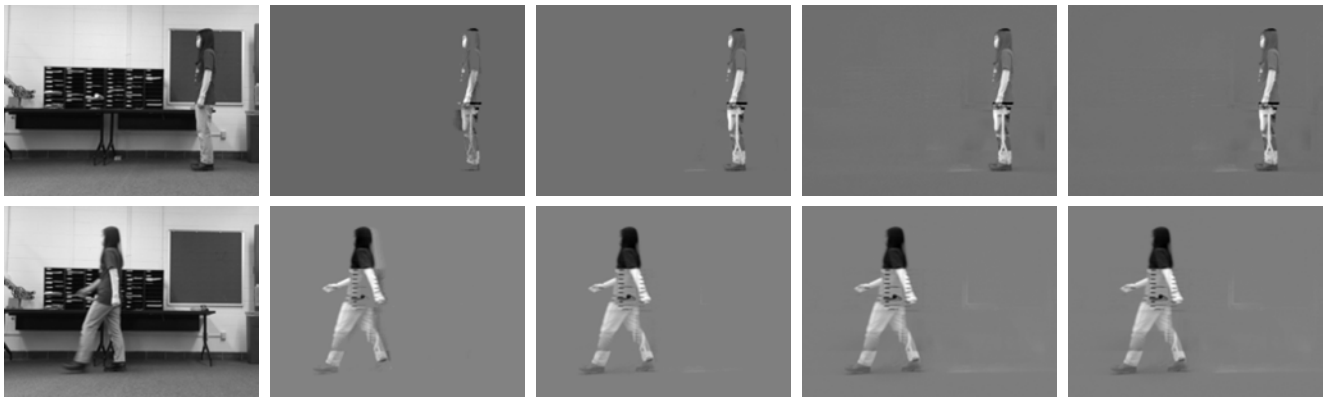

Fig. 2. Detected objects for the walking video via various methods. The two rows correspond to the first and the last video frames, respectively.

TABLE I

QUANTITATIVE COMPARISON FOR THE WALKING VIDEO

	RE	PSNR	Pr	Re	Fm
LAGO	0.0377	33.67	0.9796	0.4673	0.6328
SPCP	0.0182	39.99	0.9777	0.6354	0.7703
SPGL1	0.0148	41.81	0.9682	0.7180	0.8246
Alg. 1	0.0145	41.95	0.9688	0.7187	0.8252

TABLE II

QUANTITATIVE COMPARISON FOR THE ARM MOTION VIDEO

	RE	PSNR	Pr	Re	Fm
LAGO	0.0151	20.36	0.9617	0.8305	0.8913
SPCP	0.0132	20.44	0.9666	0.8471	0.9029
SPGL1	0.0132	35.65	0.9665	0.8610	0.9107
Alg. 1	0.0104	35.67	0.9704	0.8234	0.8909

B. Experiment 2: Arm Movement Video

In the second experiment, an arm movement video was recorded when the student volunteer rotated her forearm and hand around the elbow joint slowly. The tested video is generated by removing motionless frames and cropping the region of interest, which consists of 32 frames and each frame has 180×180 pixels. The visual comparison of foreground and background for all results are shown in Fig. 3 and Fig. 2, respectively. In Table II, we compare the qualities of the recovered background and foreground. Notice that there is movement still left on the left of the LAGO background and foreground results while some speckle noise exist in the SPCP foreground. Both SPGL1 and our approach can separate the foreground and the background clearly.

In terms of running time, SPCP takes the minimum running time (0.1 s) while SPGL1 based on Newton's iteration takes about 50 seconds. Both LAGO and our algorithm run about 5 seconds and the graph Laplacian construction can be fast using a small number of neighbors. Overall, our method can keep a good balance in running time and detection accuracy. This phenomenon also applies to the first experiment.

V. CONCLUSIONS AND FUTURE WORK

Moving object detection is one of the most fundamental tasks in video processing with a wide spectrum of applications, particularly in human-robot interaction. In the case of limited lightening conditions and/or time-varying illuminations, it becomes extremely challenging to separate a moving foreground with shadow from a static background. One classical type of methods is to segment each single frame into foreground and background. However, it usually loses the temporal smoothness and suffers from the intensive computation. In this work, we propose a novel dual-graph regularized motion detection approach. Specifically, we exploit the spatiotemporal geometry of the foreground by constructing the spatial and the temporal graph Laplacians, and adopt a weighted nuclear norm regularizer based on the error function to utilize adaptive low-rankness of the background. The proposed algorithm is derived by applying the ADMM framework. Numerical results have shown our method outperforms the other related ones on realistic data sets. In the future, we will develop fast methods based on the low-rank tensor decompositions and separate the shadow from the detected moving object under sophisticated lightening environments.

Fig. 3. Recovered backgrounds of the arm motion video via various methods.

