In this correspondence we have not addressed the problem of constructing actual codebooks. Information theory indicates that, in principle, one can construct a codebook by drawing each component of each codeword independently, using the distribution obtained from the Blahut algorithm. This procedure is not in general practical. Practical ways to construct codewords may be found in the extensive literature on vector quantization (see, e.g., the tutorial paper by R. M. Gray [19] or the book [20]). It is not clear at this point if codebook constructing methods from the vector quantizer literature are practical in the setting of this correspondence. Alternatively, one can trade complexity and performance and construct a scalar quantizer. In this case, the distribution obtained from the Blahut algorithm may be used in the Max–Lloyd algorithm [21], [22].

REFERENCES

[1] U. Grenander, *General Pattern Theory*. Oxford, U.K.: Oxford Univ. Press, 1994.

[2] A. Srivastava, M. I. Miller, and U. Grenander, "Multiple target direction of arrival tracking," *IEEE Trans. Signal Processing*, vol. 43, pp. 1282–1285, May 1995.

[3] M. I. Miller, A. Srivastava, and U. Grenander, "Conditional-mean estimation via jump-diffusion processes in multiple target tracking/recognition," *IEEE Trans. Signal Processing*, vol. 43, pp. 2678–2690, Nov. 1995.

[4] M. I. Miller, U. Grenander, J. A. O'Sullivan, and D. L. Snyder, "Automatic target recognition organized via jump-diffusion algorithms," *IEEE Trans. Image Processing*, vol. 6, pp. 1–17, Jan. 1997.

[5] A. Srivastava, M. Miller, and U. Grenander, "Ergodic algorithms on special euclidean groups for atr," *Progr. Syst. Contr.: Syst. Contr. the Twenty-First Century*, vol. 22, pp. 327–350, 1997.

[6] A. Srivastava, "Inferences on transformation groups generating patterns on rigid motions," Ph.D. dissertation, Washington Univ., Dept. Elec. Eng., St. Louis, MO, Aug. 1996.

[7] M. Artin, *Algebra*. Englewood Cliffs, NJ: Prentice Hall, 1991.

[8] U. Grenander, M. I. Miller, and A. Srivastava, "Hilbert–Schmidt lower bounds for estimators on matrix lie groups for atr," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 790–802, Aug. 1998.

[9] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[10] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[11] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compressions*. Englewood Cliffs, NJ: Prentice-Hill, 1971.

[12] J. A. O'Sullivan, "Alternating minimization algorithms: from Blahut–Arimoto to Expectation-Maximization," in *Codes, Curves, and Signals: Common Threads in Communications*, A. Vardy, Ed. Boston, MA: Kluwer, 1998, pp. 173–192.

[13] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 460–473, July 1972.

[14] T. M. Cover, "An algorithm for maximizing expected log investment return," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 369–373, Mar. 1984.

[15] I. Csiszár, "On the computation of rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 122–124, Jan. 1974.

[16] I. Csiszár and G. Tusnady, "Information geometry and alternating minimization procedures," *Statist. Decisions, Suppl. Issue 1*, pp. 205–237, 1974.

[17] D. T. Gillespie, "The Monte Carlo Method of Evaluating Integrals," Naval Weapons Ctr., China Lake, CA, Tech. Rep., Feb. 1975.

[18] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*. London, U.K.: Methuen, 1964.

[19] R. M. Gray, "Vector quantization," *IEEE ASSP Mag.*, pp. 4–28, Apr. 1984.

[20] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.

[21] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 129–137, Mar. 1982 (reprint of Bell Labs. Tech. Note, 1957).

[22] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 7–12, Mar. 1960.

An Information-Theoretic Approach to Spectral Variability, Similarity, and Discrimination for Hyperspectral Image Analysis

Chein-I Chang, *Senior Member, IEEE*

***Abstract*—A hyperspectral image can be considered as an image cube where the third dimension is the spectral domain represented by hundreds of spectral wavelengths. As a result, a hyperspectral image pixel is actually a column vector with dimension equal to the number of spectral bands and contains valuable spectral information that can be used to account for pixel variability, similarity, and discrimination. In this correspondence, we present a new hyperspectral measure, Spectral Information Measure (SIM), to describe spectral variability and two criteria, spectral information divergence and spectral discriminatory probability, for spectral similarity and discrimination, respectively. The spectral information measure is an information-theoretic measure which treats each pixel as a random variable using its spectral signature histogram as the desired probability distribution. Spectral Information Divergence (SID) compares the similarity between two pixels by measuring the probabilistic discrepancy between two corresponding spectral signatures. The spectral discriminatory probability calculates spectral probabilities of a spectral database (library) relative to a pixel to be identified so as to achieve material identification. In order to compare the discriminatory power of one spectral measure relative to another, a criterion is also introduced for performance evaluation, which is based on the power of discriminating one pixel from another relative to a reference pixel. The experimental results demonstrate that the new hyperspectral measure can characterize spectral variability more effectively than the commonly used Spectral Angle Mapper (SAM).**

