Dictionary-based Optical Filter Selection for Multi-Application Spectral Signature Classification

Jun Ke, Kathrin Berkner, M. Dirk Robinson and David G. Stork

Ricoh Innovations, 2882 Sand Hill Rd, Suite 115
Menlo Park, CA 94025-7054
jke@ece.arizona.edu,{berkner,dirkr,stork}@rii.ricoh.com

Abstract: We describe a method for selecting filter sets for simultaneously optimizing the classification rates in two separate spectral signature classification problems. Our system's performance is comparable to traditional hyper-or multispectral classifiers, but uses fewer filters.

OCIS codes: (110.1758) Computational imaging, (110.4234) Multispectral and hyperspectral imaging, (110.3055) Information theoretical analysis

1. Introduction

Hyperspectral imaging (HSI) systems have been widely used in remote sensing since the 1970s. Because data acquisition is often expensive through satellite-mounted systems, most HSI systems were general purpose, that is, designed to collect data for a variety of analysis applications. As a result, the amount of collected data is often much larger than that required for a single application. To reduce the amount of unneeded data, various approaches have been proposed for transmitting data in selected bands only for further processing after the data acquisition. Algorithms for automated band selection having been proposed based on minimizing reconstruction error using PCA or ICA [1], or minimizing classification error using Fisher Discriminant Analysis or Mutual Information [2]. Another approach is to acquire data only at spectral bands informative to the task at hand. Spectral filters used in those systems are typically application dependent and need to be redesigned for each new application.

We propose a new filter set selection method for spectral signature classification for reducing the probability of detection error of multiple applications jointly. Two hypotheses are assumed for each detection application, such as the spectral measurement originating from dry or fresh corn leaves and for dry or fresh rice leaves. We investigate the classification performance with the designed filters in an experiment using the LOPEX03 database [3].

2. Multi-application spectral signature classification problem

The measurements in a system with M spectral filters obey

$$y_m = \int x(\lambda) r_m(\lambda) s(\lambda) d\lambda + \nu \quad m = 1, 2, \cdots, M, \tag{1}$$

where y_m is the m^{th} measurement, $x(\lambda)$ is the source spectral radiance of the surface, $r_m(\lambda)$ is the m^{th} filter transmittance function, $s(\lambda)$ is the spectral response function of the detector, and ν is the noise at the detector [4]. Because there is no continuous spectrum $x(\lambda)$ available for numerical simulation, we use narrow-band hyperspectral samples $\overline{x}(\lambda_n), \overline{r}(\lambda_n), \overline{s}(\lambda_n)$ collected at wavelength $\lambda_n, n = 1, ...N$ instead. Then the measurement y_m can be represented as $y_m = \sum_{n=1}^{N} \overline{x}(\lambda_n) \overline{f}_m(\lambda_n) + \nu$, where $\overline{f}_m(\lambda_n) = \overline{r}_m(\lambda_n)\overline{s}(\lambda_n)$. We express these system measurements as

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{n}, \tag{2}$$

where $\mathbf{y} \in \mathbb{R}^{M\times 1}$ is the measurement vector, $\mathbf{F}$ the filter response function matrix with $\overline{f}_m(\lambda_n)$ as its $(m, n)^{th}$ element, and $\mathbf{x} \in \mathbb{R}^{N\times 1}$ and $\mathbf{n} \in \mathbb{R}^{M\times 1}$ are the source spectral radiance vector and the detector noise vector, respectively.

