Compressed-Sensed-Domain
L_1-PCA Video Surveillance

Ying Liu, *Member, IEEE*, and Dimitris A. Pados, *Senior Member, IEEE*

Abstract—We consider the problem of foreground and background extraction from compressed-sensed (CS) surveillance videos that are captured by a static CS camera. We propose, for the first time in the literature, a principal component analysis (PCA) approach that computes directly in the CS domain the low-rank subspace of the background scene. Rather than computing the conventional L_2-norm-based principal components, which are simply the dominant left singular vectors of the CS-domain data matrix, we compute the principal components under an L_1-norm maximization criterion. The background scene is then obtained by projecting the CS measurement vector onto the L_1 principal components followed by total-variation (TV) minimization image recovery. The proposed L_1-norm procedure directly carries out low-rank background representation without reconstructing the video sequence and, at the same time, exhibits significant robustness against outliers in CS measurements compared to L_2-norm PCA. An adaptive CS-L_1-PCA method is also developed for low-latency video surveillance. Extensive experimental studies described in this paper illustrate and support the theoretical developments.

Index Terms—Background and foreground extraction, compressed sensing, compressive sampling, convex optimization, feature extraction, L_1 principle component analysis, singular value decomposition, total-variation minimization, video surveillance.

I. INTRODUCTION

IN VIDEO surveillance, video signals are captured by cameras and transmitted to a processing center where video streams are monitored and analyzed for moving objects and/or other anomalies. Since conventional video coding requires high processing power, compressed-sensing (CS) based video streaming is attracting significant interest to reduce the required computational complexity and energy consumption. CS theory deals with sub-Nyquist-rate sampling of sparse signals of interest [1]–[3]. Rather than collecting an entire Nyquist ensemble of signal samples, CS performs signal acquisition by a small number of (random [3] or deterministic [4], for example) linear measurements. Successful signal reconstruction relies on effective sparse signal representation and appropriate recovery algorithms such as convex optimization [5], linear regression [6],

[7], or greedy procedures [8]. Wireless video surveillance via compressed sensing can capture and compress video signals simultaneously through simple linear operations, therefore highly reducing data acquisition time and power consumption [9].

In video surveillance, of particular interest is the ability to detect anomalies or moving objects that stand out from the background. To tackle this problem, the usual approach is via background subtraction. Basic non-statistical background modeling methods estimate and update the background by running average [10] or temporal median filtering [11] and classify a new pixel as foreground if the distance between the estimated background and the new pixel is above a predefined threshold. Statistical approaches model each background pixel by a probability density function (pdf) learned over a set of training frames, such as running Gaussian average [12]. To account for background containing animated textures (such as sea waves or trees shaken by the wind), multimodel pdfs have been proposed, for instance, the Gaussian mixture model (GMM) [13], [14] or kernel density estimation (KDE) [15] that models the pdf of a background pixel by the sum of Gaussian kernels centered at the most recent n background values (with n in the order of 100). A new pixel is then classified as foreground if its pdf value falls below a threshold. Since each Gaussian kernel describes just one sample point, KDE has very high computational complexity. Another method for multimodel pdf estimation is mean-shift [16]–[18], which is an effective gradient-ascent technique able to detect the main modes of the true pdf directly from the sample data. Due to its iterative nature, mean-shift also has high computational complexity and is not immediately applicable to modeling background pdfs at the pixel level. Sequential KDE [19] reduces complexity by initializing the modes of the background pdf with a mean-shift procedure and using mode propagation to update the modes.

In addition to the above classical background subtraction techniques, more sophisticated algorithms have been developed in recent years for challenging video surveillance scenarios [20]–[25]. In particular, the self-organizing background subtraction (SOBS) [20], multilayer codebook-based background subtraction (MCBS) [22], and multibackground generation [23] schemes are designed to accommodate dynamic scenes such as moving backgrounds and gradual illumination variations. Besides, algorithms developed in [24], [25] generate effective background models for variable bit rate video-based traffic monitoring systems.

Another line of research on background subtraction for video surveillance is based on low-rank subspace approximation. In contrast to the aforementioned pixel-level background modeling schemes, such schemes are block-level or image-level

and extensively explore spatial correlations. The standard approach is L_2-norm based principal component analysis (L_2-PCA), such as block-level one-dimensional PCA [26] or frame-level two-directional two-dimensional PCA $((2D)^2\text{PCA})$ [27]. L_2-PCA, in general, seeks a low-rank subspace to represent the background scene, but is easily affected by moving objects in the foreground scene, i.e., "outliers" that are numerically distant from the background. The reason is that L_2-PCA is calculated based on squared-error metrics that can give highly exaggerated values due to the squaring operation. In recent years, there has been a growing interest in robust PCA methods to deal with the problem of outliers in principal-component design [28]–[35]. In [28]–[31], subspace decomposition is performed under an L_1-error minimization criterion. In [32], non-negative matrix factorization is performed via Manhattan distance minimization (MahNMF), which aims at robustly estimating the low-rank part and sparse part of a non-negative matrix in the presence of outliers. The robust PCA method developed in [33] performs low-rank background and sparse foreground decomposition by minimizing a weighted sum of the nuclear-norm of the low-rank component and the ℓ_1-norm of the sparse component. Such a robust PCA idea is also adopted in DECOLOR [34], which in addition uses Markov random-field (MRF) modeling to improve the accuracy of detecting contiguous outliers. A CS version of robust PCA (CS-RPCA) is developed in [35] for CS surveillance videos, but offers good reconstruction quality only when a large number of frames is available (introducing, therefore, large latency to the decoding monitoring system). Overall, the robustness of all aforementioned methods comes at high computational cost due to the employed convex optimization programs.

Recently, there has been a growing documented effort to calculate robust subspace components by explicit L_1 projection maximization [36]–[38]. The resulting principal components are called L_1 principal components. The work in [36] presented a suboptimal iterative algorithm for the computation of one L_1 principal component and [37] presented an iterative algorithm for suboptimal joint computation of $d \geq 1$ L_1 principal components. In [38], for the first time in the literature, algorithms for exact calculation of L_1 principal components are developed.

While the algorithms for optimal L_1-PCA in [38] find successful applications in the original signal space, in this work we propose for the first time a direct CS-measurement-domain L_1-PCA algorithm and apply the procedure to compressed-sensed surveillance video processing. For a surveillance video sequence, the low-rank property is preserved in the CS domain if each frame of the video is captured by the same compressed-sensing matrix. Hence, L_1-PCA can be performed directly on the collected CS measurement vectors. Since the CS measurement vectors lie in a reduced dimensional space compared to the original pixel-domain data, the computational complexity for CS-L_1-PCA is dramatically lower. In the experimental studies that we present in this paper, we not only demonstrate that CS-L_1-PCA followed by regular CS image recovery can successfully extract the background scene from a surveillance video, but also illustrate the advantages of CS-L_1-PCA over CS-L_2-PCA when CS measurements are corrupted by outliers/faulty data. Furthermore, we develop an adaptive method based on the pro-

posed CS-L_1-PCA scheme which utilizes the most recently extracted background information to update the low-rank subspace with only a few new frames, therefore enabling low-latency video surveillance.

The remainder of this paper is organized as follows. In Section II, we present our notation and establish the building blocks of our proposed procedure; that is, exact computation of L_1-PCs and compressed-sensed image recovery based on total-variation (TV) minimization. In Section III, the proposed CS-L_1-PCA algorithm is developed and the overall foreground and background separation scheme is described in detail. Extensive experimental results and performance analysis are presented in Section IV. Finally, a few conclusions are drawn in Section V.

II. BUILDING BLOCKS OF PROPOSED ALGORITHM

A. Exact Computation of L_1 Principal Components

Consider an observation data matrix $\mathbf{X} \in \mathbb{R}^{D \times N}$ that consists of a low-rank component $\mathbf{L} \in \mathbb{R}^{D \times N}$ and a perturbation matrix $\mathbf{E} \in \mathbb{R}^{D \times N}$, i.e.

$$\mathbf{X} = \mathbf{L} + \mathbf{E}. \tag{1}$$

L_2-PCA refers to the problem of seeking the best rank-d ($d \leq \min\{D, N\}$) representation of $\mathbf{L}$ by solving

$$\mathcal{P}_1^{L_2} : (\mathbf{R}_{L_2}, \mathbf{S}_{L_2}) = \arg \min_{\substack{\mathbf{R} \in \mathbb{R}^{D \times d}, \mathbf{R}^T\mathbf{R}=\mathbf{I}_d \\ \mathbf{S} \in \mathbb{R}^{N \times d}}} \|\mathbf{X} - \mathbf{R}\mathbf{S}^T\|_2 \tag{2}$$

which is equivalent to the L_2 projection (energy) maximization problem

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

The optimal $\mathbf{R}_{L_2}$ solution (same in both $\mathcal{P}_1^{L_2}$ and $\mathcal{P}_2^{L_2}$) is simply the d dominant-singular-value left singular vectors of the data matrix $\mathbf{X}$.

