Weighted Low Rank Approximation for Background Estimation Problems

Aritra Dutta
King Abdullah University of Science and Technology (KAUST)
d.aritra2010@knights.ucf.edu

Xin Li
University of Central Florida
xin.li@ucf.edu

Abstract

Classical principal component analysis (PCA) is not robust when the data contain sparse outliers. The use of the ℓ_1 norm in the Robust PCA (RPCA) method successfully eliminates this weakness of PCA in separating the sparse outliers. Here we propose a weighted low rank (WLR) method, where a simple weight is inserted inside the Frobenius norm. We demonstrate how this method tackles often computationally expensive algorithms that rely on the ℓ_1 norm. As a proof of concept, we present a background estimation model based on WLR, and we compare the model with RPCA method and with other state-of-the-art algorithms used for background estimation. Our empirical validation shows that the weighted low-rank approximation we propose here can perform as well as or better than that of RPCA and other state-of-the-art algorithms.

1. Introduction

In image processing, rank-reduced signal processing, computer vision, and in many other engineering applications the classical principal component analysis (PCA) is a useful tool [19]. However, it might lead to a degraded construction in some cases because it can not preserve any structure of the data matrix. In 1987, Golub *et al*. [14] were the first to consider a *constrained* low rank approximation problem of matrices to address this fundamental flaw in PCA: Given $A = (A_1 \ A_2) \in \mathbb{R}^{m\times n}$ with $A_1 \in \mathbb{R}^{m\times k}$ and $A_2 \in \mathbb{R}^{m\times(n-k)}$, find $A_G = (\tilde{B}_1 \ \tilde{B}_2)$ such that

$$(\tilde{B}_1 \ \tilde{B}_2) = \arg \min_{\substack{B=(B_1 \ B_2)\\ B_1=A_1\\ \text{rank}(B)\le r}} \|A-B\|_F^2, \tag{1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of matrices. Golub *et al*. required a block of columns, A_1, of A must be preserved when one looks for a low rank approximation of $(A_1 \ A_2)$. As in the standard low rank approximation (which is equivalent to PCA), the constrained low-rank approximation problem of Golub *et al*. has a closed form solution.

Inspired by (1) and motivated by applications in which A_1 may contain noise, we require $\|A_1 - B_1\|_F$ small instead of asking for $B_1 = A_1$. This leads us to consider the following problem: Let $\eta > 0$, find $(\hat{B}_1 \ \hat{B}_2)$ such that

$$(\hat{B}_1 \ \hat{B}_2) = \arg \min_{\substack{B=(B_1 \ B_2)\\ \|A_1-B_1\|_F\le\eta\\ \text{rank}(B)\le r}} \|A-B\|_F^2. \tag{2}$$

Or, for a large parameter λ, consider

$$\min_{\substack{B=(B_1 \ B_2)\\ \text{rank}(B)\le r}} \{\lambda^2\|A_1-B_1\|_F^2 + \|A_2-B_2\|_F^2\}. \tag{3}$$

As it turns out, (3) can be solved in a closed form as a special case of weighted low-rank approximation with a rank-one weight matrix by the use of the singular value decomposition (SVD) of the given matrix $(\lambda A_1 \ A_2)$ [24, 25]. Using the closed form solutions, one can verify that the solution to (1) is the limit case of the solutions to (3) as $\lambda \to \infty$. Thus, (1) can be viewed as a special case of (3) when $\lambda = \infty$. Note that, problem (3) can also be cast as a special case of structured low rank problems with element-wise weights [1, 35, 36]. More specifically, we observe that (3) is contained in the following more general point-wise weighted low rank (WLR) approximation problem [12, 13, 24, 25, 32]:

$$\min_{\substack{X=(X_1 \ X_2)\\ \text{r}(X)\le r}} \|\left((A_1 \ A_2) - (X_1 \ X_2)\right) \odot (W_1 \ W_2)\|_F^2, \tag{4}$$

where $W = (W_1 \ W_2) \in \mathbb{R}^{m\times n}$ is a weight matrix and $\odot$ denotes the Hadamard product.

The idea of working with a weighted norm is very natural in solving many engineering problems. The weighted low rank approximation problem was first studied with W being an indicator weight to deal with the missing data case and then with more general weight in machine learning, collaborative filtering, 2-D filter design, and computer vision [7, 9, 18, 22, 23, 26, 30, 32, 33, 38]. Working with (4) can be challenging because no closed form solution exists [7, 23, 24, 25, 32].

1.1. RPCA and GFL for Background Estimation

In the past decade, matrix decomposition has been one of the most prevalently used methods for background estimation [2, 4, 31]. Given a sequence of n video frames with each frame converted into a vector $\mathbf{a}_i \in \mathbb{R}^m$, $i = 1, 2, ..., n$, the data matrix $A = (\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n) \in \mathbb{R}^{m\times n}$ is the concatenation of all the frame vectors. Because the background is not expected to change much throughout the

frames when the camera motion is small, it is assumed to be low-rank [27]. At the same time, the foreground is usually sparse if its size is relatively small compared with the frame size [8, 39, 21]. Therefore, naturally, one considers a matrix decomposition problem by decomposing A as the sum of its background and foreground:

$$A = B + F,$$

where $B, F \in \mathbb{R}^{m\times n}$ are the background and foreground matrices, respectively. The robust principal component analysis (RPCA) exploits this idea in [8, 21, 39] and solves the background estimation problem by assuming the background frames, B, to have a low-rank structure and the foreground, $A - B$, sparse:

$$\min_{B} \|A - B\|_{\ell_1} + \lambda\|B\|_*, \tag{5}$$

where $\|\cdot\|_{\ell_1}$ and $\|\cdot\|_*$ denote the ℓ_1 norm and the nuclear norm (sum of the singular values) of matrices, respectively. But the RPCA model cannot take advantage of any possible extra information on the background. Recently in [40], Xin *et al*. proposed a stronger model called generalized fused Lasso (GFL) for the situation where pure background frames are given as a supervised learning method. Based on the assumption that if some pure background frames are given, then the data matrix A can be written into $A = (A_1\ A_2)$, where A_1 contains the given pure background frames, Xin *et al*. [40] proposed the following model of the unknown matrices B and F: with $B = (B_1\ B_2)$ and $F = (F_1\ F_2)$ partitioned in the same way as in A, find B and F satisfying

