High-throughput hyperspectral infrared camera

Jonathan M. Mooney, Virgil E. Vickers, Myoung An, and Andrzej K. Brodzik

Rome Laboratory, Hanscom Air Force Base, Massachusetts 01731

Received January 24, 1997; revised manuscript received April 11, 1997; accepted April 11, 1997

A spectral imager constructs a three-dimensional (two spatial and one spectral) image from a series of two-dimensional images. We discuss a technique for spectral imaging that multiplexes the spatial and spectral information on a staring focal plane and then demultiplexes the resulting imagery to obtain the spectral image. The spectral image consists of 100 × 100 spatial pixels and 25 spectral bands. The current implementation operates over the 3–5-μm band, but can easily be applied to other spectral regions. This approach to spectral imaging has high optical throughput and is robust to focal plane array nonuniformities. A hardware description, the mathematical development, and experimental results are presented. © 1997 Optical Society of America [S0740-3232(97)00411-0]

1. INTRODUCTION

Often objects of similar intensity have different spectral signatures that can be used to distinguish between them. This work is motivated by our desire to perform spectral discrimination in the midwave infrared. From previous experience we know that infrared scenes consist of low-contrast details on a large background pedestal, and that focal plane array (FPA) nonuniformities and low optical throughput often limit the performance of infrared spectral imagers.[1] Our goal is to address these potential performance-limiting issues by incorporating both a throughput and a multiplex advantage into a spectral imager and at the same time address the effect of FPA nonuniformity on spectral imaging.

Typical state-of-the-art spectral imagers image the spectrum of a slit onto a two-dimensional (2D) FPA; the slit is scanned over the object to create a three-dimensional (3D) spectral image.[2] Since only one slit is imaged at a time, the out-of-slit photons are rejected, hence these spectral imagers are inherently inefficient. Further, pixel-to-pixel nonuniformities on the FPA can corrupt the spectral imagery and reduce the ultimate performance.

The spectral imager we study is similar to previously reported chromotomographic devices,[3–7] all of which image a dispersed view of the entire scene over all of the spectral bands and then computationally reconstruct the chromatic image. In chromotomography the color imagery is considered to be pseudo three dimensional. The 3D spectral image is obtained by use of tomographic imaging techniques similar to those used in medical imaging. While chromotomography is computationally demanding, it has the advantage that a slit field stop is not required. We anticipate that the spectral imager performance will reflect the larger optical throughput afforded by a larger field stop. The actual performance and potential multiplex advantage (or disadvantage) depend on how well the reconstruction algorithms perform.

Our technique differs from other chromotomographic approaches in several ways: We use a direct vision prism to disperse the light; we have derived a reconstruction method that is uniquely suited to our hardware configuration;[8,9] we have developed some new ideas for dealing with some of the limitations of chromotomography, and we have applied principal component analysis to the characterization of our experimental data. Here we describe the hardware, the reconstruction technique, and the limitations of our approach. We demonstrate infrared spectral imaging and demonstrate infrared spectral discrimination. We will also demonstrate that the imager is robust to the FPA nonuniformities that often limit infrared imager performance.

2. HARDWARE SCHEMATIC REPRESENTATION

Figure 1 is a schematic representation of our approach. It consists of a telescope, a direct-vision prism, and a camera. A direct-vision prism consists of two prisms that are arranged such that one wavelength passes through undeviated while other wavelengths are deviated along a line. The direct vision prism is mounted in a bearing so that it can be rotated around the optical axis. As the prism is rotated the spectral features in the image will trace out circles where the radius of the circles are wavelength dependent. To produce a spectral image, a series of frames of camera data are obtained with the prism in different angular orientations, and then a computer is used to extract the spectral image from the measured data.

If in Fig. 1 the field stop were replaced by a slit, the prism mount were fixed, and a mechanism for linearly scanning the scene added, then the modified system would be a standard scanned-slit spectral imager. In both cases the throughput is determined by the f-number and the field stop area. Since the f-numbers are the same and the field stop area is much larger for the angularly multiplexed configuration, it has higher throughput.

3. RECONSTRUCTION ALGORITHM

Other work in tomographic spectral imaging has relied heavily on the tomographic reconstruction techniques used in medical imaging. Here we develop a technique

0740-3232/97/112951-11$10.00

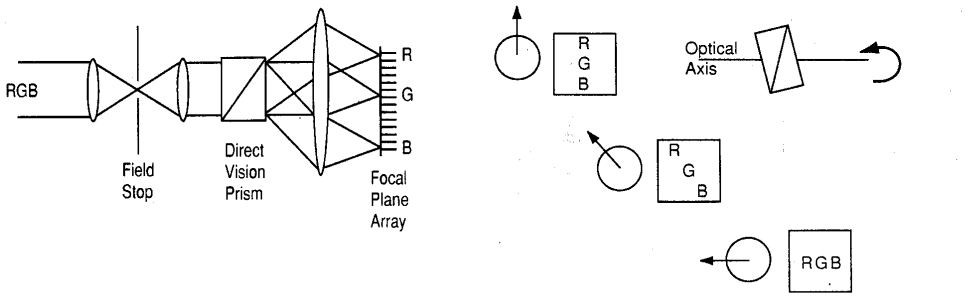

Fig. 1. Schematic representation of the spectral imager and its operation. On the left, the direct-vision prism is shown spreading red, green, and blue light across the FPA. On the right, the effect of the prism rotation on the overlapping red, green, and blue images is illustrated. The circles and arrows represent the prism orientation.

for inverting the system matrix directly. Direct matrix inversion is typically not used for tomography for several reasons, including that[10,11] (1) the matrix is prohibitively large, (2) a unique solution may not exist, and (3) no solution may exist, owing to noise. Our inversion technique avoids these difficulties by first block diagonalizing the matrix to reduce the size and then using singular value decomposition to address matrix singularities.

We note that our system matrix is quite similar to that of ectomography.[12–14] It is likely that the matrix inversion technique that we develop here could be useful there as well. In both cases all of the projections fall on the same plane, and both suffer from limited viewing angle (the concept of limited viewing angle is discussed in Section 4). Other possible applications of our block diagonalization technique are diffractive-optic-type imaging spectrometers[15,16] and 3D microscopes.[17]

Assume that we want to construct a three-color image of a scene $c(x, y, \lambda)$. We divide the scene into three spectral bands, $c_1(x, y)$, $c_2(x, y)$, and $c_3(x, y)$ (e.g. red, green, and blue in Fig. 1). The action of the prism–camera combination on the spectral image is to disperse each of the colors relative to one another. For example, with reference to Fig. 1, $c_2(x, y)$ (green) is undeviated, while for the first prism orientation $c_1(x, y)$ and $c_3(x, y)$ (red and blue) are deviated up and down, respectively. As the prism is rotated, the angular orientation of the dispersion will change while its magnitude remains unchanged.