When the perturbation matrix $\mathbf{E}$ may contain extreme outlier values (faulty measurements in $\mathbf{X}$), L_1-PCA in the following form pursues a more robust subspace representation of $\mathbf{L}$ than L_2-PCA

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

Indeed, in the presence of faulty measurements/outliers in $\mathbf{X}$, $\mathbf{R}_{L_1}$ in (4) is likely to be closer to the correct true subspace than $\mathbf{R}_{L_2}$ in (3). The d columns of $\mathbf{R}_{L_1}$ in (4) are the so-called d L_1 principal components that describe the rank-d subspace in which $\mathbf{L}$ lies. As shown in [38], 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 $d = 1$, Problem $\mathcal{P}^{L_1}$ reduces to

$$\mathcal{P}^{L_1} : \mathbf{r}_{L_1} = \arg \max_{\substack{\mathbf{r} \in \mathbb{R}^D \\ \|\mathbf{r}\|_2=1}} \|\mathbf{X}^T\mathbf{r}\|_1 \tag{5}$$

and we can rewrite $\mathcal{P}^{L_1}$ as

$$\max_{\substack{\mathbf{r} \in \mathbb{R}^D \\ \|\mathbf{r}\|_2=1}} \|\mathbf{X}^T \mathbf{r}\|_1 = \max_{\substack{\mathbf{r} \in \mathbb{R}^D \\ \|\mathbf{r}\|_2=1}} \max_{\mathbf{b} \in \{\pm 1\}^N} \mathbf{b}^T \mathbf{X}^T \mathbf{r} \tag{6}$$

$$= \max_{\mathbf{b} \in \{\pm 1\}^N} \max_{\substack{\mathbf{r} \in \mathbb{R}^D \\ \|\mathbf{r}\|_2=1}} \mathbf{r}^T \mathbf{X}\mathbf{b} \tag{7}$$

$$= \max_{\mathbf{b} \in \{\pm 1\}^N} \|\mathbf{X}\mathbf{b}\|_2. \tag{8}$$

The optimal solution for (8) can be obtained by exhaustive search in the space of the binary antipodal vector $\mathbf{b}$. Since $\|\mathbf{X}\mathbf{b}\|_2 = (\mathbf{b}^T \mathbf{X}^T \mathbf{X}\mathbf{b})^{1/2} = (-\mathbf{b}^T \mathbf{X}^T \mathbf{X}(-\mathbf{b}))^{1/2}$, if $\mathbf{b}^*$ is an optimal solution, then $-\mathbf{b}^*$ is also an optimal solution. Hence, we can always set, the first, say, coordinate of $\mathbf{b}$ to $\mathbf{b}(1) = 1$ and the complexity for exhaustive search on the other $N-1$ coordinates of $\mathbf{b}$ is 2^{N-1}. Computation of $\|\mathbf{X}\mathbf{b}\|_2$ needs DN multiplications, therefore the overall complexity for solving $\mathcal{P}^{L_1}$ is $2^{N-1}DN$.

When the rank of the nominal signal is $d > 1$, the problem $\mathcal{P}^{L_1}$ can be solved by

$$\max_{\substack{\mathbf{R} \in \mathbb{R}^{D \times d} \\ \mathbf{R}^T \mathbf{R} = I_d}} \|\mathbf{X}^T \mathbf{R}\|_1 \tag{9}$$

$$= \max_{\substack{\mathbf{R} \in \mathbb{R}^{D \times d} \\ \mathbf{R}^T \mathbf{R} = I_d}} \max_{\mathbf{B} \in \{\pm 1\}^{N \times d}} \mathrm{tr}(\mathbf{R}^T \mathbf{X}\mathbf{B}) \tag{10}$$

$$= \max_{\mathbf{B} \in \{\pm 1\}^{N \times d}} \|\mathbf{X}\mathbf{B}\|_* \tag{11}$$

where $\|\cdot\|_*$ stands for nuclear norm. By Proposition 4 of [38], to find exactly the optimal L_1-norm projection operator $\mathbf{R}_{L_1}$ in (4) we can perform the following steps:

1) solve (11) to obtain $\mathbf{B}_{\mathrm{opt}}$;
2) perform singular value decomposition (SVD) on $\mathbf{X}\mathbf{B}_{\mathrm{opt}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$; and
3) return $\mathbf{R}_{L_1} = \mathbf{U}_{:,1:d}\mathbf{V}^T$.

The complexity of the above algorithm is dominated by Step 1, which includes the exhaustive search on the binary matrix $\mathbf{B}^{N \times d}$ with complexity $\mathcal{O}(2^{Nd})$ and SVD per iteration of complexity $\mathcal{O}(\min\{D^2 d, Dd^2\})$. Therefore, the overall complexity for finding d L_1 principal components via exhaustive search is $\mathcal{O}(2^{Nd} \min\{D^2 d, Dd^2\})$. For any fixed data dimension D, a polynomial-time algorithm is developed in [38] to solve optimally (11) with complexity $\mathcal{O}(N^{\mathrm{rank}(\mathbf{X})d-d+1}), \mathrm{rank}(\mathbf{X}) \leq D$. In [39], a fast greedy approximation algorithm was proposed to solve (11) with complexity $\mathcal{O}(\min\{ND^2, N^2 D\} + N^2(d + 2) + ND)$.

B. Compressed-Sensed Video Recovery via Total-Variation (TV) Minimization

In this section, we briefly review video frame acquisition by compressive sampling and recovery using sparse gradient constraints (TV minimization). If we consider the tth frame $\mathbf{X}_t \in \mathbb{R}^{m \times n}$ of the video sequence and $\mathbf{x}_t \in \mathbb{R}^D, D = mn$, represents vectorization of $\mathbf{X}_t$ via column concatenation, then CS measurements of $\mathbf{X}_t$ are collected in the form of

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{x}_t, \quad t = 1, 2, \ldots \tag{12}$$

with a linear measurement matrix $\mathbf{\Phi}_{P \times D}, P \ll D$. Under the assumption that images are mostly piece-wise smooth, it is natural to consider utilizing the sparsity of the spatial gradient of $\mathbf{X}_t$ for CS frame reconstruction [5], [40]–[46]. If $x_{i,j}^t$ denotes the pixel in the ith row and jth column of $\mathbf{X}_t$, the horizontal and vertical gradients at $x_{i,j}^t$ are defined, respectively, as

$$D_{h;\mathrm{ij}}[\mathbf{X}_t] = \begin{cases} x_{i,j+1}^t - x_{i,j}^t, & j < n \\ 0, & j = n \end{cases}$$

and

$$D_{v;\mathrm{ij}}[\mathbf{X}_t] = \begin{cases} x_{i+1,j}^t - x_{i,j}^t, & i < m \\ 0, & i = m \end{cases}.$$

The discrete spatial gradient of $\mathbf{X}_t$ at pixel $x_{i,j}^t$ can be interpreted as the 2D vector

$$D_{\mathrm{ij}}[\mathbf{X}_t] = \begin{pmatrix} D_{h;\mathrm{ij}}[\mathbf{X}_t] \\ D_{v;\mathrm{ij}}[\mathbf{X}_t] \end{pmatrix} \tag{13}$$

and the anisotropic 2D-TV of $\mathbf{X}_t$ is simply the sum of the magnitudes of this discrete gradient at every pixel

$$\mathrm{TV}_{2\mathrm{D}}(\mathbf{x}_t) \triangleq \sum_{\mathrm{ij}} (|D_{h;\mathrm{ij}}[\mathbf{X}_t]| + |D_{v;\mathrm{ij}}[\mathbf{X}_t]|) \tag{14}$$

$$= \sum_{\mathrm{ij}} \|D_{\mathrm{ij}}[\mathbf{X}_t]\|_{\ell_1}.$$

To reconstruct $\mathbf{X}_t$, we can solve the convex program

$$\widehat{\mathbf{x}}_t = \arg\min_{\mathbf{x}_t} \mathrm{TV}_{2\mathrm{D}}(\mathbf{x}_t) \quad \text{subject to} \quad \mathbf{y}_t = \mathbf{\Phi}\mathbf{x}_t. \tag{15}$$

However, in practical situations the measurement vector $\mathbf{y}_t$ may be corrupted by noise. Then, CS acquisition of $\mathbf{x}_t$ can be formulated as

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{x}_t + \mathbf{n}_t \tag{16}$$

where $\mathbf{n}_t$ is the noise vector. To recover $\mathbf{x}_t$, we can use 2D-TV minimization as in (15) and formulate the following unconstrained optimization problem

$$\widehat{\mathbf{x}}_t = \arg\min_{\mathbf{x}_t}[\mu \mathrm{TV}_{2\mathrm{D}}(\mathbf{x}_t) + \frac{1}{2}\|\mathbf{y}_t - \mathbf{\Phi}\mathbf{x}_t\|_{\ell_2}^2] \tag{17}$$

where μ is a non-negative weight controlling the sparsity level.

