Selecting an Optimized COTS Filter Set for Multispectral Plenoptic Sensing

Timothy Doster, Colin C. Olson, Erin Fleet, Michael Yetzbacher
U.S. Naval Research Laboratory - Applied Optics Branch
4555 Overlook Ave., SW, Washington, DC
{timothy.doster,colin.olson,erin.fleet,michael.yetzbacher}@nrl.navy.mil

Abstract

A 16-band plenoptic camera allows for the rapid exchange of filter sets via a 4x4 filter array on the lens's front aperture thus allowing an operator to quickly adapt to a different locale or threat intelligence. Typically, such a system incorporates a default set of 16 equally spaced, non-overlapping, flat-topped filters. Knowing the operating theater or the likely targets of interest it becomes advantageous to tune the filters; we propose a differential evolution approach to search over a set of commercial off-the-shelf (COTS) filters for an optimal collection of filters. We examine two independent tasks: general spectral sensing and target detection. For general spectral sensing, we utilize compressive sensing and find filters that generate codings which minimize the reconstruction error. For target detection, we select filters to optimize the separation between the background and a set of targets. We compare the results obtained using the selected COTS filters to the default filter set and full spectral resolution hyperspectral (HS) filter set for target detection and general spectral sensing on a previously obtained HS image.

1. Introduction

Multispectral (MS) and hyperspectral (HS) imaging sensors diverge from traditional panchromatic or RGB sensors by offering a much finer sampling (i.e., more bands) of the continuous electromagnetic (EM) spectrum; in the case of MS tens of bands and in the case of HS hundreds of bands. This finer sampling is fundamental for the two major uses of MS/HS technology, separating known targets from complex backgrounds and general spectral sensing (e.g, anomaly detection) when there is no available target information. MS and HS technology are now widely used in military and civilian realms for numerous applications ranging from anomaly detection to natural resource exploration. Due to the long development cycles, fundamental optical technology involved and great cost of these MS/HS imaging systems the tuning of the spectral filters for a particular task or set of targets would be technologically difficult and detrimental to a wide range of other tasks. The traditional imaging paradigm calls for the separation of the design of the imaging system and post-processing algorithms.

This separation between the two designs can be seen in the large number of post processing algorithms which select optimal collections of spectral bands or transform the data into a feature space representation. These two classes of algorithms, band selection and feature extraction, both seek to find an optimal representation of recorded high dimensional data for a specific task which is usually of a lower dimensionality. High dimensionality in data is problematic because of the *curse of dimensionality* [3] or the empty space phenomenon [27]. The *curse of dimensionality* describes situations where the complexity of a problem grows exponentially with the number of dimensions and the empty space phenomenon illustrates that when data is described by a few observations then the high-dimensional space becomes sparse. Both these observations lead to classifiers overfitting the data which reduces the generality. Band selection algorithms perform this dimensionality reduction by finding an optimal subset of the recorded bands or by grouping together sets of adjacent bands. Popular band selection techniques include Mutual Information-based selection [14], Kullback-Leibler divergence-based selection [20], Bhattacharya distance-based grouping [10], classification-based grouping [15], and Sequential Forward Search (SFS) [26]. Feature selection algorithms perform this dimensionality reduction by mapping the original recorded spectrums into a feature space; this mapping need not be linear. Popular techniques here include Principle Component Analysis (PCA) [25], Laplacian Eigenmaps (LE) [4], Maximum Noise Fraction [13], and Decision Boundary Feature Extraction (DBFE) [16].

Recently, a novel MS plenoptic camera from Surface Optics Corp., which we will discuss in Section 2, was developed which allows for the rapid exchange of optical filters via an array on the lens. In this new imaging system a filter set can be tuned for a specific mission and thus the data col-

lected will be better able to address the mission task. Also, by tuning the filters, only the relevant features of the scene are captured eliminating the need for post-processing algorithms (e.g., band selection or feature extraction). When the particular mission or locale changes the filter set can easily be exchanged for a more relevant one without the need to any adjustments to the larger optical system.

In this paper, utilizing the plenoptic camera framework, we propose to select a subset of available COTS filters for two distinct imaging tasks. In Section 2 we will describe the novel plentoptic imaging system which serves as the physical environment for our developed numerical methods. In Section 3 we describe the COTS filter collection which we will seek to find an optimal subset of. In Section 4 we describe the optimization process for the filter selection which involves defining the fitness functions for both sensing tasks and the differential evolution (DE) optimization strategy. In Section 5 we conduct a numerical experiment on an existing HS image comparing the optimized filter set to the default plenoptic filters and the full spectral resolution HS filter set. In Section 6 we provide some conclusions.

2. Plenoptic Camera

For this research we will focus on designing filters for the novel Surface Optics Corp. VNIR plenoptic camera (SOC716) [8]. The SOC716 is a 16-band full motion video spectral imager with interchangeable filters contained in a mosaic filter array (MFA) that can be mounted on the lens. As a standard plenoptic camera, the micro lens array (MLA), located in close vicinity of the focal plane, image the objective aperture onto the focal plane array (FPA), creating what is nominally a 4x4 sampling of the aperture with square sub-apertures. Each lenslet forms a conjugate real image of the MFA on the focal plane. Therefore, each *super-pixel* samples nominally the same spatial content with 16 different aperture functions. The FPA is aligned with these replicated images so that each FPA pixel receives light from only one filter of the MFA. The MLA has a 640x540 format and is rigidly mounted to the FPA. The default filter set has sixteen independent and contiguous spectral channels made up of approximately flat-top bandpass filters with center wavelengths ranging from 450 to 950 nm.

