EFFICIENT SPARSITY-BASED INVERSION FOR PHOTON-SIEVE SPECTRAL IMAGERS WITH TRANSFORM LEARNING

Ulas Kamaci⋆ Fatih C. Akyon⋆ Tunc Alkanat⋆ Figen S. Oktem⋆

* Dept. of Electrical Eng., METU, 06800 Cankaya, Ankara, Turkey

ABSTRACT

We develop an efficient and adaptive sparse reconstruction approach for the recovery of spectral images from the measurements of a photon-sieve spectral imager (PSSI). PSSI is a computational imaging technique that enables higher resolution than conventional spectral imagers. Each measurement in PSSI is a superposition of the blurred spectral images; hence, the inverse problem can be viewed as a type of multi-frame deconvolution problem involving multiple objects. The transform learning-based approach reconstructs the spectral images from these superimposed measurements while simultaneously learning a sparsifying transform. This is performed using a block coordinate descent algorithm with efficient update steps. The performance is illustrated for a variety of measurement settings in solar spectral imaging. Compared to approaches with fixed sparsifying transforms, the approach is capable of efficiently reconstructing spectral images with improved reconstruction quality.

1. INTRODUCTION

Spectral imaging, capturing light in both spatial and spectral dimensions, is a fundamental diagnostic technique with widespread application in physics, chemistry, medicine, biology, astronomy, and remote sensing [1]. Due to the intrinsic limitation of two-dimensional (2D) detectors in capturing inherently 3D spectral data, spectral imaging techniques conventionally rely on a spatial or spectral scanning process, which renders them inefficient in terms of light throughput and data acquisition time. Moreover, since these techniques purely rely on physical systems, there are inherent physical limitations on their temporal, spatial, and spectral resolutions. Together, these drawbacks limit the use and effectiveness of the conventional techniques.

To overcome these limitations, spectral imaging techniques based on computational imaging emerge as effective approaches. Photon-sieve spectral imaging (PSSI) is such an approach that enables higher resolution than conventional spectral imagers [2, 3]. Other computational spectral imaging techniques include computed tomography imaging spectrometry [4], compressive coded aperture spectral imaging (CASSI) [5], and compressive hyperspectral imaging by separable operators [6]. In all these approaches, the 3D data cube is represented in terms of voxels and then reconstructed from some indirect measurements.

Sparse models play an important role in computational spectral imaging due to significant degree of structure and redundancy present in the high-dimensional spectral data [1]. While most of the earlier works use fixed analytical transforms such as discrete cosine transform (DCT) and wavelets for the purpose of sparsification, recently, learning sparse models adaptively from the data appears to be more effective.

Learning-based techniques involve either learning dictionaries [7, 8] or sparsifying transforms [9, 10]. Dictionary learning-based techniques have been exploited in solving various inverse problems in spectral imaging including unmixing [11], denoising and inpainting [12], superresolution and classification [13–15]. A Bayesian dictionary learning framework has also been used for reconstructing the spectral datacube from CASSI compressive measurements [16]. Because dictionary learning based sparse reconstruction generally have higher computational complexity for large scale problems, in this work, we focus on transform learning approach for the data-adaptive inversion of the PSSI measurements.

Each measurement in PSSI consists of superimposed blurred spectral images, and hence, the inverse problem is a non-standard multi-frame deconvolution problem involving multiple objects. Our approach adaptively learns a joint sparsifying transform and simultaneously reconstructs the spectral images from these superimposed measurements. This is performed using a block coordinate descent algorithm (i.e. alternating minimization) with efficient update steps. The performance of the reconstruction algorithm is illustrated for an application in solar spectral imaging. The results demonstrate improvements in the reconstruction quality as compared to an earlier approach with fixed analytical transforms, for a variety of measurement schemes and noise levels.