III. CS-DOMAIN L_1-PCA FOR COMPRESSED-SENSED SURVEILLANCE VIDEO

A. CS Surveillance Video Acquisition

In this paper, we consider a practical CS surveillance video acquisition system that performs pure, direct compressed sensing of each video frame. In the simple compressive video encoding block diagram shown in Fig. 1, each frame $\mathbf{X}_t$ of size $m \times n$ is viewed as a vectorized column $\mathbf{x}_t \in \mathbb{R}^D, D = mn$, where t is the frame index. Compressive sampling is performed by projecting $\mathbf{x}_t$ onto a $P \times D$ $(P < D)$ measurement matrix $\mathbf{\Phi}$

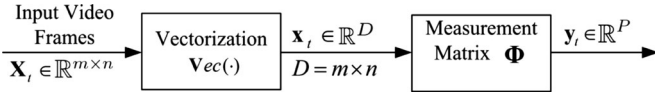$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{x}_t, \quad t = 1, 2, \ldots \tag{18}$$

Fig. 1.   Compressed-sensing (CS) video encoder.

where $\mathbf{\Phi}$ is generated by randomly permuting the columns of an order-k, $k \geq D$ and multiple-of-four, Walsh-Hadamard (WH) matrix followed by arbitrary selection of P rows from the k available WH rows (if $k > D$, only D arbitrary columns are utilized). This class of WH measurement matrices has the advantage of easy implementation (antipodal ± 1 entries), fast transformation, and satisfactory reconstruction performance as we will see later on. A richer class of matrices can be found in [47], [48]. For practical implementation, $\mathbf{\Phi}$ is generated once and fixed for all frames in the video sequence. The resulting CS measurement vectors $\mathbf{y}_t, t = 1, 2, \ldots$, are then transmitted to the decoder.

B. L_1-PCA for Background Extraction

We organize an ensemble of N CS measurement vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ in matrix form $\mathbf{Y} \triangleq [\mathbf{y}_1 \ \mathbf{y}_2 \ \ldots \ \mathbf{y}_N]$ modeled by

$$\mathbf{Y} = \ \mathbf{\Phi X} + \mathbf{O} \tag{19}$$

$$= \ \mathbf{\Phi}(\mathbf{L} + \mathbf{E}) + \mathbf{O} \tag{20}$$

$$= \ \mathbf{Y_L} + \mathbf{\Phi E} + \mathbf{O} \tag{21}$$

where $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \ldots \ \mathbf{x}_N]$ is the matrix consisting of the N corresponding video frames. If the utilized video frames are relatively time-lapsed (non-successive) such that their foreground scenes contain moving objects with low correlations, $\mathbf{X}$ can be viewed as a sum of the low-rank background $\mathbf{L}$ and the sparse moving objects $\mathbf{E}$. Assuming that $\mathbf{L}$ is a rank-d matrix, the CS-domain observed data matrix $\mathbf{Y_L} \triangleq \mathbf{\Phi L}$ is also of rank d and represents the compressed-sensed background scene. To deal in addition with the practical issue of possible faulty data, we assume that the observed CS measurements may be corrupted by outliers $\mathbf{O}$ due to acquisition failures. By applying L_1-PCA to $\mathbf{Y}$, we extract a background-scene subspace basis

$$\mathbf{R}_{L_1} = \arg \max_{\substack{\mathbf{R} \in \mathbb{R}^{P \times d} \\ \mathbf{R}^T \mathbf{R} = \mathbf{I}_d}} \|\mathbf{Y}^T \mathbf{R}\|_1. \tag{22}$$

The complexity of solving (22) is $\mathcal{O}(2^{Nd} \min\{P^2 d, Pd^2\})$ by the exhaustive search algorithm and $\mathcal{O}(N^{\mathrm{rank}(\mathbf{Y})d - d + 1})$ by the polynomial-time algorithm [38]. Compared to pixel-domain L_1-PCA computation described in (9), the complexity is significantly reduced since the vector length is reduced from D to P due to compressed sensing. By projecting the observed CS measurements $\mathbf{Y}$ onto $\mathbf{R}_{L_1}$, we can obtain the compressed-sensed low-rank component

$$\widehat{\mathbf{Y}}_{\mathbf{L}}^{L_1} = \mathbf{R}_{L_1} \mathbf{R}_{L_1}^T \mathbf{Y}. \tag{23}$$

Afterwards, the background scene can be reconstructed by performing CS recovery on the columns of $\widehat{\mathbf{Y}}_{\mathbf{L}}^{L_1}$, i.e., $\widehat{\mathbf{y}}_{\mathbf{L},t}^{L_1}, t = 1, 2, \ldots, N$. Here, we propose and use TV minimization

introduced in Section II of the following form:

$$\widehat{\ell}_t = \arg \min_{\ell_t} \mu \mathrm{TV}_{2D}(\ell_t) + \frac{1}{2}\|\widehat{\mathbf{y}}_{\mathbf{L},t}^{L_1} - \mathbf{\Phi}\ell_t\|_2^2. \tag{24}$$

For comparison purposes, in parallel to the above developments, we introduce L_2-norm based CS-domain PCA calculation (SVD) by

$$\mathbf{R}_{L_2} = \arg \max_{\substack{\mathbf{R} \in \mathbb{R}^{P \times d} \\ \mathbf{R}^T \mathbf{R} = \mathbf{I}_d}} \|\mathbf{Y}^T \mathbf{R}\|_2. \tag{25}$$

Similar to (23), the background scene can be obtained by projecting $\mathbf{Y}$ onto the L_2 principal components

$$\widehat{\mathbf{Y}}_{\mathbf{L}}^{L_2} = \mathbf{R}_{L_2} \mathbf{R}_{L_2}^T \mathbf{Y} \tag{26}$$

followed by TV minimization as in (24). Since L_2-PCA is sensitive to outlier values, the performance of CS-L_2-PCA is anticipated to be inferior to CS-L_1-PCA in the presence of faulty/corrupted data and sensitive to the selected rank value d.

C. Moving Objects Extraction

To extract the sparse moving objects in the foreground, we perform frame-by-frame CS reconstruction in the form of

$$\widehat{\mathbf{x}}_t = \arg \min_{\mathbf{x}} \left[\mu \mathrm{TV}_{2D}(\mathbf{x}) + \frac{1}{2}\|\mathbf{y}_t - \mathbf{\Phi x}\|_2^2 \right]. \tag{27}$$

With the recovered video frames $\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_1 \ \widehat{\mathbf{x}}_2 \ \ldots \ \widehat{\mathbf{x}}_N] \in \mathbb{R}^{D \times N}$ the sparse foreground can be recovered as

$$\widehat{\mathbf{e}}_t = \widehat{\mathbf{x}}_t - \widehat{\ell}_t \tag{28}$$

for $t = 1, 2, \ldots, N$.

D. Adaptive CS-L_1-PCA

To find an accurate CS-domain low-rank subspace representation of the background scene by solving (22) [or (25)], low correlation among the moving objects across CS frames $\mathbf{y}_1$ to $\mathbf{y}_N$ is required. At the same time, low-latency video surveillance requires background and foreground separation shortly after the CS measurements are received. Therefore, at each time slot at the monitoring decoder, the CS measurements of only a small number of successive frames can be assumed available for processing. In this case, a slowly moving object may not change position significantly and thus may be captured as part of the background. In this section, we introduce a method to improve the accuracy of the calculated CS-domain low-rank background subspace via recursive/adaptive CS-frame processing.

Assume that k background frames are already identified by processing preceding frames. Their CS-domain representations calculated by (22), (23) form a matrix defined by

$$\widehat{\mathbf{Y}}_{\mathbf{L},\mathrm{pre}}^{L_1} \triangleq [\widehat{\mathbf{y}}_{\mathbf{L},1}^{L_1} \ \ \widehat{\mathbf{y}}_{\mathbf{L},2}^{L_1} \ \ \cdots \ \ \widehat{\mathbf{y}}_{\mathbf{L},k}^{L_1}] \in \mathbb{R}^{P \times k}. \tag{29}$$

We then collect new CS measurements $\mathbf{Y} = [\mathbf{y}_{k+1} \ \mathbf{y}_{k+2} \ \cdots \ \mathbf{y}_{k+N}] \in \mathbb{R}^{P \times N}$ to form a current group of N frames and solve the following combined problem:

$$\mathbf{R}_{L_1,\mathrm{adapt}} = \arg \max_{\substack{\mathbf{R} \in \mathbb{R}^{P \times d} \\ \mathbf{R}^T \mathbf{R} = \mathbf{I}_d}} \|[\widehat{\mathbf{Y}}_{\mathbf{L},\mathrm{pre}}^{L_1} \ \mathbf{Y}]^T \mathbf{R}\|_1. \tag{30}$$

TABLE I
COMPLEXITY ANALYSIS

	CS-L_1-PCA	CS-L_2-PCA
$\mathbf{R}_{L_1}$ by (22)	$\mathcal{O}(2^{Nd} \min\{P^2 d, Pd^2\})$ exhaustive search [38] or	
$\mathbf{R}_{L_2}$ by (25)	$\mathcal{O}(N^{\mathrm{rank}(\mathbf{Y})d-d+1})$ polynomial-time [38] or $\mathcal{O}(\min\{NP^2, N^2 P\} + N^2(d+2) + NP)$ "fast" [39]	$\mathcal{O}(\min\{P^2 N, PN^2\})$
$\widehat{\mathbf{Y}}_{\mathbf{L}}^{L_1}$ by (23)	$\mathcal{O}(2PNd)$	$\mathcal{O}(2PNd)$
$\widehat{\mathbf{Y}}_{\mathbf{L}}^{L_2}$ by (26)		
$\widehat{\ell}_t$ by (24)	$\mathcal{O}(4Pmn \times \mathrm{max\ iteration})$	$\mathcal{O}(4Pmn \times \mathrm{max\ iteration})$
$\widehat{\mathbf{x}}_t$ by (27)		

Afterwards, the background scene can be reconstructed/updated by projecting the columns of $\mathbf{Y}$ onto the adaptive L_1 principal components $\mathbf{R}_{L_1,\mathrm{adapt}}$ of (30) followed by TV minimization CS image recovery. Initially, $\mathbf{R}_{L_1,\mathrm{adapt}}$ may be an inaccurate approximation of the CS-domain low-rank subspace of the background. As the adaptation proceeds, $\mathbf{R}_{L_1,\mathrm{adapt}}$ becomes a progressively better representation of the pursued subspace. Furthermore, as the background changes, $\mathbf{R}_{L_1,\mathrm{adapt}}$ follows the changes accordingly. Hence, adaptive CS-L_1-PCA as described not only reduces latency, but also allows the computed subspace to adapt quickly to changes in the video background.

