Iterative Re-weighted L_1-Norm Principal-Component Analysis[†]

Ying Liu*, Dimitris A. Pados[§], Stella N. Batalama[§], and Michael J. Medley[‡]

*Department of Electrical Engineering, The State University of New York at Buffalo, Buffalo, NY 14260

E-mail: yl72@buffalo.edu

[§]Department of Computer and Electrical Engineering and Computer Science & I-SENSE

Florida Atlantic University, Boca Raton, FL 33431

E-mail: {dpados, sbatalama}@fau.edu

[‡]Air Force Research Laboratory/RITF, Rome, NY 13441

E-mail: michael.medley@rl.af.mil

***Abstract*—We consider the problem of principal-component analysis of a given set of data samples. When the data set contains faulty measurements/outliers, the performance of classic L_2 principal-component analysis (L_2-PCA) deteriorates drastically. Instead, L_1 principal-component analysis (L_1-PCA) offers outlier resistance due to the L_1-norm maximization criterion it adopts to compute the principal subspace. In this work, we present an iterative re-weighted L_1-PCA method (IRW L_1-PCA) that generates a sequence of L_1-norm subspaces. In each iteration, the data set comformity of each sample is measured by the L_1 subspace calculated in the previous iteration and used to weigh the data sample before the L_1 subspace update. The approach automatically suppresses outliers in each iteration resulting in increasingly accurate subspace calculation. We provide convergence analysis and compare the proposed algorithm against benchmark algorithms in the literature. Experimental studies demonstrate the superiority of the proposed IRW L_1-PCA procedure.**

***Index Terms*—Faulty data, feature extraction, L_1-norm, robust principal component analysis, eigenvector decomposition, outliers.**

I. Introduction

Principal component analysis (PCA) is a prevalent method for dimensionality reduction and low-rank subspace approximation. Conventional L_2-norm-based principal component analysis (L_2-PCA), however, is easily affected by "outlier" values that are numerically distant from the nominal low-rank signal. To deal with the problem of outliers in principal-component design there has been a growing interest in robust PCA methods [1]-[7]. In [1]-[4], subspace decomposition is performed under an L_1-error minimization criterion. In [5], non-negative matrix factorization is performed via Manhattan distance minimization (MahNMF), which robustly estimates the low-rank part and the sparse part of a non-negative matrix. The robust PCA method (RPCA) developed in [6] performs low-rank sparse decomposition by minimizing a weighted sum of the nuclear-norm of the low-rank component and the L_1-norm of the sparse component. The GoDec algorithm developed in [7] performs low-rank and sparse decomposition as well, by alternatingly solving for the low-rank and sparse components. An accelerated method is proposed in [7] via bilateral random projection (BRP).

Recently, there has been a growing documented effort to calculate robust subspaces by explicit L_1 projection maximization [8]-[11]. The resulting principal components are called L_1 principal components. The work in [8] presented a suboptimal iterative algorithm for the computation of one L_1 principal component and [9] presented an iterative algorithm for suboptimal joint computation of $d \geq 1$ L_1 principal components. In [10], for the first time in the literature, algorithms for exact calculation of L_1 principal components are developed. Later, in [11] an approximate algorithm is developed for fast computation of the L_1 principal components. The L_1-PCA method has been successfully applied to a wide range of research fields such as direction of arrival (DoA) estimation [12] and robust face recognition [13], [14]. Most recently, compressed-sensed-domain L_1-PCA methods were developed for low-rank background scene and sparse foreground moving objects extraction from compressed-sensed surveillance video sequences [15], [16].

Nevertheless, existing L_1-PCA methods in [8]-[16] adopt "one-shot" processing. That is, for a given data set with potential outliers, the L_1-PCA algorithm is applied only once to compute from the explicit data the L_1 subspace. For severely contaminated data sets, such one-shot L_1 subspace computation can still be away from the true nominal signal subspace of interest.