Since an image displacement can be represented as a convolution with a displaced delta function, the data recorded by the computer will be[8]

$$\begin{aligned} r_1(x, y) &= p_{1,1}(x, y) ** c_1(x, y) + p_{1,2}(x, y) ** c_2(x, y) \\ &\quad + p_{1,3}(x, y) ** c_3(x, y) \\ r_2(x, y) &= p_{2,1}(x, y) ** c_1(x, y) + p_{2,2}(x, y) ** c_2(x, y) \\ &\quad + p_{2,3}(x, y) ** c_3(x, y) \\ r_3(x, y) &= p_{3,1}(x, y) ** c_1(x, y) + p_{3,2}(x, y) ** c_2(x, y) \\ &\quad + p_{3,3}(x, y) ** c_3(x, y) \\ r_4(x, y) &= p_{4,1}(x, y) ** c_1(x, y) + p_{4,2}(x, y) ** c_2(x, y) \\ &\quad + p_{4,3}(x, y) ** c_3(x, y), \end{aligned} \tag{1}$$

where $r_m(x, y)$ is the data recorded for frame m at location (x, y), $p_{m,k}(x, y)$ is the point-spread function of spectral band k recorded on image m, and $**$ indicates a 2D convolution. Each of the $r_1(x, y)$, $r_2(x, y)$, $r_3(x, y)$, and $r_4(x, y)$ correspond to the recorded image for a different angular orientation of the direct vision prism.

Here each of the $p_{m,k}(x, y)$ is a displaced delta function and is given by

$$\begin{aligned} p_{m,k}(x, y) = \delta[&x - (k - k_0)\Delta x \cos(\phi_m), \\ &y - (k - k_0)\Delta x \sin(\phi_m)], \end{aligned} \tag{2}$$

where Δx is the pixel dimension on the FPA, k_0 is the index of the undeviated color, and ϕ_m is the angular orientation of the prism on frame m. As the prism is rotated, the imagery in spectral band k_0 is unchanged, while the imagery from spectral band k orbits about that of k_0 on a circular path with radius $(k - k_0)\Delta x$. In Eq. (2) we have assumed that the pixels on the FPA are square, and we have defined the spectral band width as the change in wavelength required to change the dispersive displacement by one pixel on the FPA.

We wish to solve the system of equations in Eq. (1) for $c_k(x, y)$. First we take the spatial 2D Fourier transform of each equation:

$$\begin{aligned} R_1(\xi, \zeta) &= P_{1,1}(\xi, \zeta)C_1(\xi, \zeta) + P_{1,2}(\xi, \zeta)C_2(\xi, \zeta) \\ &\quad + P_{1,3}(\xi, \zeta)C_3(\xi, \zeta), \\ R_2(\xi, \zeta) &= P_{2,1}(\xi, \zeta)C_1(\xi, \zeta) + P_{2,2}(\xi, \zeta)C_2(\xi, \zeta) \\ &\quad + P_{2,3}(\xi, \zeta)C_3(\xi, \zeta), \\ R_3(\xi, \zeta) &= P_{3,1}(\xi, \zeta)C_1(\xi, \zeta) + P_{3,2}(\xi, \zeta)C_2(\xi, \zeta) \\ &\quad + P_{3,3}(\xi, \zeta)C_3(\xi, \zeta), \\ R_4(\xi, \zeta) &= P_{4,1}(\xi, \zeta)C_1(\xi, \zeta) + P_{4,2}(\xi, \zeta)C_2(\xi, \zeta) \\ &\quad + P_{4,3}(\xi, \zeta)C_3(\xi, \zeta), \end{aligned} \tag{3}$$

where we use capital letters to indicate that we have Fourier transformed, and ξ and ζ are the spatial frequency variables.

The form of Eq. (3) can be reduced to

$$\begin{pmatrix} R_1(\xi,\zeta) \\ R_2(\xi,\zeta) \\ R_3(\xi,\zeta) \\ R_4(\xi,\zeta) \end{pmatrix} = \begin{bmatrix} P_{1,1}(\xi,\zeta) & P_{1,2}(\xi,\zeta) & P_{1,3}(\xi,\zeta) \\ P_{2,1}(\xi,\zeta) & P_{2,2}(\xi,\zeta) & P_{2,3}(\xi,\zeta) \\ P_{3,1}(\xi,\zeta) & P_{3,2}(\xi,\zeta) & P_{3,3}(\xi,\zeta) \\ P_{4,1}(\xi,\zeta) & P_{4,2}(\xi,\zeta) & P_{4,3}(\xi,\zeta) \end{bmatrix} \times \begin{pmatrix} C_1(\xi,\zeta) \\ C_2(\xi,\zeta) \\ C_3(\xi,\zeta) \end{pmatrix}, \quad (4)$$

which is simply a large array of vector-matrix multiplications

$$\mathbf{R}(\xi,\zeta) = \boldsymbol{P}(\xi,\zeta)\mathbf{C}(\xi,\zeta). \quad (5)$$

Note that our large system of equations has been block diagonalized; we have a vector-matrix multiplication at each spatial frequency rather than one large vector-matrix multiplication. This block diagonalization is a key computational consideration for our hardware realization and subsequent data processing. For the hardware implemented, we solve for 50 240 × 240 spectral images from a sequence of 80 240 × 240 input images. Hence we invert 57,600 50 × 80 matrices rather than one 4,608,000 × 2,880,000 matrix. Since $\boldsymbol{P}(\xi,\zeta)$ is system-dependent and not scene-dependent, once computed, its inverse can be stored and need not be recomputed until the system is changed.

For the continuous case each of the $P_{m,k}(\xi,\zeta)$ is given by

$$P_{m,k}(\xi,\zeta) = \exp\{-2\pi i[\xi\cos(\phi_m) + \zeta\sin(\phi_m)] \times (k - k_0)\Delta x\}. \quad (6)$$

$P_{m,k}(\xi,\zeta)$ can be rewritten as

$$P_{m,k}(\xi,\zeta) = P_{m,k}(\boldsymbol{\nu}) = \exp(-2\pi i\boldsymbol{\nu}\cdot\boldsymbol{\rho}), \quad (7)$$

where $\boldsymbol{\rho}$ is the vector $[(k - k_0)\Delta x\cos(\phi_m), (k - k_0)\Delta x\sin(\phi_m)]$. It can also be rewritten as

$$P_{m,k}(\xi,\zeta) = (\exp\{-2\pi i[\xi\cos(\phi_m) + \zeta\sin(\phi_m)]\Delta x\})^{(k-k_0)}, \quad (8)$$

where it is evident that each $\boldsymbol{P}(\xi,\zeta)$ is Vandermonde[18] ($\boldsymbol{P}$ has elements $[p_m]^k$). If the algorithm is to be implemented on a computer, the expression for $P_{m,k}(\xi,\zeta)$ is complicated by the fact that $\boldsymbol{\rho}$ may not fall exactly on a pixel center. We address this issue by assuming that the effect on the four pixels nearest $\boldsymbol{\rho}$ is bilinearly weighted relative to their distance from $\boldsymbol{\rho}$. This is equivalent to assuming that the optical point-spread function is a uniform square the same size as a pixel and that the pixels have 100% active area.

To this point we have neglected the effects of noise. If we assume that the noise is additive, then Eq. (5) becomes

$$\mathbf{R}(\xi,\zeta) = \boldsymbol{P}(\xi,\zeta)\mathbf{C}(\xi,\zeta) + \mathbf{N}(\xi,\zeta), \quad (9)$$

where $\mathbf{N}(\xi,\zeta)$ is the 2D discrete spatial Fourier transform of the additive noise vector.

