Feature-Specific Difference Imaging

Shikhar Uttam, *Member, IEEE*, Nathan A. Goodman, *Senior Member, IEEE*, and Mark A. Neifeld, *Member, IEEE*

Abstract—Difference images quantify changes in the object scene over time. In this paper, we use the feature-specific imaging paradigm to present methods for estimating a sequence of difference images from a sequence of compressive measurements of the object scene. Our goal is twofold. First is to design, where possible, the optimal sensing matrix for taking compressive measurements. In scenarios where such sensing matrices are not tractable, we consider plausible candidate sensing matrices that either use the available *a priori* information or are nonadaptive. Second, we develop closed-form and iterative techniques for estimating the difference images. We specifically look at ℓ_2- and ℓ_1-based methods. We show that ℓ_2-based techniques can directly estimate the difference image from the measurements without first reconstructing the object scene. This direct estimation exploits the spatial and temporal correlations between the object scene at two consecutive time instants. We further develop a method to estimate a generalized difference image from multiple measurements and use it to estimate the sequence of difference images. For ℓ_1-based estimation, we consider modified forms of the total-variation method and basis pursuit denoising. We also look at a third method that directly exploits the sparsity of the difference image. We present results to show the efficacy of these techniques and discuss the advantages of each.

Index Terms—Compressive sensing (CS), difference images, feature-specific imaging (FSI), ℓ_1-reconstruction, ℓ_2-reconstruction.

I. INTRODUCTION

THE IDEA OF computing differences between images to better perform a given task is ubiquitous in research literature. Applications range from watermarking [1], [2] and material inspection [3] to video compression [4], biomedicine [5], astronomy [6], and change detection in remote sensing [7]. In this paper, using the concept of difference images, we present techniques to estimate a sequence of temporal changes in an object scene of interest by taking compressive measurements of the scene. Difference images result from subtracting the object scene at two time instants and therefore capture the changes

in the scene over time. The novelty here is that instead of conventionally imaging the scene, compressive projective measurements along with linear and nonlinear estimation techniques are used to estimate the changes by exploiting various properties such as spatiotemporal cross correlation and difference-image sparsity. The projective measurement of the object scene involves linearly mapping a higher dimensional object space to a lower dimensional measurement space leading to real-time data compression[1] and the resulting benefit in reduced sensor cost. This approach can be contrasted with the traditional approach of conventional imaging, where the goal is to obtain a pretty picture and then extract relevant information as a post-processing step. Another benefit is the improved detector-noise-limited measurement fidelity and, consequently, improved estimation performance particularly in the low signal-to-noise ratio (SNR) regime. This improvement is due to two related reasons: The first is we consider knowledge-enhanced measurements (or features), where we incorporate available *a priori* scene information to define the lower dimensional space where the object scene is projected. Second, due to this lower dimensional projection, the same number of object-scene photons are incident on a smaller number of photodetectors compared with the conventional case, leading to improved measurement fidelity for features that matter the most.

This notion of using task-specific features informed by *a priori* knowledge is referred to as feature-specific imaging (FSI) in the literature. The work by Neifeld and Shankar [8] was the first formalization of this idea. The work itself was motivated by earlier work in computational imaging related to hardware–software codesign [9], development of information metrics [10]–[13], and nontraditional [14] and novel imaging systems [15]–[17]. Here, our objective is twofold. First is to design, where possible, the optimal projection space for estimating difference images and, in scenarios where this is not possible, to propose plausible candidates. Second is to develop closed-form and iterative techniques for estimating the difference images. Toward this end, we consider estimation error minimization based on the ℓ_2 and ℓ_1 norms. For an $N \times 1$ vector $\mathbf{x}$, the ℓ_2 and ℓ_1 norms are defined as $\|\mathbf{x}\|_{\ell_1} = (\Sigma_{n=1}^{N} x_n^2)^{0.5}$ and $\|\mathbf{x}\|_{\ell_2} = \Sigma_{n=1}^{N} |x_n|$, respectively.

The ℓ_2-based linear estimation method provides an efficient closed form of the difference-image estimation operator [18]. We also show that ℓ_2-based estimation allows us to directly estimate the difference image without first reconstructing the object scene at the consecutive time instants. We show that an immediate consequence of direct estimation is the ability to exploit the spatiotemporal cross correlation between the object scene at the consecutive time instants. We further generalize the definition of the difference image to include the difference between

---

[1]We make the notion of compression precise in Section II.

the object scene at nonadjacent time instants and show how successive compressive measurements over the corresponding time interval can be used to directly estimate this generalized difference image.

We next study the ℓ_1-based estimation as it allows us to exploit the sparsity of the difference image. We set up the ℓ_1-based estimation problem as a linear inverse problem with different regularizers representing different points of view of the estimation problem. The first regularizing condition simply imposes the sparsity constraint on the difference image. We then consider a modified form of a total-variation (TV) regularizer. When we make the reasonable assumption that the image is a function of bounded variation (BV), TV is a natural measure used to capture edge discontinuities, i.e., an important feature in difference images. Lastly, we look at overcomplete representations by considering a modified form of basis pursuit (BP) denoising (BPDN). We empirically show that by using either available *a priori* information or that learned from training data, we get better performance than the compressive sensing (CS) paradigm of a sparsifying dictionary incoherently coupled with a random sensing matrix (e.g., Gaussian or Bernoulli/Radermacher matrices).

CS was introduced through a series of papers by Candés, Romberg, Tao [19]–[23], and Donoho [24]. To make the CS idea concrete, let us consider signal $\mathbf{x} \in \mathbb{R}^N$ that we sense (or measure) by projecting it onto the set of vectors $\boldsymbol{\phi}_m \in \mathbb{R}^N$ to get the following measurements:

$$y_m = \langle \boldsymbol{\phi}_m, \mathbf{x} \rangle, \quad m = 1, \ldots, M. \tag{1}$$

For a conventional imager, $\boldsymbol{\phi}_m$ will comprise the standard Euclidean basis with $M = N$ yielding a traditional image of the object scene. On the other hand, if the $\boldsymbol{\phi}_m$ values are, e.g., the discrete Fourier basis, then we take the frequency measurements of the object, as in magnetic resonance imaging. Using matrix notation, we can write (1) as $\mathbf{y} = \boldsymbol{\Phi}\mathbf{x}$. Given the set of measurements $\mathbf{y}$, the goal is to reconstruct signal $\mathbf{x}$. CS is interested in solving this problem when the system is underdetermined ($M \ll N$), i.e., the number of measurements is much less than the native dimensionality of the signal. Using data acquisition terminology, we take the undersampled measurements of the signal. In general, this problem is ill-posed and has infinitely many solutions. However, if we know that signal $\mathbf{x}$ is K-sparse in some basis, then CS proves that it is possible to recover $\mathbf{x}$ by making $M = O(K \log(N/K))$ measurements and by reconstructing with greedy algorithms or convex optimization methods. K-sparse means that the signal has only K nonzero coefficients in some basis $\boldsymbol{\Psi}$. A more realistic scenario is to consider a compressible signal with K largest coefficients containing most of the signal information. In this case, CS is able to recover the K coefficients. The main difference between CS and FSI is in the nature of the sensing matrix employed. In CS, the sensing matrix is a random matrix (Gaussian or Bernoulli/Radermarcher) satisfying the restricted isometry property. In FSI, on the other hand, the sensing matrix is designed using prior knowledge of the object scene. Either paradigm can be appropriate for a given situation, but if the particular exploitation task and the prior scene information are known, then the FSI paradigm can be very useful.

This paper and our earlier initial work [25] are the first application of the FSI paradigm to estimating difference images from compressive scene measurements. Recently, based on the CS idea, Wakin *et al.* [26] have proposed a single-pixel camera that sequentially takes compressive measurements of the scene. These measurements were then used to reconstruct the scene. Using this camera and incorporating difference images of the evolving background model and the test images, Cevher *et al.* [27] used CS theory to develop an interesting method to recover foreground innovations. Elsewhere [28], Cevher *et al.* used Ising support model to capture the clustered nature of foreground objects to reduce the number of compressive measurements required for robust recovery of the background-subtracted sparse image using a lattice-matching pursuit greedy algorithm. The use of a CS-based random sensing matrix allows the authors to exploit computer-vision ideas to perform background subtraction [29]. Our emphasis here is on the FSI sensing-matrix design in conjunction with various difference-image estimation techniques. Consequently, we do not focus on identifying foreground innovations within the difference images themselves. Another distinction is that employing random measurements allows Cehver *et al.* [27] to relate the object scene with the measurements through the central limit theorem. In our case, due to the structure that the sensing matrix possesses, such an association has not yet been established. Such an association would allow us to develop a sensing-matrix update model similar to a background update model employed in background subtraction. Here, we take the first steps in employing FSI ideas to difference-image estimation and defer sensing-matrix updates to future work.

In Section II, we give a formal definition of the difference image. In Section III, we detail our linear (ℓ_2)-based estimation method, whereas in Section IV, the ℓ_1-based estimation problem is discussed. Finally, in Section V, we present our results. We conclude in Section VI.

II. DIFFERENCE IMAGE

We define the difference image to be the residual image resulting from subtracting the object scene at two consecutive time instants from each other. Fig. 1 illustrates the basic idea behind difference images. Fig. 1(a) and (b) shows a scene with moving targets at two consecutive time instants. Fig. 1(c) shows the difference image resulting from subtracting imaged scene 1 from imaged scene 2. Let $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$ be the object scenes at consecutive time instants t_k and t_{k+1}, respectively. Then, we define the kth difference image to be

$$\Delta\mathbf{x}_k = \mathbf{x}_{k+1} - \mathbf{x}_k. \tag{2}$$

Furthermore, generalizing this definition, we define the generalized difference image between the object scene at kth and $(k + L)$th time instants as

$$\Delta\mathbf{x}_{kL} = \mathbf{x}_{k+L} - \mathbf{x}_k. \tag{3}$$

We will use this definition of the generalized difference image in Section III-G when we discuss a method to directly estimate
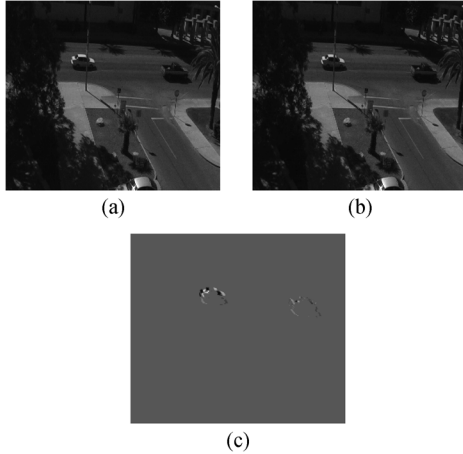
Fig. 1. (a) Object scene $\mathbf{x}_k$ at time instant t_k. (b) Object scene $\mathbf{x}_{k+1}$ at time instant t_{k+1}. (c) Resulting truth difference image $\Delta\mathbf{x} = \mathbf{x}_{k+1} - \mathbf{x}_k$.

a generalized difference image by taking compressive measurements over the time period L.

Compressive measurements reduce the dimensionality of the measured data compared with the native dimensionality of the scene. Inherent in this comparison is the discretization of the object space with respect to a certain object resolution δ_r. Let the object scene be K by S unit distance. Then, the pixel dimension of the scene is $k = K/\delta_r$ by $s = S/\delta_r$. Mathematically, the discretized scene is generally expressed in a vector form with dimension $ks \times 1$. Setting $N = ks$, we think of the scene as an $N \times 1$ vector. Therefore, N would be the dimension of the sensor array if conventional imaging were being used. Henceforth, when we talk about reduced dimensionality of compressive measurements, it will be with respect to this maximal conventional imaging dimension.

Measurements of the scene are taken with respect to a certain measurement basis whose matrix representation we refer to as the sensing matrix $\mathbf{\Phi}$. Compressive imaging systems optically project (inner product) scene $\mathbf{x}$ onto each row (measurement basis) ϕ_i, where $i = 1, \ldots, M(< N)$, of $\mathbf{\Phi}$ resulting in measurements $y_i = \phi_i \mathbf{x}$, where $i = 1, \ldots, M$. Many potential optical architectures can be designed for taking these compressive measurements. For a more detailed discussion of these optical architectures, we refer the reader to [30].