We note that the proposed imaging paradigms which we will present in this paper are agnostic to the actual imaging system and can all be easily extended to a number of other MS imagers including: filter wheel [6], camera array [31], and filtered active illumination [9] MS imaging systems. In addition our techniques are not limited to a specific spectral region.

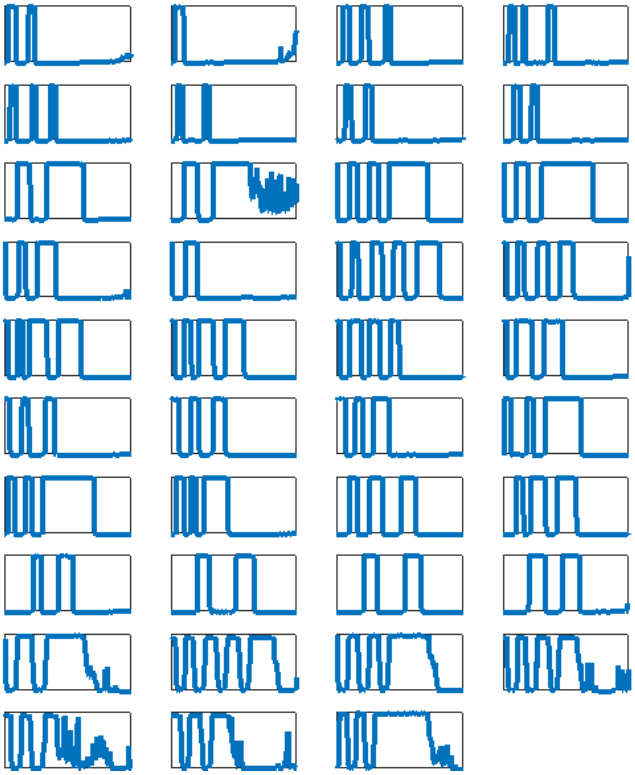

Figure 1: Collection of COTS filters. Each filter is plotted with wavelength on the x-axis ranging from 450-950 nm and transmittance on the y-axis from 0-100%.

3. COTS Filters

To alleviate the high cost of producing custom optical filters as was done in [12] we choose to utilize a collection of COTS optical filters. These COTS filters are a fraction of the cost to purchase and don't require the lengthy optical design process necessary to produce custom filters. We choose a collection of 39 Semrock multi-band fluorescence single substrate bandpass filters to comprise our collection of filters. Due to their manufacture method these filters have high transmission with steep, well-defined edges and good blocking between bandpasses. For each filter we download the predicted transmittance values over the wavelength range of the FPA on the plenoptic camera ($450-950$ nm) at a resolution of 0.2nm. The collection of filters chosen can be seen in Figure 1. Once the set of 16 filters has been selected they can be diced and arranged in the 4×4 gridAfter spectral cross-talk calibration (which can also be done after a mission) the system will acquire video rate 640×540 16-band MS data.

4. Filter Optimization

We view the multi-filter specification problem as a multi-objective design where one can vary the relative importance of various task-specific objectives. It is instructive to think of these design objectives as residing at two ends of a scale as shown in Figure 2. We have the freedom to select the spectral filters such that the sensor can better emphasize one or the other objective depending on their relative weightings, γ and $(1-\gamma)$, $\gamma \in [0,1]$, in the design performance function, P. Here we consider weightings that balance between target detection performance (P_T) and spectral resolution performance (P_R) such that:

$$P = \gamma P_T + (1-\gamma) P_R. \tag{1}$$

Our filter selection method allows for the specification of a set of spectral filters which can be tuned between emphasizing either of these task-specific objectives. The filter search involves solving an optimization problem in high-dimensional space. The formal optimization problem (shown here, without loss of generality, as a maximization) can be written as

$$\mathcal{F}^* = \max_{\mathcal{F}} g(\mathcal{F}|D), \tag{2}$$

where the goal is to find a set of filters, $\mathcal{F}^*$, from our filter collection, that maximize the objective function for a given dataset. We will utilize a DE search strategy to efficiently search the space which we describe in Section 4.1. To accomplish the general spectral resolution task we will need the theory of compressive sensing (CS) which we describe in Section 4.2. In Section 4.3 we define the fitness function for the target detection task-specific objective and in Section 4.4 we define the fitness function for the general spectral recovery task-specific objective.