To solve for $\mathbf{C}(\xi,\zeta)$, we need to invert $\boldsymbol{P}(\xi,\zeta)$. Even though efficient algorithms exist for inverting the Vandermonde matrix $\boldsymbol{P}(\xi,\zeta)$, two factors lead us to use the singular-value-decomposition[19] (SVD): $\boldsymbol{P}(\xi,\zeta)$ is rectangular, and for small values of ξ and ζ it will be singular. SVD factors an M × K matrix into a product of an M × K column-orthonormal matrix $\boldsymbol{U}$, a K × K diagonal matrix $\boldsymbol{W}$ with real nonnegative elements, and a K × K orthonormal matrix $\boldsymbol{V}^\dagger$,

$$\boldsymbol{P} = \boldsymbol{U}\boldsymbol{W}\boldsymbol{V}^\dagger. \quad (10)$$

Since $\boldsymbol{U}$ and $\boldsymbol{V}$ are column orthonormal, we find that

$$\boldsymbol{U}^\dagger\boldsymbol{U} = \boldsymbol{V}^\dagger\boldsymbol{V} = \mathbf{1}, \quad (11)$$

where the dagger indicates Hermitian adjoint (transpose conjugate). If all the $w_{i,i}$ are nonzero, the inverse of the diagonal matrix $\boldsymbol{W}$ is illustrated by

$$\begin{bmatrix} 1/w_{1,1} & & \\ & \ddots & \\ & & 1/w_{K,K} \end{bmatrix}\begin{bmatrix} w_{1,1} & & \\ & \ddots & \\ & & w_{K,K} \end{bmatrix} = \begin{bmatrix} \mathbf{1} \end{bmatrix}. \quad (12)$$

We find that for an M × K matrix $\boldsymbol{P}$ we can make a pseudoinverse $\boldsymbol{V}\boldsymbol{W}^{-1}\boldsymbol{U}^\dagger$, such that

$$\mathbf{1} = \boldsymbol{V}\boldsymbol{W}^{-1}\boldsymbol{U}^\dagger\boldsymbol{P}. \quad (13)$$

In the presence of noise we estimate $\mathbf{C}(\xi,\zeta)$ by $\tilde{C}(\xi,\zeta)$, where

$$\tilde{C}(\xi,\zeta) = \boldsymbol{V}(\xi,\zeta)\boldsymbol{W}^{-1}(\xi,\zeta)\boldsymbol{U}^\dagger(\xi,\zeta)\mathbf{R}(\xi,\zeta), \quad (14)$$

where only the matrix $\boldsymbol{W}(\xi,\zeta)$ is potentially singular, which occurs when one of its diagonal elements is zero. In this application the noise in the recorded image data $r(x, y, \phi_m)$ is effectively divided by $w_{i,i}(\xi,\zeta)$, so if the value of any of the $w_{i,i}(\xi,\zeta)$ is small, the noise will be greatly amplified. To circumvent this problem, we replace $\boldsymbol{W}^{-1}(\xi,\zeta)$ with the diagonal matrix $\tilde{\boldsymbol{W}}^{-1}(\xi,\zeta)$, with elements

$$\tilde{w}_{i,i}(\xi,\zeta) = \frac{w_{i,i}(\xi,\zeta)}{w_{i,i}^2(\xi,\zeta) + \epsilon^2}. \quad (15)$$

The optimum value of ϵ is data, noise, and spatial frequency dependent; however, we typically use a constant value on the order of unity.

The effect of singular matrices can be further investigated by substituting for $\boldsymbol{R}(\xi,\zeta)$ in Eq. (14); we also drop the explicit (ξ,ζ) reference to simplify the expression:

$$\tilde{C} = \boldsymbol{V}[\tilde{\boldsymbol{W}}^{-1}\boldsymbol{W}\boldsymbol{V}^\dagger\mathbf{C} + \tilde{\boldsymbol{W}}^{-1}\boldsymbol{U}^\dagger\mathbf{N}]. \quad (16)$$

The columns of $\boldsymbol{V}$ that correspond to the small values of $w_{i,i}$ define the null space of our imaging system. From Eq. (16) we see that in the null space both the scene information and the noise are attenuated. Sections 4 and 5 analyze the null space and describe our treatment of it.

We digress to introduce a notational variation that will be useful later. In the preceding development we showed that the measured data $\mathbf{r}$ could be described mathematically in terms of Fourier transforms and matrix multiplications of the chromatic data $\mathbf{c}$, where we organized the various 2D images into 1D vectors. If we assume that all of the housekeeping associated with reordering the data is understood, it is possible to summarize Eq. (1) (with additive noise) as

$$\mathbf{r} = \mathcal{F}^{-1}\mathcal{P}\mathcal{F}\mathbf{c} + \mathbf{n}, \quad (17)$$

where $\mathcal{F}$ is the 2D Fourier transform operator for the reordered inputs $\mathbf{c}$ and outputs $\mathbf{r}$, and $\mathcal{P}$ is the block-diagonalized assembly of matrices $\boldsymbol{P}$. Though we are being cavalier with the details, this condensed notation highlights the important issues and ignores distracting indexing. We can extend Eq. (17) to include the SVD of $\boldsymbol{P}$,

$$\mathbf{r} = \mathcal{F}^{-1}\mathcal{U}\mathcal{W}\mathcal{V}^{\dagger}\mathcal{F}\mathbf{c} + \mathbf{n}. \tag{18}$$

We will find Eq. (18) useful in our discussion of filling in the missing cone.

4. CONE OF MISSING INFORMATION

In general, when one of the $w_{i,i}(\xi, \zeta)$ values becomes too small, there is insufficient information to determine how the measured signal should be distributed among the spectral bands; information has been lost. Both Descour and Dereniak[6] and Bernhardt[7] demonstrated that the missing information is similar to the missing cone in some medical imaging techniques[20] (the illustrations in Ref. 6 are particularly illuminating).

The origin of the missing cone is best understood in the context of the Radon transform combined with the central slice theorem.[21] From the central slice theorem, we know that the 2D Fourier transform of a projection is equal to a slice through the origin of the 3D Fourier transform of the spectral image. By rotating the prism, we obtain different slices through the origin, each for a different projection direction but with the same angle θ between the projection plane and the optical axis (θ is defined relative to the normal to the plane). In the limit of an infinite number of different projections, the projections will define a cone, where the half-angle of the cone vertex is θ. Since none of the projections provide any information about the image inside this cone, it is referred to as a missing cone.

Figure 2 illustrates that the image through the direct-vision prism is analogous to a projection through the pseudovolume defined by two spatial dimensions and one spectral dimension. From Fig. 2 we can also identify the cone (triangle in the figure) of missing information. To estimate the impact of the missing cone on our imaging system, we calculate the fractional volume of the missing cone, which is determined by the projection angle θ. Our interpretation of the projection angle is governed by (1) our definition of adjacent spectral bands as displaced by one pixel on the focal plane and (2) our assumption that the volume object elements (voxels) are cubic. The definition is easily justified since it is consistent with scanned slit configurations. The assumption is completely arbitrary but has the benefit of symmetric simplicity and leads to a missing cone half-angle of 45 deg.