The FSI architectures perform incoherent imaging. Incoherent imaging systems are linear in intensity, thus limiting the sensing-matrix entries to positive values. However, the optimal mathematically derived sensing matrix can have negative entries. To bridge the gap between practice and theory, we need a way to physically implement the bipolar entries of the sensing matrix without violating the positivity requirement. One such method is dual-rail signaling consisting of two complementary arms. One arm implements $\mathbf{\Phi}_+$ (the positive entries of $\mathbf{\Phi}$ are kept, whereas the negative ones are set to zero) to get $\mathbf{y}_+ = \mathbf{\Phi}_+ \mathbf{x}$, and the second arm implements $\mathbf{\Phi}_-$ (the absolute values of the negative entries are kept, whereas the positive ones are set to zero) to get $\mathbf{y}_- = \mathbf{\Phi}_- \mathbf{x}$. The resulting measurement $\mathbf{y}$ is then given by

$\mathbf{y} = \mathbf{y}_+ - \mathbf{y}_- = \mathbf{\Phi}_+ \mathbf{x} - \mathbf{\Phi}_- \mathbf{x} = (\mathbf{\Phi}_+ - \mathbf{\Phi}_-)\mathbf{x} = \mathbf{\Phi}\mathbf{x}$. More flexibility can be added to this basic setup, as discussed in [8].

Another constraint on the sensing matrix comes from the passive nature of the optical architecture. An imager cannot increase the number of photons collected by the photo detector. In other words, the total number of photons entering the optical system is the same as the number leaving it. This condition manifests itself as the photon count constraint, which says that the absolute maximum column sum of $\mathbf{\Phi}$ (or the induced 1 norm of $\mathbf{\Phi}$) is 1. To ensure this constraint is met for the sensing matrix being considered, the sensing matrix has to be normalized by its maximum column sum. Let $c = \max_j \{\sum_i |\Phi_{ij}|\}$. Then, the sensing matrix satisfying the photon count constraint is given by $\mathbf{\Phi} = \mathbf{\Phi}/c$.

From here, we assume that we have the capability to implement optically the sensing matrix satisfying the photon count constraint.

III. LINEAR ℓ_2-BASED ESTIMATION

A. Data Model

Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be the object scene at the first two consecutive time instants t_1 and t_2, respectively. Following the explanation in the previous section, the scene at both time instants is assumed to be discretized and is represented as a vector of size $N \times 1$. Let us also define $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ to be the two corresponding optical sensing matrices of size $M \times N$. The M rows imply taking M measurements of the object scene. Thus, the sensing matrices can be thought of as $M \times N$ projection matrices that project the scene from an N-dimensional space to an M-dimensional subspace. Using these sensing matrices, we take measurements of the scene at the two time instants. The data model is given by

$$\mathbf{y}_1 = \mathbf{\Phi}_1 \mathbf{x}_1 + \mathbf{n}_1 \qquad (4)$$
$$\mathbf{y}_2 = \mathbf{\Phi}_2 \mathbf{x}_2 + \mathbf{n}_2 \qquad (5)$$

where $\mathbf{n}_1$ and $\mathbf{n}_2$ represent the sensor AWGN noise with the noise variance σ^2 and zero mean. Our goal is to estimate the difference image $\Delta\mathbf{x}_1$, given measurements $\mathbf{y}_1$ and $\mathbf{y}_2$, by finding the estimation operator that minimizes the ℓ_2 norm of the error between the truth difference image and the estimated difference image. If we take a Bayesian approach to the linear model in (4) and (5), then minimizing the ℓ_2 norm is the same as minimizing the Bayesian MSE (BMSE). The Bayesian assumption allows us to represent the scene as a stochastic process and, as a consequence, allows us to incorporate the (spatial) autocorrelation and (spatiotemporal) cross-correlation information between the scene at the two time instants in the estimation operator. Using estimation theory terminology, we call the estimation operator the linear MMSE (LMMSE) estimation operator.

B. Indirect Image Reconstruction

Before we present our method, we briefly discuss a possible approach—in line with the classical use of MMSE operators—for estimating difference images. We call this method intermediate image reconstruction (IIR). As illustrated in
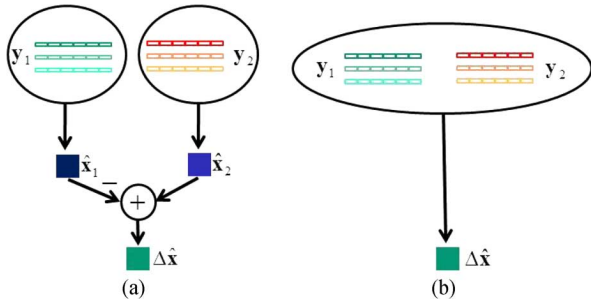
Fig. 2.   (a) IIR. (b) Direct difference image estimation.

Fig. 2(a), it involves reconstructing each object scene separately from its respective measurements and then subtracting these intermediate stage reconstructions to get the estimated difference image.

Reconstructing the object scene at both time instants means that (4) and (5) can be separated into two standalone problems. We define the reconstructed object scenes for the two time instants as

$$\hat{\mathbf{x}}_1 = \mathbf{F}_1 \mathbf{y}_1 \qquad \hat{\mathbf{x}}_2 = \mathbf{F}_2 \mathbf{y}_2 \qquad (6)$$

where $\mathbf{F}_i$ values, with $i = 1, 2$, are the linear reconstruction operators. For each i, we separately minimize the BMSE as

$$J(\mathbf{F}_i) = E\left[\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_{\ell_2}^2\right] \qquad (7)$$

with respect to $\mathbf{F}_i$.

The resulting reconstruction operators $\mathbf{F}_1$ and $\mathbf{F}_2$ are given by the well-known MMSE equation as follows:

$$\mathbf{F}_i = (\mathbf{R}_{\mathbf{x}}^{-1} + \mathbf{\Phi}_i^T \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{\Phi}_i)^{-1} \mathbf{\Phi}_i^T \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{y}, \qquad i = 1, 2 \quad (8)$$

where $\mathbf{R}_{\mathbf{x}}$ is the autocorrelation matrix of the object scene and $\mathbf{R}_{\mathbf{n}}$ is the noise covariance matrix. We always assume that we have already subtracted off the mean from the scene. If this is not the case, we can trivially modify (8) to account for the mean. If we make the additional assumption that the first two moments completely describe the scene statistics, then (8) will be the optimal solution. These assumptions, however, are rarely true in practice. Despite this restriction, as shown in Section V, it turns out that LMMSE operators are good and computationally efficient estimation operators.

Given the reconstruction operators, the indirectly estimated difference image is given by

$$\mathbf{\Delta}\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1 = \mathbf{F}_2 \mathbf{y}_2 - \mathbf{F}_1 \mathbf{y}_1. \qquad (9)$$

C. DDIE

The intermediate step involving object-scene reconstruction in the IIR method is an unnecessary step. If we remove that step by reconstructing the difference image directly from the measurements $\mathbf{y}_1$ and $\mathbf{y}_2$, we can better estimate the truth difference image. The reason is that we can incorporate not only the spatial correlation between pixels (autocorrelation of the scene) but also the temporal correlation (cross correlation between the

scene at the two time instants). We define the estimated difference image as

$$\mathbf{\Delta}\hat{\mathbf{x}}_1 = \mathbf{W}_1 \mathbf{y}_1 + \mathbf{W}_2 \mathbf{y}_2 \qquad (10)$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are the jointly optimized estimation operators. We call this the direct difference-image estimation (DDIE) technique. It is visualized in Fig. 2(b). We start by looking at the simplest case of no sensor noise.

D. DDIE: Noise Absent

Our DDIE approach makes an initial assumption of perfect knowledge of the scene at the instant we start taking measurements. From a system's perspective, this is a reasonable assumption to make. For example, the initial knowledge can be obtained from a sensor that has been on the scene for a long period of time. Therefore, for the time instant t_1, we assume perfect knowledge of the scene ($\mathbf{\Phi}_1$ is an identity matrix), and from t_2 onward, we begin taking compressive measurements of the scene. We can now rewrite the data models (4) and (5) as

$$\mathbf{y}_1 = \mathbf{I}\mathbf{x}_1 \qquad \mathbf{y}_2 = \mathbf{\Phi}_2 \mathbf{x}_2. \qquad (11)$$

The BMSE we have to minimize is

$$J(\mathbf{W}_1, \mathbf{W}_2) = E\left[\|\mathbf{\Delta}\mathbf{x} - \mathbf{\Delta}\hat{\mathbf{x}}\|_{\ell_2}^2\right]. \qquad (12)$$

Differentiating (12) with respect to $\mathbf{W}_1$ and $\mathbf{W}_2$ and equating the two derivatives to zero, we get the jointly optimized estimation operators to be

$$\mathbf{W}_1 = \left((\mathbf{R}_{21} - \mathbf{R}_{11}) - \mathbf{R}_\delta \mathbf{\Phi}_2^T \left(\mathbf{\Phi}_2 \mathbf{R}_\delta \mathbf{\Phi}_2^T\right)^{-1} \mathbf{\Phi}_2\right) \mathbf{R}_{21} \mathbf{R}_{11}$$
$$(13)$$

$$\mathbf{W}_2 = \mathbf{R}_\delta \mathbf{\Phi}_2^T \left(\mathbf{\Phi}_2 \mathbf{R}_\delta \mathbf{\Phi}_2^T\right)^{-1} \qquad (14)$$

where $\mathbf{R}_\delta = \mathbf{R}_{22} - \mathbf{R}_{12}^T \mathbf{R}_{11}^{-1} \mathbf{R}_{12}$, with $\mathbf{R}_{21} = \mathbf{R}_{12}^T$, is the (spatiotemporal) cross-correlation matrix between the scene at two consecutive time steps and $\mathbf{R}_{11}$ and $\mathbf{R}_{22}$ are the (spatial) autocorrelation matrices of the scene at the two time instants.

Now that we know the reconstruction operators, it is possible to find the optimal sensing matrix (in the ℓ_2 sense). In fact, it is given by

$$\mathbf{\Phi}_2 = [\mathbf{X} \quad \mathbf{0}] \mathbf{Q}^T \qquad (15)$$

where $\mathbf{Q}$ is the matrix of the eigenvectors of $\mathbf{R}_\delta$ and $\mathbf{X}$ is any rank-M orthonormal matrix. This is an expected result as finding a sensing matrix that minimizes the MSE between the truth and the estimated difference images in the absence of noise is analogous to finding a matrix that maximizes the projection variance of the object scene, and this is the principal component (PC) solution. Looking at (15), we see that $[\mathbf{X} \ \mathbf{0}]$ picks out the first M eigenvectors of $\mathbf{R}_\delta$ to form an M-dimensional subspace where the projection variance is maximized. Since $\mathbf{X}$ is a rank-M orthonormal matrix, we get a rotated M-dimensional subspace. As a simple case, the orthonormal matrix can be an identity matrix; in which case, the eigenvectors are the PCs. It

is interesting to note that the optimal sensing-matrix solution involves the eigenvectors of $\mathbf{R}_\delta$. $\mathbf{R}_\delta$ can be interpreted in the following way: Given the spatial autocorrelation of scene 1 and the spatiotemporal cross correlation between scenes 1 and 2, $\mathbf{R}_\delta$ contains the extra information we get from the spatial autocorrelation of scene 2. The optimal sensing matrix selects the directions that maximize this information.

