

Influence of band selection and target estimation error on the performance of the matched filter in hyperspectral imaging

Jean Minet, Jean Taboury, François Goudail, Michel Pealat, Nicolas Roux, Jacques Lonnoy, Yann Ferrec

▶ **To cite this version:**

Jean Minet, Jean Taboury, François Goudail, Michel Pealat, Nicolas Roux, et al.. Influence of band selection and target estimation error on the performance of the matched filter in hyperspectral imaging. Applied optics, 2011, 50 (22), pp.4276–4285. 10.1364/AO.50.004276 . hal-00700321

HAL Id: hal-00700321
https://hal.science/hal-00700321

Submitted on 25 May 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Influence of band selection and target estimation error on the performance of the matched filter in hyperspectral imaging

Jean Minet,[1,*] Jean Taboury,[1] François Goudail,[1] Michel Péalat,[2] Nicolas Roux,[2] Jacques Lonnoy,[2] and Yann Ferrec[3]

[1]Laboratoire Charles Fabry, Institut d'Optique, University Paris-Sud, CNRS, 2 Avenue Augustin Fresnel, F-91127 Palaiseau Cedex, France

[2]Sagem, Groupe Safran, 23 Avenue Carnot, 91344 Massy, France

[3]ONERA/DOTA, Chemin de la Hunière, 91761 Palaiseau Cedex, France

*Corresponding author: jean.minet@institutoptique.fr

Received 2 March 2011; revised 20 May 2011; accepted 24 May 2011; posted 26 May 2011 (Doc. ID 142699); published 21 July 2011

The matched filter is a widely used detector in hyperspectral detection applications because of its simplicity and its efficiency in practical situations. We propose to estimate its performance with respect to the number of spectral bands. These spectral bands are selected thanks to a genetic algorithm in order to optimize the contrast between the target and the background in the detection plane. Our band selection method can be used to optimize not only the position but also the linewidth of the spectral bands. The optimized contrast always increases with the number of selected bands. However, in practical situations, the target spectral signature has to be estimated from the image. We show that in the presence of estimation error, the maximum number of bands may not always be the best choice in terms of detection performance. © 2011 Optical Society of America

OCIS codes: 100.2000, 110.4234.

1. Introduction

Spectral imaging consists in forming several images of the same scene with different spectral content. Hyperspectral imagers can combine hundreds of spectral bands with millions of spectral pixels. They have the potential to detect small (even subpixel) spectral targets that differ only very slightly from the background [1]. However, hyperspectral detection applications suffer from two major drawbacks. First, the computation time of detection algorithms is an important issue, especially when dealing with high spectral resolution. Second, imperfect information about the target spectral signature may significantly alter the detection performance. In this paper, we provide a band selection method that reduces the spectral dimension of the acquired hypercube in order to optimize the detection performance. We also discuss the effect of estimation error in the target signature on the detection performance. We show that in the presence of estimation error, the optimal detection performance may be obtained on a reduced number of well-chosen spectral bands.

In order to quantify the detection performance, we have to select a detection algorithm. During the past two decades, a large number of hyperspectral detection algorithms have been developed [2–5]. New sophisticated algorithms are derived every year from complex models of spectral variability. According to Manolakis *et al.* [6], the small performance gains attained by those algorithms may be irrelevant in practical applications, whereas the use of simple detectors, such as the matched filter, may provide acceptable performance in real-world hyperspectral

0003-6935/11/224276-10$15.00/0
© 2011 Optical Society of America

imaging problems. In this paper, we have selected the matched filter to detect targets in real hyperspectral images because of its simplicity and its efficiency in practical situations. The detection performance of the matched filter is evaluated in terms of the contrast between the target and the background in the detection plane.

However, even the matched filter can be computationally intensive, especially when dealing with a high spectral resolution. To perform a real-time detection, it is then interesting to reduce the spectral dimension of the image before running the detector. This method is commonly referred as band selection [7–9]. Band selection is an optimization problem that aims at finding the best set of K spectral bands among M available spectral bands with respect to a defined criterion. In this paper, we use a genetic algorithm to optimize the position and the linewidth of the spectral bands yielding the best detection performance [10].

Band selection is not only motivated by issues such as computation time. High dimensional spaces (high number of bands) are mostly empty, and they can then be projected to a lower dimensional subspace without losing significant information in terms of discriminability between the target and the background [11]. In many applications, one has inaccurate information about the spectral signature of the target. The target signature estimation error then degrades the detection performance. When detector parameters have to be estimated from a finite number of training samples, detection performance can even decrease as the number of bands increases [12]. This is known in a general context as the Hughes phenomenon [13]. We show that under simple assumptions about the statistical distribution of the estimation error, the mean performance loss can be approximated as a simple decreasing function of the number of bands. Band selection results on real hyperspectral images show that the optimal detection performance may be obtained on a reduced number of well-chosen spectral bands.

In Section 2 we review the matched filter and define our criterion of detection performance. Section 3 describes our band selection method and provides results based on real hyperspectral images. In Section 4 we discuss the effect of estimation error in the target signature on the detection performance.

2. Matched Filter

A. Derivation of the Matched Filter by Optimization of the Neyman–Pearson Criterion

The matched filter is a widely used detector in hyperspectral detection applications because of its simplicity and its efficiency in practical situations. This detector can be proven to be optimal in the Neyman–Pearson sense under certain statistical assumptions about the target and nontarget classes (γ_1 and γ_0, respectively). Given an observed spectrum $\mathbf{x} = (x_1, \ldots, x_K)^T$, where K is the number of bands, we want to assign it to one of the two classes γ_0 and γ_1. We consider the detection problem specified by the following hypotheses:

$$\mathbf{x} \sim \begin{cases} \mathcal{N}(\mathbf{m}_0, \Gamma) & \text{under } \gamma_0 \text{ (target absent)} \\ \mathcal{N}(\mathbf{m}_1, \Gamma) & \text{under } \gamma_1 \text{ (target present)} \end{cases}, \tag{1}$$

where the target and background (nontarget) classes follow multivariate normal distribution with different mean vectors $\mathbf{m}_1$ and $\mathbf{m}_0$ and identical covariance matrix $\mathbf{\Gamma}$. The computation of the likelihood ratio leads to the detector

$$y = D_{\mathrm{mf}}(\mathbf{x}) = (\mathbf{m}_1 - \mathbf{m}_0)^T \mathbf{\Gamma}^{-1} (\mathbf{x} - \mathbf{m}_0), \tag{2}$$

where the superscript T refers to the transpose operation. This detector, which is called the matched filter, is optimal in the Neyman–Pearson sense, i.e., it maximizes the probability of detection for a given probability of false alarm [14]. Under the statistical assumptions of Eq. (1), the output of the matched filter is normally distributed because it is a linear combination of normal random variables. One can show that

$$y = D_{\mathrm{mf}}(\mathbf{x}) \sim \begin{cases} \mathcal{N}(0, \Delta^2) & \text{under } \gamma_0 \\ \mathcal{N}(\Delta^2, \Delta^2) & \text{under } \gamma_1 \end{cases}, \tag{3}$$

where $\Delta^2 = (\mathbf{m}_1 - \mathbf{m}_0)^T \mathbf{\Gamma}^{-1} (\mathbf{m}_1 - \mathbf{m}_0)$ is the square Mahalanobis distance between the target and the background class γ_0. For a given output $D_{\mathrm{mf}}(\mathbf{x})$ of the matched filter, the decision is made by comparing y to a certain threshold η. The spectrum $\mathbf{x}$ will be assigned to the target class γ_1 if $D_{\mathrm{mf}}(\mathbf{x}) > \eta$. The probability of false alarm P_{FA} and the probability of detection P_D can then be obtained by the relations

$$\begin{cases} P_{FA} = \int_\eta^{+\infty} p(y|\gamma_0) \mathrm{d}y, \\ P_D = \int_\eta^{+\infty} p(y|\gamma_1) \mathrm{d}y, \end{cases} \tag{4}$$

where $y = D_{\mathrm{mf}}(\mathbf{x})$ and $p(y|\gamma_i)$ is the probability density function of y under γ_i. The trade-off between the probability of detection and the probability of false alarm is fixed by the threshold η. As the conditional probability densities $p(y|\gamma_0)$ and $p(y|\gamma_1)$ depends only on the parameter Δ^2 [Eq. (3)], the detection performance is entirely determined by the square Mahalanobis distance Δ^2 between the target and the background class.