***Index Terms*—Hyperspectral image, spectral angle mapper, spectral discriminatory entropy, spectral discriminatory power, spectral discriminatory probability, spectral information divergence, spectral information measure.**

I. INTRODUCTION

The advent of high spectral resolution airborne and satellite sensors improves the capability for collection of ground targets in the fields of agriculture, geology, geography, law enforcement, military and defense, etc. [1]. One major advantage of a hyperspectral sensor over a multispectral sensor is that the former images a scene using as many as 200 contiguous bands as opposed to the latter that only uses tens of discrete bands. Therefore, hyperspectral image data permit the expansion of detection and classification activities to targets previously unresolved in multispectral images. Due to improved spectral resolution and large scene coverage, a hyperspectral pixel is generally a mixture of different materials present in the pixel with various abundance fractions. These materials absorb or reflect within each spectral band. As a consequence, spectral characterization becomes crucial in hyperspectral image analysis. However, because of atmospheric effects, the spectral information of a pixel varies during data acquisition. In order to account for spectral variability, similarity, and discrimination, we present in this correspondence an information-theoretic spectral measure, referred to as Spectral Information Measure (SIM). SIM models the spectral band-to-band variability as a result of uncertainty caused

Manuscript received August 31, 1999; revised March 10, 2000. The material in the correspondence was presented in part at the International Geoscience and Remote Sensing Symposium '99, Hamburg, Germany, June 28–July 2, 1999.

The author is with the Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250 USA (e-mail: cchang@umbc.edu).

Communicated by J. M. F. Moura, Guest Editor.

Publisher Item Identifier S 0018-9448(00)06069-7.

by randomness. It considers each pixel as a random variable and defines the desired probability distribution by normalizing its spectral histogram to unity. With this interpretation, SIM not only can describe the randomness of a pixel, but can generate also high-order statistics of the pixel based on its spectral signature. This is particularly useful for hyperspectral image analysis. Since each hyperspectral image pixel is composed of hundreds of spectral bands, the spectral information provided by this pixel is generally very valuable in material discrimination, detection, classification, and identification. However, this advantage also comes at a price; many unknown signal sources may be also collected by the sensor [2]. SIM is designed to capture such uncertainty and unpredictability. The higher the data dimensionality, the more effective is the SIM. In order to determine the spectral similarity between two pixels, a SIM-based Spectral Information Divergence (SID) is developed, based on the concept of divergence [3], [4] to measure the discrepancy of probabilistic behaviors between the spectral signatures of the two pixels. Compared to the Spectral Angle Mapper (SAM), which has been widely used in the past [5], SID is more effective in preserving spectral properties. In many applications, it is required to identify a target in an unknown image scene via a spectral library (database). To meet this need, a criterion is introduced to calculate the spectral discriminatory probabilities of the signatures in the library and to provide the likelihood of identifying the target. Finally, in order to compare the relative discriminatory power between two spectral measures, another criterion is proposed, based on the ratio of the spectral similarities of a pixel to another pixel relative to a reference pixel.

II. Spectral Information Measure (SIM)

For a given hyperspectral pixel (vector) $\boldsymbol{x} = (x_1, \cdots, x_L)^T$, each component x_l is a pixel of band B_l acquired at a particular wavelength λ_l. To simplify notation, we also use x_l to represent its spectral signature in the form of either radiance or reflectance. Suppose that $\{\lambda_l\}_{l=1}^{L}$ is a set of L wavelengths, each of which corresponds to a spectral band channel. Then $\boldsymbol{x}$ can be modeled as a random variable by defining an appropriate probability space (Ω, Σ, P) associated with it where Ω is a sample space, Σ is an event space, and P is a probability measure. In order to define a legitimate probability measure P for $\boldsymbol{x}$, we first assume that all component x_l's in $\boldsymbol{x}$ are nonnegative. This is generally a valid assumption due to the nature of radiance or reflectance. With this assumption, we can define a probability measure P for $\boldsymbol{x}$ by