E. DDIE: Noise Present

In the presence of noise, the optimal sensing operators must be modified. Due to noise, the correlation information and $\mathbf{R}_\delta$ are affected. The optimal LMMSE estimation operators in the presence of noise are

$$\mathbf{W}_1 = \Big((\mathbf{R}_{21} \quad - \mathbf{R}_{11}) - (\mathbf{R}_{\boldsymbol{\alpha}} - \mathbf{R}_{\boldsymbol{\beta}})\boldsymbol{\Phi}_2^T$$
$$\times \Big(\boldsymbol{\Phi}_2\mathbf{R}_\alpha\boldsymbol{\Phi}_2^T + \mathbf{R}_{n_2} \Big)^{-1} \boldsymbol{\Phi}_2\mathbf{R}_{21} \Big) (\mathbf{R}_{11} - \mathbf{R}_{n_1})^{-1}$$
$$(16)$$

$$\mathbf{W}_2 = (\mathbf{R}_{\boldsymbol{\alpha}} - \mathbf{R}_{\boldsymbol{\beta}})\boldsymbol{\Phi}_2^T \Big(\boldsymbol{\Phi}_2\mathbf{R}_\alpha\boldsymbol{\Phi}_2^T + \mathbf{R}_{n_2} \Big)^{-1} \qquad (17)$$

where

$$\mathbf{R}_{\boldsymbol{\alpha}} = \mathbf{R}_{22} - \mathbf{R}_{21}(\mathbf{R}_{11} + \mathbf{R}_{n_1})^{-1}\mathbf{R}_{12}$$
$$\mathbf{R}_{\boldsymbol{\beta}} = \mathbf{R}_{12} - \mathbf{R}_{11}(\mathbf{R}_{11} + \mathbf{R}_{n_1})^{-1}\mathbf{R}_{12}. \qquad (18)$$

Here, the no-noise case $\mathbf{R}_\delta$ is modified to $\mathbf{R}_{\boldsymbol{\alpha}} - \mathbf{R}_{\boldsymbol{\beta}}$. Matrix $\mathbf{R}_{\boldsymbol{\beta}}$ reflects the loss in correlation information in the presence of noise. If the noise were zero, then $\mathbf{R}_{\boldsymbol{\beta}}$ would go to zero and $\mathbf{R}_{\boldsymbol{\alpha}} - \mathbf{R}_{\boldsymbol{\beta}}$ would be identical to $\mathbf{R}_\delta$. However, in the presence of noise, there is a reduction in the available correlation information, and this reduction is quantified by $\mathbf{R}_{\boldsymbol{\beta}}$.

In the presence of noise, finding an optimal sensing matrix is mathematically intractable. As a consequence, we look at a few plausible candidate sensing matrices.

F. Sensing Matrices

PCA: We start by looking at two kinds of PCs. For the first case, we let the rows of the sensing matrix $\boldsymbol{\Phi}$ be the eigenvectors of the spatial autocorrelation matrix. To a small extent, this is similar to the solution for the no-noise case in (15) if we were to let $\mathbf{X}$ be an $M \times M$ identity matrix. However, this case only considers the spatial correlation and ignores the temporal correlation. To utilize the temporal correlation information, we also consider the difference PCs (DPCs). DPCs are the PCs of the difference image. We compute them from the spatiotemporal correlation matrix of the difference images defined as $\mathbf{R}_{\Delta\mathbf{x}_1} = \mathrm{E}[(\mathbf{x}_2 - \mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)^T] = 2\mathbf{R}_{11} - \mathbf{R}_{12} - \mathbf{R}_{21}$. Since $\mathbf{R}_{\Delta\mathbf{x}_1}$ is a symmetric matrix, its spectral factorization will give us the DPCs. There is a twofold advantage to DPCs. First, they implicitly use both spatial and temporal correlation information. Second, as we are trying to reconstruct the difference images and not the object scene itself, the PCs of the difference image are more suitable than PCs.

PCA waterfilling: PCA is a suboptimal solution in the presence of noise as it does not adjust the energies (eigenvalues) of the eigenvectors with changing SNR. We remedy this by considering weighted PCA, which redistributes the total available energy among the different eigenvectors while accounting for noise. This redistribution is achieved by maximizing the mutual information $I(\mathbf{x}; \mathbf{y})$ between scene $\mathbf{x}$ and measurement $\mathbf{y}$, assuming they are $N \times 1$ and $M \times 1$ random vectors, respectively. We briefly discuss the suboptimality of PCA and then give the weighted solution.

Let $\mathbf{R_x}$ be the correlation matrix of scene $\mathbf{x}$. Then, $\mathbf{R_x} = \mathbf{U\Lambda U}^T$ gives the eigendecomposition of $\mathbf{R_x}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues $(\lambda_i, i = 1, \ldots, N)$ in decreasing order and columns of $\mathbf{U}$ are the corresponding eigenvectors. Now, let noise be added, and let the noise covariance matrix be given by $\sigma^2\mathbf{I}$. Note that the eigenvectors in $\mathbf{U}$ are also the eigenvectors for white noise. As a result, in the presence of noise, the eigenvalues are given by $\boldsymbol{\Lambda} + \sigma^2\mathbf{I}$. Therefore, the presence of noise simply adds the noise variance to all the eigenvalues without adapting the eigenspectrum to the given SNR.

The PC sensing matrix $\boldsymbol{\Phi}_{pc} = (U(1 : N, 1 : M))^T$ makes measurement $\mathbf{y}$, which lies in the subspace spanned by the first M eigenvectors. We now consider the modified sensing matrix $\boldsymbol{\Phi}_{wpc} = \mathbf{D_w}(U(1 : N, 1 : M))^T$, where $\mathbf{D_w} = diag(w_1, \ldots, w_M)$. We are still in the subspace spanned by the first M eigenvectors, but now, the diagonal elements w_i, where $i = 1, \ldots, M$, control the weighting given to each eigenvector. We first maximize $I(\mathbf{x}; \mathbf{y})$, given an unknown but fixed $\boldsymbol{\Phi}_{wpc}$, and then use it to compute the weights w_i, where $i = 1, \ldots, M$. By maximizing the mutual information over the input distribution of $\mathbf{x}$, we get

$$I(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{M} \log\left(\frac{w_i^2\lambda_i + P_n}{P_n} \right) \qquad (19)$$

where the maximizing input distribution is a multivariate Gaussian. We know that a real-world object scene is not normally distributed, but nevertheless, we show in Section V that we still get marked improvement over PCs. In the ideal scenario of the scene being normally distributed, this solution will be optimal. We assume the logarithm base to be 2.

To find the optimal weights w_i, $i = 1, \ldots, M$, we differentiate (19) with respect to w_i^2 under constraint $\Sigma_{i=1}^{M} w_i^2 = E$, where E is the total energy in the object scene. Using the Lagrange multiplier and requiring $w_i^2 \geq 0$, the weights are

$$w_i^2 = \left(\frac{1}{\zeta} - \frac{P_n}{\lambda_i} \right)^+. \qquad (20)$$

We choose the value of the Lagrange multiplier ζ such that $\Sigma_{i=1}^{M}((1/\zeta) - (P_n/\lambda_i))^+ = E$. From (20), we see that the weights assigned to the different eigenvectors are a function of λ_i/P_n and not just λ_i. According to (20), we must put the available energy where λ_i/P_n is large. This is the waterfilling solution [31]. We perform waterfilling for both PCs and DPCs resulting in sensing matrices, i.e., waterfilled PCs (WPCs) and DPCs (WDPCs), respectively.

Optimal solution: Since it is not mathematically tractable to find an optimal sensing matrix in the presence of noise, we also numerically search for the optimal solution. We perform this search using stochastic tunneling [32], i.e., a Monte Carlo-based technique.
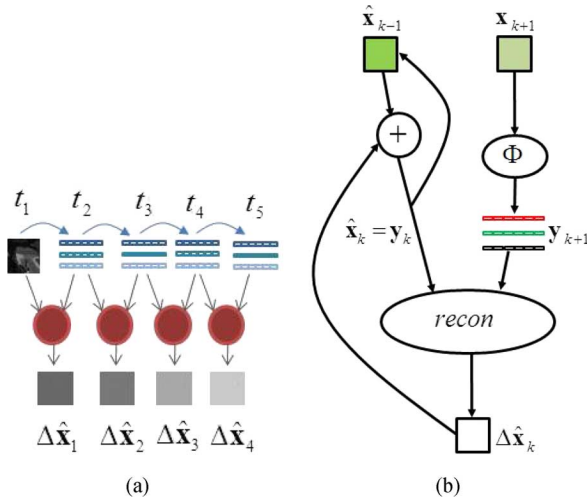
Fig. 3. Multistep DDIE. (a) Perfect knowledge of the scene is assumed at the time instant t_1, and measurements are made from t_2 on. Difference image is estimated by propagating forward the object-scene knowledge. This propagation is indicated by curved arrows from left to right. (b) At each pair of consecutive time instants t_k and t_{k+1}, (21) is implemented to estimate the difference image $\mathbf{\Delta x}_k$ and propagate the scene knowledge.

G. Multistep DDIE and LFGDIE

As depicted in Fig. 3(a), we assume knowledge of the scene at the first time instant and, from then on, take measurements of the scene. Our model allows us to use a different sensing matrix $\mathbf{\Phi}$ at every successive time instant. For simplicity, however, we assume $\mathbf{\Phi}_k = \mathbf{\Phi}$, where $k > 1$. Consequently, we have the $\{\mathbf{x}_1, \mathbf{\Phi x}_2, \mathbf{\Phi x}_3, \ldots, \mathbf{\Phi x}_L, \ldots\} = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_L, \ldots\}$ sequence of measurements. Until now, we have looked at the DDIE method for estimating the difference image between the object scene at the first two time instants. We now extend the DDIE method to estimate the $\{\mathbf{\Delta}\hat{x}_1, \mathbf{\Delta}\hat{x}_2, \ldots, \mathbf{\Delta}\hat{x}_{L-1}, \ldots\}$ sequence of difference images from the $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_L, \ldots\}$ sequence of measurements. We call this strategy the multistep DDIE. We also present a different approach to estimating the difference-image sequence by jointly using measurements taken over multiple time instants.

The DDIE method assumes knowledge of scene 1 and takes measurements of scene 2 to estimate the difference image. Therefore, in the multistep DDIE, given measurements of the object scenes at t_k and t_{k+1}, we need the knowledge of the scene at t_k. The multistep DDIE acquires this knowledge by propagating forward the knowledge of the scene at t_1 [see Fig. 3(a)]. The forward propagation is done using the following recursive equation:

$$\mathbf{\Delta}\hat{x}_k = \text{recon}(\hat{x}_k, \mathbf{y}_{k+1}), \quad k > 1$$
$$= \text{recon}\left((\hat{x}_{k-1} + \mathbf{\Delta}\hat{x}_{k-1}), \mathbf{y}_{k+1}\right), \quad k > 1. \quad (21)$$

where recon refers to the DDIE method discussed in Section III-E. For $k = 1$, we replace $\hat{x}_1$ with $\mathbf{x}_1$ because we assume the perfect knowledge of the scene at t_1. Equation (21) takes the estimate of the scene at t_k and measurements at t_{k+1} and estimates the difference image $\mathbf{\Delta}\hat{x}_k$. It then propagates forward the knowledge of the scene at t_k by computing

the estimate $\hat{x}_{k+1} = \hat{x}_k + \mathbf{\Delta}\hat{x}_k$ at t_{k+1}. This is illustrated in Fig. 3(b).

We refer to our second approach as the Lth-frame generalized difference-image estimation (LFGDIE). Given $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_L\}$, the LFGDIE directly estimates the generalized difference image $\mathbf{\Delta x}_{1L}$ between the object scene at t_1 and t_L.