B. Derivation of the Matched Filter by Contrast Optimization

In practical situations, the normal assumptions [Eq. (1)] about the statistical distributions of the target and background classes may not be verified. When the distributions of the target and the background classes deviate from the model of Eq. (1), the optimum detector in the Neyman–Pearson sense is not linear. However, we can show that the matched filter remains the optimum linear detector according

to a criterion of contrast between the target and the background.

We do not assume here the normality of the target and background statistical distributions. We only assume that the mean of the target, as well as the mean of the background and its covariance matrix, are defined as follows:

$$\begin{cases} \mathbf{m}_1 & = \mathrm{E}\{\mathbf{x}|\gamma_1\}, \\ \mathbf{m}_0 & = \mathrm{E}\{\mathbf{x}|\gamma_0\}, \\ \mathbf{\Gamma} & = \mathrm{E}\{(\mathbf{x}-\mathbf{m}_0)(\mathbf{x}-\mathbf{m}_0)^T|\gamma_0\}. \end{cases} \tag{5}$$

We consider linear detectors of the form

$$y_{\mathbf{w}} = D_{\mathbf{w}}(\mathbf{x}) = \mathbf{w}^T(\mathbf{x}-\mathbf{m}_0), \tag{6}$$

where $\mathbf{w}$ is a K dimension vector. As the spectrum $\mathbf{x}$ is not normally distributed, the output $y_{\mathbf{w}}$ of the detector is not normally distributed. However, we can easily obtain the following statistics:

$$\begin{cases} \mathrm{E}\{y_{\mathbf{w}}|\gamma_1\} & = \mathbf{w}^T(\mathbf{m}_1-\mathbf{m}_0), \\ \mathrm{E}\{y_{\mathbf{w}}|\gamma_0\} & = 0, \\ \mathrm{var}\{y_{\mathbf{w}}|\gamma_0\} & = \mathrm{E}\{\mathbf{w}^T(\mathbf{x}-\mathbf{m}_0)(\mathbf{x}-\mathbf{m}_0)^T\mathbf{w}|\gamma_0\}, \\ & = \mathbf{w}^T\mathbf{\Gamma}\mathbf{w}. \end{cases} \tag{7}$$

We want to find the vector $\mathbf{w}$ that maximizes the contrast $C_{\mathbf{w}}$ of $y_{\mathbf{w}}$ between the target and the background defined as the squared distance between the mean of the target and the mean of the background in units of background variance:

$$C_{\mathbf{w}} = \frac{[\mathrm{E}\{y_{\mathbf{w}}|\gamma_1\}-\mathrm{E}\{y_{\mathbf{w}}|\gamma_0\}]^2}{\mathrm{var}\{y_{\mathbf{w}}|\gamma_0\}} = \frac{[\mathbf{w}^T(\mathbf{m}_1-\mathbf{m}_0)]^2}{\mathbf{w}^T\mathbf{\Gamma}\mathbf{w}}. \tag{8}$$

As the contrast $C_{\mathbf{w}}$ does only depend on the direction of $\mathbf{w}$, the optimization problem can be written

$$\mathbf{w}_{\mathrm{mf}} = \arg \max_{\mathbf{w}} \mathbf{w}^T\Delta\mathbf{m} \text{ subject to } \mathbf{w}^T\mathbf{\Gamma}\mathbf{w} = 1. \tag{9}$$

This optimization problem can be solved using a Lagrangian formulation [15]. One obtains that $\mathbf{w}_{\mathrm{mf}} = \mathbf{\Gamma}^{-1}(\mathbf{m}_1-\mathbf{m}_0)$ is a solution of the optimization problem. The detector becomes

$$y = D_{\mathbf{w}_{\mathrm{mf}}}(\mathbf{x}) = (\mathbf{m}_1-\mathbf{m}_0)^T\Gamma^{-1}(\mathbf{x}-\mathbf{m}_0). \tag{10}$$

The matched filter is thus the optimal linear detector with respect to the contrast criterion C. The contrast C of the output of the matched filter is equal to the squared Mahalanobis distance between the target and the background:

$$\begin{aligned} C &= C_{\mathbf{w}_{\mathrm{mf}}} \\ &= \frac{[(\mathbf{m}_1-\mathbf{m}_0)^T\mathbf{\Gamma}^{-1}(\mathbf{m}_1-\mathbf{m}_0)]^2}{(\mathbf{m}_1-\mathbf{m}_0)^T\mathbf{\Gamma}^{-1}\mathbf{\Gamma}\mathbf{\Gamma}^{-1}(\mathbf{m}_1-\mathbf{m}_0)} \\ &= (\mathbf{m}_1-\mathbf{m}_0)^T\mathbf{\Gamma}^{-1}(\mathbf{m}_1-\mathbf{m}_0) = \Delta^2. \end{aligned} \tag{11}$$

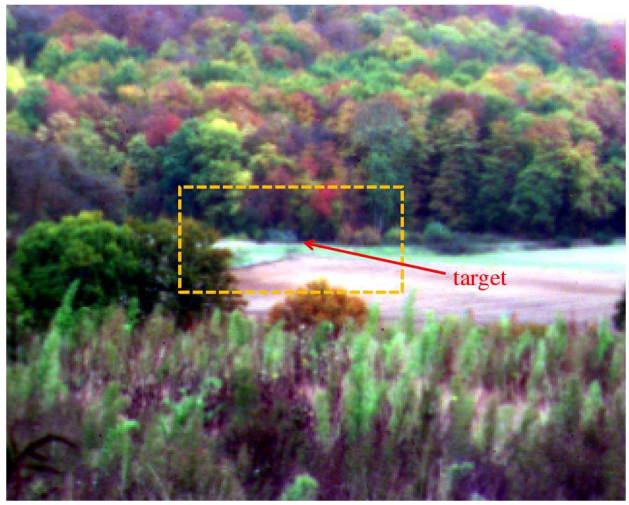


Fig. 1. (Color online) RGB representation of the multispectral image of the observed scene.

Without the assumptions of Eq. (1), the criterion of contrast C (or equivalently the squared Mahalanobis distance Δ^2) is not sufficient to characterize entirely the detection performance. However, it can be viewed as a simple measure of the detection performance that takes into account the first-order statistics of the target class and the second-order statistics of the background class.