In this paper, we propose an iterative re-weighted L_1-PCA method. Given a fixed data set that potentially contains outliers, the procedure iteratively generates a sequence of improved L_1 subspaces. In each iteration, nominal compliance of each sample is inferred by its position relative to the L_1 subspace calculated in the previous iteration and translated to a "weight." Samples with higher weights tend to be nominal samples and samples with lower weights are more likely to be

[†]This work was supported in part by the National Science Foundation under Grant ECCS-1462341 and AFRL under Grant FA8750-14-1-0074. Approved for Public Release; Distribution Unlimited: 88ABW-2017-3084, 26JUN17. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of AFRL or NSF.

the outliers. Weighted L_1-PCA calculation is then carried out in which the contribution of outlying samples in the data set is suppressed resulting in an improved L_1-subspace. The sample weights converge as the iteration number increases and the iterative algorithm terminates when the weights in the current and previous iteration are deemed close enough.

The remainder of this paper is organized as follows. In Section II, we introduce necessary background on regular L_1-PCA. In Section III, the proposed iterative re-weighted L_1-PCA algorithm is developed. In Section IV, experimental studies are provided to demonstrate the effectiveness of the proposed algorithm. Finally, a few conclusions are drawn in Section V.

II. Background on L_1-norm Principal-Component Analysis

Consider N real-valued samples $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$ of dimension D that form the $D \times N$ data matrix

$$\mathbf{X} = [\mathbf{x}_1 \;\; \mathbf{x}_2 \;\; \cdots \mathbf{x}_N]. \tag{1}$$

In the common version of the low-rank approximation problem (L_2-PCA) one seeks to describe (approximate) the data matrix $\mathbf{X}$ by a rank-r product $\mathbf{PQ}^{\mathrm{T}}$ where $\mathbf{P} \in \mathbb{R}^{D\times r}$, $\mathbf{Q} \in \mathbb{R}^{N\times r}$, $r \leq \min\{D, N\}$. Given the observation data matrix $\mathbf{X}$, L_2-PCA minimizes the sum of the element-wise squared error between the original matrix $\mathbf{X}$ and its rank-r representation $\mathbf{PQ}^T$ in the form of Problem $\mathcal{P}_1^{L_2}$ defined below,

$$\mathcal{P}_1^{L_2} : (\mathbf{P}_{L_2}, \mathbf{Q}_{L_2}) = \arg \min_{\substack{\mathbf{P}\in\mathbb{R}^{D\times r}, \mathbf{P}^T\mathbf{P}=\mathbf{I}_r \\ \mathbf{Q}\in\mathbb{R}^{N\times r}}} \|\mathbf{X} - \mathbf{PQ}^{\mathrm{T}}\|_2. \tag{2}$$

Problem $\mathcal{P}_1^{L_2}$ is equivalent to

$$\mathcal{P}_2^{L_2} : \mathbf{P}_{L_2} = \arg \max_{\substack{\mathbf{P}\in\mathbb{R}^{D\times r} \\ \mathbf{P}^T\mathbf{P}=\mathbf{I}_r}} \|\mathbf{X}^{\mathrm{T}}\mathbf{P}\|_2 \tag{3}$$

the solution of which is given by the r dominant singular-value left singular vectors of the original data matrix $\mathbf{X}$.

By minimizing the sum of squared errors, L_2 principal-component calculation becomes sensitive to extreme error value occurrences caused by the presence of outlying samples in the data matrix (samples that are numerically distant from the nominal data, appear only few times in the data matrix and are not to appear under normal system operation upon design). Motivated by this observed drawback of L_2-subspace signal processing, subspace decomposition approaches that are based on the L_1 norm were proposed for robust low-rank subspace computation. Replacing the L_2-norm in Problem $\mathcal{P}_2^{L_2}$ by L_1-norm, L_1-PCA calculates principal components in the form of

$$\mathcal{P}^{L_1} : \mathbf{P}_{L_1} = \arg \max_{\substack{\mathbf{P}\in\mathbb{R}^{D\times r} \\ \mathbf{P}^T\mathbf{P}=\mathbf{I}_r}} \|\mathbf{X}^{\mathrm{T}}\mathbf{P}\|_1. \tag{4}$$