We study a detection problem for two applications A and B with measurements $\mathbf{y_A}$ and $\mathbf{y_B}$. Each detection application $i = A, B$ has two hypotheses, $H_{0,j}$ and $H_{1,j}$ for spectral signatures $\mathbf{x}_{j,k}, j = 0, 1, i = A, B$, that is:

$$H_{j,A} : \mathbf{y}_A = \mathbf{F}\mathbf{x}_{j,A} + \mathbf{n}_A \qquad H_{j,B} : \mathbf{y}_B = \mathbf{F}\mathbf{x}_{j,B} + \mathbf{n}_B.$$

The probability of error for misclassification in a single application $i \in \{A, B\}$ is

$$P_{e_i} = P[H_{0,i}]P[\mathbf{y} \in H_{1,i}|H_{0,i}] + P[H_{1,i}]P[\mathbf{y} \in H_{0,i}|H_{1,i}], \tag{3}$$

where $P[H_{j,i}]$ is the probability that the j^{th} hypothesis is true for application i, and we assume equal probabilities $P[H_{0,i}] = P[H_{1,i}] = 1/2$. The overall classification error we seek to minimize is

$$P_e = \max_{i \in \{A,B\}} \{P_{e_i}\}, \tag{4}$$

that is, the classification error on the "worst" application.

We restrict the filters to be a subset of a dictionary of L commercially available filters represented in a $(L \times N)$ matrix $\mathbf{D}$ with each row representing one filter transmittance $\mathbf{d}_\ell = [d_\ell(1)\ d_\ell(2)\ \cdots\ d_\ell(N)]^t$, where $\ell = 1, \ldots, L$. Filter selection is represented by a selection matrix $\mathbf{W}$ of size $(M \times L)$ with entries $w_{m\ell} \in \{0,1\}$ and restrictions $\sum_\ell w_{m\ell} = 1$ and $\sum_m w_{m\ell} = 1$. Multiplying the dictionary $\mathbf{D}$ by $\mathbf{W}$ results in M different filters being selected from the dictionary. With these notations we rewrite the hypothesis as

$$H_{j,A} : \mathbf{y}_A = \mathbf{WD}\mathbf{x}_{j,A} + \mathbf{n}_A \qquad H_{j,B} : \mathbf{y}_B = \mathbf{WD}\mathbf{x}_{j,B} + \mathbf{n}_B \quad \text{for} \quad j \in \{0,1\}.$$

The explicit expression of P_{e_i} depends on signal and the noise models. We model the object signals as multivariate Gaussians, i.e., $\mathbf{x}_{j,A} \sim N(\boldsymbol{\mu}_{j,A}, \mathbf{C}_{j,A})$ and $\mathbf{x}_{j,B} \sim N(\boldsymbol{\mu}_{j,B}, \mathbf{C}_{j,B})$ for $j = 0, 1$. The detector noise is assumed to be white Gaussian noise with the same variance σ for both applications, i.e., $\mathbf{n}_A, \mathbf{n}_B \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. The closed-form solutions for a single application error probability P_{e_i}, using $Q(y) = \frac{1}{\sqrt{2\pi}} \int_y^\infty e^{-\frac{1}{2}y^2} dy$ is [5] :

$$P_{e_i}(\mathbf{W}) = Q\left(\frac{1}{2}\sqrt{(\boldsymbol{\mu}_{x_{1,i}} - \boldsymbol{\mu}_{x_{0,i}})^t (\mathbf{WD})^t (\mathbf{WDC}_{x,i}(\mathbf{WD})^t + \sigma^2\mathbf{I})^{-1} \mathbf{WD}(\boldsymbol{\mu}_{x_{1,i}} - \boldsymbol{\mu}_{x_{0,i}})}\right). \tag{5}$$

Utilizing the fact that the function $Q(x)$ is *monotonically decreasing*, the filter set selection task evaluated using the joint probability $P_e = \max_{i \in \{A,B\}}\{P_{e_i}(\mathbf{W})\}$ from Eq. 4 can be formulated as the following optimization problem:

$$\text{maximize} \quad \min_{i \in A,B} g_i(\mathbf{WD}) \quad \text{subject to} \quad w_{m,l} \in \{0,1\}, \ \sum_{l=1}^{L} w_{m,l} = 1 \tag{6}$$

with optimization variable $\mathbf{W}$ and $g_i(\mathbf{WD}) = (\boldsymbol{\mu}_{x_{1,i}} - \boldsymbol{\mu}_{x_{0,i}})^t (\mathbf{WD})^t (\mathbf{WDC}_{x,i}(\mathbf{WD})^t + \sigma^2\mathbf{I})^{-1} \mathbf{WD}(\boldsymbol{\mu}_{x_{1,i}} - \boldsymbol{\mu}_{x_{0,i}})$. Several search strategies can be applied to solve this combinatorial optimization problem. In the experiments described below we use a simple greedy search algorithm for the proof of concept for filter selection. More complex search methods could also be adapted to be applied to our framework [6, 7].