4.1. Differential Evolution

In general, the optimization problem in Equation 2 is non-convex (i.e., characterized by the presence of multiple optima) and therefore we find it useful to employ a global optimization algorithm (search routine) that is capable of (but cannot guarantee) convergence to a global (rather than local) optimum. In particular, we employ DE [29], an algorithm that, by mimicking the evolutionary processes from biology, iteratively improves a set of candidate solutions, X, according to the objective (fitness) function, $g(\cdot)$. During each iteration, the algorithm proposes new candidate solutions, X', by the mutation and cross-over of members of the population (which is composed of previous candidate solutions). These new proposed candidate solutions and the current candidate solutions are evaluated by the fitness function and only the top performing candidate solutions are retained for the next iteration. This process continues until

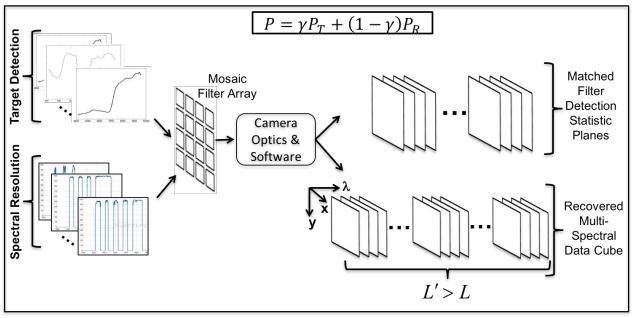

Figure 2: Conceptual diagram of the range of camera operational modes for a $L = 16$ filter (4×4 array) plenoptic camera. P is the total performance of the sensor, P_T is the detection performance, P_R quantifies spectral resolution and reconstruction performance, and γ is a weighting term. Target detection performance is exclusively favored when $\gamma = 1$. Spectral resolution is favored when $\gamma = 0$, yielding a spectral CS sensor that can recover $L' > L$ spectral bands.

a stopping condition based on the number of iterations or the output of the fitness function. The basic DE algorithm is stated in Algorithm 1. For our particular task-specific optimization problems we constrain the population to include binary vectors with a sum of 16 and length equal to the number of filters in the collection (39). We constrain the mutation and cross-over to produce solutions which live in the desired population set by performing modulus operations and randomly including or deleting filters to ensure a population members which sum to 16.

4.2. Compressive Sensing

CS is a technique used to recover a signal with fewer samples than that required by the Shannon-Nyquist Sampling theorem [28] under the restriction that the signal is sparse and the sensing matrix is incoherent relative to a selected signal model [7, 11]. For a recorded signal Y, a sensing matrix Φ, a signal model or basis (frame) Ψ; a CS recovery seeks to find $\hat{\alpha}$, which minimizes the underdetermined system:

$$\hat{\alpha} = \min_{\alpha} \left\{ \|Y - \Phi\Psi\alpha\|_2^2 + \tau\|\alpha\|_1 \right\}, \tag{3}$$

where τ is a sparseness parameter. Having found the minimizer, $\hat{\alpha}$, an estimate of the true spectrum can be recovered by applying the signal model,

$$\hat{p} = \Psi\hat{\alpha}. \tag{4}$$

For notational purposes let $CS(\Phi Y, \Psi)$ be the outcome of an algorithm which returns an estimate of the underlying true signal. The problem described in Equation 3 can be solved for with numerous techniques, e.g,. linear programming [5], Bergman iteration methods [22], Matching

Algorithm 1 Differential Evolution

Initialize:
Create a random population of P candidate solutions, X, which are contained in the $\mathcal{D}$-dimensional search space. Choose fitness function $g : \mathbb{R}^{\mathcal{D}} \mapsto \mathbb{R}$, cross over probability $CR \in [0, 1]$, and differential weight $W_F \in [0, 2]$.
for n = 1 to Stop **do**
  for $x \in X$ **do**
    Choose $a, b, c \in X \setminus x$ such that they are distinct
    $z = a + W_F \times (b - c)$
    Pick a random integer index $R \in [1, \mathcal{D}]$
    Choose r, a $\mathcal{D} \times 1$ vector, from $U(0, 1)$
    for $d = 1$ to $\mathcal{D}$ **do**
      $x'(d) = \begin{cases} z(d) & d = R \text{ or } r(d) \leq CR \\ x(d) & r(d) > CR \end{cases}$
    end for
    Constrain x to be a valid population member
    if $g(x') > g(x)$ **then**
      Replace x with x' in the population
    end if
  end for
end for

Pursuit [18] and Orthogonal Matching Pursuit [23]. For simplicity we will utilize linear programming with equality constraints.

4.3. Target Detection Fitness Function

Let $p \in I$ be the N-dimensional spectral pixel from a training dataset I of size N_I. Let $\mathcal{M}(p|\mathcal{F})$ be a function which applies the encoded candidate filters, $\mathcal{F}$, to a pixel p to yield the M-dimensional pixel p':

$$p' = \mathcal{M}(p|\mathcal{F}) = \mathcal{F}p. \tag{5}$$

Let us also denote a set of indicator functions $\mathcal{T} = \{\mathcal{T}_i\}_{i=1}^{N_T}$, where N_T is the number of distinct targets,

$$\mathcal{T}_i(p) = \begin{cases} 1 & p \text{ is a member of target class } i, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

with associated target spectral signatures $T_i \in \mathbb{R}^N$.