$$P(\{\lambda_j\}) = p_j = x_j \Big/ \sum_{l=1}^{L} x_l \ . \tag{1}$$

The vector $\boldsymbol{p} = (p_1, p_2, \cdots, p_L)^T$ is the desired probability vector resulting from the pixel vector $\boldsymbol{x}$. With this interpretation of (1), any pixel $\boldsymbol{x} = (x_1, \cdots, x_L)^T$ can be viewed as a single information source with its statistics governed by $\boldsymbol{p} = (p_1, p_2, \cdots, p_L)^T$, which can be used to describe the spectral variability of a pixel. For instance, we can define its mean

$$\mu(\boldsymbol{x}) = \sum_{l=1}^{L} p_l x_l$$

variance

$$\sigma^2(\boldsymbol{x}) = \sum_{l=1}^{L} p_l (x_l - \mu(\boldsymbol{x}))^2$$

third central moment

$$\kappa_3(\boldsymbol{x}) = \sum_{l=1}^{L} p_l (x_l - \mu(\boldsymbol{x}))^3$$

fourth central moment

$$\kappa_4(\boldsymbol{x}) = \sum_{l=1}^{L} p_l (x_l - \mu(\boldsymbol{x}))^4$$

and so on. Like the Taylor's series, which is widely used to approximate a deterministic function, we can also use the moment generation equation to fully describe the probabilistic behavior of the spectral signature of each hyperspectral image pixel where all moments can be obtained via (1). Using (1) we can further define the entropy of each hyperspectral image pixel $\boldsymbol{x}$

$$H(\boldsymbol{x}) = -\sum_{l=1}^{L} p_l \log p_l$$

which can be used to describe the uncertainty or randomness present in the pixel $\boldsymbol{x}$.

III. Spectral Information Divergence (SID)

In this section, we will derive a SIM-based spectral similarity measure to capture the spectral correlation between two pixels. Assume that $\boldsymbol{y} = (y_1, \cdots, y_L)^T$ is another pixel with probability vector $\boldsymbol{q} = (q_1, q_2, \cdots, q_L)^T$ where $q_j = y_j / \sum_{l=1}^{L} y_l$. From information theory [3] we can define the lth band self-information of $\boldsymbol{x}$ and $\boldsymbol{y}$ as follows:

$$I_l(\boldsymbol{x}) = -\log p_l \ \text{ and } \ I_l(\boldsymbol{y}) = -\log q_l, \ \text{ respectively.} \tag{2}$$

Using (2), the relative entropy of $\boldsymbol{y}$ with respect to $\boldsymbol{x}$ can be defined by

$$\begin{aligned} D(\boldsymbol{x}\|\boldsymbol{y}) &= \sum_{l=1}^{L} p_l D_l(\boldsymbol{x}\|\boldsymbol{y}) = \sum_{l=1}^{L} p_l \left(I_l(\boldsymbol{y}) - I_l(\boldsymbol{x})\right) \\ &= \sum_{l=1}^{L} p_l \log\left(\frac{p_l}{q_l}\right), \end{aligned} \tag{3}$$

$D(\boldsymbol{x}\|\boldsymbol{y})$ in (3) is also known as the Kullback–Leibler information measure, directed divergence, or cross-entropy [4]. By means of (3) we can define a symmetric hyperspectral measure, referred to as Spectral Information Divergence (SID) by

$$\mathrm{SID}(\boldsymbol{x}, \boldsymbol{y}) = D(\boldsymbol{x}\|\boldsymbol{y}) + D(\boldsymbol{y}\|\boldsymbol{x}) \tag{4}$$

that can be used to measure the spectral similarity between two pixels $\boldsymbol{x}$ and $\boldsymbol{y}$. SID offers a new look at spectral similarity by making use of relative entropy to account for the spectral information provided by each pixel.

IV. Criteria for Spectral Discrimination

The SIM, described in Section II, is a single-pixel measure, and SID in Section III is a criterion to measure the similarity between two pixels. When there are more than two pixels, discriminating one from another becomes more interesting. In what follows, three criteria are proposed for spectral discrimination.