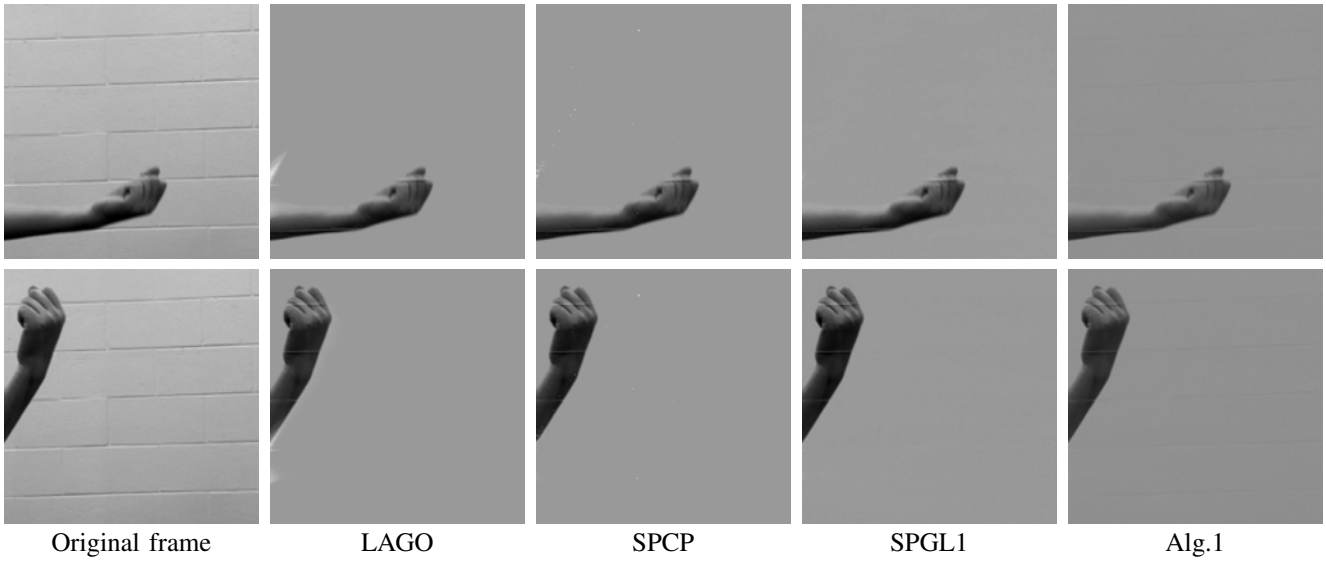

Fig. 4. Detected objects for the arm motion video. The two rows correspond to the first and the last video frames, respectively.

ACKNOWLEDGMENTS

The research of Qin is supported by the NSF grant DMS-1941197, and the research of Shen, Zhu and Xie is supported by the University of Kentucky College of Engineering Young Alumni Philanthropy Council Funding.

REFERENCES

[1] A. M. Zanchettin, P. Rocco, S. Chiappa, and R. Rossi, "Towards an optimal avoidance strategy for collaborative robots," *Robotics and Computer-Integrated Manufacturing*, vol. 59, pp. 47–55, 2019.

[2] M. K. STEİN and J. Kaivo-Oja, "Collaborative robots: Frontiers of current literature," *Journal of Intelligent Systems: Theory and Applications*, vol. 3, no. 2, pp. 13–20, 2020.

[3] S. Sajedi, W. Liu, K. Eltouny, S. Behdad, M. Zheng, and X. Liang, "Uncertainty-assisted image-processing for human-robot close collaboration," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4236–4243, 2022.

[4] V. V. Unhelkar, P. A. Lasota, Q. Tyroller, R.-D. Buhai, L. Marceau, B. Deml, and J. A. Shah, "Human-aware robotic assistant for collaborative assembly: Integrating human motion prediction with planning in time," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2394–2401, 2018.

[5] T. Bouwmans, "Subspace learning for background modeling: A survey," *Recent Patents on Computer Science*, vol. 2, no. 3, pp. 223–234, 2009.

[6] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011.

[7] S. Javed, S. Ho Oh, A. Sobral, T. Bouwmans, and S. Ki Jung, "Background subtraction via superpixel-based online matrix decomposition with structured foreground constraints," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 90–98.

[8] S. Javed, A. Mahmood, T. Bouwmans, and S. K. Jung, "Spatiotemporal low-rank modeling for complex scene background initialization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 6, pp. 1315–1329, 2016.

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

[10] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 2862–2869.

[11] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5, pp. 877–905, 2008.

[12] W. Guo, Y. Lou, J. Qin, and M. Yan, "A novel regularization based on the error function for sparse recovery," *Journal of Scientific Computing*, vol. 87, no. 1, pp. 1–22, 2021.

[13] J. Shen, H. Xu, and P. Li, "Online optimization for max-norm regularization," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[14] N. Srebro and A. Shraibman, "Rank, trace-norm and max-norm," in *International Conference on Computational Learning Theory*. Springer, 2005, pp. 545–560.

[15] D. Driggs, S. Becker, and A. Aravkin, "Adapting regularized low-rank models for parallel architectures," *SIAM Journal on Scientific Computing*, vol. 41, no. 1, pp. A163–A189, 2019.

[16] E. Van Den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2009.

[17] A. Aravkin, S. Becker, V. Cevher, and P. Olsen, "A variational approach to stable principal component pursuit," in *Conference on Uncertainty in Artificial Intelligence (UAI)*, July 2014.