$\mathbf{P}_{L_1}$ in (4) is likely to be closer to the true nominal rank-r subspace than L_2-PCA. The r columns of $\mathbf{P}_{L_1}$ in (4) are the so-called r L_1 principal components that describe the rank-r subspace in which $\mathbf{X}$ lies. As shown in [10], exact calculation of the L_1 principal components in Problem $\mathcal{P}^{L_1}$ can be recast as a combinatorial problem. In short, when the rank of the nominal signal is $r = 1$, Problem $\mathcal{P}^{L_1}$ reduces to

$$\mathbf{p}_{L_1} = \arg \max_{\substack{\mathbf{p}\in\mathbb{R}^{D} \\ \|\mathbf{p}\|_2=1}} \|\mathbf{X}^{\mathrm{T}}\mathbf{p}\|_1, \tag{5}$$

which can be reformulated as

$$\begin{aligned}
\max_{\substack{\mathbf{p}\in\mathbb{R}^{D} \\ \|\mathbf{p}\|_2=1}} \|\mathbf{X}^T\mathbf{p}\|_1 &= \max_{\substack{\mathbf{p}\in\mathbb{R}^{D} \\ \|\mathbf{p}\|_2=1}} \max_{\mathbf{b}\in\{\pm1\}^N} \mathbf{b}^T\mathbf{X}^{\mathrm{T}}\mathbf{p} && (6)\\
&= \max_{\mathbf{b}\in\{\pm1\}^N} \max_{\substack{\mathbf{p}\in\mathbb{R}^{D} \\ \|\mathbf{p}\|_2=1}} \mathbf{p}^{\mathrm{T}}\mathbf{X}\mathbf{b} && (7)\\
&= \max_{\mathbf{b}\in\{\pm1\}^N} \|\mathbf{X}\mathbf{b}\|_2. && (8)
\end{aligned}$$

The optimal solution for (8) can be obtained by exhaustive search in the space of the binary antipodal for example, vector $\mathbf{b}$ with complexity $\mathcal{O}(2^{N-1}DN)$.

When the rank of the nominal data is $r > 1$, problem $\mathcal{P}^{L_1}$ can be rewritten as [10]

$$\begin{aligned}
& \max_{\substack{\mathbf{P}\in\mathbb{R}^{D\times r} \\ \mathbf{P}^{\mathrm{T}}\mathbf{P}=\mathbf{I}_r}} \|\mathbf{X}^{\mathrm{T}}\mathbf{P}\|_1 && (9)\\
= & \max_{\substack{\mathbf{P}\in\mathbb{P}^{D\times r} \\ \mathbf{P}^{\mathrm{T}}\mathbf{P}=\mathbf{I}_r}} \max_{\mathbf{B}\in\{\pm1\}^{N\times r}} \mathrm{tr}(\mathbf{P}^{\mathrm{T}}\mathbf{X}\mathbf{B}) && (10)\\
= & \max_{\mathbf{B}\in\{\pm1\}^{N\times r}} \|\mathbf{X}\mathbf{B}\|_* && (11)
\end{aligned}$$

where $\|\cdot\|_*$ stands for nuclear norm. To find exactly the optimal L_1-norm projection operator $\mathbf{P}_{L_1}$ in (9) we can perform the following steps [10].

1) Solve (11) to obtain $\mathbf{B}_{\mathrm{opt}}$.
2) Perform singular value decomposition (SVD) on $\mathbf{XB}_{\mathrm{opt}} = \mathbf{U\Sigma V}^{\mathrm{T}}$.
3) Return $\mathbf{P}_{L_1} = \mathbf{U}_{:,1:r}\mathbf{V}^{\mathrm{T}}$.

If we solve (11) by exhaustive search, the overall complexity of the above procedure for finding r L_1 principal components will be $\mathcal{O}(2^{Nr}\min\{D^2r, Dr^2\})$. For any fixed data dimension D, a polynomial-time algorithm is developed in [10] to solve optimally (11) with complexity $\mathcal{O}(N^{\mathrm{rank}(\mathbf{X})r-r+1})$, $\mathrm{rank}(\mathbf{X}) \leq D$. In [11], a fast approximation algorithm was proposed to solve (11) with complexity $\mathcal{O}(\min\{ND^2, N^2D\} + N^2(r+2) + ND)$.

III. Proposed Iterative Re-weighted L_1 Principal Component Analysis

The regular L_1-PCA problem in (4) seeks a rank-r subspace from the data matrix $\mathbf{X} \in \mathbb{R}^{D\times N}$ by one-shot calculation. Although the adopted L_1-norm maximization is less affected by outliers compared to L_2-norm maximization in L_2-PCA in (3), the produced L_1 subspace $\mathbf{P}_{L_1}$ can still be away from the true nominal signal low-rank subspace. In this section, we propose an iterative method that generates a sequence of improved L_1 subspaces for the same data matrix $\mathbf{X}$.