Another perspective of the missing cone is obtained by analyzing the singular values of $\mathcal{P}$. When the singular values are small, information is lost. We expect the small singular values to define a conical shape. Figure 3 is a graphic representation of the values of $w_{i,i}(\xi, \zeta)$ illustrating where in the frequency domain information is lost. The dark regions in Fig. 3 are due to the missing cone. Although the physical interpretation is identical to that in Ref. 20, the mathematical treatment here differs from that commonly found in the literature, since this treatment uses a hybrid of Fourier-determined and SVD-determined basis functions. Figure 3 also supports our conclusion that the cone angle is 90 deg: at the highest chromatic frequency, the cone diameter covers ± Nyquist, as we expect for the configuration illustrated in Fig. 2.

An intuitive understanding of the missing cone can be obtained by considering the sequence of frames recorded for a monochromatic source of infinite extent that is perfectly uniform. Since the source is uniform and of infinite extent, the sequence of frames will be unaffected by the rotation of the prism (if we consider a system with no field stop). The same sequence of recorded frames would be recorded for any wavelength monochromatic source or

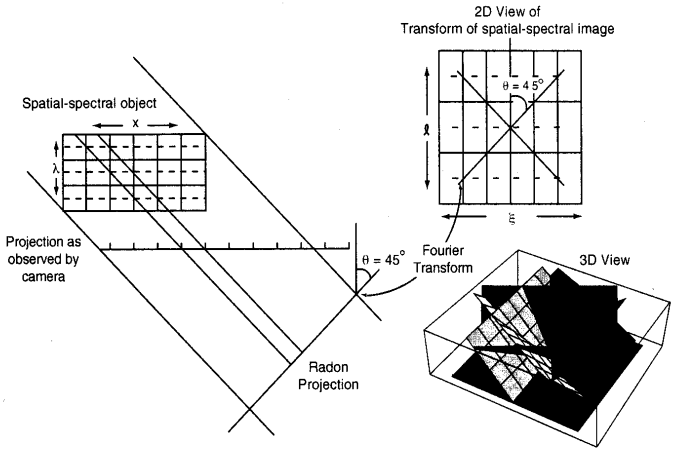

Fig. 2. Correspondence of our recorded data to Radon projections. The projection angle follows from the definition that adjacent spectral bands are displaced by one pixel on the FPA and the assumption that the object volume elements are equilateral. The spectral image can be reconstructed in the frequency domain by using the central slice theorem (shown in both two and three dimensions).

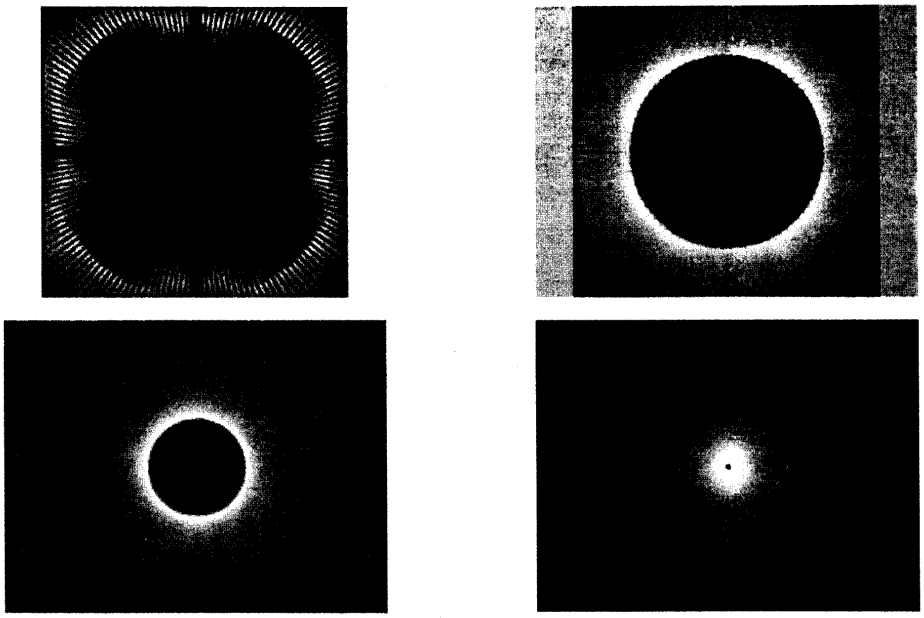

Fig. 3. Illustration of the $w_{i,j}(\xi,\zeta)$ values as a function of spatial frequency. Each of these images corresponds to a horizontal slice through the 3D representation of the spatial–spectral transform image on the right side of Fig. 2. Zero spatial frequency is in the center of each image. Bright and dark regions correspond to large and small values of $w_{i,j}(\xi,\zeta)$, respectively. The sequence from left to right, top to bottom corresponds to descending chromatic-frequency data. The lower right image corresponds to image features that change slowly as a function of wavelength (low chromatic frequencies). The upper-left image corresponds to image features that change rapidly as a function of wavelength (high chromatic frequencies). The dark region in the center of each image corresponds to the spatial–chromatic frequencies where information has been lost.

for any broadband extended source that was perfectly uniform. Therefore there is no way to distinguish the spectral content of a spatially uniform source algorithmically, and the zero spatial frequency information is lost.

Next consider a system with a slit field stop one pixel wide, with the prism dispersion oriented normal to the slit. Since the field stop is only one pixel wide, spatial and spectral information do not overlap, and the slit object is easily reconstructed from one projection without a missing cone. We could also use our matrix formalism to reconstruct the object; however, in this case the reconstruction would suffer from a missing cone. We are left to resolve why the missing cone is present with one reconstruction and not with the other from what appears to be the same set of data.

The difference arises because in the former case the information about the field stop is used in the reconstruction, whereas in the latter it is not. The field-stop information can be introduced into the central slice reconstruction by convolving the Fourier transform of the field stop with the projection data. The slice goes from being infinitely thin to having a finite thickness. For an infinitely thin slit the slice becomes infinitely thick and the missing cone is filled, or for a discrete system a slit one pixel wide would completely fill the missing cone.

The field stop $\mathcal{S}$ could be included in the matrix formalism as

$$\mathbf{r} = \mathcal{F}^{-1}\mathcal{W}\mathcal{W}^{\dagger}\mathcal{F}\mathcal{S}\mathbf{c} + \mathbf{n}, \tag{19}$$

where $\mathcal{S}$ defines the field stop and is a square diagonal matrix with zeros and ones on the diagonal. Since both $\mathcal{W}$ and $\mathcal{S}$ are singular, the pseudoinverse must treat them together. This fact destroys the block diagonal structure of the matrix and makes direct inversion of the matrix intractable for most applications. However, in Section 5 we will show that an iterative technique can be used to include the field-stop information in the reconstructed image without losing the block diagonal structure of the matrix.

5. FILLING IN THE MISSING CONE

After the direct reconstruction, the missing cone occupies 26% of the object volume in the frequency domain. Clearly this is an unacceptably large fraction of missing information. In this section we discuss several methods for reducing the fraction of missing information.