$$\min_{\substack{B,F\\ B_1=A_1}} \operatorname{rank}(B) + \|F\|_{gfl},$$

where $\|\cdot\|_{gfl}$ denotes a norm that is a combination of ℓ_1 norm and a local spatial total variation norm (to encourage connectivity of the foreground). To make the problem more tractable, Xin *et al*. further specialized the above model by assuming $\operatorname{rank}(B) = \operatorname{rank}(B_1)$. Since $B_1 = A_1$ and A_1 is given, so $r := \operatorname{rank}(B_1)$ is also given and thus, we can re-write the model of [40] as a special case of the following:

$$\min_{\substack{B=(B_1\ B_2)\\ \operatorname{rank}(B)\le r\\ B_1=A_1}} \|A - B\|_{gfl}. \tag{6}$$

Clearly, except in different norms, problem (6) is a constrained low rank approximation problem as in (1).

1.2. Main Contributions

In this paper, we propose an algorithm to solve (4) as a standalone problem for a special family of weights. As a proof of concept, we present a background estimation model by using our WLR algorithm because it seems to be a natural fit to the problem. In addition, we compare the performance of our proposed model with that of RPCA and of GFL algorithms in estimation of backgrounds that contains static and dynamic components. Our goal is to show how a properly weighted Frobenius norm can be made robust to tackle the outliers similar to the ℓ_1 norm. To comprehensively review the most recent and classical algorithms that solve the background estimation problem, we refer the reader to [2, 3, 4, 31].

We show that one can use WLR to find a more robust and efficient approach to solve the background estimation problem as compared to the RPCA [21] and GFL [40] algorithms if one uses a special weighted version of low rank approximation (4) and learns the weight (or, more precisely, the frame indices of weight as explained in Section 3). Our proposed model is not only as efficient as RPCA and GFL algorithms, but also it does not require any prior information as needed in [40]. More specifically, we show that (1) our weighted Frobenius norm minimization can replace the computationally expensive ℓ_1 minimization as in RPCA and GFL algorithms and achieve a superior or at least comparable performance, that (2) our model allows frames that are close to the background to be used without requiring prior knowledge of the pure background frames, that (3) these approximate background frames are not given to us but learned from the data, and that (4) when compared with other state-of-the-art background and foreground estimation algorithms, our method may provide a better background estimation.

2. An Algorithm for WLR

In this section, we propose an algorithm to solve (4) for a special family of weights when $W = (W_1\ W_2)$ with $W_2 = \mathbb{1}$, matrix with entries equal to 1. Note that if $W_1 = \mathbb{1}$ as well, then we are back to PCA. For convenience, let $\mathrm{r}(X_1) = k$. Then any X_2 such that $\mathrm{r}(X_1\ \ X_2) \le r$ can be given in the form

$$X_2 = X_1C + BD,$$

for some arbitrary matrices $B \in \mathbb{R}^{m\times(r-k)}$, $D \in \mathbb{R}^{(r-k)\times(n-k)}$, and $C \in \mathbb{R}^{k\times(n-k)}$. Denote $F(X_1, C, B, D) = \|(A_1 - X_1)\odot W_1\|_F^2 + \|A_2 - X_1C - BD\|_F^2$. Therefore, problem (4) with $W = (W_1\ \ \mathbb{1})$ is further reduced to:

$$\min_{X_1,C,B,D} F(X_1, C, B, D). \tag{7}$$

Note that, for the special choice of the weight matrix, with a block structure $(X_1\ \ B)\begin{pmatrix} I_k & C\\ 0 & D\end{pmatrix}$, problem (7) can be written alternatively in the framework of alternating weighted least squares algorithm in [25]. Here we directly solve (7) by using a fast and simple numerical procedure.

Problem (7) can be numerically solved by using an alternating strategy [5] of minimizing F with respect to each component iteratively:

$$\begin{cases} (X_1)_{p+1} = \arg\min_{X_1} F(X_1, C_p, B_p, D_p),\\ C_{p+1} = \arg\min_{C} F((X_1)_{p+1}, C, B_p, D_p),\\ B_{p+1} = \arg\min_{B} F((X_1)_{p+1}, C_{p+1}, B, D_p),\\ \text{and, } D_{p+1} = \arg\min_{D} F((X_1)_{p+1}, C_{p+1}, B_{p+1}, D). \end{cases}$$

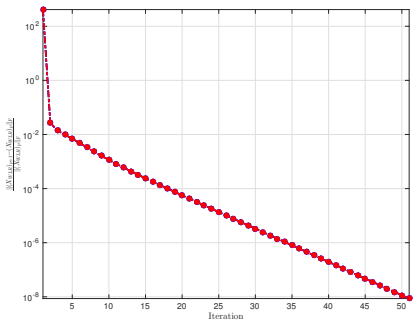


Figure 1: Iterations vs. relative error on Stuttgart video sequence: *Basic* scenario, $A \in \mathbb{R}^{5120\times 600}$.

Each sub-problem above can be solved explicitly as described in Algorithm 1. Let $(X_{WLR})_p$ be our approximation to A at pth iteration and define $E_p = \|(X_{WLR})_{p+1} - (X_{WLR})_p\|_F$. For a threshold $\epsilon > 0$ the stopping criteria of our algorithm at the pth iteration is $E_p < \epsilon$ or $\frac{E_p}{\|(X_{WLR})_p\|_F} < \epsilon$ or if it reaches the maximum iteration. Figure 1 shows iteration p vs. relative error plot for Algorithm 1 on Stuttgart video sequence which suggests the convergence of WLR. A more detailed convergence analysis is given in [10, 13].