C. Example of Target Detection Using the Matched Filter

We acquired multispectral images of a scene using a snapshot spectral imager. Each multispectral image consists of 20 bands in the VIS-NIR domain. Figure 1 shows a red-green-blue (RGB) representation of the considered multispectral image. We drew a dashed 3×12 pixel rectangle at the spatial location of the target, as shown in Fig. 2. This rectangle defines the region of interest ROI_1 of the target. Concerning the background, we define its region of interest ROI_0 as the whole image (including the target). We define the mean spectral signatures $\mathbf{m}_0$ and $\mathbf{m}_1$ of the background and the target respectively, and the covariance matrix $\mathbf{\Gamma}$ of the background as

$$\begin{cases} \mathbf{m}_1 & = \langle\mathbf{x}\rangle_1, \\ \mathbf{m}_0 & = \langle\mathbf{x}\rangle_0, \\ \mathbf{\Gamma} & = \langle(\mathbf{x}-\mathbf{m}_0)(\mathbf{x}-\mathbf{m}_0)^T\rangle_0, \end{cases} \tag{12}$$



Fig. 2. (Color online) Definition of the region of interest of the target.

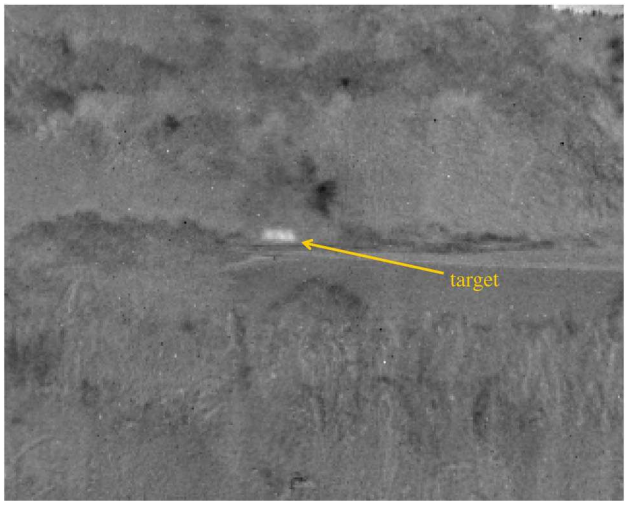


Fig. 3. (Color online) Detection plane obtained with the matched filter. The contrast C between the target and the background is 33.35.

where $\langle\cdot\rangle_j$ represents the mean on the region of interest ROI_j. Figure 3 represents the detection plane obtained when applying the matched filter $y = D_{\mathrm{mf}}(\mathbf{x}) = (\mathbf{m}_1 - \mathbf{m}_0)^T\mathbf{\Gamma}^{-1}(\mathbf{x} - \mathbf{m}_0)$ on the multispectral image. The target appears with a contrast $C = (\langle y\rangle_1 - \langle y\rangle_0)^2/\mathrm{Var}(y)_0 = 33.35$.

3. Band Selection

A. Description of the Band Selection Method

The matched filter involves the computation of the operation $y = \mathbf{w}_{\mathrm{mf}}^T(\mathbf{x} - \mathbf{m}_0)$ for each spectrum of interest $\mathbf{x}$, which requires K multiplications and $2(K-1)$ additions, where K is the number of spectral bands. Hyperspectral images typically consist of hundreds of bands. This can be an issue for processing images when dealing with real-time applications. One can then be interested in applying the matched filter on a small set of well-chosen bands. Those bands have to be chosen with respect to some criterion. This criterion is defined relatively to the application. Criteria have been developed for data representation [7], classification [9,16,17], or target detection applications [8,15,18]. We propose to select the bands to maximize the contrast C between the target and the background on the output of the matched filter.

We represent the process of filtering a spectrum by a set of K spectral filters by the operation $\mathbf{x}' = \mathbf{R}^T\mathbf{x}$ [8], where $\mathbf{x}$ is the initial spectrum (M element vector), $\mathbf{x}'$ is the final spectrum (K element vector), and $\mathbf{R} = (\mathbf{r_1}\ldots\mathbf{r_K})$ is the $M \times K$ filtering matrix as shown in Fig. 4. Its ith column $\mathbf{r_i}$ is an M element vector representing the spectral profile of the ith filter of the set. One can then use the K-band spectral image to detect a spectral target with the matched filter. We have $\mathbf{m}_1' = \mathbf{R}^T\mathbf{m}_1$, $\mathbf{m}_0' = \mathbf{R}^T\mathbf{m}_0$, and $\Gamma' = \mathbf{R}^T\mathbf{\Gamma}\mathbf{R}$. The contrast $C = C(\mathbf{R})$ between the target and the background in the detection plane obtained by the matched filter then becomes

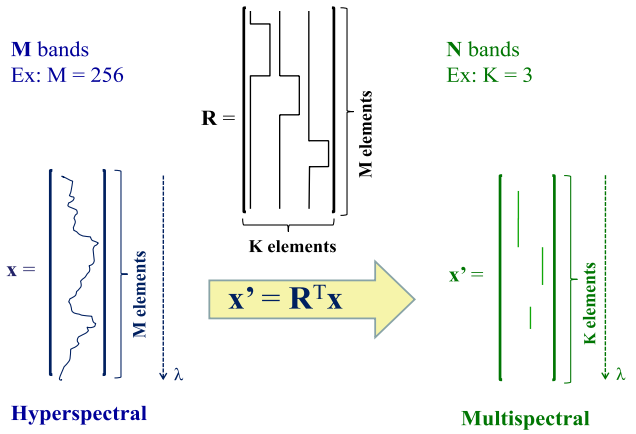


Fig. 4. (Color online) Filtering process.

$$C(\mathbf{R}) = (\mathbf{m}_1 - \mathbf{m}_0)^T\mathbf{R}(\mathbf{R}^T\mathbf{\Gamma}\mathbf{R})^{-1}\mathbf{R}^T(\mathbf{m}_1 - \mathbf{m}_0). \quad (13)$$

We restrain the band selection to a set $\mathcal{R}$ of p spectral filters. We denote as $\mathcal{R}^K$, the set of filtering matrices composed of K filters of $\mathcal{R}$. The band selection problem then consists in selecting the best filtering matrix $\mathbf{R}_{\mathbf{max}}$ among the set $\mathcal{R}^K$ with respect to the contrast criterion $C(\mathbf{R})$:

$$\mathbf{R}_{\mathbf{max}} = \arg\max_{\mathbf{R}\in\mathcal{R}^K} C(\mathbf{R}). \quad (14)$$

The optimization process can be proved to be a NP-hard problem. It means that one has to test each element of the set $\mathcal{R}^K$ to find the optimal solution. If we restrict the problem to a set of 100 spectral filters ($p = 100$), the rigorous optimization of the best 10 filters ($K = 10$) requires more than 10^{13} calculations of the criterion $C(\mathbf{R})$. We then have to find a mean for estimating an approximate optimum within a restricted time. There are at least two ways of approaching this problem. The first way consists in approximating the resolution of the exact problem by heuristic methods as sequential selection algorithms. The second way consists in finding an approximate problem for which we can obtain an exact solution. This can be done by convex optimization techniques [15].