In target detection mode ($\gamma = 1$) our goal is to select a filter that increases the probability that pixels belonging to each of the target classes will be correctly assigned to their appropriate class while simultaneously limiting the number of false alarm detections associated with background pixels that are erroneously labeled target pixels. We use the Adaptive Cosine Estimator (ACE) [19] as a similarity measure to quantify how closely each transformed pixel, p', matches a given target spectrum, T_i, given a filter set $\mathcal{F}$:

$$m_i(p|T_i, \mathcal{F}) = \frac{(\Omega_{TF})^2}{\Omega_{TT}\Omega_{FF}}, \tag{7}$$

where,

$$\Omega_{TF} = (\mathcal{M}(T_i|\mathcal{F}) - \mu_{BG})^T \Sigma_{BG}^{-1} (\mathcal{M}(p|\mathcal{F}) - \mu_{BG}), \tag{8}$$
$$\Omega_{TT} = (\mathcal{M}(T_i|\mathcal{F}) - \mu_{BG})^T \Sigma_{BG}^{-1} (\mathcal{M}(T_i|\mathcal{F}) - \mu_{BG}), \tag{9}$$
$$\Omega_{FF} = (\mathcal{M}(p|\mathcal{F}) - \mu_{BG})^T \Sigma_{BG}^{-1} (\mathcal{M}(p|\mathcal{F}) - \mu_{BG}), \tag{10}$$

for a known target signature T_i, background spectral mean of all the transformed pixels, $\mu_{BG} = \frac{1}{N_I} \sum_{p \in I} \mathcal{M}(p|\mathcal{F})$, and background covariance, $\Sigma_{BG} = \frac{1}{N_I - 1} \sum_{p \in I} (\mathcal{M}(p|\mathcal{F}) - \mu_{BG})^T (\mathcal{M}(p|\mathcal{F}) - \mu_{BG})$.

Let us now measure the mean, μ_i^{target}, and standard deviation, σ_i^{target}, of the ACE statistics for each target signature:

$$\mu_i^{\text{target}} = \frac{1}{t_i} \sum_{p \in I} m_i(p|T_i, \mathcal{F}) \mathcal{T}_i(p), \tag{11}$$

$$\sigma_i^{\text{target}} = \sqrt{\frac{1}{t_i - 1} \sum_{p \in I} \left| m_i(p|T_i, \mathcal{F}) - \mu_i^{\text{target}} \right|^2} \mathcal{T}_i(p), \tag{12}$$

where $t_i = \sum_{p \in I} \mathcal{T}_i(p)$. Similar calculations can be made for $\mu_i^{\text{background}}$ and $\sigma_i^{\text{background}}$ by considering $1 - \mathcal{T}_i(p)$.

To avoid the problem of saturating target separability from the background we use a linear discriminant technique to define the target detection fitness function:

$$g_T(\mathcal{F}) = \sum_{i=1}^{N_T} \frac{\mu_i^{\text{target}} - \mu_i^{\text{background}}}{(\sigma_i^{\text{target}})^2 + (\sigma_i^{\text{background}})^2}. \tag{13}$$

Other measures for quantifying target detection quality such as mutual information [14] or the Battacharya distance [10] are possible.

4.4. General Spectral Recovery Fitness Function

Each channel of the imager in CS mode ($\gamma = 0$) provides a single CS measurement composed of weighted linear combinations of spectral sub-bands for each pixel. The goal in this case is to design filters that allow reconstruction of spectral pixels with a relative resolution finer than the B/L resolution offered by a naive filter design based on uniform, disjoint spectral sampling of a fixed spectral band of size B. This operational mode requires significant computational resources as each spectral pixel must be individually reconstructed by solving the sparsity-regularized CS reconstruction program. We improve our ability to observe a general spectrum at the expense of specific target detection performance and increased computational load.

To optimize the spectral resolution we again define $p \in I$ to be the N-dimensional spectral pixel from a training dataset I of size N_I. Using $\mathcal{M}(p|\mathcal{F})$ as defined in

equation (5) and selecting $L_R = 16$ filters to produce an L_R-dimensional feature vector, p', we can then estimate $\hat{p} \in \mathbb{R}^N$ using the CS recovery equations (3) and (4). In short, our complete forward model for spectral recovery is given by:

$$\mathcal{M}_R(p|\mathcal{F}) = CS(\mathcal{M}(p|\mathcal{F}), \Psi) \tag{14}$$

and our objective function becomes:

$$g_R(\mathcal{F}) = \frac{1}{\sum_{p \in I} \|\mathcal{M}_R(p|\mathcal{F}) - p\|_2^2}, \tag{15}$$

which quantifies how well we recover each N-dimensional pixel in our data set given L_R-dimensional samples where $L_R < N$.

We now need a signal model Ψ which will sparsely represent our spectral signal. Here we choose to use end-members learned from the Vertex Component Analysis (VCA) algorithm [21] for the signal model (Ψ); this choice is made for convenience of implementation and that VCA has shown to provide good sparse representations of HS data. Broadly speaking there are two types of signal models: fixed and learned. Fixed models include mathematical transformations such as the discrete cosine transform (DCT), discrete wavelet transform (DWT) and discrete Fourier transform (DFT) [1]. Learned signal models offer potentially sparser representations but must be tuned to a particular dataset; learned signal model creation algorithms include non-negative matrix factorization (NNMF) [24], K-singular value decomposition (K-SVD) [2], method of optimal directions (MOD) [17], generalized principle component analysis (GPCA) [30] and end-member algorithms [32]. The method presented in this paper does not require a specific signal model, only one which can sparsely express the sensor data.