E. Complexity Analysis

The CS-L_1-PCA method described in Section III-B requires: (i) Computation of the CS-domain L_1 principal components $\mathbf{R}_{L_1}$ by (22); (ii) projection of the observed CS measurements $\mathbf{Y}$ onto $\mathbf{R}_{L_1}$ by (23); (iii) reconstruction of the background frames from $\widehat{\mathbf{Y}}_{\mathbf{L}}^{L_1}$ via TV minimization by (24); and (iv) recovery of the foreground by (27) and (28). For CS-L_2-PCA, the L_2 principal components are computed by (25), followed by similar steps as those for CS-L_1-PCA. Detailed computational complexity of every step is presented in Table I.[1] The complexity of the adaptive CS-L_1-PCA algorithmic version described in Section III-D is given by replacing in Table I N by $N + k$ where k represents the number of pre-processed preceding frames.
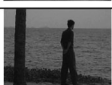
IV. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

In this section, we demonstrate the effectiveness of the proposed algorithmic developments on a wide variety of test surveillance sequences frequently used in the literature. Table II lists the utilized sequences and describes their properties. The sequences not only differ in object size and object number, with both indoor and outdoor environments presented, but also present well-known challenges for video surveillance systems, including:

1) light changes (the background learning should adapt to sudden illumination changes);
2) bootstrapping (background should be accurately extracted even in the absence of a complete and free of moving objects training set at the beginning of the sequence); and

[1]As a representative example of experimental relative computational time, when CS-L_1-PCA and CS-L_2-PCA are executed on an Intel i5-2550K 3.40 GHz platform with input the PETS2001 video sequence (see Table II), the expended time per frame is 1.805 sec and 1.741 sec, respectively.

TABLE II
VIDEO SEQUENCE BENCHMARKS

Sequence Title	Sampled Frame	Sequence Type	Image Size	Background Property	Object Size	Object Number
PETS2001		Outdoor	120×160	Static	Small	Few
Airport		Indoor	144×176	Static	Medium	Many
Daniel_light		Indoor	120×160	Illuminance Change	Large	Few
Fountain		Outdoor	128×160	Dynamic	Small	Medium
WaterSurface		Outdoor	128×160	Dynamic	Large	Few

3) dynamic background (the moving objects should be correctly detected even when part of the background scene is moving).

In the following, it is seen by qualitative and quantitative results that the proposed CS-L_1-PCA approach can cope with the above mentioned issues in background extraction and achieve successful state-of-the-art moving object extraction in various types of videos taken with static CS cameras.

A. Non-Adaptive CS-L_1-PCA Background Extraction

We first perform experiments to illustrate and evaluate the non-adaptive CS-L_1-PCA background extraction approach developed in Sections III-A–C. We demonstrate that CS-L_1-PCA outperforms CS-L_2-PCA in both absence and presence of CS measurement outliers, therefore the L_1-norm metric in (4), (22) appears more appropriate than the L_2-norm metric in (3), (25) in identifying a low-rank subspace for background scene representation. In this manuscript, we include results for two of the testing sequences, PETS2001 and Airport. For each video sequence, $N = 10$ frames are selected to form a video volume. Each frame is compressed-sensed independently using the same randomly permutated partial Walsh-Hadamard matrix. The number of CS
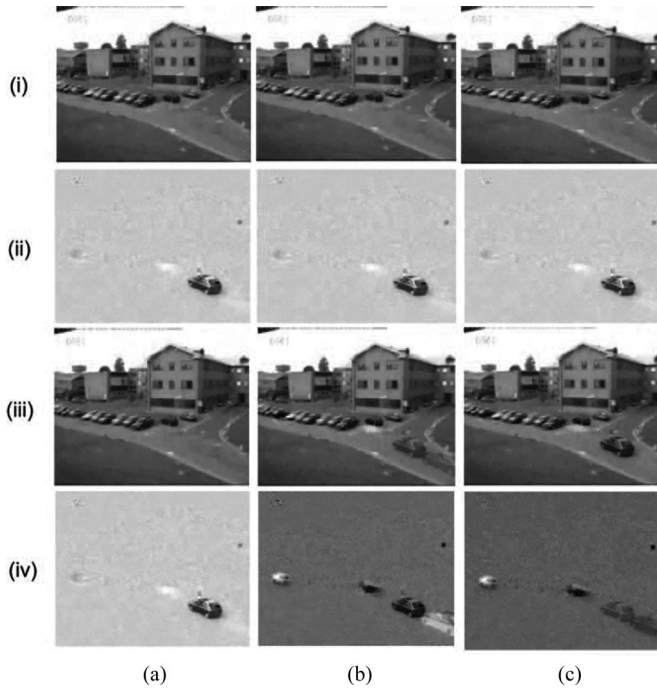
Fig. 2.   *PETS2001* sequence (clean CS measurements): CS-L_1-PCA reconstructed background and moving objects (row (i) and (ii), respectively) and CS-L_2-PCA reconstructed background and moving objects (row (iii) and (iv), respectively) with $d = 1, 2,$ or 3 principal components (columns (a), (b), and (c), respectively).
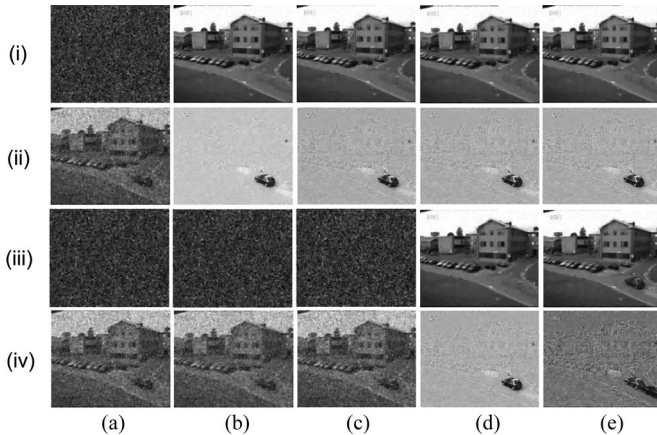

Fig. 3.   *PETS2001* sequence (75% outliers in CS measurements of three randomly selected frames): CS-L_1-PCA reconstructed background and moving objects (row (i) and (ii), respectively) and CS-L_2-PCA reconstructed background and moving objects (row (iii) and (iv), respectively) with $d = 1, 2, 3, 4,$ or 5 principal components (columns (a), (b), (c), (d), and (e), respectively).

measurements per frame is 37.5% of the total number of pixels in the video frame.

Figs. 2–5 display the extracted background scenes and foreground moving objects for the two testing sequences. In particular, Fig. 2 displays results for the representative *PETS2001* frame shown in Table II when the CS measurements *do not contain outliers*. Rows (i) and (ii) are the results of CS-L_1-PCA[2] background and foreground, respectively, extraction with
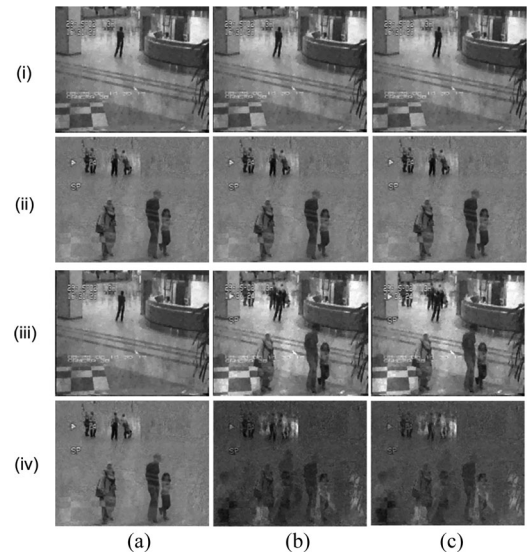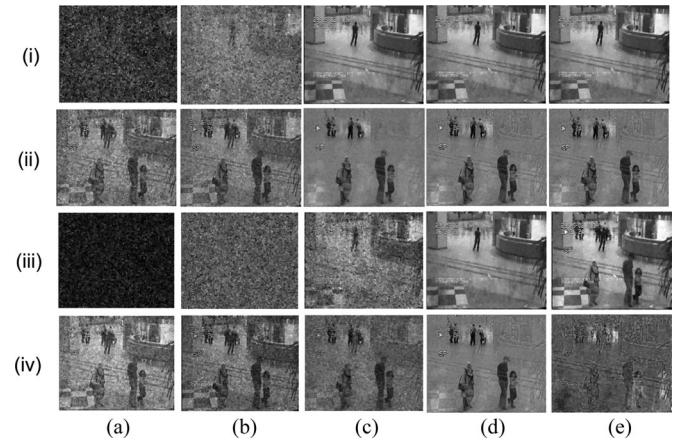

Fig. 4.   *Airport* sequence (clean CS measurements): CS-L_1-PCA reconstructed background and moving objects (row (i) and (ii), respectively) and CS-L_2-PCA reconstructed background and moving objects (row (iii) and (iv), respectively) with $d = 1, 2,$ or 3 principal components (columns (a), (b), and (c), respectively).