Sequential forward selection (SFS) and sequential backward selection (SBS) are simple methods able to find suboptimal solutions in a reasonable time [19]. Further refinements, such as sequential forward floating selection, sequential backward floating selection, and steepest ascent [20], have been proposed to overcome the limitations of SFS and SBS. Metaheuristics, such as branch and bound algorithms, tabu search [21] or genetic algorithms [22], have been used for feature selection in pattern recognition literature. Kudo made a comparative study of most of the previous algorithms for large-scale feature selection [23]. We propose to solve our problem heuristically by means of a genetic algorithm [24]. Genetic algorithms are based on the analogy with

evolutionary theory, which claims that the population of a species tends to keep the genes that are the most adapted with respect to its environment. In our problem, the species is composed of filtering matrices $\mathbf{R}$ that need to maximize their associated contrast $C(\mathbf{R})$. Our algorithm aims to evolve a population of individuals from generation to generation in a random process in which the best individuals are favored to give birth to the next generation. The principle of our genetic algorithm is depicted in Fig. 5. The algorithm is composed of four basics steps:

1. Initialization: an initial population of N individuals is randomly chosen among the set S_n of all possible individuals (filtering matrices). Each individual is represented by a chromosome c (bit sequence) that codes a given filtering matrix $\mathbf{R} = f(\mathsf{c})$. Figure 6 provides the scheme $\mathbf{R} = f(\mathsf{c})$ used to code two types of filters (single-band and bandpass filters).
2. Evaluation: each individual is evaluated by its corresponding contrast $C(\mathbf{R}) = C(f(\mathsf{c}))$. Only the best N individuals are kept; the remaining individuals are thrown away.
3. Selection: N couples of individuals are randomly selected such that each individual's probability of selection increases with its corresponding contrast.
4. Reproduction: each selected couple gives birth to a novel individual. The reproduction process is illustrated in Fig. 7. In the first step, called recombination, we first select the bits shared by both parents and then draw randomly the remaining bits among the bits shared by only one parent. The second step is a mutation process, which consists in randomly changing a small number of bits (the total number of ones remains constant) to prevent the algorithm from converging to a local optimum. The N new individuals are then added to the population. The algorithm then goes to step 2 unless the maximum number of generations $i_{\max}$ is reached. In the latter case, the best individual of the population defines our band selection result.

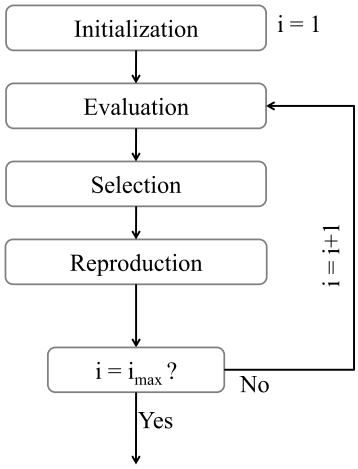


Fig. 5. Evolutionary cycle.

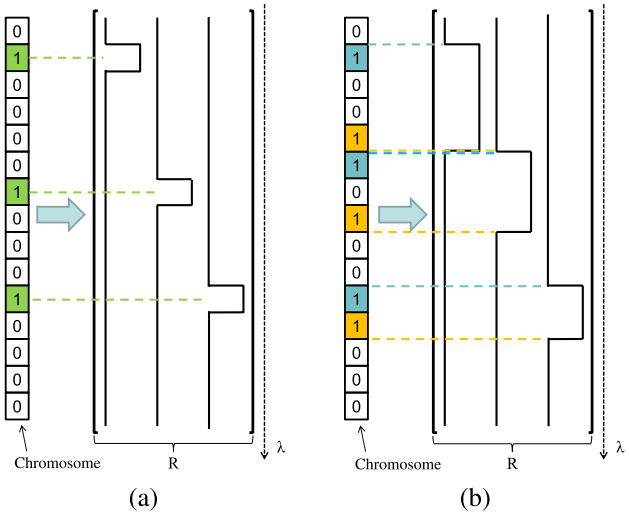


Fig. 6. (Color online) Chromosome and its corresponding filtering matrix $\mathbf{R}$ in the case of (a) ideal single-band filters and (b) bandpass filters.

B. Examples of Band Selection with Real Hyperspectral Imagery

We apply our band selection method to a 256-band hyperspectral image acquired using a Specim ImSpector QE V10E hyperspectral imager. On this image, we define two regions of interest, ROI_0 and ROI_1, corresponding, respectively, to the background and the target of interest. The contrast C between the target and the background calculated on the 256-band hyperspectral image is 150.41. We apply our algorithm with two different types of filters: single-band filters and bandpass filters (Fig. 6). Figure 8 represents the four best single bands selected by the algorithm. Figure 9 represents the four best bandpass filters selected by the algorithm.

Figure 10 represents the optimized contrast versus the number of selected bands for the two different types of filters. The optimized contrast is higher in the case of bandpass filters because the algorithm was able to optimize not only the position of the filters but also their linewidth.

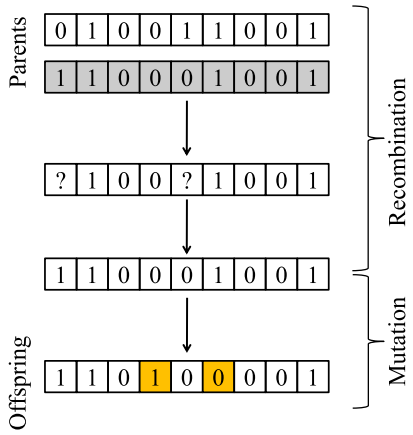


Fig. 7. (Color online) Reproduction process.

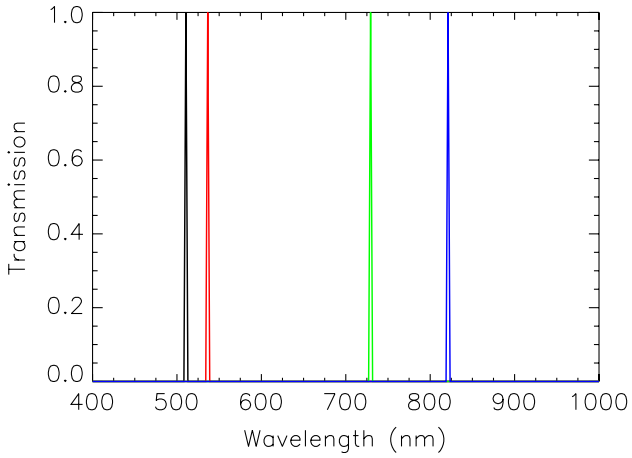


Fig. 8. (Color online) Profiles of the four best single-band filters. Each peak is a symbolic representation of the instrumental function of the Specim ImSpector QE V10E hyperspectral imager.

C. Performance of the Genetic Algorithm Compared with Two Other Algorithms

We compare our genetic algorithm with two other simple algorithms: a Monte Carlo algorithm and a SFS algorithm. The comparison is done in the case of single-band filters. The Monte Carlo algorithm selects randomly P sets of K bands and computes the P contrasts corresponding to each of the P sets. The set of bands that gives the maximum contrast is chosen as the solution of the optimization problem. As P tends to infinity, the Monte Carlo algorithm will converge to the optimal combination of K bands. Unfortunately, the convergence appears to be very slow. Our genetic algorithm can be seen as a Monte Carlo algorithm in which the memory of previous draws is used to orient the future draws. As a result, the genetic algorithm converges much faster to the optimal solution, as shown in Fig. 11.

The SFS algorithm is a simple and fast optimization technique. It starts from an empty set and iteratively adds to the set the best band among those that have not yet been chosen. Figure 12 compares the

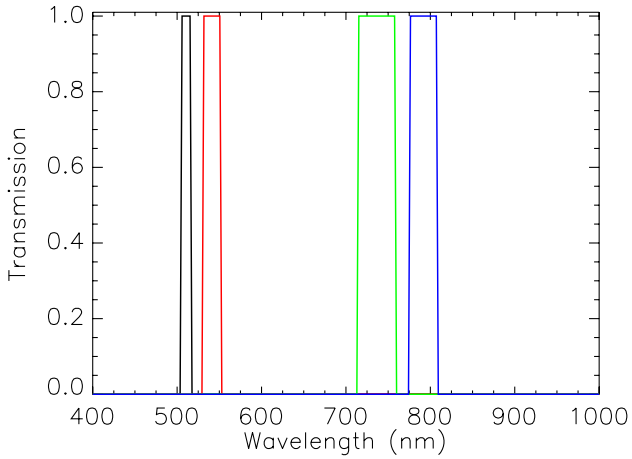


Fig. 9. (Color online) Profiles of the four best bandpass filters.