A. Spectral Discriminatory Probability

In many applications we are required to identify a target pixel extracted from an unknown image scene using an existing spectral library or database Δ. In this case, it is of interest to know what is the likelihood of the pixel in question being identified as one of the signature spectra in Δ. Let $\{\boldsymbol{s}_j\}_{j=1}^{J}$ be J spectral signatures in Δ and $\boldsymbol{t}$ be a target pixel to be identified using Δ. Let $m(\cdot, \cdot)$ be any given hyperspectral measure. We define the spectral discriminatory probabilities of all $\boldsymbol{s}_j$'s in Δ with respect to $\boldsymbol{t}$ as follows:

$$p^m_{\boldsymbol{t},\Delta}(i) = \frac{m(\boldsymbol{t}, \boldsymbol{s}_i)}{\sum_{j=l}^{J} m(\boldsymbol{t}, \boldsymbol{s}_j)}, \qquad \text{for } i = 1, 2, \cdots, J \tag{5}$$

where $\sum_{j=1}^{J} m(\boldsymbol{t}, \boldsymbol{s}_j)$ is a normalization constant determined by $\boldsymbol{t}$ and Δ. The resulting probability vector

$$\boldsymbol{p}^m_{\boldsymbol{t},\Delta} = \left(p^m_{\boldsymbol{t},\Delta}(1), p^m_{\boldsymbol{t},\Delta}(2), \cdots, p^m_{\boldsymbol{t},\Delta}(J)\right)^T$$

TABLE I
FOUR MOMENTS GENERATED BY SIM FOR FIVE SIGNATURES IN FIG. 1

	μ_1	μ_2	μ_3	μ_4
Blackbrush	0.2670	0.0786	0.0245)	0.0079
Creosote	0.4091	0.1941	0.0982	0.0512
Drygrass	0.5310	0.2905	0.1628	0.0929
Redsoil	0.4202	0.1828	0.0812	0.0366
Sagebrush	0.3979	0.1774	0.0846	0.0419

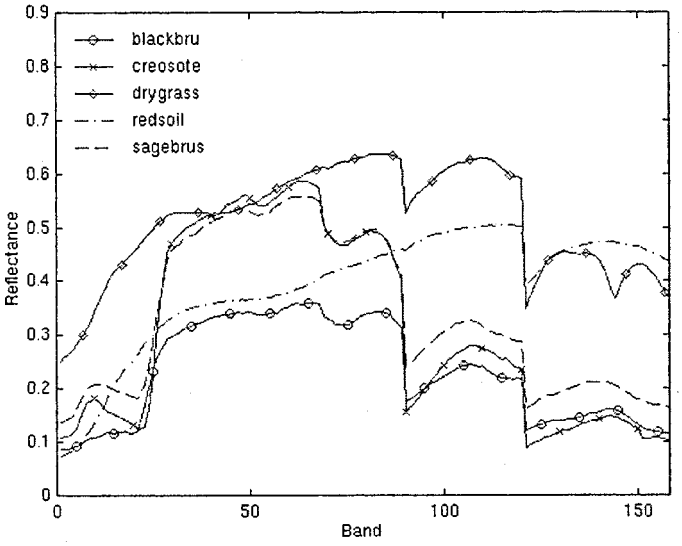


Fig. 1. Spectral reflectances of AVIRIS five signatures.

is called the spectral discriminatory probability vector of Δ with respect to $\boldsymbol{t}$. Using (5), the target pixel $\boldsymbol{t}$ can be identified by selecting the one with the smallest spectral discriminatory probability because $\boldsymbol{t}$ and the selected one have the minimum spectral discrimination.

B. Spectral Discriminatory Power

If we are given two spectral similarity measures, how do we evaluate which of the two is more effective? To meet this need, a criterion, spectral discriminatory power, is developed. It is designed based on the power of discriminating one pixel from another relative to a reference pixel $\boldsymbol{d}$. Assume that $m(\cdot, \cdot)$ is any given hyperspectral measure. Let $\boldsymbol{d}$ be the spectral signature of a reference pixel and $\boldsymbol{s}_i$, $\boldsymbol{s}_j$ be the spectral signatures of two pixels. We define spectral discriminatory power of $m(\cdot, \cdot)$ by

$$\mathrm{PW}^m(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{d}) = \max\left\{\frac{m(\boldsymbol{s}_i, \boldsymbol{d})}{m(\boldsymbol{s}_j, \boldsymbol{d})}, \frac{m(\boldsymbol{s}_j, \boldsymbol{d})}{m(\boldsymbol{s}_i, \boldsymbol{d})}\right\}. \tag{6}$$