Fig. 5.   *Airport* sequence (50% outliers in CS measurements of three randomly selected frames): CS-L_1-PCA reconstructed background and moving objects (row (i) and (ii), respectively) and CS-L_2-PCA reconstructed background and moving objects (row (iii) and (iv), respectively) with $d = 1, 2, 3, 4,$ or 5 principal components (columns (a), (b), (c), (d), and (e), respectively).

$d = 1, 2,$ and 3 principal components (columns (a), (b), and (c), correspondingly). Rows (iii) and (iv) repeat the study for CS-L_2-PCA. It is observed that in the absence of outliers, both CS-L_1-PCA and CS-L_2-PCA correctly extract the background and the moving objects with one principal component ($d = 1$). When d increases (rank mismatch), CS-L_1-PCA maintains superior performance, while CS-L_2-PCA rapidly deteriorates.

Fig. 3 repeats the study of Fig. 2 for the same data set with corrupted CS measurements. In the experiment, 75% of the CS measurements of three randomly selected frames are corrupted by outliers.[3] The presence of outliers in three frames modifies/

---

[2] For L_1-PCA calculation, the fast algorithm developed in [39] is adopted.

[3] If the ith CS measurement of the tth frame, $\mathbf{y}_t(i)$, is selected to be corrupted by an outlier, then in the experiments $\mathbf{y}_t(i)$ is replaced by $-3.5 \max_{1 \leq t \leq N} |\mathbf{y}_t(i)|$.

increases the effective SVD rank of the background from $d = 1$ to 4. Naturally, when we use $d = 1$, both L_1- and L_2-PCA cannot recover the background/foreground scenes. When $d \geq 2$, L_1-PCA shows remarkable resistance to outliers and recovers the background and the moving objects well with L_1 rank choice $d = 2$ or above ($d = 3, 4$, or 5). On the other hand, L_2-PCA needs specifically $d = 4$ principal components (number of corrupted frames plus one) to recover the low-rank outlier corrupted background scene and its performance decreases quickly when $d > 4$.

For increased credibility of the study, a similar experiment is performed on the *Airport* sequence (Figs. 4 and 5). In Fig. 5, 50% of the CS measurements of three randomly selected frames are corrupted by outliers as in the study of Fig. 3. Similar conclusion to the studies of Figs. 2 and 3 can be drawn. CS-L_1-PCA offers superior robustness in rank selection and background/moving objects extraction in both clean and outlier corrupted video sequences.

B. Adaptive CS-L_1-PCA Background Extraction

Next, we compare the adaptive CS-L_1-PCA background extraction scheme developed in Section III.D with non-adaptive CS-L_1-PCA and three state-of-the-art video surveillance procedure from the literature (MahNMF [32], DECOLOR [34], and CS-RPCA [35]).

For each test sequence in Table II, we initialize the adaptive CS-L_1-PCA process with the first $N_0 = 20$ compressively sensed frames. The most recent $k \leq N_0 = 20$ extracted CS-domain background representations are grouped to form the matrix $\widehat{\mathbf{Y}}_{\mathbf{L},\text{pre}}^{L_1} \in \mathbb{R}^{P \times k}$ in (29). Then, new CS measurements are collected over the next $N = 5$ successive frames, with two of them corrupted by outlier values as in Figs. 3 and 5. We append the new five CS frames to $\widehat{\mathbf{Y}}_{\mathbf{L},\text{pre}}^{L_1}$ and solve (30) to obtain the new principal components $\mathbf{R}_{L_1,\text{adapt}}$; (23) and (24) are solved to update the background/foreground scenes. We continue, subsequently, the process with a "sliding-forward window" of $N = 5$ successive frames (again two of them always corrupted by outlier values).

Fig. 6 displays the background and foreground extracted at multiple distinct time slots $t = 27, 30, 39, 43, 47$ of *PETS2001* with $d = 2$ principal components by non-adaptive CS-L_1-PCA [rows (ii) and (iii)], adaptive CS-L_1-PCA with $k = 5$ [rows (iv) and (v)], and adaptive CS-L_1-PCA with $k = 15$ [rows (vi) and (vii)]. It is observed that the adaptive methods ($k = 5, 15$) generate better background and foreground scenes compared to the non-adaptive method. Since *PETS2001* is a relatively easy sequence with static background, small object number and size, and somewhat low correlation among consecutive frames, adaptive CS-L_1-PCA methods with $k = 5$ and $k = 15$ have similar performance [rows (iv)–(vii)]. However, when applied to challenging sequences, adaptive CS-L_1-PCA with a higher k value shows appealing performance improvement compared to that with a small k value as seen in Figs. 7–10. Specifically, Fig. 7 shows the results for the *Airport* sequence that contains multiple moving objects and requires "bootstrapping", i.e., object-free frames are unavailable at the beginning of the sequence to train
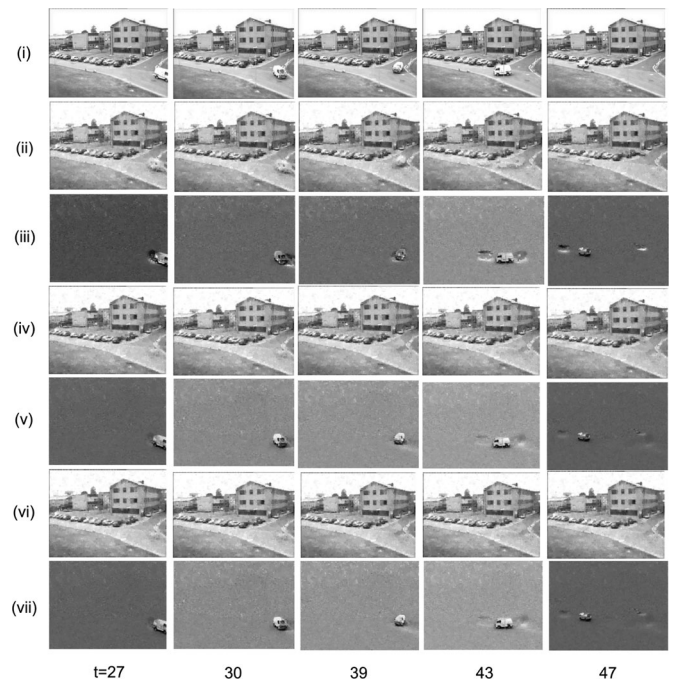