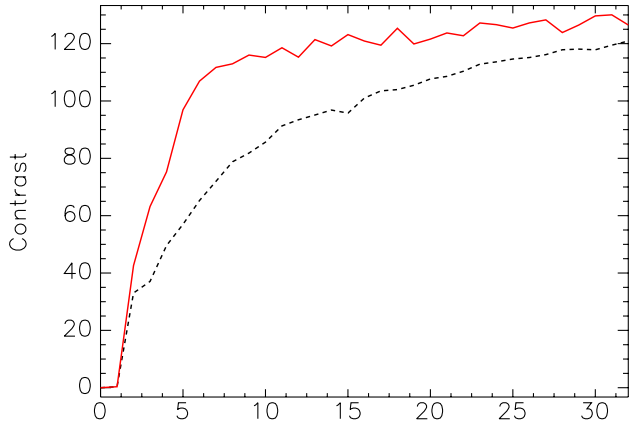


Fig. 10. (Color online) Contrast between the target and the background in the detection plane versus the number of selected bands. The black dotted curve corresponds to the optimized contrast obtained with single-band filters; the red solid curve corresponds to the optimized contrast obtained with bandpass filters.

performance of our genetic algorithm with the SFS algorithm. Between 4 and 10 selected bands, our algorithm outperforms the SFS algorithm. This illustrates the fact that the SFS algorithm may provide a suboptimal solution.

Table 1 displays the CPU times in seconds spent in running various band selection algorithms; all computation was done using IDL on a Windows platform with an Intel Core 2 Duo CPU running at 2.20 GHz and RAM of 3 GB. The running times were computed for a selection of 10 single-band filters. The running time of the genetic algorithm is approximately proportional to the product $i_{\max} \times N_{\text{elements}}$, where $i_{\max}$ is the number of generations and N_{elements} is the number of elements per generation. One can then tune those parameters to make a trade-off between running time and optimization performance. For

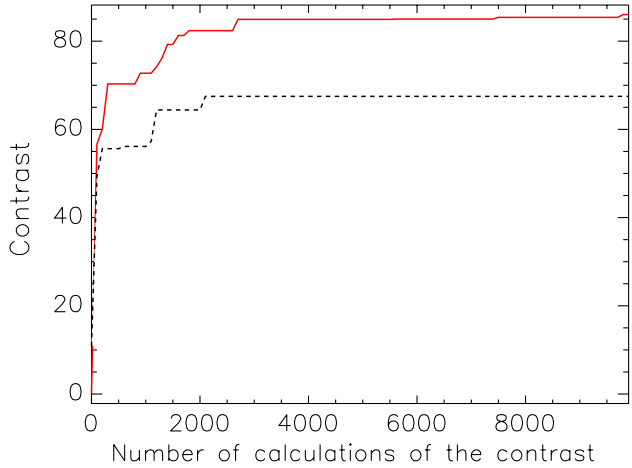


Fig. 11. (Color online) Maximum contrast obtained versus the number P of calculations of the contrast for two different algorithms: genetic algorithm (red solid curve) and Monte Carlo algorithm (black dotted curve). The optimization is done for $K = 10$ bands.

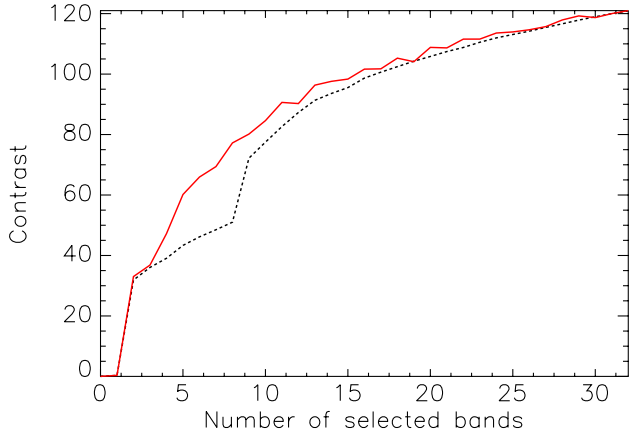


Fig. 12. (Color online) Contrast versus the number of selected single bands for two optimization algorithms: SFS algorithm (black dotted curve) and genetic algorithm (red solid curve). The genetic algorithm runs 100 generations of 100 elements ($P = 10,000$ calculations of the contrast).

similar running times, the SFS algorithm and the genetic algorithm seem to be on par. On the one hand, the genetic algorithm has the advantage of being tunable and adaptable to various kinds of filters. On the other hand, the SFS algorithm provides deterministic solutions and also intrinsically provides a ranking of each selected band.

D. Robustness of Band Selection Against Different Background Environments

The band selection results presented above have been obtained for one target against a specific background (image #1: birch foliage). It is of interest to discuss the robustness of band selection for the same target against different background environments. We acquired two other hyperspectral images of two different scenes containing the same target on different background environments (image #2: cedar foliage and image #3: grass). We can then discuss the robustness of the band selection against these different backgrounds. Let us consider those three images ($i = 1, 2, 3$). For each $i \in \{1, 2, 3\}$, we can compute the statistics $\mathbf{m}_{i,1}$, $\mathbf{m}_{i,0}$, and $\mathbf{\Gamma}_i$. Let us define the normalized contrast $\bar{C}_i(\mathbf{R})$ as

Table 1. CPU Running Time of the Band Selection Algorithms for a Selection of 10 Single Bands[a]

Algorithm	$i_{\max}$	N_{elements}	CPU Time (s)	Contrast C
SFS	—	—	6.06	77.48
Genetic	10	10	0.31	55.65
Genetic	20	20	1.26	69.56
Genetic	50	50	8.00	80.86
Genetic	100	100	32.21	87.56

[a]The genetic algorithm was tested for different values of two parameters: $i_{\max}$, the number of generations, and N_{elements}, the number of elements per generation. For the genetic algorithm, CPU times and contrasts have been averaged on 10 runs of the algorithm.

$$\bar{C}_i(\mathbf{R}) = \frac{(\mathbf{m}_{i,1} - \mathbf{m}_{i,0})^T \mathbf{R}(\mathbf{R}^T \mathbf{\Gamma}_i \mathbf{R})^{-1} \mathbf{R}^T (\mathbf{m}_{i,1} - \mathbf{m}_{i,0})}{(\mathbf{m}_{i,1} - \mathbf{m}_{i,0})^T \mathbf{\Gamma}_i^{-1} (\mathbf{m}_{i,1} - \mathbf{m}_{i,0})}. \tag{15}$$

For each $i \in \{1, 2, 3\}$, we can then find the best filtering matrix $\mathbf{R}_{\max}^i$ with respect to the contrast $\bar{C}_i = \bar{C}_i(\mathbf{R})$:

$$\mathbf{R}_{\max}^i = \arg \max_{\mathbf{R} \in \mathcal{R}^K} \bar{C}_i(\mathbf{R}). \tag{16}$$