The $\mathrm{PW}^m(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{d})$ defined by (6) provides an index of spectral discrimination capability of a specific hyperspectral measure $m(\cdot, \cdot)$ between any two spectral signatures $\boldsymbol{s}_i$, $\boldsymbol{s}_j$ relative to $\boldsymbol{d}$. Obviously, the higher the $\mathrm{PW}^m(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{d})$ is, the better discriminatory power $m(\cdot, \cdot)$ has. In addition, $\mathrm{PW}^m(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{d})$ is symmetric and bounded below by one, i.e., $\mathrm{PW}^m(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{d}) \geq 1$ with equality if and only if $\boldsymbol{s}_i = \boldsymbol{s}_j$.

C. Spectral Discriminatory Entropy

Since

$$\boldsymbol{p}^m_{\boldsymbol{t},\Delta} = (p^m_{\boldsymbol{t},\Delta}(1), p^m_{\boldsymbol{t},\Delta}(2), \cdots, p^m_{\boldsymbol{t},\Delta}(J))^T$$

given by (5) is the spectral discriminatory probability vector of $\boldsymbol{t}$ using a spectral library Δ, we can further define the spectral discriminatory entropy of Δ with respect to $\boldsymbol{t}$ by

$$H^m(\boldsymbol{t}; \Delta) = -\sum_{j=1}^{J} p^m_{\boldsymbol{t},\Delta}(j) \log p^m_{\boldsymbol{t},\Delta}(j). \tag{7}$$

Equation (7) provides the uncertainty measure of identifying $\boldsymbol{t}$ using the spectral signatures in Δ. A smaller $H^m(\boldsymbol{t}; \Delta)$ indicates a better chance to identify $\boldsymbol{t}$.

V. EXPERIMENTS USING HYPERSPECTRAL DATA

In order to evaluate the proposed SIM-based criteria, we compare it to two commonly used spectral similarity measures, Euclidean distance defined by

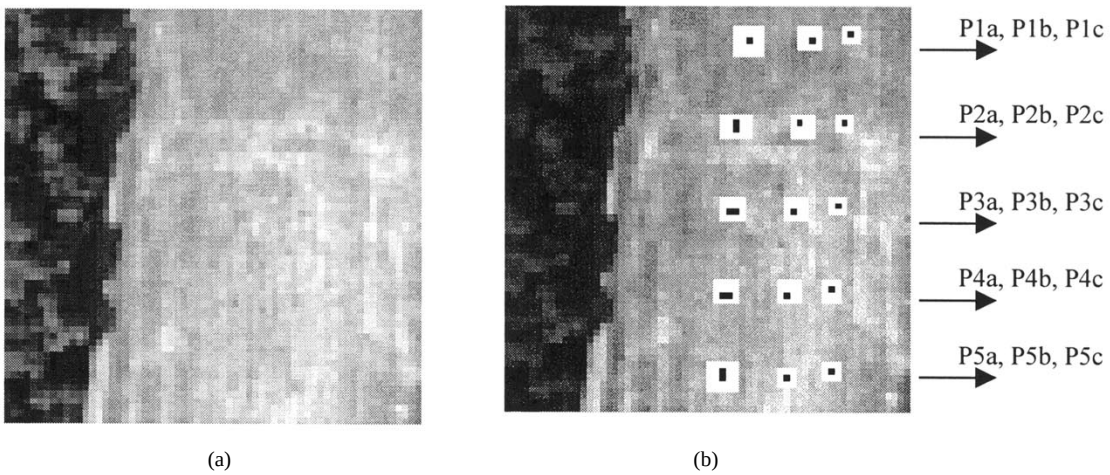
$$\mathrm{ED}(\boldsymbol{s}_i, \boldsymbol{s}_j) = \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \equiv \left[\sum_{l=1}^{L} (s_{il} - s_{jl})^2\right]^{1/2} \tag{8}$$