The first and easiest way to reduce the missing volume is to filter the data.[13] Figure 4 illustrates the effect of reducing the chromatic resolution by a factor of 2. In this case a rectangle function filter is used to pass only the low chromatic frequency data (the figures and discussion in Ref. 13 are illuminating). The effect of this filter is to reduce the fraction of missing information to 6.5%.

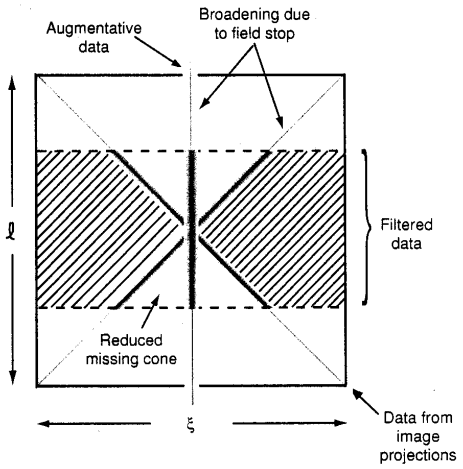

Fig. 4. Schematic representation of how low-pass filtering, augmenting the data, and use of the field-stop prior knowledge can be used to reduce the fractional volume of the missing cone. Low-pass filtering reduces the chromatic resolution but preferentially filters out missing information, the augmentative data fills in the zero spatial-frequency axis information, and the field-stop prior knowledge broadens all of the data samples.

The second approach is to use the field stop to fill in the missing cone.[22] As we discussed in Section 4, direct methods for including the field-stop information fail because of the large computational and memory requirements of inverting the matrix $\mathscr{W}\mathscr{T}\mathscr{F}\mathscr{S}$ directly. We avoid this problem by using an iterative method similar to those in Refs. 11 and 23. The idea is to find a solution that is consistent both with our measured data and with our knowledge of the field stop. We begin by constructing a constraint $\mathscr{T}$ such that $\mathscr{T}\tilde{c}$ is an altered version of our estimate $\tilde{c}$ that is consistent with our knowledge of the field stop. Then we transform $\mathscr{T}\tilde{c}$ into the frequency domain and replace the data values in the missing cone with the values as modified by the constraint. We repeat the process until some halt criterion is reached. In general, the operator $\mathscr{T}$ need not be linear so that we can utilize constraints such as nonnegativity.

The initial inverse gives the result

$$\tilde{c}_0 = \mathscr{F}^{-1}\mathscr{T}\tilde{\mathscr{W}}^{-1}\mathscr{H}^{\dagger}\mathscr{F}\mathbf{r}. \tag{20}$$

Subsequent iterations use the preceding results and the constraint to fill in the missing cone

$$\tilde{c}_j = \mathscr{F}^{-1}\mathscr{T}[\mathscr{L}\tilde{\mathscr{W}}^{-1}\mathscr{H}^{\dagger}\mathscr{F}\mathbf{r} + \mathscr{R}\mathscr{T}^{\dagger}\mathscr{F}\mathscr{T}\tilde{c}_{j-1}], \tag{21}$$

where $\mathscr{L}$ and $\mathscr{R}$ are generalizations of $\boldsymbol{L}(\xi, \zeta)$ and $\boldsymbol{R}(\xi, \zeta)$, which are diagonal with elements

$$L_{i,i}(\xi, \zeta) = \frac{w_{i,i}^2(\xi, \zeta)}{w_{i,i}^2(\xi, \zeta) + \beta^2}, \tag{22}$$

and

$$R_{i,i}(\xi, \zeta) = 1 - L_{i,i}(\xi, \zeta), \tag{23}$$

and β is a constant similar to ϵ. From Eqs. (22) and (23) we see that $\mathscr{L}$ and $\mathscr{R}$ function as a switch that selects between the left-hand and right-hand terms in Eq. (21). Large values of $w_{i,i}(\xi, \zeta)$ select the left-hand term (measured data selected outside the missing cone), and small values of $w_{i,i}(\xi, \zeta)$ select the right-hand term (constrained solutions selected inside the missing cone). In Fig. 4 we see that the effect of the field stop is to broaden the data samples.

The third way to fill in the missing cone is to augment the data set with an alternative spectral technique. In particular we have found it useful to reimage the entrance pupil onto the field stop and extract the on-axis component of the missing cone from the step response at the edge of the field stop. This additional spectral information can be incorporated into the constraint $\mathscr{T}$ in Eq. (21). When combined with the field-stop constraint, the information on axis is broadened, and our image reconstruction very rapidly converges.

A clever technique that uses a zoom lens to fill in the missing cone has been proposed by Bernhardt.[7] We did not use Bernhardt's technique, because an infrared zoom lens with sufficient variation in power was not available. We believe that some relief on the zoom-lens requirements can be achieved by incorporating the methods described immediately above, but a demonstration of that has been left for future work.

An additional technique for filling in the missing cone will be introduced in Section 9 when we discuss the principal components analysis of our data. We have not yet investigated its utility and postpone its discussion until after we have introduced principal components analysis. Investigation of additional methods for reducing the fractional volume of the missing cone is an area of continuing research.

6. IMPLEMENTATION

We verified the operation of the spectral imager by using a PtSi infrared camera imaging over the 3–5-μm band. The camera operates at $f/2.7$ and uses a 320×244 FPA obtained from FLIR Systems. We chose this FPA because it was noninterlaced and had a fill factor of more than 80%. A Diversified Optical Products infrared direct-vision prism was used as the disperser. The IR camera, drive electronics, and 12 bit/pixel real-time consecutive frame data acquisition electronics were all built in house.

The strong dependence of PtSi quantum efficiency on wavelength $\eta(\lambda)$ has both negative and positive consequences. While complicating calibration of the incident photon flux, it compensates for the large imbalance between the photon flux at short and long wavelengths. This compensated balance simplifies measurement of object emissivity and makes PtSi an appealing candidate for spectral imaging applications where measurement of object emissivity rather than measurement of incident flux is required.

A field stop 100×100 pixels on a side was mounted at the real image in the telescope. In this infrared imaging application the field stop is warm and emissive, so it provides the prior knowledge that it is uniform, but its intensity is nonzero and the spectral distribution is unknown. Since we are most interested in spectral discrimination,

we chose to measure the scene spectrum relative to the field stop. We could use this instrument radiometrically by calibrating the field stop, but as yet we have avoided this complication.

The format of the FPA dictated the size of the discrete Fourier transform. We cropped the two top and bottom rows and the 40 left and right columns to use a 240 × 240 Fourier transform by applying the Good–Thomas prime factor algorithm described in Ref. 24. The software to invert the data was written in house in C for Solaris. The matrix decomposition was performed once for a given configuration, and the result was stored in a file on disk. Subsequent inversions accessed the data on disk. The size of the file was reduced by using matrix symmetry properties. If we assume that $\phi_m = 2\pi m/M$, that $0 \leq m < M$, that M is divisible by 4, that $0 \leq k < K$, and that m, M, k, k_0 and K are integers, then the following symmetry properties hold:

$$P_{m,k}(\xi, \zeta) = \begin{cases} P^*_{m,2k_0-k}(\xi, \zeta) \\ P^*_{m+M/2,k}(\xi, \zeta) \\ P^*_{m,k}(-\xi, -\zeta) \\ P_{m+M/4,k}(\zeta, -\xi) \\ P_{M/2-m,k}(-\xi, \zeta) \end{cases} . \tag{24}$$

In effect the decomposition of only one octant need be stored. The symmetry was also used to reduce the data from complex to real. The precision of the stored data was short integer (16 bits). The size of the file for the current configuration was approximately 100 Mbytes. We believe that an additional reduction by a factor of 2 is possible, but we have left it for future work.