A. Algorithm

We consider the calculation of r principal components $\mathbf{P}_{L_1} \in \mathbb{R}^{D\times r}$, $D > r > 1$. Initially, the direct L_1 subspace is computed via (4) and denoted by $\mathbf{P}_{L_1}^{(0)}$. Next, the distance of each sample $\mathbf{x}_n$ from subspace $\mathbf{P}_{L_1}^{(0)}$ is defined as the L_2 error between $\mathbf{x}_n$ and its rank-r surrogate

$$d_n^{(1)} = \|\mathbf{x}_n - \mathbf{P}_{L_1}^{(0)}\mathbf{P}_{L_1}^{(0)^{\mathrm{T}}}\mathbf{x}_n\|_2, \quad n = 1, ..., N. \tag{12}$$

We expect large $d_n^{(1)}$ if $\mathbf{x}_n$ is an "outlier" and small $d_n^{(1)}$ if $\mathbf{x}_n$ is a nominal sample. Therefore, the nominal-likeness (weight) of each sample can be measured as the reciprocal of its L_2 distance from the subspace, i.e.,

$$w_n^{(1)} = (d_n^{(1)})^{-1}, \quad n = 1, ..., N, \tag{13}$$

followed by normalization,

$$\widetilde{w}_n^{(1)} = \frac{w_n^{(1)}}{\sum_{n=1}^{N} w_n^{(1)}}, \quad n = 1, ..., N. \tag{14}$$

When computing the L_1 subspace, data samples with larger nominal-likeness (weight) should contribute more and samples with smaller nominal-likeness (weight) should be suppressed such that the resulting calculated L_1 subspace is more accurate. In this direction, we propose that each data sample $\mathbf{x}_n$ is weighed by $\widetilde{w}_n^{(1)}$. We form a weight matrix

$$\widetilde{\mathbf{W}}^{(1)} = \begin{bmatrix} \widetilde{w}_1^{(1)} & 0 & 0 & \cdots \\ 0 & \widetilde{w}_2^{(1)} & 0 & \cdots \\ & \vdots & & \\ 0 & 0 & \cdots & \widetilde{w}_N^{(1)} \end{bmatrix} \tag{15}$$

and update the L_1 subspace by

$$\mathbf{P}_{L_1}^{(1)} = \arg \max_{\substack{P\in\mathbb{R}^{D\times r} \\ \mathbf{P}^T\mathbf{P}=\mathbf{I}_r}} \|(\mathbf{X}\widetilde{\mathbf{W}}^{(1)})^{\mathrm{T}}\mathbf{P}\|_1. \tag{16}$$

Generalizing, in the $(k + 1)$th iteration new weights are computed using the L_1 subspace $\mathbf{P}_{L_1}^{(k)}$ computed at the kth iteration, i.e.

$$\begin{aligned} d_n^{(k+1)} &= \|\mathbf{x}_n - \mathbf{P}_{L_1}^{(k)}\mathbf{P}_{L_1}^{(k)^{\mathrm{T}}}\mathbf{x}_n\|_2, 1 \le n \le N, && (17) \\ w_n^{(k+1)} &= (d_n^{(k+1)})^{-1}, && (18) \\ \widetilde{w}_n^{(k+1)} &= \frac{w_n^{(k+1)}}{\sum_{n=1}^{N} w_n^{(k+1)}}, && (19) \\ \widetilde{\mathbf{W}}^{(k+1)} &= \begin{bmatrix} \widetilde{w}_1^{(k+1)} & 0 & 0 & \cdots \\ 0 & \widetilde{w}_2^{(k+1)} & 0 & \cdots \\ & \vdots & & \\ 0 & 0 & \cdots & \widetilde{w}_N^{(k+1)} \end{bmatrix}. && (20) \end{aligned}$$