To obtain the LFGDIE operators, we first define the LFGDIE data model as

$$\mathbf{y}_1 = \mathbf{I x}_1 + \mathbf{n}_1 \quad (22)$$
$$\mathbf{y}_2 = \mathbf{\Phi x}_2 + \mathbf{n}_2 \quad (23)$$
$$\cdots$$
$$\mathbf{y}_L = \mathbf{\Phi x}_L + \mathbf{n}_L \quad (24)$$

where $\mathbf{I}$ is the identity sensing matrix symbolizing complete knowledge of the initial scene. Note that this model is an extension of (4) and (5) because, here, we consider multiple measurements. The estimated generalized difference image is then given by

$$\mathbf{\Delta}\hat{x}_{1L} = \sum_{i=1}^{L} \mathbf{W}'_i \mathbf{y}_i \quad (25)$$

where $\mathbf{W}'_i$ values, with $i = 1, \ldots, L$, are the jointly optimized estimation operators. Rewriting (25) in the matrix form, we have

$$\mathbf{\Delta}\hat{x}_{1L} = \begin{bmatrix} \mathbf{W}'_1 & \mathbf{W}'_p & \mathbf{W}'_L \end{bmatrix} \begin{bmatrix} \mathbf{y}_1^T & \mathbf{y}_p^T & \mathbf{y}_L^T \end{bmatrix}^T \quad (26)$$

where $\mathbf{W}'_p = \begin{bmatrix} \mathbf{W}'_2 & \mathbf{W}'_3 & \cdots & \mathbf{W}'_{L-1} \end{bmatrix}$ and $\mathbf{y}_p = \begin{bmatrix} \mathbf{y}_2^T & \mathbf{y}_3^T & \cdots & \mathbf{y}_{L-1}^T \end{bmatrix}^T$. Let us also define $\mathbf{x}_p = \begin{bmatrix} \mathbf{x}_2^T & \mathbf{x}_3^T & \cdots & \mathbf{x}_{L-1}^T \end{bmatrix}^T$, $\mathbf{R}_{11} = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^T]$, $\mathbf{R}_{L1} = \mathrm{E}[\mathbf{x}_L \mathbf{x}_1^T] = \mathbf{R}_{1L}^T$, $\mathbf{R}_{p1} = \mathrm{E}[\mathbf{x}_p \mathbf{x}_1^T] = \mathbf{R}_{1p}^T$, $\mathbf{R}_{pL} = \mathrm{E}[\mathbf{x}_p \mathbf{x}_L^T] = \mathbf{R}_{Lp}^T$, and $\mathbf{\Phi}_p = \mathbf{I}_p \otimes \mathbf{\Phi}$, where $\mathbf{I}_p$ is a $p \times p$ identity matrix.

Minimizing the BMSE between $\mathbf{\Delta x}_{1L}$ and $\mathbf{\Delta}\hat{x}_{1L}$ by differentiating it with respect to $\mathbf{W}'_1$, $\mathbf{W}'_p$, and $\mathbf{W}'_L$, and equating the derivatives to zero, the reconstruction operators are

$$\mathbf{W}'_1 = \left(\mathbf{R}_{L1} - \mathbf{R}_{11} - \mathbf{W}'_p \mathbf{\Phi}_p \mathbf{R}_{p1} - \mathbf{W}'_L \mathbf{\Phi}_L \mathbf{R}_{L1}\right)$$
$$\times (\mathbf{R}_{11} + \mathbf{R}_{n_1})^{-1} \quad (27)$$

$$\mathbf{W}'_p = \left(\mathbf{R}_{\alpha p} + \mathbf{R}_{\alpha L} \mathbf{\Phi}_L^T \left(\mathbf{\Phi}_L \mathbf{R}_{L \gamma L} \mathbf{\Phi}_L^T + \mathbf{R}_{n_L}\right)^{-1} \mathbf{\Phi}_L \mathbf{R}_{\beta p}\right)$$
$$\times \mathbf{\Phi}_p^T \left(\mathbf{\Phi}_p \mathbf{R}_{\Phi_L} \mathbf{\Phi}_p^T + \mathbf{R}_{n_p}\right)^{-1} \quad (28)$$

$$\mathbf{W}'_L = (\mathbf{R}_{\alpha L} + \mathbf{W}'_p \mathbf{\Phi}_p \mathbf{R}_{\beta L})$$
$$\times \mathbf{\Phi}_L^T \left(\mathbf{\Phi}_L \mathbf{R}_{L \gamma L} \mathbf{\Phi}_L^T + \mathbf{R}_{n_L}\right)^{-1} \quad (29)$$

where

$$\mathbf{R}_{\alpha L} = \mathbf{R}_{LL} - \mathbf{R}_{1L} - (\mathbf{R}_{L1} - \mathbf{R}_{11})$$
$$\times (\mathbf{R}_{11} + \mathbf{R}_{n_1})^{-1} \mathbf{R}_{1L} \quad (30)$$

$$\mathbf{R}_{\alpha p} = \mathbf{R}_{Lp} - \mathbf{R}_{1p} - (\mathbf{R}_{L1} - \mathbf{R}_{11})$$
$$\times (\mathbf{R}_{11} + \mathbf{R}_{n_1})^{-1} \mathbf{R}_{1p} \quad (31)$$

$$\mathbf{R}_{\beta L} = \mathbf{R}_{p1} (\mathbf{R}_{11} + \mathbf{R}_{n_1})^{-1} \mathbf{R}_{1L} - \mathbf{R}_{pL} \quad (32)$$

$$\mathbf{R}_{\beta p} = \mathbf{R}_{L1} (\mathbf{R}_{11} + \mathbf{R}_{n_1})^{-1} \mathbf{R}_{1p} - \mathbf{R}_{Lp} \quad (33)$$

$$\mathbf{R}_{\gamma L} = \mathbf{R}_{L1} (\mathbf{R}_{11} + \mathbf{R}_{n_1})^{-1} \mathbf{R}_{1L} \quad (34)$$

$$\mathbf{R}_{\gamma p} = \mathbf{R}_{p1}(\mathbf{R}_{11} + \mathbf{R}_{n_1})^{-1}\mathbf{R}_{1p} \tag{35}$$

$$\mathbf{R}_{L\gamma L} = \mathbf{R}_{LL} - \mathbf{R}_{\gamma L} \tag{36}$$

$$\mathbf{R}_{\Phi_L} = \mathbf{R}_{pp} - \mathbf{R}_{\gamma p} - \mathbf{R}_{\beta L}\boldsymbol{\Phi}_L^T\left(\boldsymbol{\Phi}_L\mathbf{R}_{L\gamma L}\boldsymbol{\Phi}_L^T + \mathbf{R}_{n_L}\right)^{-1}$$
$$\times \boldsymbol{\Phi}_L\mathbf{R}_{\beta p}. \tag{37}$$

It is interesting to note that when $L = 2$, i.e., $p = 0$, $\mathbf{R}_{L\gamma L} = \mathbf{R}_\alpha$, $\mathbf{R}_{\alpha L} = \mathbf{R}_\alpha - \mathbf{R}_\beta$, $\mathbf{W}_p'$ disappears, and $\mathbf{W}_1' = \mathbf{W}_1$ and $\mathbf{W}_L' = \mathbf{W}_2$. Therefore, the LFGDIE method reduces to the DDIE method. It is when $L > 2$ that we see the benefit of employing LFGDIE method. To see this, let $\mathbf{x}_1$ and $\mathbf{x}_L$ be the object scenes at time instants t_1 and t_L. Then, the generalized difference image is given by $\boldsymbol{\Delta}\mathbf{x}_{1L} = \mathbf{x}_L - \mathbf{x}_1$. By rewriting this generalized difference image, we get

$$\boldsymbol{\Delta}\mathbf{x}_{1L} = (\mathbf{x}_L - \mathbf{x}_{L-1}) + (\mathbf{x}_{L-1} - \mathbf{x}_{L-2})$$
$$+ \cdots + (\mathbf{x}_3 - \mathbf{x}_2) + (\mathbf{x}_2 - \mathbf{x}_1)$$
$$= \sum_{i=1}^{L-1} \boldsymbol{\Delta}\mathbf{x}_i \tag{38}$$

where the right-hand side is a pairwise sum of difference images of the scene at successive time instants. The LFGDIE estimate of the generalized difference image $\boldsymbol{\Delta}\mathbf{x}_{1L}$ is $\boldsymbol{\Delta}\hat{\mathbf{x}}_{1L} = \widehat{\sum_{i-1}^{L-1} \boldsymbol{\Delta}\mathbf{x}_i}$, which requires joint estimation of all the successive pairwise difference images $\boldsymbol{\Delta}\mathbf{x}_i$. This joint estimation exploits the spatial and temporal cross correlations between the scene at all L time steps, as is manifested in (27)–(37). The multistep DDIE method on the other hand estimates $\sum_{i-1}^{L-1} \boldsymbol{\Delta}\hat{\mathbf{x}}_i$, which exploits only pairwise cross correlation between the scene at two successive time steps. This ability of the LFGDIE method to perform joint estimations leads to its superior performance over the multistep DDIE as we show in Section V.

IV. ℓ_1-BASED ESTIMATION

The advantage of the linear ℓ_2-based estimation lies in its ability to provide closed-form linear estimation operators that minimize the MSE over an entire ensemble of object scenes. Difference images, however, are sparse, and the ℓ_2-based difference-image estimation does not exploit this characteristic. Moreover, the use of a spatiotemporal correlation matrix, despite good results discussed in the results section, assumes image stationarity, which is a crude assumption. We therefore extend our study to look at the ℓ_1-based estimation of the difference images. We are motivated by a few reasons; each of which looks at the problem from a different perspective. First, as mentioned above, difference images are sparsely represented in pixel space (a finite-dimensional Euclidean space), and exploiting this sparsity for difference-image estimation seems to be a natural extension of the image restoration problem. Second, modeling optical images as functions of BV has been successfully used in image denoising and restoration. The use of an ℓ_1-based TV measure [33] in this context has been shown to estimate accurately edge features, which are important components in difference images. Third, signal decomposition using overcomplete dictionaries gives sparse signal representations with respect to atoms of these dictionaries. It has been

shown that BP gives an optimal (in ℓ_1 sense) solution for this signal decomposition problem [34]. These three approaches fit well into the ℓ_1-based estimation of the difference image.

We consider ℓ_1-based difference-image estimation problem as a linear inverse problem. The linearity comes from the forward data model being defined through the linear transform $\mathbf{D}$ (applied to input $\mathbf{s}$ in the presence of noise), i.e.,

$$\mathbf{y} = \mathbf{D}\mathbf{s} + \mathbf{n}. \tag{39}$$

The goal then of the linear inverse problem is to estimate $\mathbf{s}$, given the noisy measurements $\mathbf{y}$. The model in (39) is typical of ℓ_1-based reconstruction problems. However, the model we have in (4) and (5) is not of the same form as (39). Consequently, we rewrite (4) and (5) as

$$\mathbf{y} = \boldsymbol{\Phi}_D\mathbf{x} + \mathbf{n} \tag{40}$$

where

$$\boldsymbol{\Phi}_D = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \quad \mathbf{n} = \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \end{bmatrix}. \tag{41}$$

Incorporating a sparse representation of $\mathbf{x}$ with respect to a sparsifying dictionary $\boldsymbol{\Psi}$ in (40), we have

$$\mathbf{y} = \boldsymbol{\Phi}_D\boldsymbol{\Psi}\mathbf{s} + \mathbf{n} \tag{42}$$

where $\mathbf{s}$ is a sparse representation of $\mathbf{x}$, i.e., $\mathbf{x} = \boldsymbol{\Psi}\mathbf{s}$ (comparing (39) with (42), we have $\mathbf{D} = \boldsymbol{\Phi}_D\boldsymbol{\Psi}$.) Solution $\hat{\mathbf{s}}$ and, therefore $\hat{\mathbf{x}}$, to the linear inverse problem is then given by solving the following optimization problem:

$$\arg\min_{\mathbf{s}} \|\mathbf{y} - \boldsymbol{\Phi}_D\boldsymbol{\Psi}\mathbf{s}\|_{\ell_2}^2 + \xi R(\mathbf{s}). \tag{43}$$

Here, the ℓ_2 term is the fitness term controlling how much of a fit the solution is to the measured data, whereas $R(\mathbf{s})$ is the regularizer term controlling how much the solution meets the desired constraint. We define $R(\mathbf{s})$ to be an ℓ_1 convex regularizer; its form being decided by the three points of view we are considering. The weighting factor ξ is the regularization parameter.