4.5. Joint Task-specific Optimization

Though not done here, one could allow $\gamma \in (0, 1)$ and optimize for both problems. This would allow for the optimized detection of several targets while at the same time allowing for good general spectral sensing of other materials in the scene. Such an optimization can be accomplished by creating a joint fitness function or by designating filters to belong solely to one of the two extremal tasks. In the latter option, the filters selection would be iteratively updated, switching between the two tasks and freezing the filters which have been selected for the other task.

5. Experiment

In this section we utilize a previous collected HS data set to optimize the selection of COTS filters for the different task-specific areas. We will test both the target detection and general spectral recovery task-specific extremes and compare the COTS optimized results to the default plenoptic filter set. To search the space of filters we performed $10,000$ iterations of the DE algorithm with a population of 20, $CR = 0.5$, and $W_F = 0.5$.

5.1. Data

An AVIRIS HS cube acquired in February 2002 over the San Diego Airport was used for the experiment; a RGB representation of the scene can be seen in Figure 3. The cube was cropped to 400×400 pixels in size and spectral channels with water absorption, poor SNR, or outside of the plenoptic sensor range were removed leaving 51 bands. Three target signatures and associated target masks were selected from the scene for ROC curve analysis. The AVIRIS cube will represent the ground truth as well as what is possible with a HS sensor.

Figure 3: RGB representation of the AVIRIS cube. (Best viewed in color)

We numerically simulate how the 16-band default plentoptic filter set would sense the scene by creating flat-top, non-overlapping filters, and applying them to the data. For the COTS filters we perform a cubic interpolation to match the 51 AVIRIS wavelength centers then apply the interpolated filters.

5.2. Target Detection

Using the DE algorithm and the target detection fitness function presented in Equation 13 we found a collection of 16 COTS filters which maximized the fitness function during the DE optimization search; the fitness function evaluations can be seen in Figure 4. In Figure 5, we plot the ROC curve for each of the 3 selected targets from the AVIRIS scene. The ROC curves compare the probability of detection (p_d) to the probability of false alarm (p_{fa}). As was expected, by selecting filters from a modestly-sized set of

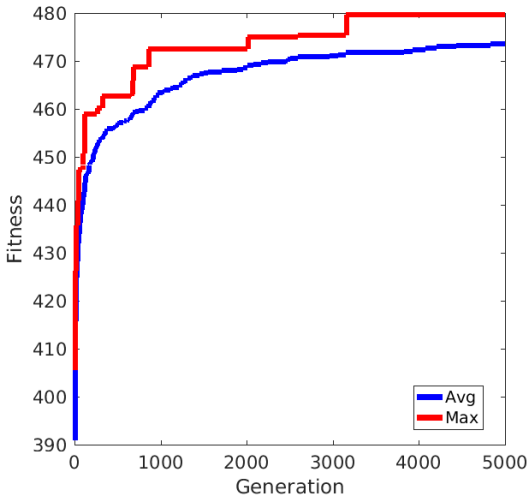

Figure 4: Evolution of fitness value for target detection task. Red is the elite solution for each generation and the blue is the population average. (Best viewed in color)

COTS filters we were able to outperform the default plenoptic filter set for all targets. Also, by tuning the filter selection we were able to outperform the baseline HS filter set, which had approximately 3.5 times better spectral resolution, fully on targets 1 and 3 and partially on target 2. The DE-based filter selection was able to perform a similar operation to the best band or feature selection algorithms listed earlier; however, since the selection was performed optically there was no need to perform post-processing on the collected data saving time and computational resources.

5.3. General Spectral Recovery

Using the DE algorithm and the general spectral recovery fitness function as presented in Equation 15 we found a collection of 16 COTS filters which which were optimal for the DE optimization search; the fitness function evaluations can be seen in Figure 6. To quantify the improvement in the spectral resolution acheived by using the DE optimized COTS filter subset over the default plentoptic filter set we will measure the distance between the recovered spectrum and the true HS spectrum. In Figure 7 we show a plot for 4 random pixels from the scene comparing the HS spectrum to each of the recovered spectrums (default and COTS) and the absolute relative error between the 16 filter spectral upsampled image $\mathcal{I}$ and the hyperspectral cube $\mathcal{H}$ for a pixel at location j:

$$\text{error}(j)_i = |\mathcal{I}(j)_i - \mathcal{H}(j)_i| \,/\mathcal{H}(j)_i, \qquad (16)$$

where the subscript indicates the spectral band $(1, 2 \ldots, b)$. We notice that the absolute relative error (shown in the dashed red and green lines with the right y-axis) for the COTS optimized filter set reconstruction is much lower then