The dispersion of the prism, the focal length of the final lens, and the bandpass of the camera established that a broadband point source was spread over 50 pixels ($K = 50$). In the current implementation $M = 80$; we have found this acceptable but suspect that it is not optimal. The operation of the instrument consists of (1) digitally recording the 80 projection images, (2) digitally recording diffused image data with the prism oriented normal to one edge of the field stop (supplementary data used to fill in the missing cone), (3) performing the direct inverse, (4) performing the iterative refinement calculations. For a camera operating at 30 frames/s the data acquisition is completed in less than 3 s. The data processing is completed in under 5 min on a Pentium Pro 200 computer.

7. EFFECT OF FOCAL PLANE ARRAY NONUNIFORMITY

Before we demonstrate the operation of the spectral imager, we would like to highlight one of its most important characteristics (especially in infrared imaging). Infrared imagery is characterized by low-contrast details on a large background pedestal. Pixel-to-pixel nonuniformities on the FPA interact with the large background pedestal to create a spatial noise pattern in the imagery. This effect is well documented[1] and played a large role in the evolution of infrared imagers. Therefore we expect FPA nonuniformity to play a large role in the development of infrared spectral imagers, and we must consider the effect of FPA nonuniformity on our measured data.

As the prism rotates, the nonuniformity is spatially fixed on the FPA and coincides with the undeviated spectral band (also fixed). Additive nonuniformities will be entirely consistent with image features in the undeviated spectral band and will be treated as such by the inversion algorithm. Therefore the undeviated spectral band will be corrupted by FPA additive nonuniformities, while the other spectral bands will not. Note that this effect is much stronger than simple averaging, because we have inverted the system matrix. The system could be designed such that the corrupted band coincides with an atmospheric absorption band or some other band of little interest in order to minimize the effect of nonuniformity on spectral discrimination.

The effect of multiplicative nonuniformities is dominated by the interaction of the nonuniformities with the large background pedestal. Since the background pedestal is independent of prism rotation, the bulk of multiplicative nonuniformities will again be mapped to the undeviated spectral band by the inversion algorithm, and a similar argument can be made for the spectral nonuniformity interaction with the background pedestal. We find that the vast majority of nonuniformity effects appear in the undeviated spectral band.

The fraction of the multiplicative and higher-order nonuniformities that interact with low-contrast image details, the higher frequency component of the $1/f$ noise (drift that occurs during data acquisition), and the interaction of the spectral nonuniformities with the low-contrast details will all appear as noise. These higher-order terms are almost always negligible. The inherent robustness of this technique to FPA nonuniformity and its high throughput are the two major technical reasons for considering this spectral imaging approach.

8. DEMONSTRATION

To demonstrate the operation of the spectral imager, we imaged a laboratory scene that contained objects with different spectral signatures. We used a controlled laboratory scene so that we could investigate the effect of the missing cone on spectral reconstruction and spectral discrimination. We note in passing that the standard technique of imaging the exit slit of a monochrometer to calibrate the spectral imager will give optimistic results, since a slit object has a lot of high-spatial-frequency energy that falls outside the missing cone. Our demonstration scene consists of extended objects, where a large fraction of the scene energy falls inside the missing cone, creating more of a challenge for our reconstruction algorithm.

A visible image of the scene is shown in Fig. 5. It consists of a differential source and a reflective plate. The differential source consists of two plates one in front of the other. The front plate has a four-bar pattern of holes so that the back plate can be seen through the holes. The back plate is held at a temperature elevated above that of the front plate so that in the infrared the source appears as four bright bars.

The reflective plate was illuminated with an incandescent lamp that was controlled by a variable autotransformer. The lamp voltage was adjusted so that in the

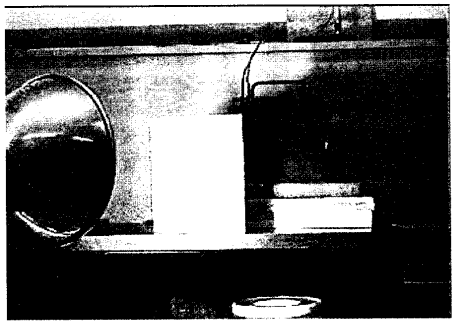

Fig. 5. Visible image of the scene used for the demonstration. The reflective plate and differential source are illuminated by an incandescent lamp. Here the lamp intensity is much greater than when the experiment was performed so that the reflected light can be seen.

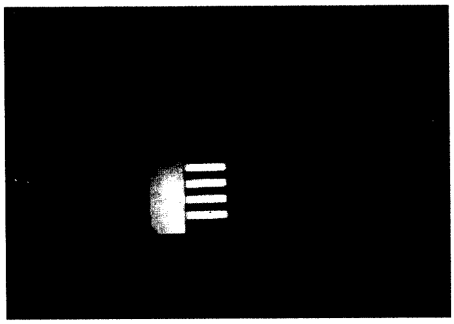

Fig. 6. Broadband infrared image of the scene used for the demonstration. Here the field stop limits the field of view to the central 100 × 100 pixels.

broadband infrared image the signal from the reflective plate matched that of the differential source. A broadband infrared image of the scene is shown in Fig. 6. We expect three major spectral signatures (the background, the hot bars, and the reflective plate) and hope to use the differences between the spectral profiles of the plate and the bars to distinguish between them.

Figures 7 and 8 illustrate one of the projections and one of the spectral reconstructions, respectively. The spectral profiles of a pixel in the plate and of a pixel in a bar are shown in Fig. 9. Since the illumination of the plate is intended to simulate solar illumination, the plate profile is labeled "Solar". We see that the Solar profile is larger at short wavelengths while the thermal profile is larger at long wavelengths and that the profiles cross at approximately 4.1 μm. The spectral notch at 4.27 μm is due to absorption by CO_2 in the atmosphere.

Our goal of spectral discrimination relies on the fact that each of the three principal components of our scene (bars, plate, and background) has a spectral signature that is distinct from that of the other two components. Since our reconstruction of the scene consists, after data filtering, of 25 spectral bands, each of the spectral signatures can be represented by a point in 25-dimensional space defined by the 25 spectral bands. Since our scene consists of only three objects and since three objects define a plane, we can reduce the dimensionality of our scene by finding the two axial components required to describe the plane on which our three object signatures fall. Identifying the principal components of a scene in this manner is called principal component analysis[25]; spectrometers that gather many more spectral bands than there are principal components are referred to as hyperspectral.[26]

In order to identify the two basis spectra that describe the principal components plane, we again utilize SVD (note that the matrices, singular values, and eigenvectors that follow are distinct from those in Sections 3–6). Each pixel's spectral profile is used to form the rows of a matrix that we decompose using SVD. The decomposition yields singular values, spectral eigenvectors (columns of $\boldsymbol{V}$), and spatial eigenvectors (columns of $\boldsymbol{U}$). The spectral eigenvectors that correspond to the two largest singular values define the 2D basis in spectral space that defines our principal components plane.