From the first point of view, difference images are sparse in the pixel basis. The pixel basis can be thought of as the standard basis for a finite-dimensional Euclidean space, where the dimension is that of the object scene. Consequently, $\boldsymbol{\Psi}$ is the identity matrix and $\mathbf{s} = \mathbf{x}$. However, to maximize sparsity, we define regularizer R to be a function of $\boldsymbol{\Delta}\mathbf{x}_1$ instead of $\mathbf{s}$ as follows:

$$R(\boldsymbol{\Delta}\mathbf{x}_1) = \|\mathbf{x}_2 - \mathbf{x}_1\|_{\ell_1} = \|\boldsymbol{\Delta}\mathbf{x}_1\|_{\ell_1}. \tag{44}$$

It is evident that this ℓ_1 regularizer enforces the sparsity constraint on the difference image by favoring values closer to 0. Note that regularizer $R(s) = \|s\|_{\ell_1}$ does not maximize the sparsity of the difference image but just minimizes the ℓ_1 norm of $\mathbf{s}$ and, consequently, is not optimal.

For the TV restoration problem, we again consider $\boldsymbol{\Psi}$ to be an identity matrix because, in this formulation, the function space

of BV is defined to be on a discrete finite support. Regularizer R for the TV problem is usually defined to be either

$$R_{\text{iso}}(\mathbf{x}) = \sum_i \sqrt{(\Delta_i^h \mathbf{x})^2 + (\Delta_i^v \mathbf{x})^2} \qquad (45)$$

or

$$R_{\text{niso}}(\mathbf{x}) = \sum_i \left| \Delta_i^h \mathbf{x} \right| + \left| \Delta_i^v \mathbf{x} \right| \qquad (46)$$

where $R_{\text{iso}}(\mathbf{x})$ and $R_{\text{niso}}(\mathbf{x})$ are the isotropic and nonisotropic discrete TV regularizers, respectively. The Δ_i^h and Δ_i^v operators are, respectively, the first-order horizontal and vertical difference operators. Instead of imposing the TV condition on $\mathbf{s}\ (= \mathbf{x})$, however, as in (44), we impose it on the difference image $\Delta \mathbf{x}_1 = \mathbf{x}_2 - \mathbf{x}_1$. Now, the regularizer is defined as either

$$R_{\text{iso}}(\Delta \mathbf{x}_1) = \sum_i \sqrt{(\Delta_i^h \Delta \mathbf{x}_1)^2 + (\Delta_i^v \Delta \mathbf{x}_1)^2} \qquad (47)$$

or

$$R_{\text{niso}}(\Delta \mathbf{x}_1) = \sum_i \left| \Delta_i^h \Delta \mathbf{x}_1 \right| + \left| \Delta_i^v \Delta \mathbf{x}_1 \right|. \qquad (48)$$

It is easy to see that a bounded TV assumption for the object scenes $\mathbf{x}_1$ and $\mathbf{x}_2$ results in $\mathbf{x}_2 - \mathbf{x}_1$ also having BV. Therefore, this modified form does not violate any TV condition.

For the sake of completeness, we also look at the difference-image estimation using an overcomplete sparsifying dictionary Ψ. Specifically, we consider the dictionary to be the symmetric biorthogonal wavelet transform and the regularizer to be $R(\mathbf{s}) = \|\mathbf{s}\|_{\ell_1}$. This results in the familiar BPDN model, which decomposes the signal as superposition of the atoms of Ψ such that the ℓ_1 norm of $\mathbf{s}$ is smallest of all possible decompositions over the dictionary. However, the set up here is slightly different from traditional BPDN in that we include the sensing matrix Φ_D in the model. In traditional BPDN, the sensing matrix is an identity matrix. This modification does not affect the fundamental problem. The classical BPDN problem can be thought of as finding the regularized denoised sparse representation of the object scene from the noisy version of the scene. The modified BPDN, on the other hand, finds the regularized denoised sparse representation of the object scene from measurements of the scene using the sensing matrix Φ_D.

Notice that, by applying (42), we use $\mathbf{s}$ to estimate $\mathbf{x} = [\mathbf{x}_1^T\ \mathbf{x}_2^T]^T$ and not $\Delta \mathbf{x}_1$. Therefore, because of the system constraint, there is the additional step of computing $\Delta \mathbf{x}_1$ from the estimated $\mathbf{x}$. Estimating $\mathbf{s}$, however, allows us to take advantage of the correlation between the object scene at the two time instants. By solving (43), with $R(\mathbf{s}) = \|\mathbf{s}\|_{\ell_1}$, we compute a joint estimate of $\mathbf{x}_1$ and $\mathbf{x}_2$ in the form of $\mathbf{s}$. Note that, although $\hat{\mathbf{x}}$ is separable into $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$, $\mathbf{s}$ is not. Similarly, in (40), we jointly exploit the scene at the two time instants by considering regularization terms (44), (47) and (48) that are functions of the difference image. Regularizer (44) sparsifies the difference image, whereas regularizers (47) and (48) minimize the TV of the difference image. In fact, by acting directly on the difference image, (40) exploits the correlation between the scene at the two time instants more strongly than (42).

Extension to estimating sequences of difference images follows directly from the multistep DDIE and, more specifically, (21). We will use the ℓ_1-based techniques in the multistep setting. In Section V, we present the performance of these three approaches.

A. Learning Sensing and Sparsifying Matrices

Given the above approaches to ℓ_1-based difference-image estimation, one question that naturally arises is whether we can learn the optimal Φ and Ψ from the available training data. Recently, Carvajalino and Sapiro [35] proposed a very interesting scheme to learn simultaneously Φ_D and Ψ using training data. The sparsifying dictionary is assumed to be overcomplete. This assumption makes it difficult for us to apply it to our current context. Consider the forward model for computing the difference image from the object scene at two consecutive time instants. Let the time instants be t_1 and t_2. Then, we have

$$\Delta \mathbf{x}_1 = \mathbf{x}_2 - \mathbf{x}_1 = [-\mathbf{I}\ \mathbf{I}] \left[\mathbf{x}_1^T\ \mathbf{x}_2^T\right]^T = [-\mathbf{I}\ \mathbf{I}]\mathbf{x}. \qquad (49)$$

Here, $[-\mathbf{I}\ \mathbf{I}]$ is a $N \times 2N$ matrix with both $\mathbf{x}_1$ and $\mathbf{x}_2$ being $N \times 1$ vectors. We know that $\Delta \mathbf{x}_1$ is sparse. Therefore, ideally, we would like to find Ψ such that, when $\mathbf{x} = \Psi \boldsymbol{\theta}$, then $\boldsymbol{\theta} = \Delta \hat{\mathbf{x}}_1$. This, however, is not possible using the algorithm proposed by Carvajalino *et al.* because Ψ will be a $2N \times N$ matrix, which is not an overcomplete dictionary. Our model has a unique characteristic that the forward representation is overcomplete, whereas the one in the other direction is not. This is unlike most CS signal models where overcompleteness of the dictionary in this other direction is exploited to achieve sparsity. We can of course compute the pseudoinverse of $[-\mathbf{I}\ \mathbf{I}]$ but that is an ℓ_2 solution. We therefore let Φ_D be as defined in (41).

V. RESULTS

We now present our results for ℓ_1- and ℓ_2-based difference-image estimation methods. We evaluate the performance using measured video imagery of an urban intersection (i.e., object scene; see Fig. 1) as the input into a simulation that models compressive measurements. We use a Panasonic PV-GS500 video camcorder to image the object scene. The reason we use conventionally imaged data as the truth data and simulate the compressive optical imaging system is to achieve flexibility in accurately implementing different sensing matrices Φ. This flexibility is required to analyze performance of our proposed sensing matrices in estimating sequence of difference images based on ℓ_1 and ℓ_2 norms. On the other hand, we now have to do computationally what a compressive imaging system will do optically. Consequently, instead of considering the entire 480×720 object scene, we reduce the problem by looking at the scene in 8×8, 16×16, and 32×32 blocks. The blocks are stitched together to reconstruct the difference image.

To compute the spatial and temporal correlations, we use a training set comprising 6000 frames of the object scene. From each 480×720 frame, we chose at random 30 blocks of sizes 8×8, 16×16, and 32×32 to give us $180\,000$ samples to compute the spatial autocorrelation matrix. To obtain the spatiotemporal cross-correlation matrix between object scenes at consecutive time instants, we select pairs of successive frames and randomly select 30 pairs of 8×8, 16×16, and 32×32 blocks
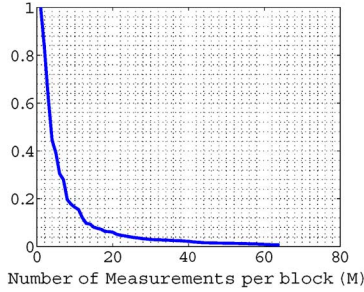
Fig. 4. Eigenspectrum computed from the sample correlation matrix of the difference images generated from the 6000 training set frames collected using Panasonic PV-GS500 video camcorder (8×8 block size).

Fig. 5. Estimated difference image for SNR $= 10$ dB and $M = 5$ using (a) WDPC, (b) Optimal, and (c) truth difference image.

from each frame pair, to give us approximately 180 000 sample pairs again. A pair of blocks consists of two blocks each drawn from the same region of the two consecutive image frames. Similarly, we extend the spatiotemporal correlations over longer time lengths for the LFGDIE method where, instead of considering two successive frames, we consider multiple consecutive frames. Once computed, these correlation matrices are stored for use in the testing stage.

The use of a spatiotemporal correlation matrix requires the assumption of wide-sense stationarity, which seldom holds in practice with the possible exception of texture images. Despite this, there is considerable literature on using second-order statistics to perform various image processing and imaging tasks [36]–[40]. These tasks attempt to reduce the nonstationarity by, e.g., camera tilt compensation or, in the case of eigenface analysis, by face centering and head orientation correction. Other methods have also been suggested for transforming nonstationary images to exhibit stationary characteristics [41], [42]. Such methods are at best approximate and not necessarily suitable or directly applicable to this paper. Therefore, in this first step toward developing an FSI-based imaging system, we empirically show that our proposed techniques yield good performance because of our ability to compress difference images by exploiting second-order statistics without artificial attempts at introducing stationarity. To illustrate this compression, we compute the eigenspectrum of the difference images generated from the 6000 training set frames. Fig. 4 shows that this eigenspectrum decays rapidly. This decay is a consequence of velocities and directions of motion being constrained in many environments of interest. These constraints can be in the form of roads, sidewalks, trails, corridors, etc. Therefore, if we think of difference images as a set of translated delta functions in 2-D, the nonuniqueness of these translations due to the constrained motion results in the eigenspectrum decay, allowing us to estimate the difference images, with high fidelity, from compressive measurements of the scene. The testing data also comprise of 6000 frames. These frames are subdivided into 100 groups; each comprising the object scene at 60 consecutive time instants. The testing set includes diverse cases of multiple targets moving at different speeds and directions. All performance plots (RMSE versus SNR and RMSE versus number of measurements per block M) in the following analysis have been averaged over the 60 time steps and the 100 groups.
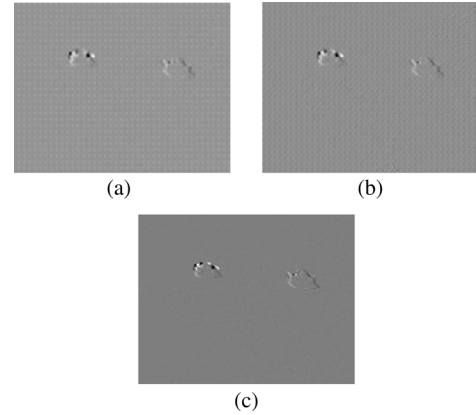
As discussed in Section III, there is no mathematically tractable optimal (ℓ_2) sensing matrix. Currently, we consider some possible choices for $\boldsymbol{\Phi}$, which were discussed in Section III. To remind the reader, they are PCs, DPCs, WPCs, WDPCs, numerically computed optimal sensing matrix (optimal) and Gaussian random sensing matrix (GPR). The GPR sensing matrix represents a set of fixed (aymptotically) basis projections that do not use any *a priori* information about the scene. The entries of the GPR matrix are Gaussian distributed with a mean of zero and a variance of one. We also consider the identity matrix (conventional) to mimic the conventional imager. We use the conventional imager for baseline performance comparison. The conventional imager always images the whole scene, i.e., it always makes $480 \times 720 = 345\,600$ measurements per frame.