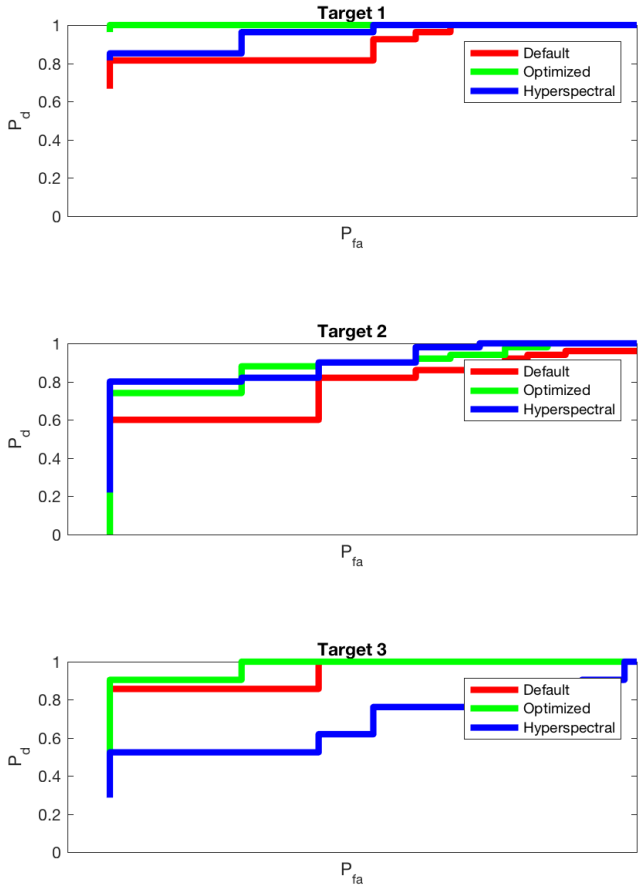

Figure 5: ROC curves for the 3 targets of interest comparing the default filter set (red), COTS optimized filter set (green), and baseline HS filter set (blue). (Best viewed in color)

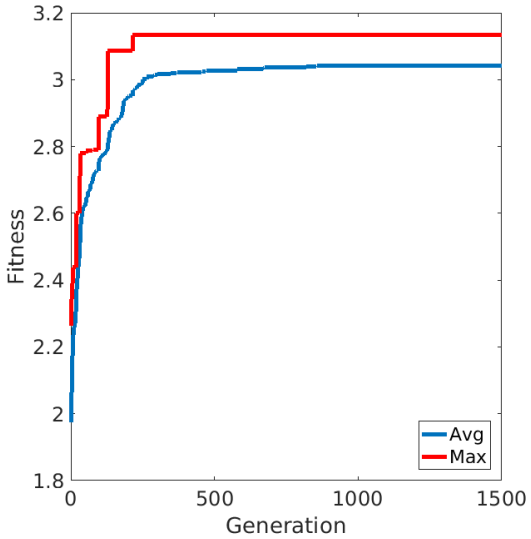

Figure 6: Evolution of fitness value for general spectral recovery task. Red is the elite solution for each generation and the blue is the population average. (Best viewed in color)

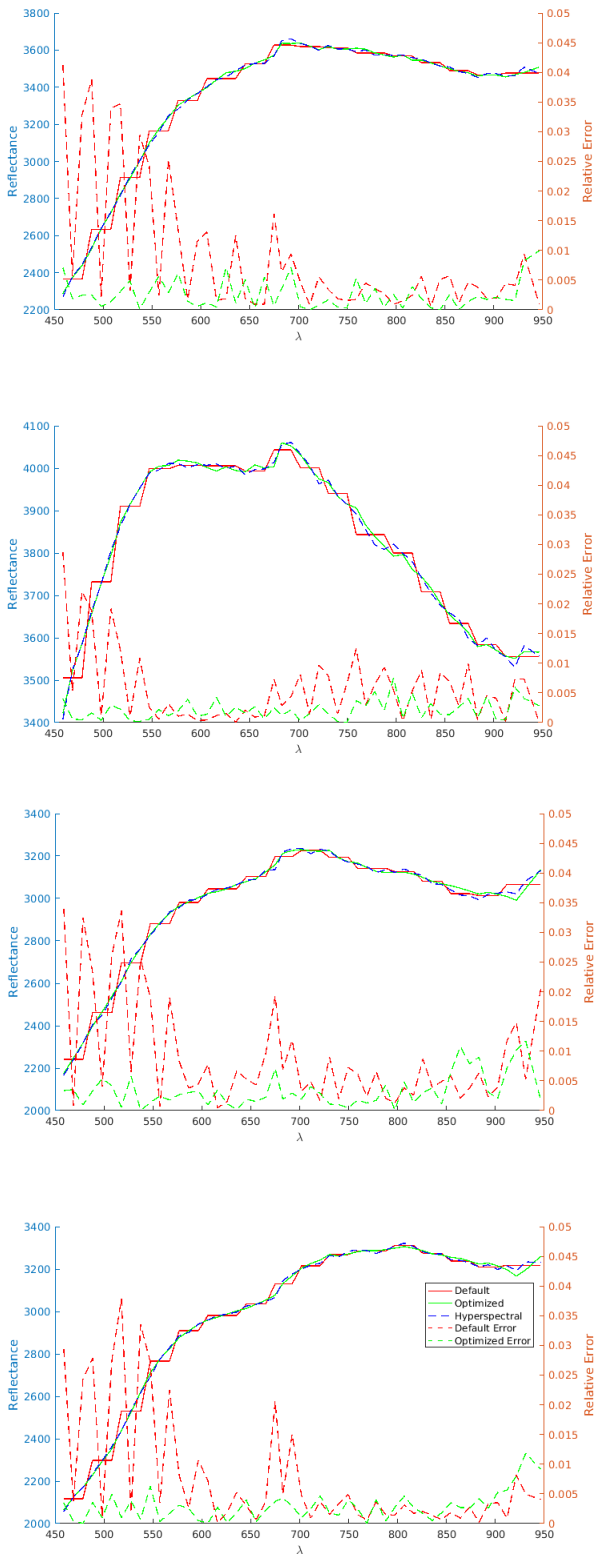

Figure 7: The HS spectral fidelity for four random pixels from the cube for the default plentoptic filter set (red), the COTS optimized filter set (green), and the HS ground truth (blue). Absolute relative error shown with right y-axis and dashed red and green lines. (Best viewed in color)