Fig. 6. *PETS2001* sequence: original frame [row (i)] of time slot $t = 27, 30, 39, 43$, and 47; non-adaptive CS-L_1-PCA reconstructed background and moving objects [rows (ii) and (iii)]; adaptive CS-L_1-PCA reconstructed background and moving objects $k = 5$ [rows (iv) and (v)]; adaptive CS-L_1-PCA reconstructed background and moving objects $k = 15$ [rows (vi) and (vii)].
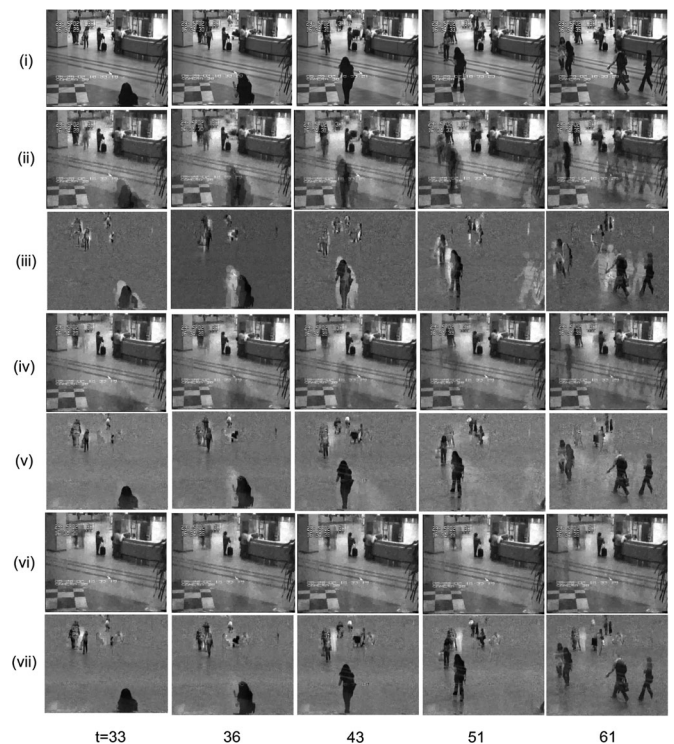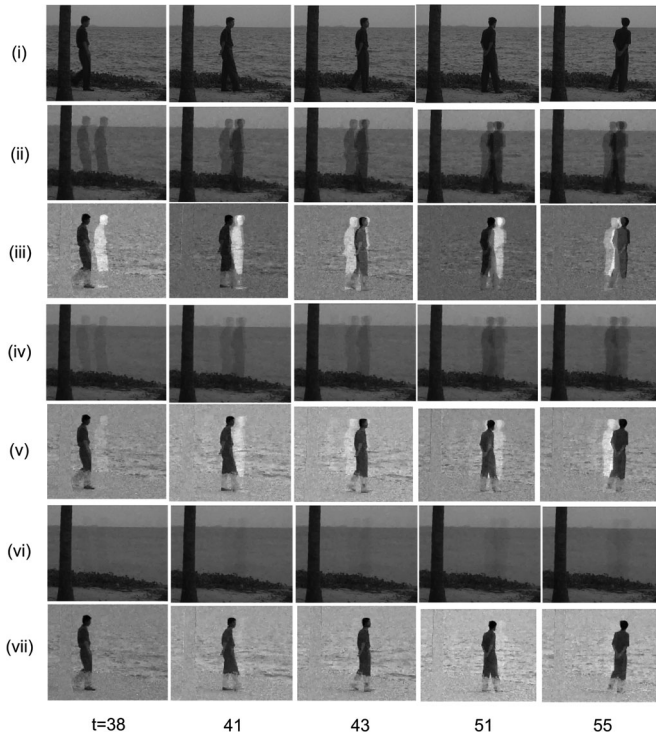

Fig. 7. *Airport* sequence: original frame [row (i)] of time slot $t = 33, 36, 43, 51$, and 61; non-adaptive CS-L_1-PCA reconstructed background and moving objects [rows (ii) and (iii)]; adaptive CS-L_1-PCA reconstructed background and moving objects $k = 5$ [rows (iv) and (v)]; adaptive CS-L_1-PCA reconstructed background and moving objects $k = 15$ [rows (vi) and (vii)].
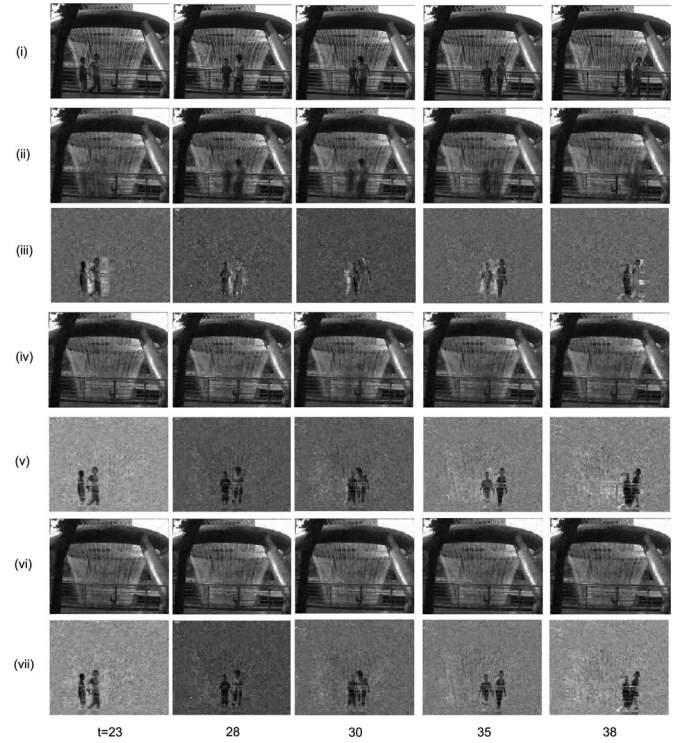
Fig. 8. *Daniel_light* sequence: original frame [row (i)] of time slot $t = 26, 30, 43, 48$, and 93; non-adaptive CS-L_1-PCA reconstructed background and moving objects [rows (ii) and (iii)]; adaptive CS-L_1-PCA reconstructed background and moving objects $k = 5$ [rows (iv) and (v)]; adaptive CS-L_1-PCA reconstructed background and moving objects $k = 15$ [rows (vi) and (vii)].

Fig. 9. *WaterSurface* sequence: original frame [row (i)] of time slot $t = 38, 41, 43, 51$, and 55; non-adaptive CS-L_1-PCA reconstructed background and moving objects [rows (ii) and (iii)]; adaptive CS-L_1-PCA reconstructed background and moving objects $k = 5$ [rows (iv) and (v)]; adaptive CS-L_1-PCA reconstructed background and moving objects $k = 15$ [rows (vi) and (vii)].

Fig. 10. *Fountain* sequence: Original frame [row (i)] of time slot $t = 23, 28, 30, 35$, and 38; non-adaptive CS-L_1-PCA reconstructed background and moving objects [rows (ii) and (iii)]; adaptive CS-L_1-PCA reconstructed background and moving objects $k = 5$ [rows (iv) and (v)]; adaptive CS-L_1-PCA reconstructed background and moving objects $k = 15$ [rows (vi) and (vii)].

the low-rank background subspace. Fig. 8 tests the algorithm on the *Daniel ₋ light* sequence with sudden illumination change due to light switching. Fig. 9 shows the results on the *WaterSurface* sequence with a slow moving person and dynamic background. Fig. 10 is the *Fountain* sequence with multiple moving objects, as well as dynamic background. For all these challenging sequences, the proposed adaptive CS-L_1-PCA method performs significantly better than its non-adaptive counterpart, in that the "ghost" phenomena in the extracted background scenes are effectively removed or mitigated. Performance is improving as we move on from $k = 5$ to $k = 15$.

For increased credibility of our experimental studies and conclusions, we also carry out quantitative analysis. Binary masks of three sample frames are generated for each of the five sequences by thresholding the adaptive CS-L_1-PCA extracted foreground images and different accuracy metrics are calculated such as *Recall, Precision*, F_1, and *Similarity* ([20] and [21]–[25]) using the generated binary masks and ground truth (Tables III–VII). *Recall*, also known as *detection rate*, gives the percentage of detected true positives (detected true foreground object pixels) as compared to the total number of true positives in the ground truth

$$\text{Recall} \triangleq \frac{tp}{tp + fn} \tag{31}$$

where *tp* is the total number of *true positives*, *fn* is the total number of *false negatives*, and $(tp + fn)$ indicates the total

TABLE III
BINARY OBJECT MASKS AND ACCURACY METRICS FOR *PETS2001*

Sampled frames			
CS-L_1-PCA Foreground			
Ground Truth			
CS-L_1-PCA			
Recall	0.890	0.895	0.900
Precision	0.859	0.756	0.787
Similarity	0.776	0.695	0.724
F_1	0.874	0.820	0.840
MahNMF [32]			
Recall	0.909	0.833	0.691
Precision	0.915	0.795	0.785
Similarity	0.838	0.685	0.581
F_1	0.912	0.813	0.735
DECOLOR [34]			
Recall	0.968	1.00	0.969
Precision	0.795	0.684	0.730
Similarity	0.774	0.684	0.713
F_1	0.873	0.812	0.833
CS-RPCA [35]			
Recall	0.802	0.796	0.707
Precision	0.954	0.895	0.832
Similarity	0.772	0.728	0.618
F_1	0.871	0.843	0.764

TABLE IV
BINARY OBJECT MASKS AND ACCURACY METRICS FOR *Airport*

Sampled frames			
CS-L_1-PCA Foreground			
Ground Truth			
CS-L_1-PCA			
Recall	0.916	0.891	0.877
Precision	0.735	0.733	0.797
Similarity	0.688	0.672	0.717
F_1	0.815	0.804	0.835
MahNMF [32]			
Recall	0.481	0.428	0.683
Precision	0.596	0.747	0.873
Similarity	0.363	0.374	0.621
F_1	0.533	0.544	0.766
DECOLOR [34]			
Recall	0.509	0.507	0.617
Precision	0.513	0.535	0.871
Similarity	0.343	0.352	0.565
F_1	0.511	0.521	0.722
CS-RPCA [35]			
Recall	0.457	0.425	0.347
Precision	0.529	0.598	0.944
Similarity	0.325	0.331	0.340
F_1	0.490	0.497	0.508

number of foreground object pixels present in the ground truth. *Recall* alone is not enough to compare different methods and is generally used in conjunction with *Precision*, also known as *positive prediction*, that gives the percentage of detected true positives as compared to the total number of foreground object pixels detected by the method

$$\text{Precision} \triangleq \frac{tp}{tp + fp}. \qquad (32)$$

Here, *fp* is the total number of *false positives* and $(tp + fp)$ indicates the total number of detected foreground pixels. Moreover, we consider the F_1 metric, also known as *Figure* of *Merit* or F-measure, that is the weighted harmonic mean of *Precision* and *Recall*

$$F_1 \triangleq \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}. \qquad (33)$$

Such measure allows to obtain a single value that can be used to "rank" different methods. Finally, we consider the pixel-based *Similarity* measure defined as

$$\text{Similarity} \triangleq \frac{tp}{tp + fn + fp}. \qquad (34)$$

We compare these quantitative metrics for the proposed adaptive CS-L_1-PCA algorithm ($k = 15$) with three state-of-the-art approaches, MahNMF [32], DECOLOR [34], and CS-RPCA [35]. To implement MahNMF [32] and DECOLOR [34] which are pixel-domain methods, we first obtain the pixel-domain reconstruction of each frame from its CS measurements by (27). Then, both algorithms are applied to each group of twenty recovered *non-corrupted* frames. The binary foreground object masks and accuracy metrics are shown in Tables III–VII.