The developed image reconstruction approach can also be extended to other computational spectral imaging modalities. Here the focus is on PSSI, since, different than the other modalities, it offers both diffraction-limited high spatial resolution as well as higher spectral resolution than conventional spectral imagers employing wavelength filters, with the additional advantage of a simple optical configuration [2].
2. IMAGING SYSTEM AND FORWARD PROBLEM
The photon-sieve spectral imaging technique uses an optical configuration that consists of a single diffractive imaging element called photon sieve [17]. Photon sieves provide an alternative to lenses and mirrors, and are obtained by replacing the open zones of Fresnel zone plates with circular holes. The PSSI system is illustrated in Fig. 1. Here \( d_s \) and \( d_k \) denote the distances from the object and \( k \)th measurement plane to the photon sieve plane \((k = 1, \ldots, K)\). We consider a polychromatic radiation from the object, which is composed of \( S \) wavelength components, each with a different wavelength \( \lambda_s \) \((s = 1, \ldots, S)\) and mutually incoherent from others [18]. Because the focal length of the photon-sieve depends on its aperture function and has a inherent point-spread function (PSF), each wavelength component is focused at a different distance from the sieve. If, for example, a measurement is taken from a plane where one spectral component overlaps with the defocused images of all other components. A total of \( K \) such measurements are obtained, for example, by using a moving detector.

![Fig. 1: The photon sieve spectral imaging (PSSI) system](image)

The image formation model that relates the intensities of individual spectral components to the measurements has been derived as follows [3]:

\[
t_k[m,n] = \sum_{s=1}^{S} x_s[m,n] * h_{d_k,\lambda_s}[m,n],
\]

where \( m, n = -N/2, \ldots, N/2 - 1 \). Here \( t_k[m,n] \) represents the \( k \)th measurement obtained at distance \( d_k \) over \( N \times N \) pixels. As seen, each measurement consists of \( S \) terms arising from \( S \) different spectral components. The term \( x_s[m,n] \) denotes the diffraction-limited, discretized intensity of the spectral component with wavelength \( \lambda_s \). Each such spectral image with wavelength \( \lambda_s \) is convolved at distance \( d_k \) with the incoherent point-spread function (PSF), \( h_{d_k,\lambda_s}[m,n], \) of the photon sieve. This sampled PSF is uniformly discretized version of its continuous form, i.e. \( h_{d_k,\lambda_s}[m,n] = h_{d_k,\lambda_s}(m\delta_x, n\delta_x) \) with \( \delta_x \) being the pixel size of the detector. The PSF of the photon sieve depends on its aperture function and has a closed-form expression given elsewhere [19].

Let the PSF \( h_{d_k,\lambda_s}[m,n] \) has \( M \times M \) support, i.e. \( h_{d_k,\lambda_s}[m,n] = 0 \) for \( m, n \notin [-M/2, M/2 - 1] \). We assume that the size of the input objects are limited to a slightly smaller region than the detector range, i.e. \( x_s[m,n] = 0 \) for \( m, n \notin [-N/2, N/2 - 1] \). With this, the convolution in Eqn. (1) can be replaced with a circular convolution of \( N \) points, which will be exploited in the development of the fast data-adaptive inversion method.

In the inverse problem framework, the following matrix-vector form obtained from the above image-formation model will be used [3]:

\[
y = Hx + n,
\]

Where \( y \) is the lexicographically ordered noisy \( k \)th measurement vector, whereas \( y \) is the overall noisy measurement vector obtained by combining all the \( K \) measurements. Similarly, the vector \( x_s \) contains the spectral image with wavelength \( \lambda_s \), and the vector \( x \) is obtained by vertically concatenating all vectors \( x_s \). The matrix \( H_{k,s} \) is an \( N^2 \times N^2 \) block circulant matrix with circular blocks corresponding to the circular convolution operation with \( h_{d_k,\lambda_s} \), and \( H \) is the overall system matrix of size \( KN^2 \times SN^2 \). Lastly, the vector \( n = [n_1^T \ldots n_K^T]^T \) is the additive noise vector with \( (n_k)_s \sim N(0, \sigma_k^2) \) representing white Gaussian noise.

3. IMAGE RECONSTRUCTION WITH LEARNING
In the inverse problem, the goal is to recover the unknown spectral images, \( x \), from their noisy, superimposed and blurred measurements, \( y \). This ill-posed problem can be viewed as a type of multiframe deconvolution problem involving multiple images. By incorporating the prior information available for the spectral images, we formulate the problem as follows:

\[
\min_x \nu \| y - Hx \|_2^2 + R(x).
\]

This is a regularized least squares problem, which can also be related to maximum posterior estimation (MAP). Here the first term controls data fidelity, whereas the second term \( R(x) \) controls how well the reconstruction matches our prior knowledge of the solution, with the scalar parameter \( \nu \) trading off between these two terms.