the default plenoptic filter set on the lower wavelengths and lower to equal on the higher wavelengths. If we look at the spectral coverage offered by the COTS filter collection in Figure 1 we can see there is weak support on the higher wavelengths. By adding additional COTS filters to the collection which have support in this higher wavelength spectral region we believe the results will further be improved. In Figure 8 we show the spectral sum of the absolute relative error for a pixel at location j:

$$\text{error}^S(j) = \sum_{i=1}^{b} \left|\mathcal{I}(j)_i - \mathcal{H}(j)_i\right| / \mathcal{H}(j)_i. \qquad (17)$$

We see that the optimized COTS filters with CS reconstruction produced a spectral upsampled image cube with lower pixel spectral errors than the default plentoptic filter set. This increased spectral fidelity is important for analysts who rely on spectral shapes to determine material compositions of scenes and to compare with spectral libraries.

6. Conclusions

In this paper we have established a system in which one can optimize the target detection performance or the general spectral recovery of a scene by choosing filters from a collection of COTS filters for use with a plenoptic camera. By utilizing the design philosophy described in this paper it is possible to achieve significant improvements in task-specific imaging as compared to the default plentoptic filter set composed of optical bandpass filters. We showed that the DE search strategy allows for quick optimization results that outperform the default set. We are currently planning a field test where we will deploy the COTS collection against the default filter set with various targets and background materials. We also seek to refine the CS recovery algorithms to utilize the spatial information in the scene and explore other signal models for improved reconstruction results.

Acknowledgements

The authors would like to thank the anonymous reviewers for their helpful comments and the Office of Naval Research (ONR) for funding this work.

References

[1] M. M. Abo-Zahhad, A. I. Hussein, and A. M. Mohamed. Compressive sensing algorithms for signal processing applications: A survey. *International Journal of Communications, Network and System Sciences*, 8(06):197, 2015.

[2] M. Aharon, M. Elad, and A. Bruckstein. rmK-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