Algorithm 1: WLR Algorithm

Input : $A = (A_1 \;\; A_2) \in \mathbb{R}^{m\times n}$ (the given matrix); $W = (W_1 \;\; \mathbb{1}) \in \mathbb{R}^{m\times n}$ (the weight), threshold $\epsilon > 0$;
Initialize: $(X_1)_0, C_0, B_0, D_0$;
while *not converged* **do**
  $E_p = A_1 \odot W_1 \odot W_1 + (A_2 - B_pD_p)C_p^T$;
  for $i = 1 : m$ **do**
    $(X_1(i,:))_{p+1} = (E(i,:))_p(\text{diag}(W_1^2(i,1)$ $W_1^2(i,2)\cdots W_1^2(i,k)) + C_pC_p^T)^{-1}$;
  end
  $C_{p+1} = ((X_1)_{p+1}^T(X_1)_{p+1})^{-1}(X_1)_{p+1}^T(A_2 - B_pD_p)$;
  $B_{p+1} = (A_2 - (X_1)_{p+1}C_{p+1})D_p^T(D_pD_p^T)^{-1}$;
  $D_{p+1} = (B_{p+1}^TB_{p+1})^{-1}B_{p+1}^T(A_2 - (X_1)_{p+1}C_{p+1})$;
  $p = p + 1$;
end
Output : $(X_1)_{p+1}, (X_1)_{p+1}C_{p+1} + B_{p+1}D_{p+1}$.

3. Background Estimation by using WLR

In this section, we propose a background estimation model that uses Algorithm 1. To use our proposed algorithm in background estimation, we first solve WLR for $W = \mathbb{1}$ (no weighted case which is just PCA) to obtain an initialization of the background and foreground: $A = B_{In} + F_{In}$, where B_{In} is a low rank approximation to A given by PCA. Next, we use B_{In} and F_{In} to learn the frame indices that are closest to the pure background. This is done heuristically (with a similar idea as in [11] (see Figure 2)) such that we find the frame indices that are close to the pure background in A. By setting a threshold $\varepsilon_1 > 0$ based on the histogram of F_{In}, we convert F_{In} into a binary matrix LF_{In}: all entries of F_{In} bigger than ε_1 are replaced by 1 and the others are replaced by 0. The matrix B_{In} is directly converted to a binary matrix LB_{In}. Next, we calculate the ratios of the frame sum (i.e. the column sum) of LF_{In} to the corresponding frame sum of LB_{In} and identify the indices with ratios less than the mode of these ratios as possible pure background frame indices. Finally, we apply WLR by putting the weight at the learned frame indices to decompose the data matrix A into background and foreground: $A = B + F$. Our experiments show that the performance depends more on the correct location (indices) of the background frames than on the values of the weight. We remark that Dutta and Li [12] and Xin *et al.* [40] used the pure background frames in their background estimation model, but the frames were already given to them. On the contrary, Algorithm 2 *learns* the background frame indices from the data, thus providing a robust background estimation model.

Algorithm 2: Background Estimation using WLR

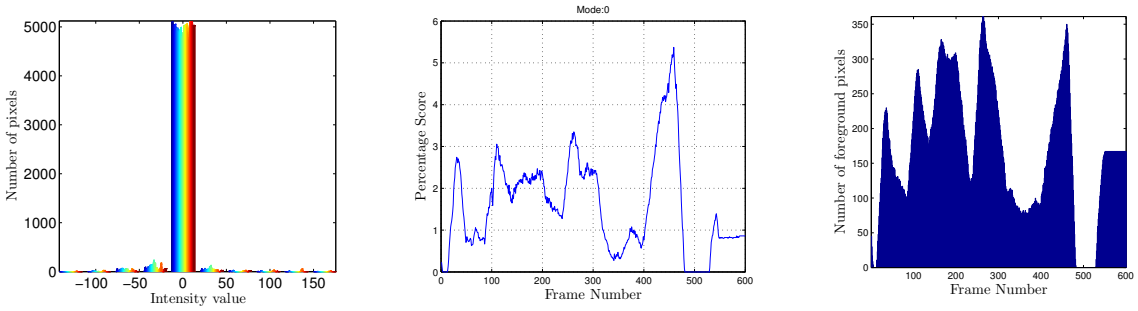
Input : $A = (A_1 \;\; A_2) \in \mathbb{R}^{m\times n}$ (the data matrix); $W = (W_1 \;\; \mathbb{1}) \in \mathbb{R}^{m\times n}$,(the weight), threshold $\epsilon > 0$, $i_1, i_2 \in \mathbb{N}$;
Run PCA to get low rank B_{In} and $F_{In} = A - B_{In}$;
Learn background frame indices S from B_{In} and F_{In};
Set $k = \left\lceil |S|/i_1 \right\rceil, r = k + i_2$;
Rearrange data: $\tilde{A}_1 = (A(:,i))_{m\times k}$, randomly chosen k frames from $i \in S$ and $\tilde{A}_2 = (A(:,i'))_{m\times(n-k)}$, i' from the remaining frames;
Apply WLR on $\tilde{A} = (\tilde{A}_1 \;\; \tilde{A}_2)$ to obtain $\tilde{X}$;
Rearrange the columns of $\tilde{X}$ similar to A to find X;
Output : X.

4. Experiments and Comparisons

In this section, we report how we compare extensively WLR with RPCA to validate its effectiveness in solving the background estimation problem. We also compare WLR with supervised and unsupervised GFL, as well as with other state-of-the-art background estimation methods.

4.1. Comparison with RPCA

In this set of experiments, we use the Stuttgart synthetic video data set [6] for rigorous qualitative and quantitative comparisons. This video data set is a computer generated video sequence that comprises both static and dy-



(a) (b) (c)

Figure 2: Learning the weighted frame indices for the *Basic* scenario. (a) Histogram to choose the threshold ϵ_1. (b) Percentage score plot for 600 frames. (c) Original binary G column sum, which indicates we are able pick up the indices correctly corresponding to the frames that have the least foreground movement. Originally, there are 53 frames in G that have less than 5 pixels. We picked up 58 frame indexes on the *Basic* scenario.