A. ℓ_2-Based Difference-Image Estimation

Fig. 5 gives an example of an estimated difference image from a sequence of difference images using the multistep DDIE. The block size is 8×8, the SNR value is 10 dB, and the number of measurements per block M is 5. For $N = 480 * 720 = 345\,600$-D object scene, $M = 5$ translates to 27 000 measurements and a compression of measured data by 92.2%. The illustrated example has been computed with WDPC and optimal sensing matrices. We see that the performance of the optimal $\boldsymbol{\Phi}$ is visually close to that of the truth difference image. More importantly, the WDPC also estimates the difference image with good results. Note that it is much easier and computationally efficient to compute the WDPC than to numerically find the optimal sensing matrix. We quantify these results by plotting the root MSE (RMSE) as a function of SNR. We define RMSE as

$$\text{RMSE} = \sqrt{\sum_i (\Delta x_i - \Delta \hat{x}_i)^2 \bigg/ \sum_i (\Delta x_i)^2}. \qquad (50)$$

The normalization is done using the truth difference image $\Delta \mathbf{x}$.

Fig. 6 plots the multistep DDIE method's RMSE performance as a function of SNR for $M = 1, 5$ and for the 8×8 block size. We see an improvement in quantitative performance with
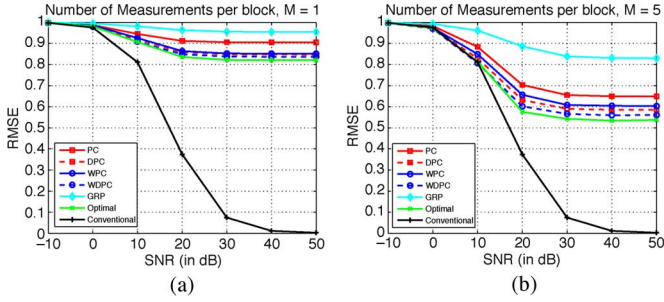
Fig. 6. RMSE versus SNR plots for the 8×8 block size. (a) $M = 1$. (b) $M = 5$. M is the number of measurements per block.
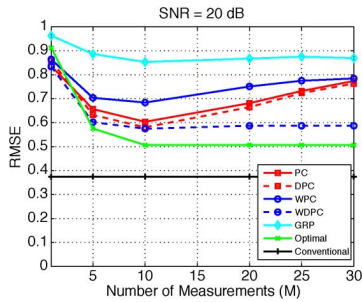

Fig. 8. (a) Performance comparison between single-step and multistep DDIE methods for $M = 5$ and SNR $= 20$ dB using WDPCs. Changing RMSE is compared over time. (b) Maximum divergence between single-step and multistep DDIE methods for varying SNR.

Fig. 7. RMSE versus number of measurements M for SNR $= 20$ dB for the 8×8 block size.

more measurements, with RMSE significantly decreasing for $M = 5$. Fig. 7, however, shows that this is not true generally for increasing measurements. With increasing M, the performance first improves and then can degrade. This behavior is a direct result of enforcing the photon count constraint. For small M, any additional measurement adds more information. However, because the total number of photons is fixed, as the number of measurements increase, additional information per measurement goes down. Eventually, the additive noise overwhelms the incremental information, and we see a degradation in performance for larger M. This is true for PC, DPC, and WPC sensing matrices. However, both WDPC and optimal sensing matrices avoid the degradation in performance with increasing measurements because they are optimized for a given SNR. Measurements are considered until they improve performance. Once information per measurement begins to be drowned out by noise, the additional measurements are ignored. There seems to be a certain discrepancy between the example estimates in Fig. 5 and the plots in Fig. 6(b). Although the visually estimated difference image looks good, the plots for $M = 5$ have a relatively high RMSE. This discrepancy is because ℓ_2 minimization minimizes the MSE over an ensemble and does not explicitly enforce sparsity. Consequently, there are small deviations (from the true pixel values) spread out over the whole estimated difference image. These small deviations from the truth, when normalized against the sparse truth difference image, bias the RMSE.

In Fig. 6, we can make a few observations about the efficacy of the various sensing matrices. As expected, waterfilling does improve performance of both PCs and DPCs by weighting the projections according to noise statistics. This is more evident
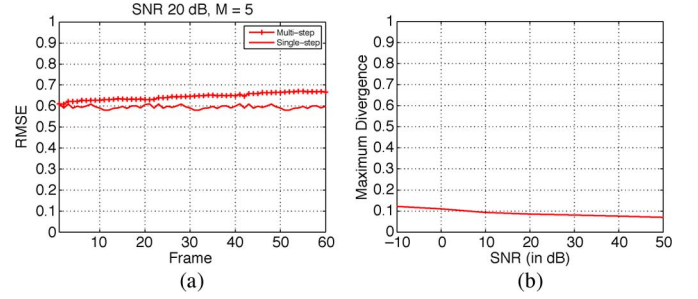
in Fig. 6(b), where we take five measurements per block. Numerically searching for the optimal sensing matrix further improves upon WPCs and WDPCs. The advantage that the latter have over a numerical search is that they are much simpler to compute for changing SNR. Therefore, searching for the optimal solution is reasonable only if improvement in performance outweighs the increased computational cost. In Fig. 5, we see that this not the case. For the sake of completeness, we have also plotted the RMSE performance of Gaussian random projections (GRPs). We are aware that there is no theoretical framework for a nonadaptive GRP to outperform sensing matrices exploiting *a priori* information. This is experimentally validated in the plots, which show a GRP to be performing the worst. Lastly, for ℓ_2-based estimation, the conventional imager outperforms all the sensing matrices. This is to be expected as minimizing the MSE alone does not give an advantage to compressive measurements over a conventional imager. We need an additional constraint, which, for our case, turns out to be sparsity. We show in the next subsection that we can actually beat the conventional imager performance when we enforce sparsity via a nonlinear estimation.However, we stress that, as seen in Fig. 5, the qualitative performance of WDPCs is visually close to that of the truth difference image. In fact, from a practical perspective, it can be used to provide a good input to a tracker.

The multistep DDIE, as defined in (21), forms a closed loop between the estimate of the scene and the difference image. As a result, there will be degradation in performance over time. To grade the performance of the multistep DDIE, we consider a clairvoyant scenario for estimating the sequence of difference images. We will refer to this special case as the single-step DDIE. Assuming we are estimating the difference image of the scene between t_k and t_{k+1}, the single-step DDIE always assumes perfect knowledge of the scene at t_k as follows:

$$\boldsymbol{\Delta}\hat{\mathbf{x}}_k = recon(\mathbf{x}_k, \mathbf{y}_{k+1}). \qquad (51)$$

Obviously, the single-step DDIE is practically infeasible. However, it does bound the performance of the multistep DDIE and, as such, allows us to evaluate the efficacy of the multistep strategy. Fig. 8(a) shows the performance comparison between the two for 60 time steps for $M = 5$ and SNR $= 20$ dB. Since single-step DDIE assumes perfect knowledge at every stage, its RMSE as a function of time is nearly constant. The
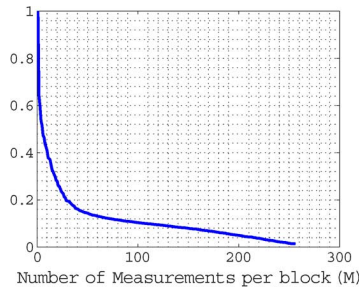
Fig. 9. Eigenspectrum computed from the sample correlation matrix of the difference images generated from the 3500 training set frames from PETS data set (16×16 block size).
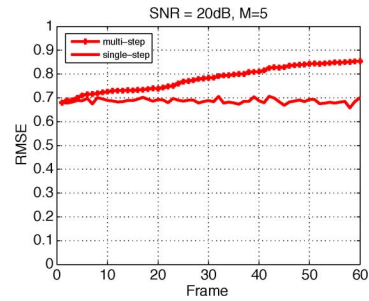

Fig. 10. PETS data set: Performance comparison between single-step and multi-step DDIE methods for $M = 5$ and SNR $= 20$ dB using WDPC. Changing RMSE is compared over time.

RMSE for the multistep DDIE is the same as the single-step DDIE at t_1. With passing time, however, the performance of multistep DDIE degrades as the RMSE drifts. However, as can be seen, the drift is slow, showing that the multistep DDIE is temporally robust. In Fig. 8(b), we plot the maximum divergence of the multistep DDIE from the ideal single-step DDIE for changing the SNR. Maximum divergence is the maximum value by which the multistep method diverges from the single-step method over the 60 time steps. The resulting plot is a line with a small slope indicating that the multistep DDIE does not diverge significantly for changing the SNR. To emphasize further empirically the temporal robustness of the multistep DDIE, we consider a second data set, i.e., the PETS 2007 data set. We use two camera views of the same object scene from this data set, each with 4001 frames. We use 3500 frames from the first camera view as the training set. We follow the exact procedure as for the previous data set to compute the spatiotemporal cross-correlation matrix (Fig. 9 illustrates the compressibility of difference images from this data set in a manner similar to the previous data set). To ensure no overlap between the training and the testing sets, the images from the remaining 501 frames in the second camera view are used for the testing set. Fig. 10 shows the performance comparison between single-step and multistep DDIE for the PETS data set for 60 time steps. We see that the RMSE is slightly more than that in Fig. 8(a), but the trend is the same. The reason for a slight increase in RMSE is that the training and testing data set are significantly different. In our primary data set, we follow a data collection strategy where the training is done on the same object scene that is of interest for target tracking purposes, i.e., the camera angle is fixed, but the training and testing data sets are defined at different nonoverlapping time intervals. The use of the PETS data set allows us to replicate our results for a different scenario where the camera must be trained on one scene and operate on another. Fig. 11 gives an example of an estimated difference image using the PETS testing data set. The first image is the truth difference image at the 50th time step of an image sequence. The corresponding multistep-DDIE-based estimated difference image using the optimal sensing matrix is shown in Fig. 11(b) for $M = 5$ and SNR $= 10$ dB. The block size is 8×8. We see the the visual quality of the estimate is not far from the truth difference image after 50 time steps.

We now look at the performance of the LFGDIE method for estimating a sequence of difference images. We claimed that the
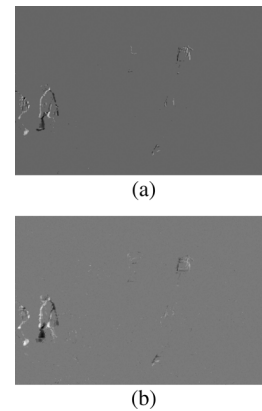

(a)

(b)

Fig. 11. PETS data set: Multistep-DDIE-based estimated difference image for $M = 5$ and SNR $= 10$ dB using the optimal sensing matrix. (a) Truth difference image. (b) Estimated difference image.