The two spatial eigenvectors that correspond to the two largest singular values define the plane on which all three object fall. In Fig. 10 each pixel is represented by a bright point on our principal components plane. In the figure we see that the pixels associated with the background, plate, and bars are distinguishable from one another. The pixels that lie between the bars group and the background group are pixels that fall on a bar edge. The

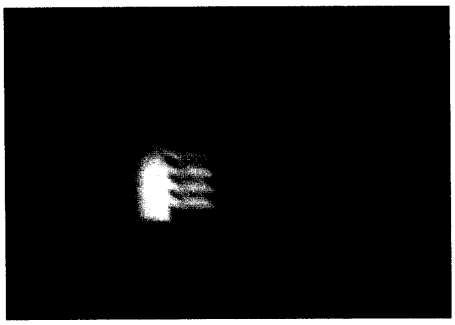

Fig. 7. One of the 80 projections. Here the blurring is due to the prism dispersion.

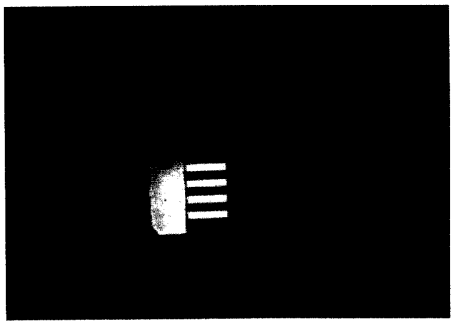

Fig. 8. One of the 50 spectral reconstructions before filtering. This reconstruction corresponds to 3.8 μm.

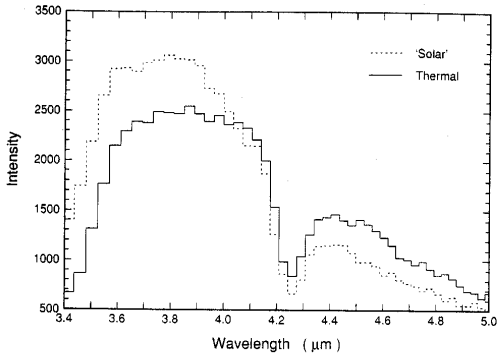

Fig. 9. A comparison of single-pixel spectral profiles from the reflective plate, 'Solar', and a hot bar, Thermal, before filtering. The spectral profiles shown are measured relative to the field stop, which we have assumed to be spectrally flat.

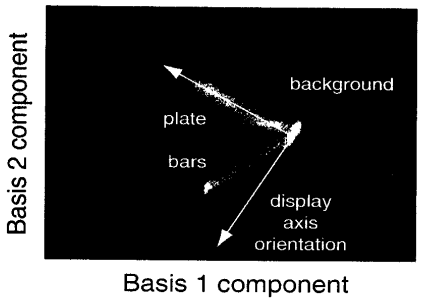

Fig. 10. Scatterplot of the spatial–spectral image data, made with the two leading terms of the SVD. The indicated display axis corresponds to the left image in Fig. 11.

Fig. 11. Spatial display of the spectrally discriminated objects. The center image is a broadband image. On the left the display axis is oriented to map the plate and the background to the same level and therefore accentuate the bars. On the right the plate is accentuated.

fraction of the pixel that is filled by the bar can be determined from its location relative to the two groups.

In Fig. 11 we have linearly combined the two spatial eigenvectors to accent the bars or the plate by rotating the display axis. The left image in Fig. 11 corresponds to the display axis as shown in Fig. 10. Both Fig. 10 and Fig. 11 show that this approach to spectral imaging is effective as a spectral discriminant.

9. ANALYSIS OF THE ITERATIVE ENHANCEMENTS

We have not yet established a meaningful metric to quantify the performance of the spectral imager for all situations. Although the high optical throughput gives us reason to be optimistic, the extent to which the missing cone will degrade performance is not clear. The nature of the missing cone implies that the performance will be data dependent, and our techniques for filling in the missing cone make the data dependence difficult to identify. In this section we address the effect of our fill-in techniques on the fidelity of our reconstructed spectral imagery.

Since the demonstration scene was designed to create data that can be represented by two eigenvectors, analysis of the remaining eigenvectors provides information about the effect of the missing cone on the reconstructed image. Further, by analyzing the singular values as a function of the iteration number, we can estimate how the quality of the reconstruction improves as we iterate.

Figure 12 is a plot of five of the singular values as a function of number of iterations. The scene information is represented by the largest two singular values; the missing cone artifacts are represented by the third and lower singular values. Noise is present in all of the eigenvectors and becomes dominant for the smaller singular values. Our assumption that the scene information is separated from the missing-cone artifacts is supported by the lack of any dependence of the largest two singular values on iteration.

In Fig. 12, we see that the third largest singular value is most affected by the iterative process; its value drops by almost a factor of 3 after four iterations. The spatial eigenvector corresponding to the third-largest singular value is shown in Fig. 13. The hypothesis that Fig. 13 represents a missing-cone artifact is supported by the fact that the dark edge around the bottom part of the plate and the bright areas between the bars are inconsistent with the physical attributes of the scene.

From this analysis we conclude that the iterative procedure has reduced the effect of the missing cone by a factor of nearly 3. Given that the filtering had reduced the fractional missing-cone volume to 6.5%, we estimate that the postiteration missing fractional volume is just above 2% for this particular scene. We could use this approach to reduce further the missing-cone artifacts for this particular scene by eliminating the third and higher eigen-

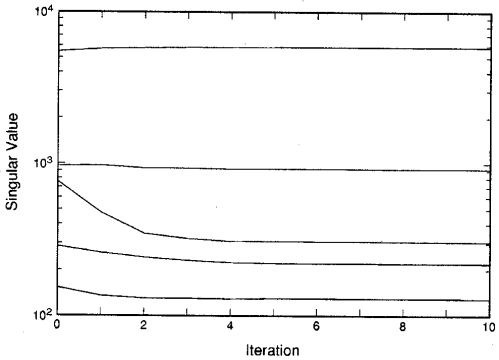

Fig. 12. Four largest and smallest singular values as a function of number of iterations. The decline of the third largest singular value for the first few iterations indicates the effectiveness of the iterative techniques for filling the missing cone.

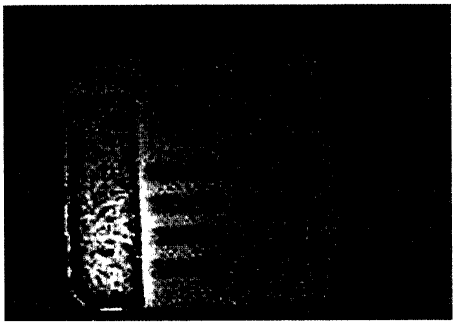

Fig. 13. Enlarged view of the spatial eigenvector corresponding to the third-largest singular value (artifact due to the missing cone) after five iterations. The ringing at the plate edges is indicative of lost information.

vectors from reconstruction altogether, but then we could no longer use the principal components analysis to evaluate the results.