Figure 3: The effect of using weights in WLR algorithm on the *Basic* scenario, frame 435. Background estimation using WLR with: (a) $(W_1)_{ij} \in [5, 10]$, (b) $(W_1)_{ij} \in [500, 1000]$. In (a) the estimated background has the blurry foreground object, but as the weight is increased, the foreground object disappears in (b).

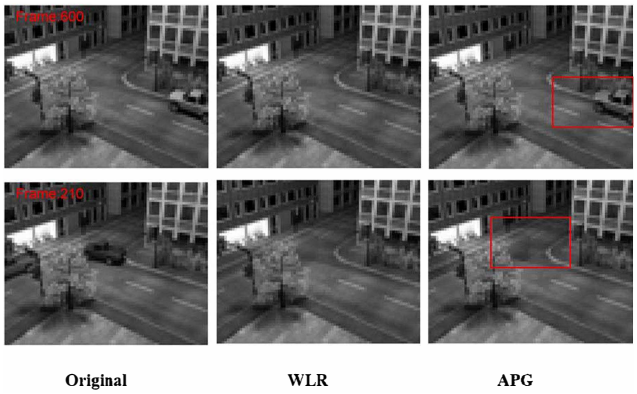


Figure 4: Background estimated by WLR and APG on *Basic* scenario. Top row shows frame 600 and bottom shows frame 210. APG can not remove the static foreground object in frame 600. In frame 210, the low-rank background estimated by APG has some black patches.

namic background/foreground objects and varying illumination in the background. We use **three** different test scenarios of the sequence [6]: (i) *Basic*: This scenario has neither noisy artifacts nor sudden illumination changes, and it is used as a general performance measure. (ii) *Noisy night:* This scenario is a low-contrast nighttime scene, with increased sensor noise and camouflage. (iii) *Light switch:* This scenario has varying illumination effects throughout the sequence. Note that each scenario has 600 frames and identical foreground and background objects. Frames 551 to 600 have static foreground, and frames 6 to 12 and 483 to 528 have no foreground. Additionally, the foreground comes with high quality ground truth mask for each video frame. To compare our WLR method with the existing RPCA algorithms, we use the inexact augmented Lagrange multiplier (iEALM) method proposed by Lin *et al.* [21], and the accelerated proximal gradient (APG) algorithm proposed by Wright *et al.* [39]. We only report on APG if iEALM has similar performance. For iEALM and APG, we set $\lambda = 1/\sqrt{\max\{m, n\}}$, and for iEALM we choose $\mu = 1.5, \rho = 1.25$ as used in [8, 21, 39]. Each frame in the test sequence is resized to 64×80 due to the memory constraint of RPCA algorithms (originally they were 600×800). The resized frames are stacked as column vectors to form the data matrix A. For the Stuttgart video sequence, we empirically choose $k = \left\lceil |S|/2 \right\rceil$ and set $r = k + 1$. Therefore, in Algorithm 2, we use $i_1 = 2$ and $i_2 = 1$. However, such assumptions do not apply to all practical scenarios. The choices of r and k are problem-dependent and highly heuristic. When calling WLR, we run Algorithm 1 for 50 iterations and choose threshold $\epsilon = 10^{-7}$.

Qualitative Analysis. We present frame 435 of the *Basic* scenario in Figure 3 to show the effect of a large weight, W_1, on the first block A_1: our weighted low-rank algorithm can perform well in background estimation with

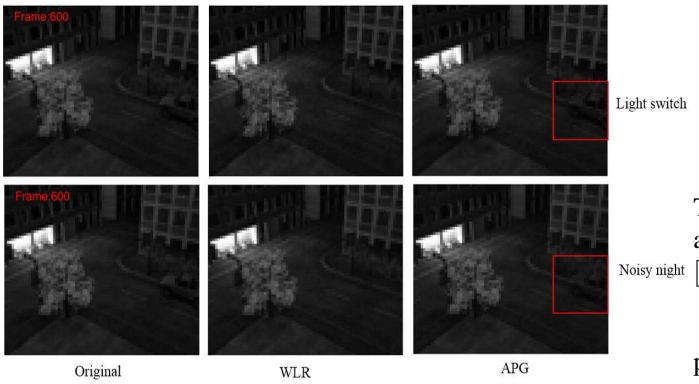


Figure 5: Background estimated by WLR and APG on *Light switch* and *Noisy night* scenario, frame 600. APG was not able to remove the static foreground, but WLR completely removed the static foreground from both scenarios.

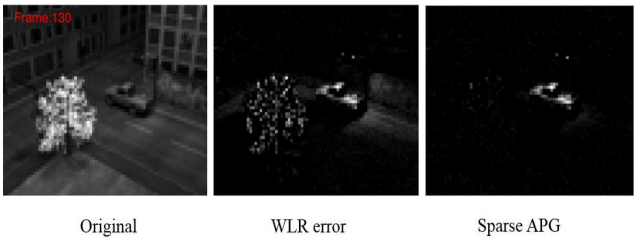


Figure 6: Foreground recovered by WLR and APG on *Light switch* scenario, frame 130. Starting from frame 125 the illumination changes suddenly. The sparse foreground recovered by APG does not capture the change in illumination. WLR captures the effect of change in illumination, irregular movements of the tree leaves, and reflections.