Subsequently, the L_1 subspace at the $(k + 1)$th iteration is updated to

$$\mathbf{P}_{L_1}^{(k+1)} = \arg \max_{\substack{P\in\mathbb{R}^{D\times r} \\ \mathbf{P}^T\mathbf{P}=\mathbf{I}_r}} \|(\mathbf{X}\widetilde{\mathbf{W}}^{(k+1)})^{\mathrm{T}}\mathbf{P}\|_1. \tag{21}$$

B. Convergence Analysis

To guarantee a convergent weight sequence for practical algorithmic implementation, we modify the weight update formula as follows. In the $(k+1)$th iteration, we first compute the ℓ_2 error (distance) for each sample as in (17). Then, we define

$$u_n^{(k+1)} = (d_n^{(k+1)})^{-1} \tag{22}$$

and update the weight based on $u_n^{(k+1)}$ by

$$w_n^{(k+1)} = \begin{cases} w_n^{(k)}(1-\beta^k), \text{ if } u_n^{(k+1)} < w_n^{(k)}(1-\beta^k), \\ u_n^{(k+1)}, \text{ if } w_n^{(k)}(1-\beta^k) \le u_n^{(k+1)} \le w_n^{(k)}(1+\beta^k), \\ w_n^{(k)}(1+\beta^k), \text{ if } u_n^{(k+1)} > w_n^{(k)}(1+\beta^k) \end{cases}$$

where $0 < \beta < 1$ is a pre-defined parameter. Intuitively, we avoid updating the weights too aggressively by restricting the new weight $w_n^{(k+1)}$ to be within a small neighborhood of the weight in the previous iteration $w_n^{(k)}$. The size of the neighborhood depends on β. Subsequently, $w_n^{(k+1)}$ is normalized as in (19), followed by weight matrix construction in (20). The convergence of the weight sequence can be verified by

$$\begin{aligned} \lim_{k\to\infty} \beta^k &= 0, && (23) \\ \lim_{k\to\infty} (w_n^{(k+1)} - w_n^{(k)}) &= 0, && (24) \\ \lim_{k\to\infty} (\widetilde{w}_n^{(k+1)} - \widetilde{w}_n^{(k)}) &= 0. && (25) \end{aligned}$$

C. Stopping Criterion

In implementing the proposed iterative algorithm, we exit the algorithm when the difference between the weight vectors at the kth and $(k + 1)$th iteration is smaller than a predefined threshold $\epsilon > 0$, that is,

$$\|\mathbf{w}^{(k+1)} - \mathbf{w}^{(k)}\| < \epsilon, \tag{26}$$

where $\mathbf{w}^{(k)} = [w_1^{(k)}, w_2^{(k)}, \cdots, w_N^{(k)}]^T$ and $\mathbf{w}^{(k+1)} = [w_1^{(k+1)}, w_2^{(k+1)}, \cdots, w_N^{(k+1)}]^T$.

IV. Applications and Experimental Studies

In this section, we assess the effectiveness of the proposed iterative re-weighted L_1-PCA (IRW L_1-PCA) algorithm through two experiments: (i) Dimensionality reduction of a 2-dimensional Gaussian data set (artificial data), and (ii) video surveillance (field data foreground extraction).