Because natural spectral images can be represented sparsely in some transform domain [1], sparsity-based regularization in the form of \( R(x) = \| Dx \|_0 \) is desired where \( D \) is some fixed sparsifying transform. Since this leads to an NP-hard problem, by convex relaxation of the \( \ell_0 \) quasi norm with the \( \ell_1 \) norm, the problem can be converted to a convex optimization problem, which can be solved using standard convex optimization techniques or fixed-point algorithms [20, 21]. One such efficient algorithm has been developed [3] based on "half-quadratic regularization" method [22] for this imaging problem.

In this work, the underlying sparsity-inducing transform is assumed unknown. This allows the sparse model to be
adapted to the specific objects being imaged. Our goal is to simultaneously enforce sparsity of the reconstructed spectral images in an adaptive transform domain and obtain reconstructions that are consistent with the measurements. For this, we use the following patch-based transform learning regularizer suggested in [10]:

\[ R(x) = \min_{W,B} \sum_{s=1}^{S} \sum_{j=1}^{N^2} ||WP_j x_s - b_{s,j}||_2^2 + \lambda Q(W) \text{ s.t. } ||B||_0 \leq \beta \]

A key distinction here is that the sparsifying transform \( W \in \mathbb{C}^{L \times L} \) is shared across all spectral images \( x_s \), since all share common statistical and structural properties. The matrix \( P_j \in \mathbb{R}^{L \times N^2} \) extracts a \( \sqrt{L} \times \sqrt{L} \) patch from the image \( x_s \) in \( \mathbb{R}^{N^2} \), and a total of \( N^2 \) overlapping wrap around patches with unity overlap stride are used for each image. The sparse code for the \( j \)th patch of the \( s \)th image is denoted by \( b_{s,j} \in \mathbb{C}^L \). The sparse code matrix \( B \in \mathbb{C}^{L \times SN^2} \) is constructed by horizontally concatenating all sparse codes. Then \( ||B||_0 = \sum_{s=1}^{S} \sum_{j=1}^{N^2} ||b_{s,j}||_0 \) is the total number of nonzero elements in the sparse codes, and \( \beta \) determines its maximum sparsity level. The additional term \( Q(W) = 0.5 ||W||_F^2 - \log |\det W| \) imposes regularization on the transform to avoid degenerate solutions.

To solve the resulting problem in (3), we utilize the alternating minimization approach which has been commonly used in the development of dictionary and transform learning algorithms. That is, in each update step, the minimization problem is solved for only one of the variables \( x, W \), and \( B \), while the other two are kept fixed. Each of these update steps (i.e., image update, transform learning, and sparse coding) takes forms that are efficient to compute.

The transform learning and sparse coding problems are similar to the ones in [10], except that a single transform is used and learned for all spectral images. The optimization problem for the transform learning step has a closed form solution involving singular value decomposition [23] once it is rewritten as follows:

\[ \min_W ||WX - B||_F^2 + 0.5\lambda ||W||_F^2 - \lambda \log |\det W| \]

where \( X \in \mathbb{R}^{L \times SN^2} \) consists of \( P_j x_s \) vectors in its columns. Sparse coding problem is also given by:

\[ \min_B ||WX - B||_F^2 \text{ s.t. } ||B||_0 \leq \beta \]

and its solution can be efficiently obtained by assigning the largest \( \beta \) elements of \( WX \) to \( B \), and zeroing the others.

The image update step is different and requires solving the following least-squares problem for \( x \):

\[ \min_x \nu ||y - H x||_2^2 + \sum_{s=1}^{S} \sum_{j=1}^{N^2} ||WP_j x_s - b_{s,j}||_2^2 \]

After a reformulation and solving the normal equation, the solution for the image update step is given by

\[ x = (\nu H^H H + G)^{-1}(c + \nu H^H y) \]

Here \( G \) is a block diagonal matrix with \( S \times S \) blocks of \( N^2 \times N^2 \) elements, with each block given by \( \Psi = \sum_{j=1}^{N^2} P_j^T W^H W P_j \). That is, \( G = IS \otimes \Psi \) where \( IS \) is \( S \times S \) identity matrix and \( \otimes \) is the Kronecker product. The vector \( c = [c_1^T \ldots c_S^T]^T \) with \( c_s = \sum_{j=1}^{N^2} P_j^T W^H b_{s,j} \).