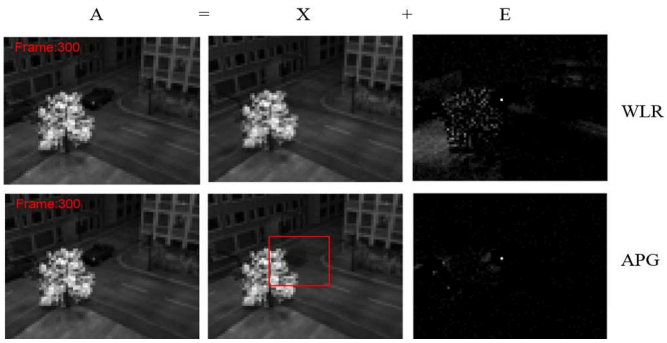


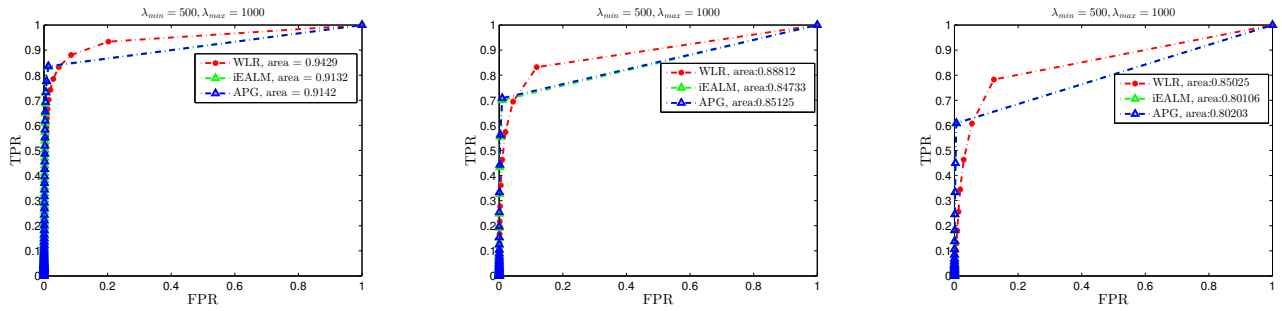
Figure 7: Background and foreground estimated by WLR and APG on *Light switch* scenario, frame 300. WLR has least MSSIM for frame 300 but it still provides a better visual quality foreground and background estimation than APG. The red bounding box in APG frame is indicating the presence of foreground patch.

Method	*Basic*	*Noisy night*	*Light switch*
WLR	**23.0676**	**24.0970**	**20.1874**
iEALM	160.251981	108.679550	173.903928
APG	107.982398	115.547544	109.976457

Table 1: Average Computational time (in seconds) for WLR and RPCA algorithms in processing 600 frames of size $[64, 80]$.

proper choice of weight. Next, in Figure 4, we present frame 210 and 600 of the *Basic* scenario. The performance of APG on frame 210 is comparable with WLR, but on frame 600, WLR outperforms APG. Finally, Figure 5 shows that WLR removes the static foreground and provides a better visual background in scenes with varying illumination and sensor noise. To conclude, when the foreground is static, with the proper choice of W, r, and k our algorithm can provide a good estimation of the background by removing the static foreground object. Figures 6 and 7 present the foreground recovered by WLR and APG on the *Light switch* scenario. We show that WLR can capture the changing illumination and irregular dynamic background movements better than APG and can provide a visually better background frame, even on Frame number 300 of *Lightswitch* scenario where WLR has least MSSIM. This can be attributed to the fact that, RPCA algorithms are based on the assumption that the low-rank component is exactly low-rank while the sparse component being exactly sparse [4, 8, 39]. Furthermore, considering the computational time of each algorithm from Table 1, WLR has minimal execution cost in producing a superior background estimation.

Quantitative Analysis. We now present different quantitative comparisons between the performance of our algorithm and that of the existing RPCA algorithms. We use three different quantitative measures for this purpose: traditionally used receiver and operating characteristic (ROC) curve, peak signal to noise ratio (PSNR), and the most advanced measure mean structural similarity index (MSSIM). Because a ground truth mask is available for each video frame, we use a pixel-based measure of F, the foreground recovered by each method to form the confusion matrix for the predictive analysis. In our case, the pixels are represented by the use of 8 bits per sample, and $\mathrm{M_I}$, the maximum pixel value of the image is 255. Therefore, a uniform threshold vector linspace(0, M_I, 100) is used to compare the pixel-wise predictive analysis between each recovered foreground frame and the corresponding ground truth frame. From the ROC curves in Figure 8, the increment in performance of WLR compared with RPCA algorithms appear to be substantial. However, the qualitative performance of the proposed weighted algorithm in all three scenarios is much superior. We attribute this to the fact that WLR removes the



(a) Basic (b) Noisy night (c) Lightswitch

Figure 8: ROC curve to compare between WLR, iEALM, and APG. The performance gain by WLR compare to APG on *Basic, Noisy night,* and *Light switch* scenarios are 3.252%, 4.3313%, and 6.012% respectively, and compare to iEALM are 3.139%, 4.8139%, and 6.141% respectively.

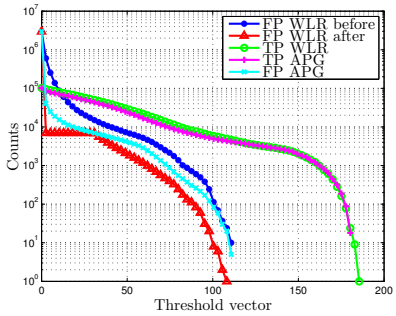


Figure 9: True positive and false positive count for WLR and APG on *Basic* scenario. The false positive count for WLR substantially drops after thresholding F by ϵ_1. On the other hand, WLR always has more or equal number of true positives as APG.