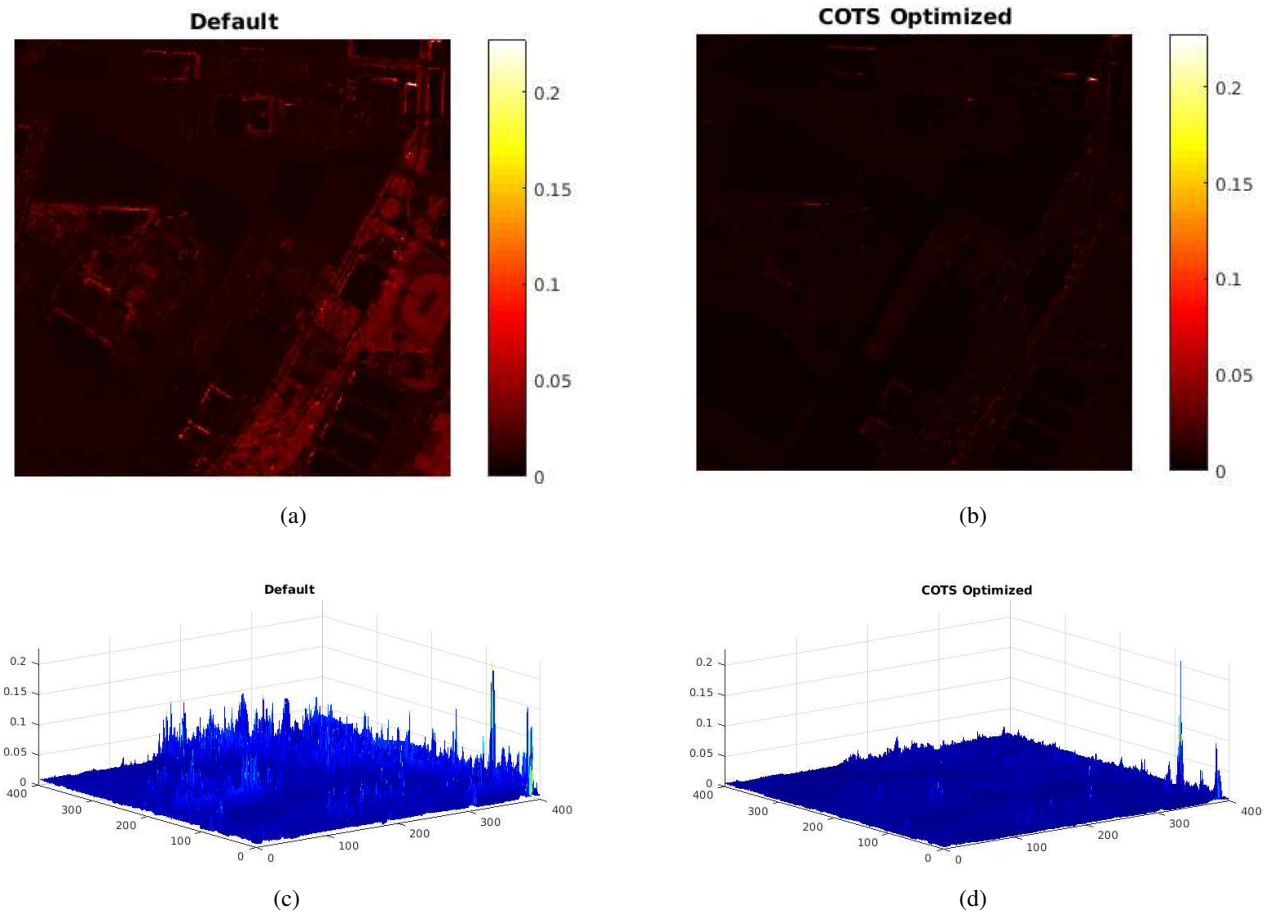

Figure 8: Absolute relative error between the full spectral resolution HS scene and the default plenoptic filter set reconstruction ((a) and (c)). Absolute relative error between the full spectral resolution HS scene and COTS optimized CS reconstruction (b) and (d). Top and bottom show same error data (colormap on top and z-axis on bottom). Note that the scales are the same. (Best viewed in color)

[3] R. E. Bellman. *Adaptive Control Processes: A Guided Tour*, volume 4. Princeton University Press, 1961.

[4] J. J. Benedetto, W. Czaja, J. Dobrosotskaya, T. Doster, K. Duke, and D. Gillis. Integration of heterogeneous data for classification in hyperspectral satellite imagery. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, Proc. SPIE,*, volume 8390, pages 8390–78. International Society for Optics and Photonics, 2012.

[5] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[6] J. Brauers, N. Schulte, and T. Aach. Multispectral filter-wheel cameras: Geometric distortion model and compensation algorithms. *IEEE Transactions on Image Processing*, 17(12):2368–2380, 2008.

[7] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[8] D. B. Cavanaugh, J. M. Lorenz, N. Unwin, M. Dombrowski, and P. Willson. VNIR hypersensor camera system. In *SPIE Optical Engineering+ Applications*, pages 74570O–74570O. International Society for Optics and Photonics, 2009.

[9] C. Chi, H. Yoo, and M. Ben-Ezra. Multi-spectral imaging by optimized wide band illumination. *International Journal of Computer Vision*, 86(2-3):140, 2010.

[10] S. De Backer, P. Kempeneers, W. Debruyn, and P. Scheunders. A band selection technique for spectral classification. *IEEE Geoscience and Remote Sensing Letters*, 2(3):319–323, 2005.

[11] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[12] T. Doster, C. C. Olson, E. Fleet, M. Yetzbacher, A. Kanaev, P. Lebow, and R. Leathers. Designing manufacturable filters for a 16-band plenoptic camera using differential evolution. In *Algorithms and Technologies for Multispectral, Hy-*

perspectral, and Ultraspectral Imagery XXIII, Proc. SPIE,. International Society for Optics and Photonics, 2017.

[13] A. A. Green, M. Berman, P. Switzer, and M. D. Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):65–74, 1988.

[14] B. Guo, S. R. Gunn, R. Damper, and J. Nelson. Band selection for hyperspectral image classification using mutual information. *IEEE Geoscience and Remote Sensing Letters*, 3(4):522–526, 2006.

[15] S. Kumar, J. Ghosh, and M. M. Crawford. Best-bases feature extraction algorithms for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1368–1379, 2001.

[16] C. Lee and D. A. Landgrebe. Feature extraction based on decision boundaries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):388–400, 1993.

[17] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[18] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[19] D. Manolakis and G. Shaw. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Processing Magazine*, pages 29–43, January 2002.

[20] A. MartÍnez-UsÓ, F. Pla, J. M. Sotoca, and P. García-Sevilla. Clustering-based hyperspectral band selection using information measures. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):4158–4171, 2007.

[21] J. M. P. Nascimento and J. M. B. Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):898–910, 2005.

[22] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[23] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 40–44. IEEE, 1993.

[24] V. P. Pauca, J. Piper, and R. J. Plemmons. Nonnegative matrix factorization for spectral data analysis. *Linear algebra and its applications*, 416(1):29–47, 2006.

[25] K. Pearson. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[26] P. Pudil, J. Novovičová, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11):1119–1125, 1994.

[27] D. W. Scott and J. R. Thompson. Probability density estimation in higher dimensions. In *Computer Science and Statistics: Proceedings of the Fifteenth Symposium on the Interface*, volume 528, pages 173–179, 1983.

[28] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[29] R. Storn and K. Price. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

[30] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (GPCA). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1945–1959, 2005.

[31] S. Wug Oh, M. S. Brown, M. Pollefeys, and S. Joo Kim. Do it yourself hyperspectral imaging with everyday digital cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2461–2469, 2016.

[32] A. Zare and K. Ho. Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. *IEEE Signal Processing Magazine*, 31(1):95–104, 2014.