LFGDIE will perform better than the multistep DDIE as it is able to exploit temporal correlation between all time instants. Fig. 12 shows that this is indeed the case. RMSE performance has improved compared with Fig. 6(b). As expected, however, the trends are still the same. The WDPC sensing matrix still outperforms all other candidate sensing matrices with the exception of the numerically searched optimal sensing matrix. Until now, we have considered the block size of 8×8. In Fig. 13, we graph the RMSE performance as a function of SNR for the 16×16 block size. We see that, for $M = 1$, there is an improvement in performance over the 8×8 block size. In fact, the performance we get for the 16×16 block with $M = 1$ is similar to the performance of the 8×8 block with $M = 5$. Notice that $M = 1$, for 16×16 block size, implies 1350 measurements for the whole scene, which translates to less than 0.4% measurements in comparison with the conventional imager. Thus, we see that, for the larger block size, we have improved performance with simultaneously higher compression ratio. Fig. 14, however, shows that the amount of improvement in performance we get by going from block size 8×8 to 16×16 is reduced when we go from block size 16×16 to 32×32. This happens because any stationarity that holds for the small block sizes of 8×8 begins to break for larger block sizes. As a result, the spatial structure represented by the sample autocorrelation and cross-correlation matrices obtained from training data does not completely represent the true correlation. The LFGDIE method also has the same trends as the multistep DDIE. The ℓ_1-based estimation on
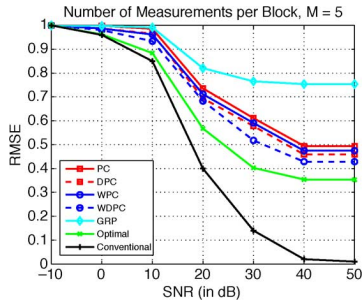
Fig. 12. RMSE versus SNR plots for the LFGDIE method. Block size is 8×8 and $M = 5$.
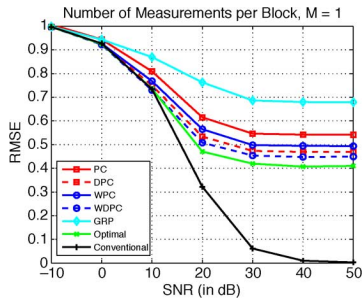

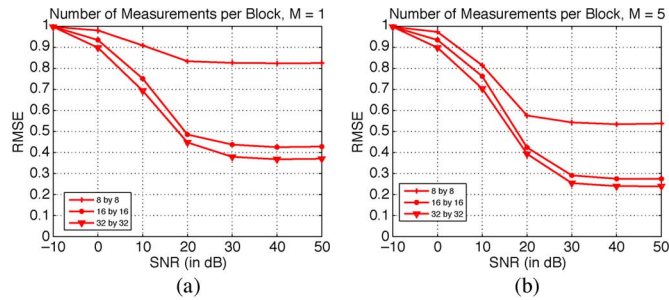Fig. 13. RMSE versus SNR performance plots for block size 16×16, for $M = 1$.

Fig. 14. Optimal-sensing-matrix based performance comparison between the three block sizes 8×8, 16×16, and 32×32 for (a) $M = 1$ and (b) $M = 5$.

Fig. 15. Estimated difference image for SNR $= 10$ dB and $M = 5$ for (a) sparsity-enforced difference image, (b) TV method, and (c) BPDN method.

Fig. 16. RMSE versus SNR curves for the 8×8 blocks and for the TV method. (a) $M = 1$. (b) $M = 5$. (c) $M = 10$. (d) $M = 20$.

the other hand does not depend on a stationarity assumption. Its goal is to find the best estimate based on the measured data that enforces the sparsity constraint. In the following subsection, we discuss the performance of ℓ_1-based estimation methods.

B. ℓ_1-Based Difference-Image Estimation

The advantage of using the ℓ_2 norm is that we get precise linear estimation operators, which allow for quick and easy computation of the estimate of the difference image. The disadvantage, however, has to do with the inability to exploit the sparsity of a difference image. Solving the convex optimization problem of (43) allows us to overcome this disadvantage.

Here, we consider the PC, DPC, WPC, WDPC, and GRP sensing matrices. All abbreviations are the same as before. Gaussian sensing matrices (i.e., GRP) have been suggested in theory of compressed sensing for measuring data because they are incoherent with all representation bases. As a result, they have become nearly universal in applications of compressed sensing for being able to reconstruct signals of interest without
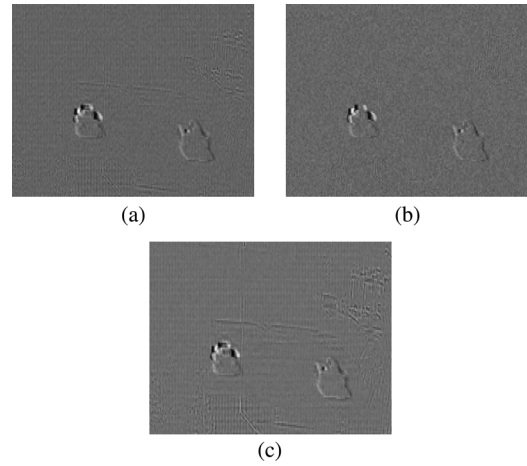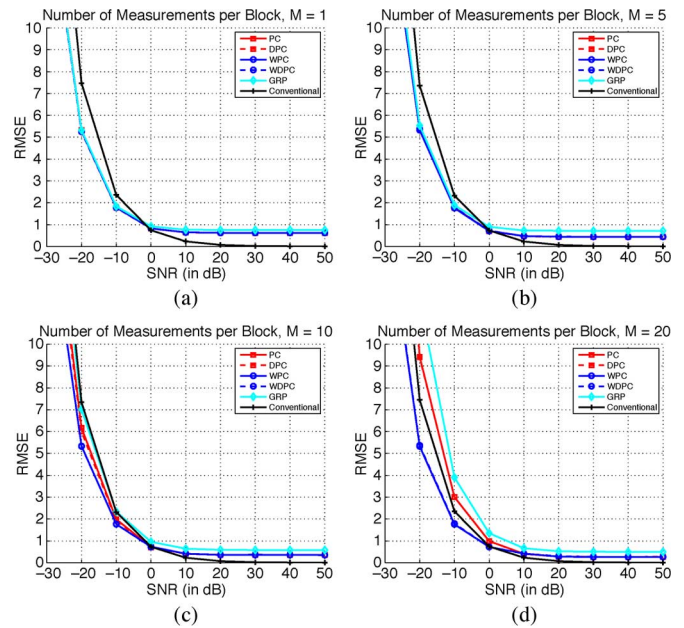
prior knowledge of the signal structure. Fig. 15 shows examples of the estimated difference image for the 8×8 block size, using the three ℓ_1-based methods discussed in Section IV. Visually, all three methods are effective, although TV method performs better than the other two. Both isotropic and nonisotropic TV regularizers have similar performance. All the results shown here are for the nonisotropic TV regularizer. In Fig. 16, we plot the RMSE performance of the TV method as a function of SNR. The block size is 8×8, and $M = 1, 5, 10, 20$. For $M = 1, 5$, all sensing matrices have similar performance. In fact, at low SNR, all sensing matrices perform better than the conventional imager. Thus, unlike ℓ_2-based estimation, ℓ_1-based methods are better able to utilize the concentration of energy into a few measurements. This is surprisingly true even for the GRP sensing matrix. At low SNR, there is a higher premium on the available energy, and as a consequence, a small number of random measurements perform better than the conventional
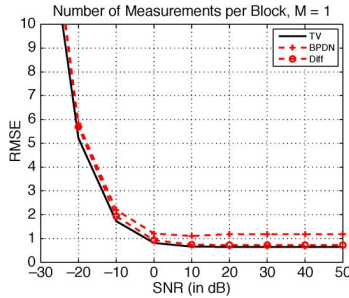
Fig. 17. WDPC-sensing-matrix RMSE versus SNR curves for the three ℓ_1-based difference image estimation methods. Block size 8×8 and $M = 1$.

imager, where the small energy is spread over all the $N = 64$ measurements. For $M = 10, 20$, the curves for different sensing matrices begin to separate. Yet, the performance of PC and DPC sensing matrices is similar. This is because the ℓ_1-based estimation does not directly estimate the difference image but, instead, jointly estimates the scene at the two time instants. As a result, we cannot take advantage of the difference image form that ℓ_2-based estimation afforded us. We see that waterfilling improves upon both PC and DPC sensing matrices, but due to the same reason, the performance of WPC and WDPC sensing matrices is also similar. The WPC and WDPC sensing matrices give the best RMSE performance.

In Fig. 17, we compare the performance of the three ℓ_1-based methods by looking at the RMSE versus SNR curves for the WDPC sensing matrix. Within the three ℓ_1 strategies, we see that the TV method performs better than the sparsity-enforced difference-image method and BPDN. However, the improvement is small, particularly when compared with the sparsity-enforced difference-image method. By minimizing the gradient of the difference image, the TV method is best able to capture the changes in intensity across the difference image. On the other hand, difference images are also sparse; hence, enforcing sparsity also gives good results. The BPDN method has a higher RMSE compared with the other two strategies. This is to be expected because, although BPDN also minimizes the ℓ_1 norm with respect to a sparsifying basis, it does not directly utilize the difference image as is done by the regularization term of the other two methods. Instead, the BPDN method computes the joint sparse representation **s** of the object scene at the two time instants. This results in reduced performance. However, as corroborated in Fig. 15, the degradation in performance is small. Plotting RMSE as a function of M shows us that we are able to beat the performance of the conventional imager. This is illustrated in Fig. 18. At low SNR and for smaller measurements, we see a wide gap in performance between all sensing matrices and the conventional imager. Note that the conventional imager always makes $720 * 480 = 345600$ measurements. As the number of measurements increase, the performance of all the sensing matrices begins to degrade. The rate of this degradation is a function of SNR. It slows down as SNR increases. When the SNR is high, there is no advantage to be obtained from better utilizing the energy as there is enough energy for all measurements, and the conventional imager performs the best. Finally, we note that sensing matrices using *a priori* information perform better
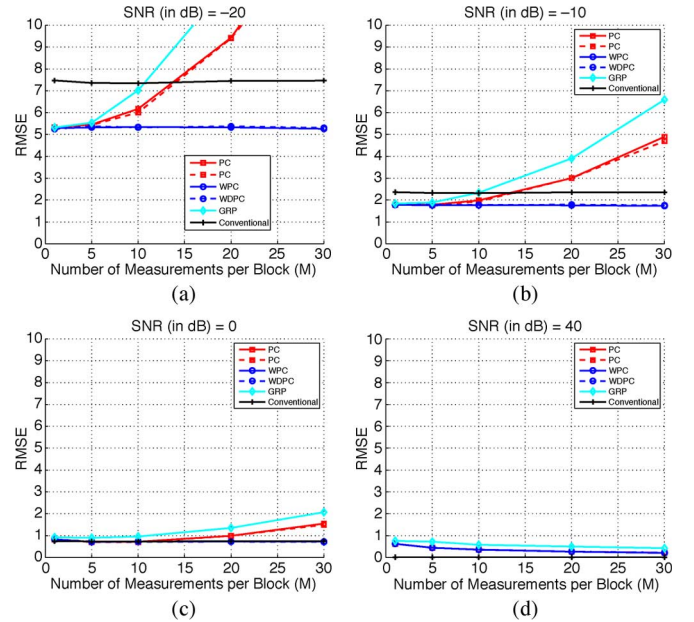

Fig. 18. RMSE versus number of measurements per block for 8×8 blocks and for the TV method. (a) SNR $= -20$ dB. (b) SNR $= -10$ dB. (c) SNR $= 0$ dB. (d) SNR $= 40$ dB.
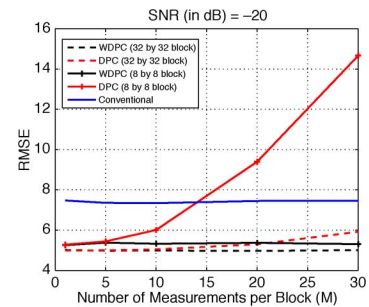

Fig. 19. Performance comparison between the ℓ_1-based TV method using 8×8 and 32×32 blocks. (a) $M = 1$.

than GRP. Looking at the RMSE versus SNR performance of GRP, we see that it has the most error at all SNRs and for any number of measurements. At the same time, RMSE versus M plots show that the rate at which the GRP performance degrades is fastest among all sensing matrices. Increasing block size leads to improved performance for a fewer number of measurements as a fraction of the total. Unlike the ℓ_2-based method, here, we do not suffer from the stationarity assumption. In Fig. 19, we plot the performance of the TV method using the WDPC sensing matrix for 8×8 and 32×32 block sizes. We see that the rate of performance degradation is significantly reduced for the 32×32 block sizes. The improved performance is mainly because the sparsity condition holds better for larger block sizes. If the block size is very small, then even a sparse image might not be sparse within that block. However, with increasing block size, we achieve sparsity and, as a result, get better performance.