We can also define the global normalized contrast $\bar{C}_{\text{glob}}(\mathbf{R})$ as

$$\bar{C}_{\text{glob}}(\mathbf{R}) = \frac{1}{3} \sum_{i=1}^{3} \bar{C}_i(\mathbf{R}). \tag{17}$$

The optimization of $\bar{C}_{\text{glob}}(\mathbf{R})$ leads to the filtering matrix $\mathbf{R}_{\max}^{\text{glob}}$:

$$\mathbf{R}_{\max}^{\text{glob}} = \arg \max_{\mathbf{R} \in \mathcal{R}^K} \bar{C}_{\text{glob}}(\mathbf{R}). \tag{18}$$

Figure 13 represents the optimization of the global normalized contrast $\bar{C}_{\text{glob}}(\mathbf{R}_{\max}^{\text{glob}})$ and the mean of the three independently optimized normalized contrast $\langle \bar{C}_i(\mathbf{R}_{\max}^i) \rangle_{i \in \{1,2,3\}}$ versus the number of selected bandpass filters. The detection performance obtained when the band selection is optimized for a trade-off between the three backgrounds is obviously inferior to the performance obtained when the band selection can be tuned to each background. However, the loss in performance can be considered as acceptable.

4. Contrast of the Matched Filter in the Presence of Target Signature Estimation Error

In Section 2 we used the information we had about the spatial location of the target to detect it on the image. Strictly speaking, we enhanced the contrast of the target, but we did not detect it, as we already

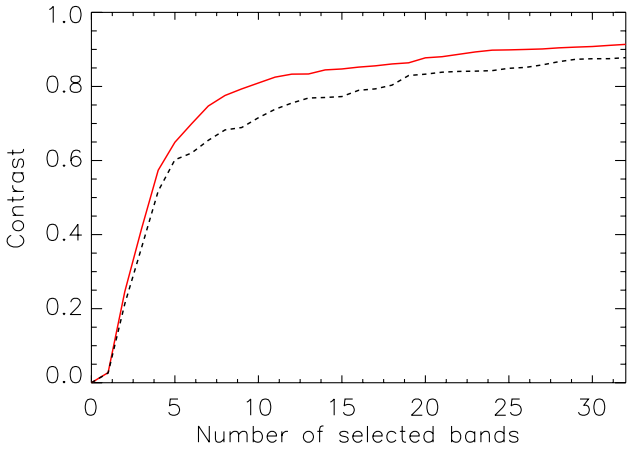


Fig. 13. (Color online) Optimization of the global normalized contrast $\bar{C}_{\text{glob}}(\mathbf{R}_{\max}^{\text{glob}})$ (black dotted curve) and mean of the three optimized normalized contrast $\langle \bar{C}_i(\mathbf{R}_{\max}^i) \rangle_{i \in \{1,2,3\}}$ (red solid curve) versus the number of selected bands.

knew its location. In real-world applications, the target spectral signature has to be estimated. This estimation can be performed through reflectance retrieval techniques [25] or from the information about previously detected targets. One then has an estimation $\hat{\mathbf{m}}_1$ of the target mean spectrum $\mathbf{m}_1$. We can define a detector using the matched filter in which the spectral mean of the target is replaced by its estimation:

$$y' = D_{\mathrm{amf}}(\mathbf{x}) = (\hat{\mathbf{m}}_1 - \mathbf{m}_0)^T \boldsymbol{\Gamma}^{-1}(\mathbf{x} - \mathbf{m}_0). \tag{19}$$

This detector is called the *adaptive matched filter* (AMF) [4]; it is of interest to compute its corresponding contrast C'. We define the whitened variables $\hat{\boldsymbol{\mu}} = \boldsymbol{\Gamma}^{-1/2}(\hat{\mathbf{m}}_1 - \mathbf{m}_0)$ and $\boldsymbol{\mu} = \boldsymbol{\Gamma}^{-1/2}(\mathbf{m}_1 - \mathbf{m}_0)$. Then, we can show that

$$\begin{cases} \mathrm{E}\{y'|\gamma_1\} & = \hat{\boldsymbol{\mu}}^T \boldsymbol{\mu}, \\ \mathrm{E}\{y'|\gamma_0\} & = 0, \\ \mathrm{var}\{y'|\gamma_0\} & = \hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\mu}}. \end{cases} \tag{20}$$

We then have

$$C' = \frac{(\hat{\boldsymbol{\mu}}^T \boldsymbol{\mu})^2}{\hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\mu}}}, \tag{21}$$

$$= \boldsymbol{\mu}^T \boldsymbol{\mu} \frac{(\hat{\boldsymbol{\mu}}^T \boldsymbol{\mu})^2}{(\hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\mu}})(\boldsymbol{\mu}^T \boldsymbol{\mu})}, \tag{22}$$

$$= C \frac{(\hat{\boldsymbol{\mu}}^T \boldsymbol{\mu})^2}{(\hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\mu}})(\boldsymbol{\mu}^T \boldsymbol{\mu})}, \tag{23}$$

where $C = (\mathbf{m}_1 - \mathbf{m}_0)^T \boldsymbol{\Gamma}^{-1}(\mathbf{m}_1 - \mathbf{m}_0) = \boldsymbol{\mu}^T \boldsymbol{\mu}$ is the contrast obtained in the absence of estimation error. Using the Cauchy–Schartz inequality, we observe that $C' \le C$. The contrast loss C'/C is equal to the squared cosine of the angle between $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}$.

Let us first consider the worst case of target signature estimation where we randomly select a background pixel that defines the estimated target signature $\hat{\mathbf{m}}_1$. In this case, the estimated target signature $\hat{\mathbf{m}}_1$ is a random vector that has the same statistical distribution as the background. The whitened variable $\hat{\boldsymbol{\mu}}$ is thus a random vector with zero mean and identity covariance matrix. Let $(\mathbf{u}_i)_{i\in[1,K]}$ with $\mathbf{u}_1 = \boldsymbol{\mu}/\|\boldsymbol{\mu}\|$ be an orthonormal basis of $\mathbb{R}^K$. The decomposition of $\boldsymbol{\mu}$ in this basis can be written $\boldsymbol{\mu} = \sum_{i=1}^K \mu_i \mathbf{u}_i = \mu_1 \mathbf{u}_1$. Similarly, $\hat{\boldsymbol{\mu}} = \sum_{i=1}^K \hat{\mu}_i \mathbf{u}_i$. We then have

$$\hat{\boldsymbol{\mu}}^T \boldsymbol{\mu} = \hat{\mu}_1 \mu_1, \tag{24}$$

$$\hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\mu}} = \sum_{i=1}^K \hat{\mu}_i^2. \tag{25}$$

Using Eq. (23), we obtain

$$C' = C \frac{\hat{\mu}_1^2}{\sum_{i=1}^K \hat{\mu}_i^2}. \tag{26}$$

Because the components $\hat{\mu}_j$ of $\hat{\boldsymbol{\mu}}$ are independent and identically distributed random variables, we have

$$\begin{aligned} \mathrm{E}\left\{\frac{\hat{\mu}_1^2}{\sum_{i=1}^K \hat{\mu}_i^2}\right\} &= \frac{1}{K}\sum_{j=1}^K \mathrm{E}\left\{\frac{\hat{\mu}_j^2}{\sum_{i=1}^K \hat{\mu}_i^2}\right\} \\ &= \frac{1}{K}\mathrm{E}\left\{\sum_{j=1}^K \frac{\hat{\mu}_j^2}{\sum_{i=1}^K \hat{\mu}_i^2}\right\} \\ &= \frac{1}{K}. \end{aligned} \tag{27}$$

One can then obtain the expectation of the contrast C':

$$\mathrm{E}\{C'\} = \frac{C}{K}. \tag{28}$$

When one has no information about the spectral signature of the target, the expectation of the contrast C' is equal to the maximum contrast C divided by the number of bands K.