and SAM defined by

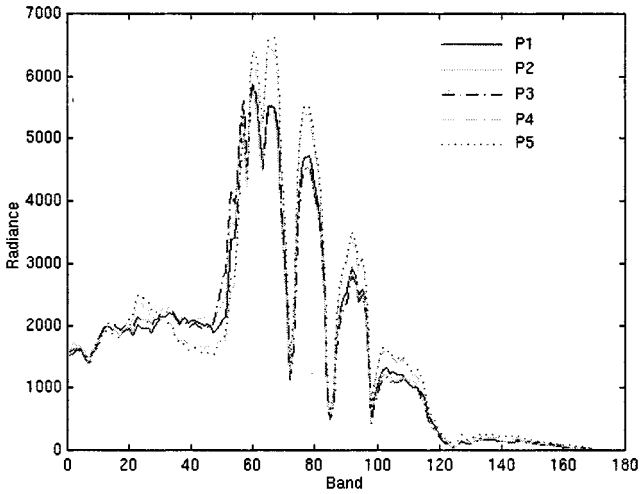
$$\mathrm{SAM}(\boldsymbol{s}_i, \boldsymbol{s}_j) = \cos^{-1}\left(\frac{\langle \boldsymbol{s}_i, \boldsymbol{s}_j \rangle}{\|\boldsymbol{s}_i\| \ \|\boldsymbol{s}_j\|}\right) \text{ (in radians)} \tag{9}$$

where $\langle \boldsymbol{s}_i, \boldsymbol{s}_j \rangle$ is the inner product of $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$. It is worth noting that if both $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ are unit vectors, it is easy to show that the relationship between Euclidean distance and SAM can be established by $\mathrm{ED}(\boldsymbol{s}_i, \boldsymbol{s}_j) = 2\sin(\mathrm{SAM}(\boldsymbol{s}_i, \boldsymbol{s}_j)/2)$. In particular, when $\mathrm{SAM}(\boldsymbol{s}_i, \boldsymbol{s}_j)$ is small, $2\sin(\mathrm{SAM}(\boldsymbol{s}_i, \boldsymbol{s}_j)/2) \approx \mathrm{SAM}(\boldsymbol{s}_i, \boldsymbol{s}_j)$. In this case, the $\mathrm{SAM}(\boldsymbol{s}_i, \boldsymbol{s}_j)$ is nearly the same as $\mathrm{ED}(\boldsymbol{s}_i, \boldsymbol{s}_j)$. Because of that, only SAM is used in the following experiments for comparative study. Two examples are considered.

Example 1—Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Reflectance Data: The data used in the following example are the same AVIRIS reflectance data used in [6] and [7]. They were five field reflectance spectra, blackbrush (indicated by open circle), creosote leaves (indicated by asterisk), dry grass (indicated by diamond), red soil (indicated by dash-dot), and sagebrush (indicated by dash) shown in Fig. 1, with spectral coverage from 0.4 μm to 2.5 μm. There were 158 bands after the water bands were removed, and all spectra were normalized to unity. For spectral variability, Table I tabulates the statistics generated by SIM for the five signatures up to four moments where μ_i is the ith moment. From Fig. 1 we can see that both creosote leaves and sagebrush have very similar spectra. This fact is also reflected in Table I where their four moments generated by SIM are very close. For spectral similarity, Table II tabulates the spectral similarity values among these five signatures where the values in the upper triangle above the diagonal line and the lower triangle below the diagonal line were produced by SAM and SID, respectively. The smaller the value for two signatures is, the more similar the two signatures are. From the table, sagebrush is closer to creosote leaves than to blackbrush. The similarity value produced by SAM for blackbrush and sagebrush was about twice as much as for



(a) (b)



(c)

Fig. 2. (a) A HYDICE image scene which contain 15 panels. (b) Spatial locations of 15 panels in Fig. 2(a) provided by ground truth. (c) Spectra of five panel signatures, P1, P2, P3, P4, and P5 generated from Fig. 2(b).

TABLE II
SIMILARITY VALUES AMONG FIVE SIGNATURES USING SAM AND SID

	Blackbrush	Creosote	Drygrass	Redsoil	Sagebrush
Blackbrush	0	0.1767 (SAM)	0.2575 (SAM)	0.4058 (SAM)	0.0681 (SAM)
Creosote	0.0497 (SID)	0	0.4213 (SAM)	0.5714 (SAM)	0.1289 (SAM)
Drygrass	0.0766 (SID)	0.2298 (SID)	0	0.2179 (SAM)	0.2968 (SAM)
Redsoil	0.1861 (SID)	0.4154 (SID)	0.0640 (SID)	0	0.4515 (SAM)
Sagebrush	0.0063 (SID)	0.0303 (SID)	0.0973 (SID)	0.2340 (SID)	0

creosote leaves and sagebrush. The similarity value produced by SID for blackbrush and sagebrush was about five times that for creosote leaves and sagebrush.