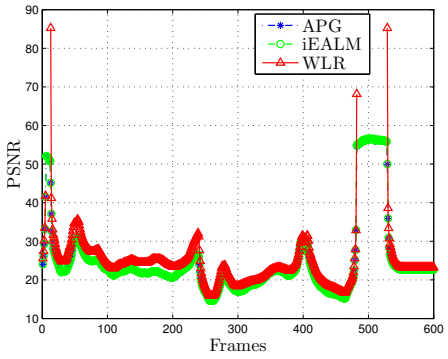


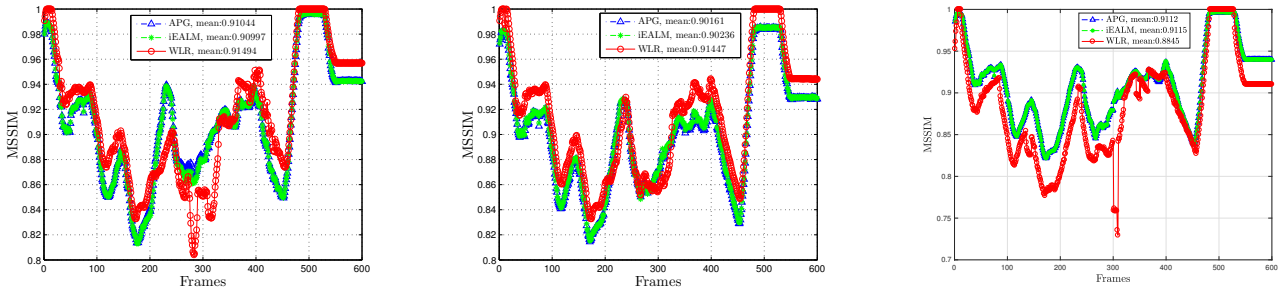
Figure 10: Frames vs. PSNR for *Basic* scenario. The mean PSNR of APG and iEALM on the *Basic* sequence are 25.0092 and 25.0551, respectively. For WLR, the frames that do not contain the foreground object have 0 MSE, resulting PSNR equal to infinity, in all three scenarios.

noise uniformly from the video sequence.

In calculating the PSNR, we perceive the information on how the high intensity regions of the image are coming through the noise, and consequently, we pay much less attention to the low intensity regions. This motivated us to remove the noisy components from the recovered foreground, F, by using the threshold ϵ_1 (see Section 3), such that we set the components below ϵ_1 in F to 0. Using this new F, we give the next two quantitative measures. PSNR is calculated using the metric: $10log_{10}\frac{\mathrm{M}_I^2}{\mathrm{MSE}}$, where $\mathrm{MSE} = \frac{1}{mn}\|F(:,i) - G(:,i)\|_2^2$. Conventionally, the higher the PSNR value, the better the reconstruction algorithm. Figure 10 indicates the PSNR of WLR is superior than the RPCA algorithms. This can be attributed to the fact that after thresholding the foreground frames recovered by WLR in all three scenarios are identical to the ground truth frames.Finally we use the mean SSIM (MSSIM) index to evaluate the overall image quality [37]. In order to calculate MSSIM of each recovered foreground video frame, we consider a 11×11 Gaussian window with standard deviation $\sigma = 1.5$. In Figure 11, we plot the MSSIM of different methods for all three scenarios. The MSSIM plot demonstrates that WLR has superior performance to the RPCA algorithms, especially when there is no foreground or static foreground exists. In Figure 12 the SSIM index map of two sample foreground video frames indicate fragmentary foreground recovered by the RPCA algorithms.

4.2. Comparison with GFL

In this section, we compare the performance of WLR with the supervised and unsupervised GFL background estimation model of Xin *et al*. [40]. Because the choice of r and k are problem specific for our model, we have provided only the quantitative comparison on the *Waving tree* scene of the Wallflower dataset [34] and *Basic* scenario of the Stuttgart dataset. Xin *et al*. used 200 pure background



(a) Basic (b) Noisy night (c) Lightswitch

Figure 11: MSSIM of different methods on all three scenarios. WLR has better MSSIM compare to the RPCA algorithms corresponding to the frames which has static foreground or no foreground. The slight deterioration of performance of WLR around frame 300 in (c) can be attributed by poor thresholding as the noise level in those frames are very high.

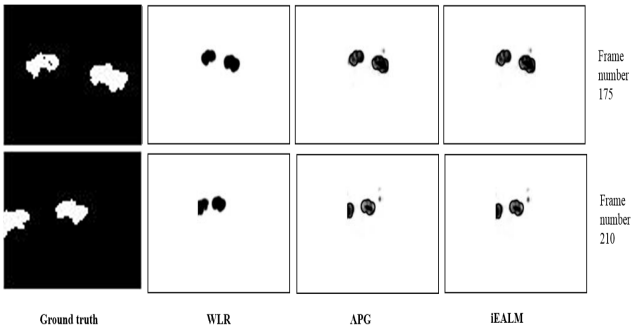


Figure 12: SSIM index map for frame 175 and 210 of the *Basic* scenrio. Left to right: Ground truth frame (size 64×80), SSIM index map (size 54×70) of WLR, APG, and iEALM. WLR has superior SSIM index map than RPCA algorithms.

Figure 13: SSIM index map of *Waving tree*. Left to right: Ground truth frame (size 64×80), WLR SSIM index map (size 54×70), and GFL SSIM index map. MSSIM for WLR and GFL are 0.5018 and 0.5014 respectively.

frames as a prior for supervised GFL. On the other hand, we used all frames of the sequence to learn the weighted frame indices and estimate the background without using the exact location of the pure background frames. SSIM index map in Figure 13 shows that both methods are very competitive.

For the *Basic* scenario of the Stuttgart dataset we run the unsupervised GFL without using the knowledge of pure background frames and resize the first 200 frames as in

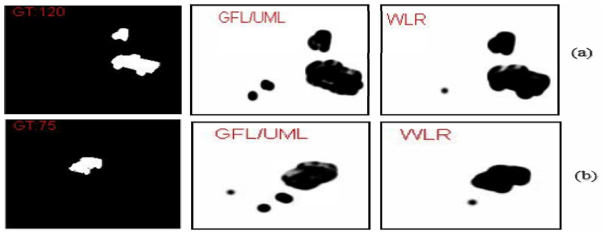


Figure 14: SSIM index map of *Basic* scenario. Left to right: Ground truth frame (size 144×176), SSIM index map (size 134×166) for WLR and GFL. MSSIM for WLR and GFL are (a) frame 120: 0.9326 and 0.9244, (b) frame 75: 0.9659 and 0.9677 respectively.

software [40]. For fair comparison we use the same data matrix for WLR. From SSIM index map in Figure 14, we see that both methods are very competitive with WLR being extraordinarily time efficient than the unsupervised GFL model. WLR takes approximately 17.75 seconds, while, on the same hardware, unsupervised GFL took 52297.39 seconds to conduct the experiment.