This image update can be efficiently computed by exploiting the properties of \( H \) and \( \Psi \) matrices. It has been shown that \( \Psi \) is a block circulant matrix with circular blocks (BCCB), and hence diagonalized by the DFT matrix \( F \) [10]. That is, \( \Psi = F^H \Gamma F \) where \( \Gamma \) is a diagonal matrix whose diagonal can be computed by taking the DFT of the first column of \( \Psi \). Moreover, each block of \( H \) is diagonalized by the DFT matrix since \( H_{k,s} \) is also a BCCB matrix. The diagonal corresponding to its diagonal matrix, \( \Lambda_{k,s} \), can be computed by taking the DFT of the first column of \( H_{k,s} \), which corresponds to the PSF \( h_{d_k,\lambda_s} \). Let \( \Lambda \) be the matrix composed of \( K \times S \) blocks given by \( \Lambda_{k,s} \) for \( k = 1, \ldots, K \) and \( s = 1, \ldots, S \).

Efficient computation of Eqn. (7) hence follows from the special structure of the matrix \( A = \nu H^H H + G \). Note that \( A \) is a matrix of \( S \times S \) blocks, with each block given by

\[ A_{i,j} = \nu \sum_{k=1}^{K} H_{k,i}^H H_{k,j} + \delta_{i,j} \Psi = F^H \Sigma_{i,j} F \]

where \( \Sigma_{i,j} = \nu \sum_{k=1}^{K} \Lambda_{k,i}^H \Lambda_{k,j} + \delta_{i,j} \Gamma \) with \( \delta_{i,j} \) being the Kronecker delta function and \( i, j = 1, \ldots, S \). Then \( A \) can be rewritten as \( A = F^H \Sigma F \) where \( \tilde{F} = I_S \otimes F \) and \( \Sigma \) is a matrix of \( S \times S \) blocks each given by \( \Sigma_{i,j} \). Since the inverse of \( A \) is given by \( \tilde{F}^{-1} \Sigma^{-1} \tilde{F} \), efficient computation of the image update step follows from rewriting Eqn. (7) as follows:

\[ x = \tilde{F}^H \Sigma^{-1} \tilde{F} (c + \nu H^H y) = \tilde{F}^H \Sigma^{-1} (\tilde{F} c + \nu A^H \tilde{F} y) \]

Here \( \tilde{F} c \) and \( \tilde{F} y \) can be computed efficiently using FFT since \( \tilde{F} c = [(F c_1)^T \ldots (F c_S)^T]^T \) and \( \tilde{F} y \) is similar. Note that \( \Sigma^{-1} \) can also be efficiently computed through a recursive block matrix inversion approach [24]. The fact that each block is a diagonal matrix simplifies this computation, resulting in only elementwise multiplication and division operations. The image update step is summarized as follows:

**Algorithm I Image Update Step**

**Inputs:** y - measurements, PSFs \( h_{d_k,\lambda_s} \), \( W \) - current adapted transform, \( b_{s,j} \) - current sparse code vectors

**Outputs:** x - updated image vector

1. Compute \( \Lambda_{k,s} \) from FFT of the corresponding PSF \( h_{d_k,\lambda_s} \).
2. Compute \( \Gamma \) from the FFT of the first column of \( \Psi \), and form \( \Sigma \).
3. Compute the inverse of \( \Sigma \) through recursive inversion.
4. Compute the image \( c_s = \sum_{j=1}^{N^2} P_j^T W^H b_{s,j} \) for \( s = 1, \ldots, S \).
   Construct \( \tilde{F} c \) from FFT(\( c_s \)).
5. Construct \( \tilde{F} y \) from FFT(\( y_s \))’s for \( k = 1, \ldots, K \).
6. Compute \( \tilde{F} c + \nu A^H \tilde{F} y \).
7. Update the images \( x \) as IFFT(\( \Sigma^{-1} (\tilde{F} c + \nu A^H \tilde{F} y) \)).
step. All of these have the same complexity of \( O(L^2SN^2) \), which is the complexity of each algorithm iteration. Computing the inverse of \( \Sigma \) recursively costs \( O(S^2N^2) \) where \( S \ll L^2 \) typically. In the numerical simulations performed in Matlab, average run time of the transform learning algorithm for \( S = 2 \) case was 40 seconds with an Intel Core i5 3.2GHz processor and 8GB RAM.