A. Simple Model for Estimation Error

Let us now consider unbiased estimations $\hat{\mathbf{m}}_1$ of the mean of the target. For instance, we can consider that $\hat{\mathbf{m}}_1 = \mathbf{m}_1 + \boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}$, where $\boldsymbol{\Omega}$ is a symmetric positive semidefinite matrix and $\boldsymbol{\epsilon}$ is a random vector with zero mean and identity covariance matrix. The estimation error $\hat{\mathbf{m}}_1 - \mathbf{m}_1$ is then a random vector with zero mean (unbiased estimator) and Ω covariance matrix. We can note that $\hat{\boldsymbol{\mu}} = \boldsymbol{\Gamma}^{-1/2}(\hat{\mathbf{m}}_1 - \mathbf{m}_0) = \boldsymbol{\mu} + \boldsymbol{\Gamma}^{-1/2}\boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}$.

In the simple case where the covariance matrix of the estimation error is proportional to the covariance matrix of the background ($\boldsymbol{\Omega} = \alpha^2\boldsymbol{\Gamma}$), the whitened vector $\hat{\boldsymbol{\mu}}$ simplifies to $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \alpha\boldsymbol{\epsilon}$ and it becomes possible to obtain a simple approximation for the expectation $\mathrm{E}\{C'\}$ of the contrast C'.

This situation occurs, for example, when the target spectral signature $\mathbf{m}_1$ is estimated from N independent samples $(\mathbf{x}_i)_{i\in[1,N]}$ of the target class:

$$\hat{\mathbf{m}}_1 = \frac{1}{N}\sum_{i=1}^N \mathbf{x}_i. \tag{29}$$

The estimation error $\hat{\mathbf{m}}_1 - \mathbf{m}_1$ is a random variable. We have under the normal assumptions of Eq. (1):

$$\mathrm{E}\{\hat{\mathbf{m}}_1 - \mathbf{m}_1\} = \mathbf{0}, \tag{30}$$

$$\mathrm{Cov}\{\hat{\mathbf{m}}_1 - \mathbf{m}_1\} = \frac{1}{N}\boldsymbol{\Gamma}. \tag{31}$$

In this case, the covariance matrix of the estimation error is proportional to the covariance matrix of the background ($\alpha^2 = 1/N$).

Let $(\mathbf{u}_i)_{i\in[1,K]}$ with $\mathbf{u}_1 = \boldsymbol{\mu}/\|\boldsymbol{\mu}\|$ be an orthonormal basis of $\mathbb{R}^K$. The decomposition of $\boldsymbol{\mu}$ in this basis can be written $\boldsymbol{\mu} = \sum_{i=1}^K \mu_i \mathbf{u}_i = \mu_1 \mathbf{u}_1$. Similarly, $\boldsymbol{\epsilon} = \sum_{i=1}^K \epsilon_i \mathbf{u}_i$. We then have

$$\hat{\boldsymbol{\mu}}^T \boldsymbol{\mu} = ((\mu_1 + \alpha\epsilon_1)\mathbf{u}_1 + \sum_{i=2}^K \alpha\epsilon_i \mathbf{u}_i)^T \mu_1 \mathbf{u}_1 = \mu_1^2 + \alpha\epsilon_1\mu_1, \tag{32}$$

$$\hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\mu}} = \|(\mu_1 + \alpha\epsilon_1)\mathbf{u}_1 + \sum_{i=2}^K \alpha\epsilon_i \mathbf{u}_i\|^2 = (\mu_1 + \alpha\epsilon_1)^2 + \sum_{i=2}^K \alpha^2\epsilon_i^2. \tag{33}$$

We finally obtain the contrast C':

$$C' = \mu_i^2 \frac{(\mu_1 + \alpha\epsilon_1)^2}{(\mu_1 + \alpha\epsilon_1)^2 + \sum_{i=2}^K \alpha^2\epsilon_i^2} = C \times \frac{(\mu_1/\alpha + \epsilon_1)^2}{(\mu_1/\alpha + \epsilon_1)^2 + \sum_{i=2}^K \epsilon_i^2} = C \times \frac{\chi^2_{1,C/\alpha^2}}{\chi^2_{1,C/\alpha^2} + \chi^2_{K-1}}, \tag{34}$$

where $\chi^2_{1,C/\alpha^2}$ is a random variable that follows a noncentral χ^2 distribution with 1 degree of freedom and noncentrality parameter C/α^2 and χ^2_{K-1} is a random variable that follows a central χ^2 distribution with $K-1$ degrees of freedom. The random variable C'/C has a noncentral beta distribution with shape parameter $a = \frac{1}{2}$, $b = \frac{K-1}{2}$, and noncentrality parameter $\lambda = C/\alpha^2$ [26]. There is no simple formula for the mean of the noncentral beta function. However, we can approach the expectation of C' by

$$\mathrm{E}\{C'\} \simeq C \times \frac{\mathrm{E}\{(\mu_1/\alpha + \epsilon_1)^2\}}{\mathrm{E}\{(\mu_1/\alpha + \epsilon_1)^2 + \sum_{i=2}^K \epsilon_i^2\}} \simeq C \times \frac{C/\alpha^2 + 1}{C/\alpha^2 + K}. \tag{35}$$

This formula gives a synthetic expression of the contrast loss as a function of α^2, which represents the variance of the target estimation error relatively to the background variance and K, the number of spectral bands. Figure 14 represents the mean contrast loss $\mathrm{E}\{C'\}/C$ versus the number of spectral bands K for different values of the parameter C/α^2. The mean contrast loss was computed using two ways: the solid lines represent the mean contrast loss estimated from 1000 random draws of $\boldsymbol{\epsilon}$; the dotted lines represent the mean contrast loss estimated from the approximation of Eq. (35). We observe that the approximation is accurate, especially when $C/\alpha^2 \gg 1$,

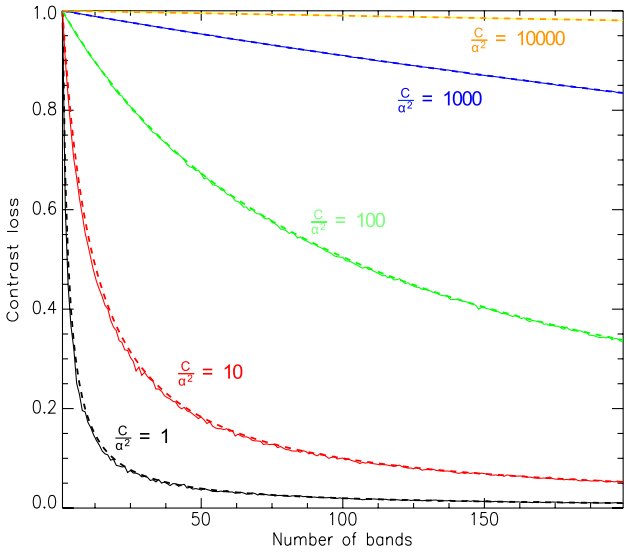


Fig. 14. (Color online) Mean contrast loss $\mathrm{E}\{C'\}/C$ versus the number of bands K for different values of the parameter C/α^2. Solid curves, contrast loss estimation from 1000 random draws. Dotted curves, $(C/\alpha^2+1)/(C/\alpha^2+K)$.

which is true for most detection applications. One can note that when $K \simeq C/\alpha^2 + 2$, the expected contrast $\mathrm{E}\{C'\}$ is about half of the maximum contrast C.