4.3. Comparison with other state-of-the-art background estimation models

We also compare WLR with other state-of-the-art robust background estimation algorithms, such as, Grassmannian robust adaptive subspace estimation (GRASTA) [17], recursive projected compressive sensing algorithm (ReProCS) [15, 16], and incremental principal component pursuit (incPCP) [28, 29] on the *Basic* scenario. Due to the limitation of space, we refer the readers to the references for an explanation of these algorithms. For GRASTA, we set the subsample percentage s at 10%, 20%, and 30% respectively, estimated rank 60, and other parameters the same as those in [17]. We use 200 background frames of the *Basic* sequence for initialization of ReProCS. incPCP algorithm uses the first video frame of the *Basic* scenario for

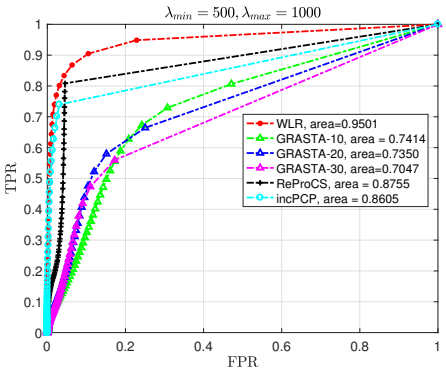


Figure 15: ROC curve to compare between WLR, GRASTA with different subsamples, ReProCS, and incPCP.

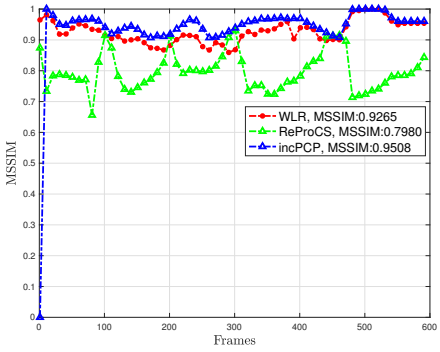


Figure 16: MSSIM to compare between WLR, ReProCS, and incPCP on *Basic* scenario.

initialization. Each frame is resized to 144×176. The ROC curves in Figure 15 shows that WLR outperforms all other methods. MSSIM presented in Figure 16 shows incPCP is slightly better or comparable to WLR. However, the qualitative result in Figure 17 shows when the foreground is static, the ℓ_1 norm in incPCP cannot capture the foreground object, resulting the presence of the static car as a part of the background. In contrast, WLR detects the static foreground.

4.4. Further Experiments on Dynamic Background

To demonstrate the power of our method on more complex data sets containing dynamic foreground, we perform extensive qualitative and quantitative analysis on the Li data set [20]. We use four sequences of the data set containing dynamic foreground. The SSIM index map on all four recovered foreground indicates that WLR performs consistently well on the video sequences containing dynamic background (see Figure 18).

5. Conclusion

Our weighted low-rank approximation algorithm is simple and fast for a special family of weights. Moreover, the

Figure 17: Frame 600, *Basic* scenario. Left to right: Original, incPCP background, WLR background, WLR SSIM index map, incPCP SSIM index map. Though incPCP has slightly better or same MSSIM compare to WLR, it fails to detect the static foreground object.

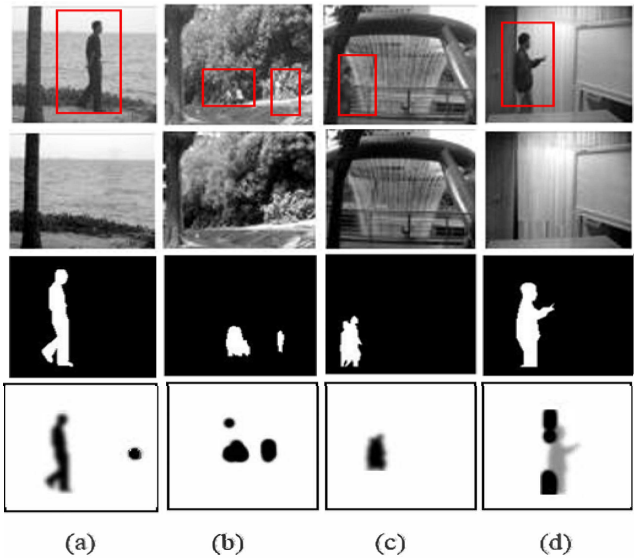


Figure 18: SSIM index map of: (a) *Water Surface*, (b) *Waving tree*, (c) *Fountain*, and (d) *Curtain*. Top to bottom: Original, background estimated by WLR, ground truth frame (size 64×80), SSIM index map (size 54×70) for WLR. The MSSIM are 0.9851, 0.9082, 0.9940, and 0.9343 respectively.

model that we devised for background estimation is efficient and robust. We demonstrated that when applied to complex video sequences, our method is more effective than the existing RPCA method and other state-of-the-art algorithms. The main motivation of the paper is not just to propose a background estimation model. Rather, we wish to make a case for a newcomer, WLR method, by demonstrating how a properly weighted Frobenius norm can be made robust to the outliers, similarly to RPCA, GFL, and to other state-of-the-art background estimation models.