A. Dimensionality Reduction

We generate a nominal data set $\mathbf{X}_{D\times N}$ of $N = 30$ two-dimensional ($D = 2$) observation points drawn from the Gaussian distribution $\mathcal{N}\left(\mathbf{0}_2, \begin{bmatrix} 10.5 & 13 \\ 13 & 30 \end{bmatrix}\right)$ as shown in Fig. 1. We assume that our data matrix is corrupted by four additional outlier measurements, $\mathbf{o}_1, \mathbf{o}_2, \mathbf{o}_3, \mathbf{o}_4$, shown in the bottom right corner of Fig. 1. For the corrupted data matrix $\mathbf{X}^{\mathrm{CRPT}} = [\mathbf{X}, \mathbf{o}_1, \mathbf{o}_2, \mathbf{o}_3, \mathbf{o}_4]$, we calculate and plot in Fig. 1 the rank-1 subspace by L_2-PCA, regular L_1-PCA, as well as the proposed IRW L_1-PCA method with number of iterations $k = 1, 4, 8, 15, 21$ ($\beta = 0.9$). For reference purposes, we also plot the true nominal-data maximum-variance direction, i.e., the dominant eigenvector of the covariance matrix $\begin{bmatrix} 10.5 & 13 \\ 13 & 30 \end{bmatrix}$. We observe that the proposed IRW L_1-PCA approach offers better estimation of the principal component than the L_2 and regular L_1-PCA methods [10]. As the number of iterations increases from $k = 1$ to $k = 21$, the IRW L_1 principal component comes closer to the true rank-1 subspace. The algorithm converges empirically at $k = 21$. To quantify the impact of the outliers, we generate 1000 new independent evaluation data points from $\mathcal{N}\left(\mathbf{0}_2, \begin{bmatrix} 10.5 & 13 \\ 13 & 30 \end{bmatrix}\right)$. In Fig. 2, we estimate the mean-square-fit-error (MSFE) $\mathbf{E}\{\|\mathbf{x} - \mathbf{p}\mathbf{p}^{\mathrm{T}}\mathbf{x}\|_2^2\}$ by $\frac{1}{1000}\sum_{i=1}^{1000} \|\mathbf{x}_i - \mathbf{p}\mathbf{p}^{\mathrm{T}}\mathbf{x}_i\|_2^2$ for $\mathbf{p}_{L_2}(\mathbf{X}^{\mathrm{CRPT}})$, $\mathbf{p}_{L_1}(\mathbf{X}^{\mathrm{CRPT}})$, and IRW $\mathbf{p}_{L_1}(\mathbf{X}^{\mathrm{CRPT}})$. Again, for reference purposes, we plot the MSFE curve for the true nominal-data maximum-variance direction. We observe that the MSFE value of the proposed IRW L_1 component decreases rapidly as the iteration number increases and converges toward the minimum at $k = 21$.

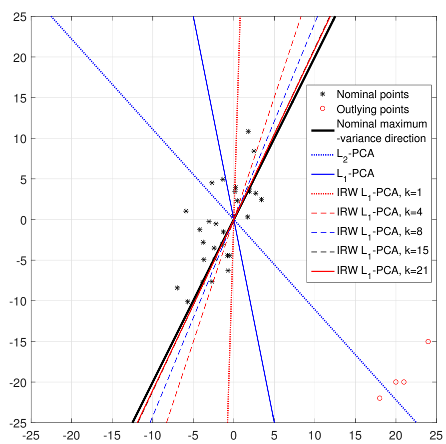

Fig. 1. Training data matrix $\mathbf{X}_{2\times 30}$ corrupted by four outlier points in bottom right with calculated rank-1 L_2, regular L_1, and IRW L_1 principal components at iteration $k = 1, 4, 8, 15$, and 21.

B. Video Surveillance

Consider a sequence of surveillance video frames $\mathbf{X}_t \in \mathbb{R}^{m\times n}$ with frame resolution of $m \times n$ pixels and time index $t = 1, ..., N$. For a surveillance video sequence, the background scene is usually static and the objective is to extract foreground moving objects. In our experiment, we perform block-by-block IRW L_1-PCA for low-rank background modeling and foreground extraction. We divide each frame $\mathbf{X}_t$ into J blocks $\mathbf{X}_t^j \in \mathbb{R}^{m_b\times n_b}$, $j = 1, ..., J$. We let $\mathbf{x}_t^j \in \mathbb{R}^D$, $D = m_b n_b$, represent vectorization of $\mathbf{X}_t^j$ via column concatenation. For each sequence of co-located blocks, $\mathbf{x}_t^j$, $t = 1, ..., N$, it is likely that the moving objects appear only in a few of these blocks, therefore we can model the

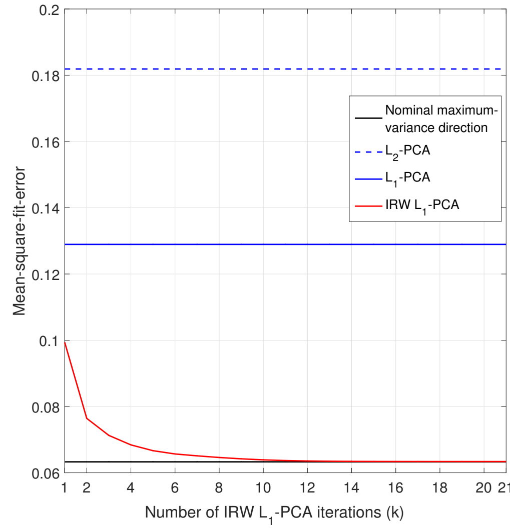

Fig. 2. Mean square-fit-error of L_2, regular L_1, and IRW L_1 principal components calculated from corrupted training data set $\mathbf{X}^{\mathrm{CRPT}}$ versus iteration index k.