This trend of improved performance with increasing block size augurs well for future practical implementation of our proposed strategies. A practical FSI system with the capability to handle different sensing matrices could optically measure the entire scene at very fast speed. An optical system would not face

the same limitations that we have in terms of simulating measurements due to a large scene.

VI. Conclusion

In this paper, we have shown ℓ_2- and ℓ_1-based techniques for estimating a sequence of difference images from a sequence of compressive measurements. We have presented qualitative and quantitative results to attest that both techniques are able to estimate successfully the difference images within the FSI paradigm. Each has its advantage. ℓ_2-based techniques give closed-form expressions for the linear estimation operators, which are easy to compute. On the other hand, ℓ_1-based methods exploit the natural sparsity of the difference image. Within ℓ_2-based techniques, we looked at the multistep DDIE and LFGDIE methods to reconstruct directly the difference image from the compressive measurements. For the ℓ_1-based estimation problem, we looked at three different approaches to the linear inverse problem and compared their performance. We showed that the modified TV method performs the best, although the method that enforced the sparsity condition performs only slightly worse. Lastly, we showed that WDPC sensing matrix had the lowest RMSE for both ℓ_2- and ℓ_1-based methods, although for the latter WPC did equally well. The performance of all sensing matrices that utilized *a priori* information was better than the nonadaptive GRP sensing matrix. In fact, from a practical perspective, depending on the SNR and the number of measurements that can be taken, anyone of them can be used to provide a decent input to a tracker.

References

[1] D. W. Xue and Z. M. Lu, "Difference image watermarking based reversible image authentication with tampering localization capability," *Int. J. Comput. Sci. Eng. Syst.*, vol. 2, no. 3, pp. 213–218, 2008.

[2] S. K. Lee, Y. H. Suh, and Y. S. Ho, "Reversible image authentication based on watermarking," in *Proc. IEEE Int. Conf. Multimedia Expo.*, 2006, pp. 1321–1324.

[3] P. B. Heffernan and R. A. Robb, "Difference image reconstruction from a few projections for nondestructive materials inspection," *Appl. Opt.*, vol. 24, no. 23, pp. 4105–4110, Dec. 1985.

[4] S. G. Kong, "Classification of interframe difference image blocks for video compression," in *Proc. SPIE*, 2002, vol. 4668, pp. 29–37.

[5] D. Hahn, V. Daum, J. Hornegger, W. Bautz, and T. Kuwert, "Difference imaging of inter- and intra-ictal spect images for the localization of seizure onset in epilepsy," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec.*, 2007, pp. 4331–4335.

[6] C. Alcock, R. A. Allsman, D. Alves, T. S. Axelrod, A. C. Becker, D. P. Bennett, K. H. Cook, A. J. Drake, K. C. Freeman, K. Griest, M. J. Lehner, S. L. Marshall, D. Minniti, B. A. Peterson, M. R. Pratt, P. J. Quinn, C. W. Stubbs, W. Sutherland, A. Tomaney, T. Vandehei, and D. L. Welch, "Difference image analysis of galactic microlensing. I. data analysis," *Astrophys. J.*, vol. 521, no. 2, pp. 602–612, Aug. 1999.

[7] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[8] M. A. Neifeld and P. Shankar, "Feature-specific imaging," *Appl. Opt.*, vol. 42, no. 17, pp. 3379–3389, Jun. 2003.

[9] S. Tucker, W. T. Cathey, and J. Edward Dowski, "Extended depth of field and aberration control for inexpensive digital microscope systems," *Opt. Exp.*, vol. 4, no. 11, pp. 467–474, May 1999.

[10] S. Prasad, "Information capacity of a seeing-limited imaging system," *Opt. Commun.*, vol. 177, no. 1–6, pp. 119–134, Apr. 2000.

[11] E. Clarkson and H. H. Barrett, "Approximations to ideal-observer performance on signal-detection tasks," *Appl. Opt.*, vol. 39, no. 11, pp. 1783–1793, Apr. 2000.

[12] W.-C. Chou, M. A. Neifeld, and R. Xuan, "Information-based optical design for binary-valued imagery," *Appl. Opt.*, vol. 39, no. 11, pp. 1731–1742, Apr. 2000.

[13] A. Ashok and M. Neifeld, "Information-based analysis of simple incoherent imaging systems," *Opt. Exp.*, vol. 11, no. 18, pp. 2153–2162, Sep. 2003.

[14] M. P. Christensen, G. W. Euliss, M. J. McFadden, K. M. Coyle, P. Milojkovic, M. W. Haney, J. van der Gracht, and R. A. Athale, "ACTIVE-EYES: An adaptive pixel-by-pixel image-segmentation sensor architecture for high-dynamic-range hyperspectral imaging," *Appl. Opt.*, vol. 41, no. 29, pp. 6093–6103, Oct. 2002.

[15] D. L. Marks, R. Stack, A. J. Johnson, D. J. Brady, and D. C. Munson, "Cone-beam tomography with a digital camera," *Appl. Opt.*, vol. 40, no. 11, pp. 1795–1805, Apr. 2001.

[16] Z. Liu, M. Centurion, G. Panotopoulos, J. Hong, and D. Psaltis, "Holographic recording of fast events on a CCD camera," *Opt. Lett.*, vol. 27, no. 1, pp. 22–24, Jan. 2002.

[17] P. Potuluri, M. Fetterman, and D. Brady, "High depth of field microscopic imaging using an interferometric camera," *Opt. Exp.*, vol. 8, no. 11, pp. 624–630, May 2001.

[18] S. M. Kay, *Fundamentals of Statistical Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[19] E. J. Candès, J. K. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory* vol. 52, no. 2, pp. 489–509, Feb. 2006 [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TIT.2005.862083

[20] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding 2strategies?," *IEEE Trans. Inf. Theory* vol. 52, no. 12, pp. 5406–5425, Dec. 2006 [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TIT.2006.885507

[21] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory* vol. 51, no. 12, pp. 4203–4215, Dec. 2005 [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TIT.2005.858979

[22] E. J. Candès, J. K. Romberg, and T. Tao, "Signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2005.

[23] E. J. Candès and J. K. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Found. Comput. Math.* vol. 6, no. 2, pp. 227–254, Apr. 2006 [Online]. Available: http://dx.doi.org/10.1007/s10208-004-0162-x

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

[25] S. Uttam, N. A. Goodman, and M. A. Neifeld, "Direct reconstruction of difference images from optimal spatial-domain projections," in *Proc. SPIE*, 2008, vol. 7096, pp. 709 608-1–709 608-6.

[26] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "An architecture for compressive imaging," in *Proc. IEEE Int. Conf. Image Process.*, 2006, pp. 1273–1276.

[27] V. Cehver, A. Sankaranarayanan, M. F. Duarte, D. Reddy, R. G. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," in *Proc. 10th Eur. Conf. Comput. Vis.*, 2008, pp. 155–168.

[28] V. Cehver, M. F. Duarte, C. Hedge, and R. Baraniuk, "Sparse signal recovery using Markov random fields," presented at the Neural Information Processing Systems (NIPS), Vancouver, BC, Canada, 2008, EPFL-CONF-151 469, unpublished.

[29] M. Piccardi, "Background subtraction techniques: A review," in *Proc. IEEE Syst., Man Cybern.*, 2004, vol. 4, pp. 3099–3104.

[30] M. A. Neifeld and J. Ke, "Optical architectures for compressive imaging," *Appl. Opt.*, vol. 46, no. 22, pp. 5293–5303, Aug. 2007.

[31] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ: Wiley, 2006.

[32] W. Wenzel and K. Hamacher, "A stochastic tunneling approach for global minimization," *Phys. Rev. Lett.*, vol. 82, no. 15, pp. 3003–3007, 1999.

[33] S. Osher, L. Rudin, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D.*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

[34] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[35] J. M. D. Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.

[36] L. Malagón-Borja and O. Fuentes, "An object detection system using image reconstruction with pca," in *Proc. Comput. Robot Vis., Can. Conf.*, 2005, pp. 2–8.

[37] M. Turk and A. Pentland, "Face recognition using Eigenfaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1991, pp. 586–591.

[38] A. Pentland, B. Moghaddam, and T. Starner, View-based and modular eigenspaces for face recognition MIT, Cambridge, MA, Tech. Rep., 1994.

[39] H. Moon and P. Phillips, "Computational and performance aspects of PCA-based face recognition algorithms," *Perception*, vol. 30, no. 3, pp. 303–321, Jan. 2001.

[40] H. S. Pal, D. Ganotra, and M. A. Neifeld, "Face recognition by using feature-specific imaging," *Appl. Opt.*, vol. 44, no. 18, pp. 3784–3794, Jun. 2005.

[41] R. Chin, "Restoration of images with nonstationary mean and autocorrelation," in *Proc. ICASSP*, Apr. 1988, vol. 2, pp. 1008–1011.

[42] R. N. Strickland, "Transforming images into block stationary behavior," *Appl. Opt.*, vol. 22, no. 10, pp. 1462–1473, May 1983.

Nathan A. Goodman (S'98–M'02–SM'07) received the B.S., M.S., and Ph.D. degrees from the University of Kansas, Lawrence, in 1995, 1997, and 2002, respectively, all in electrical engineering.

From 1996 to 1998, he was a Radio-Frequency Systems Engineer with Texas Instruments, Dallas. From 1998 to 2002, he was a Graduate Research Assistant with the Radar Systems and Remote Sensing Laboratory, University of Kansas. In 2009–2010, he was a Visiting Senior Research Engineer with the Georgia Tech Research Institute, Smyrna, GA. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Arizona, Tucson. Within the department, he directs the Laboratory for Sensor and Array Processing. His research interests include novel system and processing concepts for radar, including sparse-array and multistatic radar systems, adaptive waveform transmission for target recognition and air-to-ground surveillance, and compressed sensing for radar.

Dr. Goodman was a Technical Co-chair for the 2011 IEEE Radar Conference and will be the Finance Chair for the 2012 IEEE Sensor Array and Multichannel Signal Processing Workshop. He has served as a Reviewer for numerous journals and conferences and currently serves as a Deputy Editor-in-Chief for *Elsevier Digital Signal Processing*. He was a recipient of the Madison A. and Lila Self Graduate Fellowship upon returning to the University of Kansas for Ph.D. studies in 1998, the IEEE 2001 International Geoscience and Remote Sensing Symposium Interactive Session Prize Paper Award, and a Best Paper Award at the 2008 Army Science Conference.

Shikhar Uttam (S'00–M'10) received the B.Tech. degree from Jamia Millia Islamia, New Delhi, India, in 2002, and the M.S. and Ph.D. degrees from the University of Arizona, Tucson, in 2006 and 2010, respectively, all in electrical engineering.

From 2005 to 2010, he was a member with the Laboratory for Sensor and Array Processing, Department of Electrical and Computer Engineering, University of Arizona. He is currently a Postdoctoral Research Associate with the Biomedical Optical and Imaging Laboratory, Department of Medicine and with the Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA. His current research interests include computational imaging, quantitative phase microscopy, spectral imaging, diffraction tomography, color-based image processing, and machine learning.

Dr. Uttam is a member of SPIE and SIAM.

Mark A. Neifeld (S'84–M'91) received the B.S.E.E. degree from the Georgia Institute of Technology, Atlanta, in 1985 and the M.S. and Ph.D. degrees from the California Institute of Technology, Pasadena, in 1987 and 1991, respectively.

In 1991, he joined the Faculty of the Department of Electrical and Computer Engineering and the Optical Sciences Center, University of Arizona, Tucson. He is the author or coauthor of more than 100 journal articles and more than 250 conference papers in the areas of optical communications and storage, coding and signal processing, and optical imaging and processing systems. His current research interests include information- and communication-theoretic methods in image processing, nontraditional imaging techniques that exploit the joint optimization of optical and postprocessing degrees of freedom, coding and modulation for fiber and free-space optical communications, and applications of slow and fast light.

Dr. Neifeld is a Fellow of the Optical Society of America and a member of SPIE and APS. He has served on the organizing committees of numerous conferences and symposia. He has also been a two-term Topical Editor for *Applied Optics* and a three-time Guest Editor of special issues of *Applied Optics*.