References

[1] M. Abdolali and M. Rahmati. Multiscale decomposition in low-rank approximation. *IEEE Signal Processing Letters*, 24(7):1015–1019,

2017. 1

[2] T. Bouwmans. Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11:31 – 66, 2014. 1, 2

[3] T. Bouwmans, L. Maddalena, and A. Petrosino. Scene background initialization: A taxonomy. *Pattern Recognition Letters*, 2017. 2

[4] 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*, 23:1–71, 2017. 1, 2, 5

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. 2

[6] S. Brutzer, B. Höferlin, and G. Heidemann. Evaluation of background subtraction techniques for video surveillance. *IEEE Computer Vision and Pattern Recognition*, pages 1568–1575, 2012. 3, 4

[7] A. M. Buchanan and A. W. Fitzgibbon. Damped Newton algorithms for matrix factorization with missing data. *In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:316–322, 2005. 1

[8] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the Association for Computing Machinery*, 58(3):11:1–11:37, 2011. 2, 4, 5

[9] Y. Chen, Y. Wang, M. Li, and G. He. Augmented lagrangian alternating direction method for low-rank minimization via non-convex approximation. In *Signal, Image and Video Processing*, 2017. 1

[10] A. Dutta. Weighted low-rank approximation of matrices: Some analytical and numerical aspects, 2016. Ph.D. dissertation, Department of Mathematics, University of Central Florida, Orlando, FL. 3

[11] A. Dutta, B. Gong, X. Li, and M. Shah. Weighted singular value thresholding and its applications to background estimation, 2017. arXiv:1707.00133. 3

[12] A. Dutta and X. Li. A fast algorithm for a weighted low rank approximation. In *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, pages 93–96, 2017. 1, 3

[13] A. Dutta and X. Li. On a problem of weighted low-rank approximation of matrices. *SIAM Journal on Matrix Analysis and Applications*, 38(2):530–553, 2017. 1, 3

[14] G. H. Golub, A. Hoffman, and G. W. Stewart. A generalization of the Eckart-Young-Mirsky matrix approximation theorem. *Linear Algebra and its Applications*, 88(89):317–327, 1987. 1

[15] H. Guo, C. Qiu, and N. Vaswani. An online algorithm for separating sparse and low-dimensional signal sequences from their sum. *IEEE Transactions on Signal Processing*, 62(16):4284–4297, 2014. 7

[16] H. Guo, C. Qiu, and N. Vaswani. Practical REPROCS for separating sparse and low-dimensional signal sequences from their sum-part 1. In *IEEE International Conference on Acoustic, Speech and Signal Processing*, pages 4161–4165, 2014. 7

[17] J. He, L. Balzano, and A. Szlam. Incremental gradient on the grassmannian for online foreground and background separation in subsampled video. *IEEE Computer Vision and Pattern Recognition*, pages 1937–1944, 2012. 7

[18] 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*, 2016. 1

[19] I. T. Jolliffee. Principal component analysis, 2002. Second edition. 1

[20] L. Li, W. Huang, I.-H. Gu, and Q. Tian. Statistical modeling of complex backgrounds for foreground object detection. *IEEE Transactions on Image Processing*, 13(11):1459–1472, 2004. 8

[21] Z. Lin, M. Chen, and Y. Ma. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices, 2010. arXiv1009.5055. 2, 4

[22] W. S. Lu, S. C. Pei, and P. H. Wang. Weighted low-rank approximation of general complex matrices and its application in the design of 2-d digital filters. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 44(7):650–655, 1997. 1

[23] J. H. Manton, R. Mehony, and Y. Hua. The geometry of weighted low-rank approximations. *IEEE Transactions on Signal Processing*, 51(2):500–514, 2003. 1

[24] I. Markovsky. Low-rank approximation: algorithms, implementation, applications, 2012. Springer. 1

[25] I. Markovsky, J. C. Willems, B. D. Moor, and S. V. Huffel. Exact and approximate modeling of linear systems: a behavioral approach, 2006. SIAM. 1, 2

[26] T. Okatani and K. Deguchi. On the Wiberg algorithm for matrix factorization in the presence of missing components. *International Journal of Computer Vision*, 72(3):329–337, 2007. 1

[27] N. Oliver, B. Rosario, and A. Pentland. A Bayesian computer vision system for modeling human interactions. In *International Conference on Computer Vision Systems*, pages 255–272, 1999. 2

[28] P. Rodriguez and B. Wohlberg. Incremental principal component pursuit for video background modeling. *Journal of Mathematical Imaging and Vision*, 55(1):1–18, 2016. 7

[29] P. Rodrguez and B. Wohlberg. A matlab implementation of a fast incremental principal component pursuit algorithm for video background modeling. In *IEEE International Conference on Image Processing*, pages 3414–3416, 2014. 7

[30] D. Shpak. A weighted-least-squares matrix decomposition with application to the design of 2-d digital filters. *Proceedings of IEEE 33rd Midwest Symposium on Circuits and Systems*, pages 1070–1073, 1990. 1

[31] A. Sobral and A. Vacavant. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 122:4–21, 2014. 1, 2

[32] N. Srebro and T. S. Jaakkola. Weighted low-rank approximations. *20th International Conference on Machine Learning*, pages 720–727, 2003. 1

[33] N. Srebro, J. D. M. Rennie, and T. S. Jaakkola. Maximum-margin matrix factorization. *In Proceedings of Advances in Neural Information Processing Systems*, 18:1329–1336, 2005. 1

[34] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintainance. *Seventh International Conference on Computer Vision*, pages 255–261, 1999. 6

[35] K. Usevich and I. Markovsky. Optimization on a grassmann manifold with application to system identification. *Automatica*, 50(6):1656–1662, 2014. 1

[36] K. Usevich and I. Markovsky. Variable projection methods for affinely structured low-rank approximation in weighted 2-norms. *Journal of Computational and Applied Mathematics*, 272:430–448, 2014. 1

[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transaction on Image Processing*, 13(4):600–612, 2004. 6

[38] T. Wiberg. Computation of principal components when data are missing. *In Proceedings of the Second Symposium of Computational Statistics*, pages 229–236, 1976. 1

[39] J. Wright, Y. Peng, Y. Ma, A. Ganseh, and S. Rao. Robust principal component analysis: exact recovery of corrputed low-rank matrices by convex optimization. *Proceedings of 22nd Advances in Neural Information Processing systems*, pages 2080–2088, 2009. 2, 4, 5

[40] B. Xin, Y. Tian, Y. Wang, and W. Gao. Background subtraction via generalized fused Lasso foreground modeling. *IEEE Computer Vision and Pattern Recognition*, pages 4676–4684, 2015. 2, 3, 6, 7