static background scene as a low-rank component $\mathbf{z}_t^j$ and the foreground moving objects as an outlying component $\mathbf{s}_t^j$. That is,

$$\mathbf{x}_t^j = \mathbf{z}_t^j + \mathbf{s}_t^j, \quad t = 1, ..., N. \tag{27}$$

In matrix form representation of the jth block across N frames, $\mathbf{X}^j \triangleq [\mathbf{x}_1^j, ..., \mathbf{x}_N^j] \in \mathbb{R}^{D\times N}$ and

$$\mathbf{X}^j = \mathbf{Z}^j + \mathbf{S}^j. \tag{28}$$

To extract the low-rank background information, we carry out IRW L_1-PCA on $\mathbf{X}^j$ and obtain the rank-1 L_1 subspace $\mathbf{p}_{L_1}^j$ at convergence. Afterwards, the background blocks can be approximated by $\widehat{\mathbf{Z}}^j = \mathbf{p}_{L_1}^j \mathbf{p}_{L_1}^{j^T} \mathbf{X}^j$ and the foreground blocks can be extracted as $\widehat{\mathbf{S}}^j = \mathbf{X}^j - \widehat{\mathbf{Z}}^j$, $j = 1, ..., J$.

We test the method on the *Airport* video sequence with 96 frames, each of 144×176 pixels. We process $N = 8$ successive frames at a time. To mitigate the "blockiness" artifact, we divide each frame into $J = 370$ overlapping blocks of size 26×32 and apply the proposed IRW L_1-PCA method independently to each group of co-located blocks across 8 frames. The final background and foreground scenes are obtained by averaging the extracted background pixels (as well as the foreground pixels) for which multiple results are available.

Fig. 3 displays the background and foreground extracted at multiple distinct time slots $t = 5, 7, 66, 67$ with $r = 1$ principal component by the proposed IRW L_1-PCA, regular L_1-PCA [10], and the robust PCA method of [6]. The results show that both the regular L_1-PCA and robust PCA method suffer from severe "ghost" presence in the estimated background scene, which results in problematic foreground extraction. In contrast, IRW L_1-PCA significantly mitigates the "ghost" effect in the estimated background and offers a much clearer foreground scene.

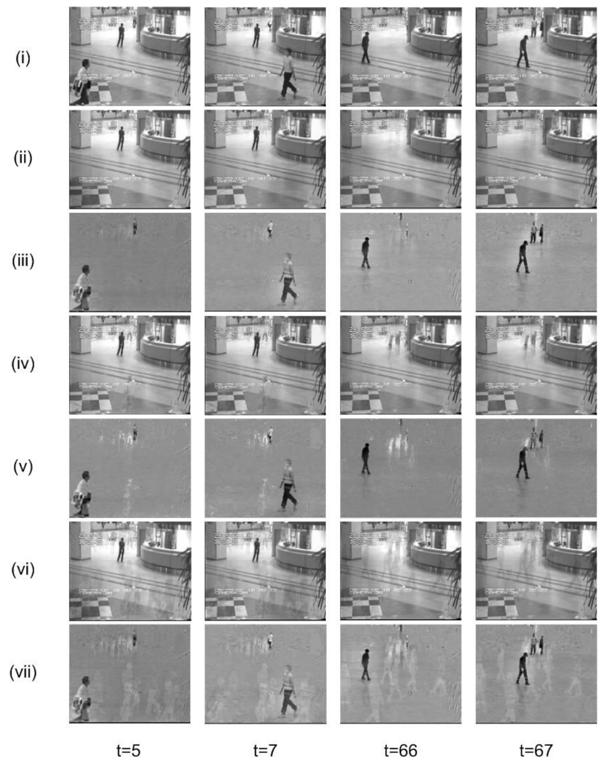

Fig. 3. *Airport* sequence: Original frame [row (i)] of time slot $t = 5, 7, 66,$ and 67; proposed IRW L_1-PCA reconstructed background and moving objects [rows (ii) and (iii)]; robust PCA [6] reconstructed background and moving objects [rows (iv) and (v)]; regular L_1-PCA [10] reconstructed background and moving objects [rows (vi) and (vii)].