3. Numerical experiments

We use surface spectral reflectance signatures of plants 400nm and 1000nm from the LOPEX93 database [3]. The two hypotheses in each classification application are the spectral signature originating from dry or fresh corn leaves (A) and dry or fresh rice leaves (B) (middle of Fig. 1).

We model the filter dictionary from interference filter specifications given at the CVI Melles Griot website [8] using the parameters CWL (center wavelength), bandwidth FWHM (full width at half maximum), band shape specified with the number of cavities (layers of coating), and the incident light propagation angle. The dictionary used in the numerical experiments contains filter responses simulated by the model using four cavities for the band shape and 0 degree for the incident angle. In total, the dictionary consists of 149 filters: 11 long pass filters, 12 short pass filters, and 126 bandpass filters. Bandwidths of the bandpass filters are in the range of 3nm to 80nm (left in Fig. 1). We use the spectral response function of a Hamamatsu NMOS linear image sensor for the sensor spectral transmittance in Eq. 1.

For a given noise level, the quantity g_i (Eq. 6) is maximized using the following simple greedy strategy: a) Set initial number of selected filters to $k = 0$, b) find row vectors $\mathbf{w}_j, \mathbf{w}_i$ of $\mathbf{W}$ such that $g_B(\mathbf{w}_j\mathbf{D}) = \max_m\{g_B(\mathbf{w}_m\mathbf{D})\}$ and $g_A(\mathbf{w}_i\mathbf{D}) = \max_m\{g_A(\mathbf{w}_m\mathbf{D})\}$ c) Choose $\mathbf{w}^* = \mathbf{w}_i$ if $|g_B(\mathbf{w}_j\mathbf{D}) - g_A(\mathbf{w}_j\mathbf{D})| > |g_B(\mathbf{w}_i\mathbf{D}) - g_A(\mathbf{w}_i\mathbf{D})|$, otherwise choose $\mathbf{w}^* = \mathbf{w}_j$. d) Eliminate the chosen filter $\mathbf{w}^*\mathbf{D}$ from the dictionary; increase k to $k+1$. e) If $k < K$, go to step b), otherwise stop.

For each candidate set of K filters, we compute the error probabilities using Eq. 5 as shown in the right of Fig. 1. The error probability for a single application only is shown (light blue and magenta) for comparison. To detect fresh or dry corn leaves, the filters selected for both applications jointly produce the same classification error as the filters selected only for the corn application (cyan and blue lines overlap) since the harder-to-classify corn application dominates the error P_e. For the easier-to-classify rice application, the decrease of performance using the joint-application filter selection compared to the single-application scenario is small. For example, selecting 10 filters for detecting both rice and corn applications, given by selection matrix $\mathbf{W}^*$, results in an worst-case error of $P_e(\mathbf{W}^*) = 10^{-0.65} \sim 0.2$ in a 20dB system, with the corn application being the harder-to-classify application. Applying the selected filters to the easier-to-classify rice application results in an error probability of $P_{e,rice}(\mathbf{W}^*) = 10^{-1.35} \sim 0.04$, increasing the error probability achieved using filters designed specifically for the rice application from $P_{e,rice}(\mathbf{W}_{rice})10^{-1.5} \sim 0.03$ only by 0.01. The two filter sets designed by optimizing the detection probabilities for each application separately contain each 10 filters that do not all overlap. That means K^* different filters with $10 < K^* \leq 20$ are selected when being designed for each application separately compared to only $K = 10$ filters in our proposed filter selection framework.