### 4. NUMERICAL RESULTS

Here we present numerical results to illustrate the performance of the developed reconstruction method for solar spectral imaging. We also compare the performance with the \( \ell_1 \)-norm based reconstruction method [3] that utilizes fixed transforms. Reconstruction results are presented for two realistic scenarios. In the first scenario, the object of interest is a polychromatic source which emits two quasi-monochromatic waves at wavelengths \( \lambda_1 = 33.4 \) nm and \( \lambda_2 = 33.5 \) nm (i.e. \( S = 2 \)). For the second scenario, a third emission at wavelength \( \lambda_3 = 33.3 \) nm also exists. For the photon sieve, a sample design in [2] for EUV solar imaging is considered, with the outer diameter of the photon sieve as 25 mm, and the diameter of the smallest hole as 5 \( \mu \)m. This results in a photon sieve with focal lengths \( f_1 = 3.742 \) m, \( f_2 = 3.731 \) m, and \( f_3 = 3.754 \) m for each wavelength. The photon sieve system takes measurements at the focal planes of each of these wavelengths (\( K = 2 \) and 3 for the first and second scenarios).

We use solar EUV images of size \( 128 \times 128 \) as the inputs to the photon sieve system. Using the forward model in (2), we generated a data set for the signal-to-noise ratios (SNRs) of 10 to 30 dB. Fig. 2 shows sample measurements at the two focal planes for 25 dB SNR case together with the contributions from each spectral component. The reconstructed images from these two measurements using \( \ell_1 \)-norm based regularization with patch-based DCT and the developed transform learning based method are shown in Fig. 3, together with the only-diffraction-limited versions of the original scenes. The average peak-signal-to-noise ratios (PSNRs) between reconstructions and diffraction-limited images are also provided in Table 1 for different SNR levels and different reconstruction methods involving DCT, patch-based DCT, and adaptively learned transform. The reconstruction PSNRs of the second scenario which reconstructs three images from three measurements are also given in Table 2 for the three different SNR values.

Table 1: Comparison of average PSNRs (dB) for different reconstruction methods and different SNRs.

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>Full DCT</th>
<th>Patch Based DCT</th>
<th>Transform Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>28.34</td>
<td>30.74</td>
<td>31.40</td>
</tr>
<tr>
<td>15</td>
<td>30.79</td>
<td>32.79</td>
<td>33.09</td>
</tr>
<tr>
<td>20</td>
<td>32.93</td>
<td>34.54</td>
<td>35.29</td>
</tr>
<tr>
<td>25</td>
<td>35.06</td>
<td>35.75</td>
<td>36.97</td>
</tr>
<tr>
<td>30</td>
<td>37.02</td>
<td>37.69</td>
<td>39.35</td>
</tr>
</tbody>
</table>

The results in Fig. 3 and Table 1 demonstrate improvements of up to 2dB in the reconstruction quality as compared to the previously developed method that uses fixed transforms (DCT or patch-based DCT) with \( \ell_1 \)-norm regularization.

### 5. CONCLUSION

We have presented an efficient, data-adaptive, sparse image reconstruction method for the high-resolution PSSI system based on transform learning. Substantial improvements in the reconstruction quality is achieved with this method as compared to approaches with fixed analytical transforms. In fact, for many SNRs of practical interest (30 dB or higher), almost diffraction-limited imaging performance can be achieved through this adaptive sparse reconstruction. More substantial improvements are expected to be achieved in compressed sensing scenarios and with more recent learning-based reconstruction methods, which is a topic of future study.
6. REFERENCES