V. Conclusion

In this work, we proposed an iterative re-weighted L_1 principal-component analysis algorithm to compute principal subspaces from data sets that may contain outliers. Instead of computing a "one-shot" L_1 subspace, the proposed procedure iteratively computes a sequence of L_1 subspaces. In every iteration, each data sample is weighed according to compliance to nominal data behavior measured by the L_1 subspace computed in the previous iteration. We evaluated the effectiveness of the proposed IRW L_1-PCA method experimentally and the results showed significantly better performance than regular L_1-PCA and state-of-the-art robust PCA methods.

References

[1] Q. Ke and T. Kanade, "Robust subspace computation using L_1 norm," *Technical Report CMU-CS-03-172*, Carnegie Mellon Univ., http://ra.adm.cs.cmu.edu/anon/usr0/ftp/usr/anon/2003/CMU-CS-03-172.pdf, Aug. 2003.

[2] Q. Ke and T. Kanade, "Robust norm factorization in the presence of outliers and missing data by alternative convex programming," in *Proc. IEEE Conf. Comput. Vision Pattern Recog. (CVPR)*, San Diego, CA, Jun. 2005, pp. 739-746.

[3] A. Eriksson and A. v. d. Hengel, "Efficient computation of robust low-rank matrix approximations in the presence of missing data using the L_1 norm," in *Proc. IEEE Conf. Comput. Vision Pattern Recog. (CVPR)*, San Francisco, CA, Jun. 2010, pp. 771-778.

[4] L. Yu, M. Zhang, and C. Ding, "An efficient algorithm for L_1-norm principal component analysis," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 1377-1380.

[5] N. Guan, D. Tao, Z. Luo, and J. S. Taylor, "MahNMF: Manhattan non-negative matrix factorization," *arXiv:1207.3438*, July 2012.

[6] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, pp. 1-37, May 2011.

[7] T. Zhou and D. Tao, "GoDec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. Int. Conf. Machine Learning*, Bellevue, WA, 2011.

[8] N. Kwak, "Principal component analysis based on L_1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, pp. 1672-1680, Sept. 2008.

[9] F. Nie, H. Huang, C. Ding, D. Luo, and H. Wang, "Robust principal component analysis with non-greedy ℓ_1-norm maximization," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, Barcelona, Spain, July 2011, pp. 1433-1438.

[10] P. P. Markopoulos, G. N. Karystinos, and D. A. Pados, "Optimal algorithms for L_1-subspace signal processing," *IEEE Trans. Signal Process.*, vol. 62, pp. 5046-5058, Oct. 2014.

[11] S. Kundu, P. P. Markopoulos and D. A. Pados, "Fast computation of the L_1-principal component of real-valued data," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP)*, Florence, Italy, May 2014, pp. 8028-8032.

[12] P. P. Markopoulos, N. Tsagkarakis, D. A. Pados, and G. N. Karystinos, "Direction finding with L_1-norm subspaces," *SPIE Compressive Sensing Conference, Defense, Security, and Sensing (SPIE DSS 2014)*, Baltimore, MD, May 2014, vol. 9109, pp. 0J1-0J11.

[13] M. Johnson and A. Savakis, "Fast L_1-eigenfaces for robust face recognition," in *Proc. IEEE Western NY Image and Signal Process. Workshop*, Rochester, NY, Nov. 2014, pp. 1-5.

[14] F. Maritato, Y. Liu, D. A. Pados, and S. Colonnese, "Face recognition with L_1-norm subspaces," in *Proc. SPIE Commercial + Scientific Sensing and Imaging*, Baltimore, MD, Apr. 2016.

[15] Y. Liu and D. A. Pados, "Compress-sensed-domain L_1-PCA video surveillance," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 351-363, Mar. 2016.

[16] M. Pierantozzi, Y. Liu, D. A. Pados, and S. Colonnese, "Video background tracking and foreground extraction via L_1-subspace updates," in *Proc. SPIE Commercial + Scientific Sensing and Imaging*, Baltimore, MD, Apr. 2016.