The approximation of Eq. (35) can then be used to evaluate the effect of estimation error on the performance of the matched filter. Figure 15 represents the expected contrast $\mathrm{E}\{C'\}$ versus the number of spectral bands K for different values of the parameter α^2. The black curve (no error) represents the optimized contrast C versus the number of bands K when there is no error in the estimation of the target ($\hat{\mathbf{m}}_1 = \mathbf{m}_1$). The contrast $C(K)$ is optimized for the hyperspectral image of Section 3 using a genetic algorithm (single-band filters). We then use $C(K)$ to estimate the expectation of contrast for different values of α^2 using the approximation of Eq. (35). Figure 15 shows that in the presence of estimation error, it may not be optimal to use the full spectral resolution as the

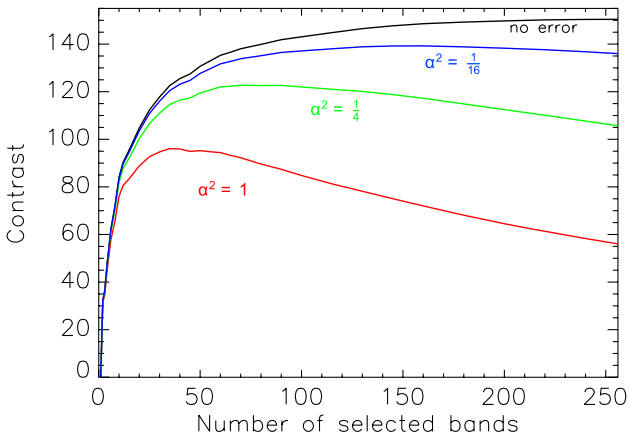


Fig. 15. (Color online) Expected contrast versus number of bands for different values of α^2 [Eq. (35)].

maximum expected contrast is reached for a reduced number of selected spectral bands.

5. Conclusion

We addressed the issue of selecting the spectral bands that optimize the detection performance of the matched filter. The spectral bands have been chosen to maximize a contrast criterion, which is equal to the Mahalanobis distance between the target and the background. Under the assumption that the target and the background follow multivariate normal distributions with different means and identical covariance matrices, this approach leads to an optimal detector in the Neyman–Pearson sense. When this assumption is not verified, the obtained detector is not optimal in the Neyman–Pearson sense, but it remains the linear detector that optimizes the contrast between the target and the background in the detection plane.

We developed a band selection method that allows selecting not only the position of the spectral bands but also their linewidth. The optimization of the cutoff wavelengths of the bandpass filters greatly improves the contrast in the detection plane. It can be very useful for the design of multispectral imagers adapted to detection applications. We proposed to select the bands using a genetic algorithm. We showed that our genetic algorithm provides a good framework to optimize various kinds of filters. We showed results of band selection from a real hyperspectral image in the case of single-band and bandpass filters. The performance of our method was compared with success with other band selection algorithms.

Finally, we discussed the effect of estimation error in the target signature on the performance of the matched filter. We showed that under certain assumptions about the estimation error, the contrast loss can be approximated by a very simple formula. This formula is a useful tool to estimate the number of bands that is needed for a given detection application.

The authors thank Rodolphe Marion of the Commissariat à l'Énergie Atomique for the acquisition of hyperspectral images. The work of J. Minet is supported by the Direction Générale de l'Armement.

References

1. J. C. Harsanyi and C. I. Chang, "Detection of subpixel signatures in hyperspectral image sequences," in *Proceedings of the American Society for Photogrammetry and Remote Sensing* (1994), pp. 236–247.
2. J. C. Harsanyi and C. I. Chang, "Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach," IEEE Trans. Geosci. Remote Sens. **32**, 779–785 (1994).
3. X. Yu, I. S. Reed, and A. D. Stocker, "Comparative performance analysis of adaptive multispectral detectors," IEEE Trans. Signal Process. **41**, 2639–2656 (1993).
4. D. G. Manolakis and G. A. Shaw, "Detection algorithms for hyperspectral imaging applications," IEEE Signal Process. Mag. **19**, 29–43 (2002).
5. D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," Lincoln Lab. J. **14**, 79–116 (2003).
6. D. G. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, "Is there a best hyperspectral detection algorithm?" SPIE Newsroom (17 June 2009).
7. J. C. Price, "Band selection procedure for multispectral scanners," Appl. Opt. **33**, 3281–3288 (1994).
8. J. Karlholm and I. Renhorn, "Wavelength band selection method for multispectral target detection," Appl. Opt. **41**, 6786–6795 (2002).
9. S. De Backer, P. Kempeneers, W. Debruyn, and P. Scheunders, "Band selection for hyperspectral remote sensing," IEEE Geosci. Remote Sens. Lett. **2**, 319–323 (2005).
10. J. Minet, J. Taboury, M. Péalat, N. Roux, J. Lonnoy, and Y. Ferrec, "Adaptive band selection snapshot multispectral imaging in the VIS/NIR domain," Proc. SPIE **7835**, 78350W (2010).
11. D. Landgrebe, "Hyperspectral image data analysis as a high dimensional signal processing problem," IEEE Signal Process. Mag. **19**, 17–28 (2002).
12. F. Goudail, N. Roux, I. Baarstad, T. Løke, P. Kaspersen, M. Alouini, and X. Normandin, "Some practical issues in anomaly detection and exploitation of regions of interest in hyperspectral images," Appl. Opt. **45**, 5223–5236 (2006).
13. G. Hughes, "On the mean accuracy of statistical pattern recognizers," IEEE Trans. Inf. Theory. **14**, 55–63 (1968).
14. S. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory* (Prentice-Hall, 1998).
15. J. Theiler and K. Glocer, "Sparse linear filters for detection and classification in hyperspectral imagery," Proc. SPIE **6233**, 62330H (2006).
16. C. I. Chang, Q. Du, T. L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens. **37**, 2631–2641 (1999).
17. E. Arzuaga-Cruz, L. O. Jimenez-Rodriguez, and M. Vélez-Reyes, "Unsupervised feature extraction and band subset selection techniques based on relative entropy criteria for hyperspectral data analysis," Proc. SPIE **5093**, 462–473 (2003).
18. C. I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," IEEE Trans. Geosci. Remote Sens. **44**, 1575–1585 (2006).
19. S. D. Stearns, B. E. Wilson, and J. R. Peterson, "Dimensionality reduction by optimal band selection for pixel classification of hyperspectral imagery," Proc. SPIE **2028**, 118–127 (1993).
20. S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," IEEE Trans. Geosci. Remote Sens. **39**, 1360–1367 (2001).
21. D. Korycinski, M. M. Crawford, and J. W. Barnes, "Adaptive feature selection for hyperspectral data analysis," Proc. SPIE **5238**, 213–225 (2004).
22. W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," Pattern Recogn. Lett. **10**, 335–347 (1989).
23. M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," Pattern Recogn. **33**, 25–41 (2000).
24. D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison-Wesley, 1989).
25. M. K. Griffin and H. K. Burke, "Compensation of hyperspectral data for atmospheric effects," Lincoln Lab. J. **14**, 29–54 (2003).
26. R. Chattamvelli and R. Shanmugam, "Computing the non-central beta distribution function," Appl. Statist. **46**, 146–156 (1997).