TABLE V
BINARY OBJECT MASKS AND ACCURACY METRICS FOR *Daniel_Light*

Sampled frames			
CS-L_1-PCA Foreground			
Ground Truth			
CS-L_1-PCA			
Recall	0.704	0.795	0.648
Precision	0.991	0.970	0.765
Similarity	0.700	0.776	0.540
F_1	0.823	0.874	0.702
MahNMF [32]			
Recall	0.670	0.746	0.533
Precision	0.988	0.976	0.984
Similarity	0.665	0.733	0.529
F_1	0.799	0.846	0.692
DECOLOR [34]			
Recall	0.061	0.998	0.337
Precision	0.983	0.627	0.915
Similarity	0.061	0.626	0.327
F_1	0.115	0.770	0.493
CS-RPCA [35]			
Recall	0.298	0.560	0.177
Precision	0.990	0.984	0.758
Similarity	0.297	0.555	0.167
F_1	0.458	0.714	0.287

TABLE VI
BINARY OBJECT MASKS AND ACCURACY METRICS FOR *WaterSurface*

Sampled frames			
CS-L_1-PCA Foreground			
Ground Truth			
CS-L_1-PCA			
Recall	0.757	0.764	0.799
Precision	0.962	0.986	0.958
Similarity	0.735	0.755	0.772
F_1	0.847	0.861	0.871
MahNMF [32]			
Recall	0.245	0.332	0.676
Precision	0.573	0.822	0.959
Similarity	0.207	0.310	0.657
F_1	0.343	0.473	0.793
DECOLOR [34]			
Recall	0.756	0.766	0.114
Precision	0.892	0.930	0.995
Similarity	0.693	0.724	0.114
F_1	0.819	0.840	0.205
CS-RPCA [35]			
Recall	0.275	0.273	0.109
Precision	0.885	0.932	0.498
Similarity	0.266	0.268	0.098
F_1	0.420	0.422	0.178

Because DECOLOR models the continuity prior of the sparse foreground objects, it can offer higher *Recall* values compared to adaptive CS-L_1-PCA for certain frames. However, it sacrifices in *Precision*. On average, MahNMF has higher *Precision* than DECOLOR, but due to lack of adaptivity it performs poorly for highly correlated frames, such as *WaterSurface* in which the object is moving slowly across frames. In addition, the pixel-domain iterative matrix factorization in MahNMF leads to the highest computation time among all schemes in comparison. To place our comparison studies in a broader context, we also contrast our findings against traditional low-rank and sparse matrix decomposition for compressed-sensed surveillance video processing, i.e., CS-RPCA in [35]. Since CS-RPCA cannot accommodate outliers in CS measurements, we apply CS-RPCA to *non-corrupted* CS measurements of the same set of video frames studied for the other three approaches. The results show that the detected foreground objects have many false negatives

even when all CS measurements are received correctly. The reason is that convex-optimization-based foreground/background scene recovery relies on a clear, strong low-rank structure across the compressed-sensed data, which is not satisfied when only a moderate number of frames is considered (twenty frames in our experiment). Indeed in [35], 100 frames are considered for successful low-rank background scene recovery from the CS measurements.

For all five sequences, among all methods in the comparison the proposed adaptive CS-L_1-PCA procedure achieves the highest average F_1 and *Similarity* value (which are considered as most comprehensive accuracy measures). In addition, the gray-scale foreground scenes extracted by adaptive CS-L_1-PCA (row (ii) in Tables III–VII) impressively display object details for visually recognizing the moving objects. To supplement the

TABLE VII
BINARY OBJECT MASKS AND ACCURACY METRICS FOR *Fountain*

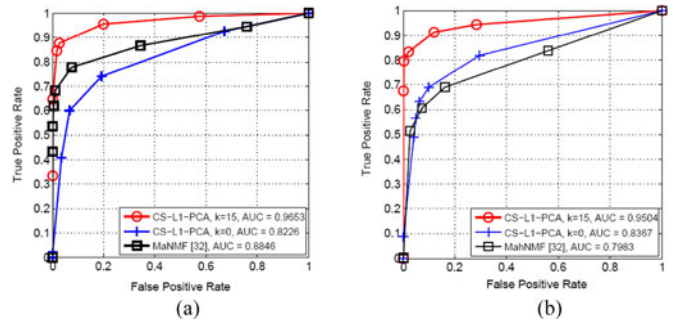|			
Sampled frames			
CS-L_1-PCA Foreground			
Ground Truth			
CS-L_1-PCA			
Recall	0.845	0.678	0.750
Precision	0.936	0.895	0.913
Similarity	0.798	0.628	0.700
F_1	0.888	0.771	0.824
MahNMF [32]			
Recall	0.790	0.433	0.666
Precision	0.970	0.977	0.967
Similarity	0.771	0.429	0.651
F_1	0.871	0.600	0.789
DECOLOR [34]			
Recall	0.932	0.358	0.829
Precision	0.729	0.978	0.769
Similarity	0.692	0.355	0.663
F_1	0.818	0.524	0.798
CS-RPCA [35]			
Recall	0.761	0.534	0.698
Precision	0.945	0.944	0.947
Similarity	0.729	0.518	0.672
F_1	0.843	0.682	0.804

Fig. 11. ROC curves and AUC values of (a) sample frame 3 in Table IV for the *Airport* sequence and (b) sample frame 1 in Table VI for the *WaterSurface* sequence.

proposed adaptive CS-L_1-PCA $(k = 15)$ procedure provides the highest AUC value for both sequences.[4]

V. CONCLUSION

We proposed a compressed-sensing-domain L_1-norm maximization principal-component-analysis scheme for compressed-sensed surveillance video processing. The algorithm computes a low-rank subspace via L_1-PCA to represent the background scene directly in the CS domain and enjoys significantly lower computational complexity than pixel-domain L_1-PCA. Background reconstruction is then performed by projecting the CS measurement vectors onto the L_1 principal components followed by regular CS image recovery (for example, total-variation minimization). Experiments demonstrate that (i) the CS-L_1-PCA algorithm performs better than L_2-norm based CS domain PCA when the CS measurements are corrupted by outliers and (ii) even in clean CS data operation, CS-L_1-PCA offers exceptional performance and robustness in background subspace-rank selection compared to CS-L_2-PCA. An adaptive CS-L_1-PCA method is developed as well for low-latency video surveillance in which previous extracted background frames are utilized to more effectively update the background subspace for new frames. Experiments show that the adaptive CS-L_1-PCA performs better than its non-adaptive counterpart especially when the number of involved previous background frames increases. Finally, qualitative and quantitative results demonstrate the effectiveness of the adaptive CS-L_1-PCA method for different type of surveillance videos compared to state-of-the-art surveillance video systems in the literature. In the context of future research, more advanced algorithms may be developed under the CS-L_1-PCA framework to deal with challenges in video surveillance such as detecting moving objects with static parts, removing shadows cast by objects to more accurately describe the object shape, and addressing the camouflage problem.

experimental studies, we provide sample videos of the detected *Airport* and *WaterSurface* foreground objects in gray-scale format alongside the corresponding original video sequences. The video streams clearly demonstrate the performance superiority of the proposed adaptive CS-L_1-PCA procedure against Mah-NMF [32], DECOLOR [34], and CS-RPCA [35]. For each sequence, it is observed that the proposed method detects well and consistently over time the foreground objects, while the other three methods have fluctuating performance and may even miss the entire object in certain frames. Finally, the receiver operating characteristic (ROC) and area under curve (AUC) metrics [49] are provided for the *Airport* and *WaterSurface* sequences in Fig. 11. We compare the ROC and AUC metrics of three schemes, adaptive CS-L_1-PCA $(k = 15)$, non-adaptive CS-L_1-PCA $(k = 0)$, and MahNMF [32]. The results show that the

---

[4]Since the operating points at high false positive rates (FPR) are unavailable for the DECOLOR [34] and CS-RPCA [35] algorithms, we do not compare the ROC and AUC metrics of those two methods.

References

[1] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

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

[3] E. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[4] K. Gao, S. N. Batalama, D. A. Pados, and B. W. Suter, "Compressive sampling with generalized polygons," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4759–4766, Oct. 2011.

[5] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, pp. 1207–1223, Aug. 2006.

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc., Ser. B*, vol. 58, pp. 267–288, 1996.