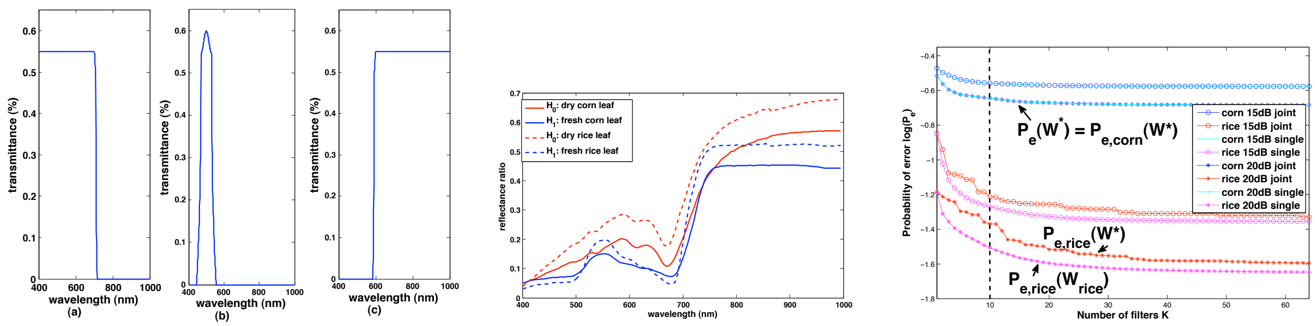

Fig. 1. Left: Modeled optical filters with 4 cavities: a) short pass filter with cut-off frequency at 650nm; b) bandpass filter with CWL = 500nm and FWHM = 80nm; c) long pass filter with cut-off frequency at 600nm. Middle: Typical spectral signatures for fresh and dry corn and rice leaves. Right: Probability of the classification error vs. number of filters in signal model in third case with one application and two applications (right).

4. Conclusions

We proposed a filter-design framework for a multi-spectral imaging system with the goal to select a set of filters that performs well for two separate two-class classification applications. Assuming a probabilistic signal model in white noise, detection probabilities were derived. Then an optimization framework was derived that selects filters to minimize the maximum of the detection probabilities for the jointly considered applications. The constraints imposed on the filters to be selected from a specific dictionary of available filters turned the filter selection process into a complex search problem. The theoretical filter selection framework was tested with experiments using spectral signatures for two vegetation classes, corn and rice, using data from the LOPEX data base. The initial experimental results indicated that the joint design of filters for a detection problem show promising performance by demonstrating a similar performance for the joint two applications case to that for each individual application separately, but utilizing less filters.

References

1. H.Wang, E. Angelopoulou, "Sensor Band Selection for Multispectral Imaging via Average Normalized Information," *Journal of Real Time Image Processing*, 2006.
2. R. Ramanath, W.E. Snyder, H. Qi, "Mosaic Multispectral Focal Plane Array Cameras," *Proceedings of the SPIE Defense and Security Symposium*, vol. 5406, pp. 701–712, 2004.
3. European Commission's Joint Research Centre, "Leaf Optical Properties Experiment 93 (LOPEX93)," Website, 1993, http://ies.jrc.ec.europa.eu/index.php?page=data-portals.
4. J. P. Kerekes, J. E. Baum, "Full-Spectrum Spectral Imaging System Analytical Model," *IEEE Transactions on Geoscience and remote sensing*, vol. 43, no. 3, pp. 571–580, 2005.
5. Steven M. Kay, *Fundamentals of Statistical Signal Processing: Volume II Detection Theory*, Prentice Hall, New Jersey, 1998.
6. S. De Backer et al., "Band Selection for Hyperspectral Remote Sensing," *IEEE Geoscience and Remote Sensing Letters*, vol. 2, pp. 319– 323, 2005.
7. S.B. Serpico, G. Moser, "Extraction of Spectral Channels From Hyperspectral Images for Classification Purposes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, pp. 484–495, 2007.
8. CVI Melles Griot, "Interference filters," Website, 2009, www.mellesgriot.com/pdf/0013.25-13.29.pdf.