It is also possible that the principal components analysis could be combined with the iterative procedure to eliminate the missing cone entirely. By modifying the constraint $\mathcal{T}$ so that $\mathcal{T}\tilde{c}$ consists of only the principal components, the data in the missing cone is forced to be consistent with principal components. If the weighting of the principal components is not ambiguous in the missing cone then the missing cone could be completely filled. Although development of these ideas remains for future work, we are optimistic about the prospects for eliminating the missing cone.

10. CONCLUSIONS

We have described and demonstrated a high-throughput approach to spectral imaging that is also robust to FPA nonuniformities. We demonstrated our approach in the midinfrared, but it is generally applicable to any spectral band where two-dimensional imaging arrays and direct-vision prisms are available. The high throughput and robustness to FPA nonuniformity mean that this approach can be used with nonuniform high-quantum-efficiency FPA's to minimize the amount of time required to collect high fidelity hyperspectral image data.

We have also shown that this approach is subject to a cone of missing information. In particular, the low-spatial-frequency/high-chromatic-frequency information is missing from the resulting spectral imagery, which leads to limited degradation of the image quality. We demonstrated that by filtering and using iterative techniques, we could reduce the fractional volume of missing information to approximately 2%. In the process of analyzing our data, we demonstrated that principal components analysis is a useful tool for identifying artifacts in image sequences and indicated that it might be useful for eliminating the missing cone. For scenes with substantial spatial structure, this is a promising approach for spectral discrimination.

ACKNOWLEDGMENTS

The authors thank Bill Ewing and Steve DiSalvo for fabricating the infrared camera; John Mead for fabricating the data acquisition interface; Linda Bouthillette for graphic assistance; and Jerry Silverman, Paul Pellegrini, Al Kahan, and Freeman Shepherd for critical reading of this manuscript. This work was supported by U.S. Air Force Office of Scientific Research AFOSR Task 2305CR under Gerry Witt.

REFERENCES

1. J. Mooney, F. Shepherd, W. Ewing, J. Murguia, and J. Silverman, "Responsivity nonuniformity limited performance of infrared staring cameras," Opt. Eng. **28**, 1151–1161 (1989).
2. R. Basedow, D. Carmer, and M. Anderson, "Hydice systems, implementation and performance," in *Imaging Spectrometry I*, M. R. Descour, J. M. Mooney, D. L. Perry, and L. Illing, eds., Proc. SPIE **2480**, 258–267 (1995).
3. G. G. Levin and G. N. Vishnyakov, "On the possibilities of chronotomography of high-speed processes," Opt. Commun. **56**, 231–234 (1985).
4. T. Okamoto, A. Takahashi, and I. Yamaguchi, "Simultaneous acquisition of spectral and spatial intensity distribution," Appl. Spectrosc. **47**, 1198–1202 (1993).
5. Y. Betremieux, T. A. Cook, D. M. Cotton, and S. Chakrabarti, "SPINR: two-dimensional spectral imaging through tomographic reconstruction," Opt. Eng. **32**, 3133–3138 (1993).
6. M. Descour and E. Dereniak, "Computed-tomography imaging spectrometer: experimental calibration and reconstruction results," Appl. Opt. **34**, 4817–4826 (1995).
7. P. A. Bernhardt, "Direct reconstruction methods for hyperspectral imaging with rotational spectrotomography," J. Opt. Soc. Am. A **12**, 1884–1901 (1995).
8. J. M. Mooney, "Spectral imaging via computed tomography," in *Proceedings of the 1994 Meeting of the Infrared Information Symposia Specialty Group on Passive Sensors* (Defense Technical Information Center, Alexandria, Va., 1994), Vol. 1, pp. 203–215.
9. J. M. Mooney, "Angularly multiplexed spectral imager," in *Imaging Spectrometry I*, M. R. Descour, J. M. Mooney, D. L. Perry, and L. Illing, eds., Proc. SPIE **2480**, 65–77 (1995).
10. H. H. Barrett and W. Swindell, *Radiological Imaging* (Academic, New York, 1981), Vol. 2, pp. 433–434.
11. K. C. Tam and V. Perez-Mendez, "Tomographical imaging with limited-angle input," J. Opt. Soc. Am. **71**, 582–592 (1981).
12. P. Edholm, G. Granlund, H. Knutsson, and C. Petersson, "Ectomography—a new radiographic method for reproducing a selected slice of varying thickness," Acta Radiol. Diagn. **21**, 433–442 (1980).
13. H. Knutsson, P. Edholm, G. Granlund, and C. Petersson, "Ectomography—a new radiographic reconstruction method—I. Theory and error estimates," IEEE Trans. Biomed. Eng. **BME-27**, 640–648 (1980).
14. C. Petersson, P. Edholm, G. Granlund, and H. Knutsson, "Ectomography—a new radiographic reconstruction method—II. Computer simulated experiments," IEEE Trans. Biomed. Eng. **BME-27**, 649–655 (1980).
15. D. M. Lyons and K. Whitcomb, "Characterization of the DIOS prototype: a diffractive optic image spectrometer," in *Imaging Spectrometry II*, M. R. Descour and J. M. Mooney, eds., Proc. SPIE **2819**, 206–217 (1996).
16. M. Hinnrichs and M. A. Massie, "Image multispectral sensing: a new and innovative instrument for hyperspectral imaging using dispersive techniques," in *Imaging Spectrometry I*, M. R. Descour, J. M. Mooney, D. L. Perry, and L. Illing, eds., Proc. SPIE **2480**, 93–104 (1995).
17. A. Erhardt, G. Zinser, D. Komitowski, and J. Bille, "Recon-

structing 3-D light-microscope images by digital image processing," Appl. Opt. **24**, 194–200 (1985).
18. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge U. Press, New York, 1988), pp. 52–54.
19. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge U. Press, New York, 1988), pp. 60–72.
20. H. H. Barrett, "Editorial: Limited-angle tomography for the nineties," J. Nucl. Med. **31**, 1688–1692 (1990).
21. H. H. Barrett and W. Swindell, *Radiological Imaging* (Academic, New York, 1981), Vol. 2, pp. 375–446.
22. A. K. Brodzik, J. M. Mooney, and M. An, "Image restoration by convex projections: application to image spectrometry," in *Imaging Spectrometry II*, M. R. Descour and J. M. Mooney, eds., Proc. SPIE **2819**, 231–242 (1996).
23. A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," IEEE Trans. Circuits Syst. **CAS-22**, 735–742 (1975).
24. R. Tolimieri, M. An, and C. Lu, *Algorithms for Discrete Fourier Transform and Convolution* (Springer, New York, 1989).
25. I. T. Jolliffe, *Principal Component Analysis* (Springer-Verlag, New York, 1986).
26. J. W. Boardman, "Analysis, understanding and visualization of hyperspectral data as a convex set in n-space," in *Imaging Spectrometry I*, M. R. Descour, J. M. Mooney, D. L. Perry, and L. Illing, eds., Proc. SPIE **2480**, 14–22 (1995).