[7] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, pp. 407–451, Apr. 2004.

[8] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[9] S. Pudlewski and T. Melodia, "Compressive video streaming: Design and rate-energy-distortion analysis," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 2072–2086, Dec. 2013.

[10] Y. Zheng and L. Fan, "Moving object detection based on running average background and temporal difference," in *Proc. IEEE Int. Conf. Intell. Syst. Knowl. Eng.*, Nov. 2010, pp. 270–272.

[11] Q. Zhou and J. Aggarwal, "Tracking and classifying moving objects from video," in *Proc. IEEE Int. Workshop Perform. Eval. Tracking Surveillance*, Dec. 2001, pp. 1–8.

[12] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, Jul. 1997.

[13] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 1999, pp. 246–252.

[14] M. Haque, M. Murshed, and M. Paul, "Improved Gaussian mixtures for robust object detection by adaptive multi-background generation," in *Proc. Int. Conf. Pattern Recog.*, Dec. 2008, pp. 1–4.

[15] A. Elgammal, D. Harwood, and L. S. Davis, "Nonparametric model for background subtraction," in *Proc. 6th Eur. Conf. Comput. Vis.—Part II*, Jun. 2000, pp. 751–767.

[16] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[17] M. Piccardi and T. Jan, "Mean-shift background image modeling," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2004, pp. 3399–3402.

[18] Y. Liu, H. Yao, W. Gao, X. Chen, and D. Zhao, "Nonparametric background generation," *J. Vis. Commun. Image Represent.*, vol. 18, pp. 253–263, Jun. 2007.

[19] B. Han, D. Comaniciu, and L. S. Davis, "Sequential kernel density approximation through model propagation: Applications to background modeling," in *Proc. Asian Conf. Comput. Vis.*, Jan. 2004, pp. 1–7.

[20] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1168–1177, Jul. 2001.

[21] S. C. Huang, "An advanced motion detection algorithm with video quality analysis for video surveillance systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 1, pp. 1–14, Jan. 2011.

[22] J. M. Guo *et al.*, "Fast background subtraction based on a multilayer codebook model for moving object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1809–1821, Oct. 2013.

[23] S. C. Huang and B. H. Do, "Radial basis function based neural network for motion detection in dynamic scenes," *IEEE Trans. Cybern.*, vol. 44, no. 1, pp. 114–125, Jan. 2014.

[24] S. C. Huang and B. H. Chen, "Highly accurate moving object detection in variable bit rate video-based traffic monitoring systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, pp. 1920–1931, Dec. 2013.

[25] S. C. Huang and B. H. Chen, "Automatic moving object extraction through a real-world variable-bandwidth network for traffic monitoring systems," *IEEE Trans. Ind. Electron.*, vol. 61, no. 4, pp. 2099–2112, Apr. 2014.

[26] B. Chen and S. Huang, "An advanced moving object detection algorithm for automatic traffic monitoring in real-world limited bandwidth networks," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 837–847, Apr. 2014.

[27] J. Seo and S. D. Kim, "Recursive on-line $(2D)^2$ PCA and its application to long-term background subtraction," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2333–2344, Dec. 2014.

[28] Q. Ke and T. Kanade, "Robust subspace computation using L_1 norm," Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-03-172, Aug. 2003, [Online]. Available: http://ra.adm.cs.cmu.edu/anon/usr0/ftp/usr/anon/2003/CMU-CS-03-172.pdf.

[29] 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. Vis. Pattern Recog.*, Jun. 2005, pp. 739–746.

[30] 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. Vis. Pattern Recog.*, Jun. 2010, pp. 771–778.

[31] 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.*, Mar. 2012, pp. 1377–1380.

[32] N. Guan, D. Tao, Z. Luo, and J. S. Taylor, "MahNMF: Manhattan non-negative matrix factorization," *CoRR*, Jul. 2012, [Online]. Available: http://arxiv.org/abs/1207.3438.

[33] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, pp. 1–37, May 2011.

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

[35] H. Jiang, W. Deng, and Z. Shen, "Surveillance video processing using compressive sensing," *Inverse Problems Imaging*, vol. 6, pp. 201–214, 2012.

[36] N. Kwak, "Principal component analysis based on L_1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1672–1680, Sep. 2008.

[37] 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.*, Jul. 2011, pp. 1433–1438.

[38] P. P. Markopoulos, G. N. Karystinos, and D. A. Pados, "Optimal algorithms for L_1-subspace signal processing," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5046–5058, Oct. 2014.

[39] 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, Signal Proc.*, May 2014, pp. 8028–8032.

[40] E. Candès and J. Romberg, "ℓ1-magic: Recovery of sparse signals via convex programming," www.acm.caltech.edu/l1magic/downloads/l1magic.pdf.

[41] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Resonance Med.*, vol. 6, pp. 1182–1195, Dec. 2007.

[42] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, "An efficient algorithm for compressed MR imaging using total variation and wavelets," in *Proc. IEEE Conf. Comp. Vis. Pattern Recog.*, Jun. 2008, pp. 1–8.

[43] C. Li, "An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing," M.S. thesis, Dept. Comp. Appl. Math., Rice Univ., Houston, TX, USA, 2009.

[44] M. R. Dadkhah, S. Shirani, and M. J. Deen, "Compressive sensing with modified total variation minimization algorithm," in *Proc. IEEE Int. Conf. Acoust, Speech, Signal Process.*, Mar. 2010, pp. 1310–1313.

[45] Y. Liu and D. A. Pados, "Decoding of framewise compressed-sensed video via interframe total variation minimization," *SPIE J. Electron. Imaging, Special Issue Compressive Sens. Imaging*, vol. 22, pp. 1–8, Apr.–Jun. 2013.

[46] Y. Liu and D. A. Pados, "Rate-adaptive compressive video acquisition with sliding-window total-variation-minimization reconstruction," in *Proc. SPIE Compressive Sens. Conf., SPIE Defense, Security, Sens.*, May 2013, pp. 1–13.

[47] H. Ganapathy, D. A. Pados, and G. N. Karystinos, "New bounds and optimal binary signature sets—Part I: Periodic total squared correlation," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 1123–1132, Apr. 2011.

[48] H. Ganapathy, D. A. Pados, and G. N. Karystinos, "New bounds and optimal binary signature sets—Part II: Aperiodic total squared correlation," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1411–1420, May 2011.

[49] J. Ke, A. Ashok, and M. A. Neifeld, "Block-wise motion detection using compressive imaging system," *Optics Commun.*, vol. 284, pp. 1170–1180, Mar. 2011.
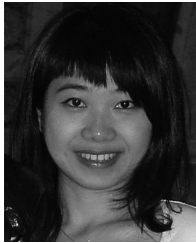
Ying Liu (S'11–M'13) received the B.S. degree in telecommunications engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2006 and the Ph.D. degree in electrical engineering from the State University of New York at Buffalo, Buffalo, NY, USA, in 2012.

She is currently a Postdoctoral Research Associate with the State University of New York at Buffalo. Her research interests include image and video processing, compressed sensing, and machine learning.

Dimitris A. Pados (M'95–SM'15) received the Diploma degree in computer science and engineering from the University of Patras, Patras, Greece, in 1989, and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, USA, in 1994.

From 1994 to 1997, he was an Assistant Professor with the Department of Electrical and Computer Engineering and the Center for Telecommunications Studies, University of Louisiana, Lafayette, LA, USA. Since August 1997, he has been with the Department of Electrical Engineering, The State University of New York at Buffalo, Buffalo, NY, USA, holding in sequence the titles of Assistant Professor, Associate Professor, Professor, and currently Clifford C. Furnas Chair Professor of Electrical Engineering. He served as Associate Chair from 2009 to 2010. He was elected University Faculty Senator four times (2004 to 2006, 2008 to 2010, 2010 to 2012, and 2016 to 2018), and served on the Faculty Senate Executive Committee from 2009 to 2010. His research interests include communication theory and adaptive signal processing with applications to signal waveform design; multiplexing, multiple-access, and interference channels; secure wireless communications; cognitive radios and networks; L1-subspace signal processing; compressed sensing; image and video processing; error correcting coding; data embedding and hiding; and antenna arrays and array radar.

Dr. Pados is a Member of the IEEE Communications, IEEE Signal Processing, IEEE Information Theory, and IEEE Computational Intelligence Societies. He served as an Associate Editor for the IEEE Signal Processing Letters from 2001 to 2004 and the IEEE Transactions on Neural Networks from 2001 to 2005. He was the recipient of the 2001 IEEE International Conference on Telecommunications Best Paper Award, the 2003 IEEE Transactions on Neural Networks Outstanding Paper Award, the 2010 IEEE International Communications Conference Best Paper Award in Signal Processing for Communications, the 2013 International Symposium on Wireless Communication Systems Best Paper Award in Physical Layer Communications and Signal Processing, and a Best of IEEE GLOBECOM 2014—Top 50 Papers distinction for articles that he coauthored with students and colleagues. He was also a recipient of the 2009 SUNY-wide Chancellor's Award for Excellence in Teaching and the 2011 University at Buffalo Exceptional Scholar—Sustained Achievement Award.