In order to see which measure is more effective, we chose blackbrush as the reference signature $\boldsymbol{d}$, creosote leaves as $\boldsymbol{s}_i$, and sagebrush as $\boldsymbol{s}_j$ since all three are similar. We computed their spectral discriminatory powers as follows:

$$\mathrm{PW}^{\mathrm{SID}}(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{d}) = \frac{\mathrm{SID}(\boldsymbol{s}_j, \boldsymbol{d})}{\mathrm{SID}(\boldsymbol{s}_i, \boldsymbol{d})} = \frac{0.0497}{0.0063} \approx 7.9 \tag{10}$$

$$\mathrm{PW}^{\mathrm{SAM}}(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{d}) = \frac{\mathrm{SAM}(\boldsymbol{s}_j, \boldsymbol{d})}{\mathrm{SAM}(\boldsymbol{s}_i, \boldsymbol{d})} = \frac{0.1767}{0.0681} \approx 2.6. \tag{11}$$

Equation (10) shows that the spectral discriminatory power of SID to distinguish blackbrush from creosote leaves is eight times better than to distinguish blackbrush from sagebrush. Compared to SAM, SID was three times more effective. In order to evaluate which measure is more effective in terms of spectral discrimination, a mixed spectral signature is randomly generated for use as a target signature $\boldsymbol{t}$. It is composed of 0.1055 blackbrush, 0.0292 creosote leaves, 0.0272 dry grass, 0.7588

TABLE III
SPECTRAL DISCRIMINATORY PROBABILITY VECTORS PRODUCED BY SAM AND SID WITH $\boldsymbol{t}$ CHOSEN TO BE A MIXTURE OF 0.1055 BLACKBRUSH, 0.0292 CREOSOTE LEAVES, 0.0272 DRY GRASS, 0.7588 RED SOIL, AND 0.0974 SAGEBRUSH

	Blackbrush	Creosote	Drygrass	Redsoil	Sagebrush	entropy
SAM	0.2212	0.3430	0.1044	0.0769	0.2546	1.1339
SID	0.1897	0.4933	0.0588	0.0112	0.2500	0.8843

TABLE IV
SIMILARITY VALUES AMONG FIVE PANEL SIGNATURES USING SAM AND SID

	P1	P2	P3	P4	P5
P1	0	0.0435 (SAM)	0.0673 (SAM)	0.1144 (SAM)	0.1240 (SAM)
P2	0.0039 (SID)	0	0.0430 (SAM)	0.1479 (SAM)	0.1567 (SAM)
P3	0.0086 (SID)	0.0033 (SID)	0	0.1652 (SAM)	0.1710 (SAM)
P4	0.0233 (SID)	0.0385 (SID)	0.0476 (SID)	0	0.0248 (SAM)
P5	0.0313 (SID)	0.0484 (SID)	0.0570 (SID)	0.0025 (SID)	0

TABLE V
SPECTRAL DISCRIMINATORY PROBABILITIES RESULTING FROM SAM AND SID

	P1	P2	P3	P4	P5	entropy
SAM	0.2520	0.3121	0.3366	0.0672	0.0322	1.3693
SID	0.2315	0.3421	0.3929	0.0287	0.0048	1.2002

red soil, and 0.0974 sagebrush. The database Δ was made up of the five signatures. Table III tabulates the spectral discriminatory probability vectors of SAM and SID. As we can see in Table I, the spectrum of red soil is very similar to that of dry grass. The ratio of SAM between $\boldsymbol{t}$ and red soil to SAM between $\boldsymbol{t}$ and dry grass was $0.1044 : 0.0769 \approx 1.36$. In comparison, SID yielded $0.0588 : 0.0112 \approx 5.25$, which was nearly four times more effective in identifying the $\boldsymbol{t}$ as red soil than SAM. Finally, Table III also shows that the spectral discriminatory entropy of SID produced the smaller entropy, 0.8843, as opposed to 1.1339 produced by SAM.

Example 2—HYperspectral Digital Imagery Collection Experiment (HYDICE) Data: Unlike the AVIRIS data studied in Example 1, the HYDICE data used in the following experiments were directly extracted from the HYDICE image scene of size 64×64 shown in Fig. 2(a). There are 15 panels located on the field and arranged in a 5×3 matrix. The low signal/high noise bands: bands 1–3 and bands 202–210; and water vapor absorption bands: bands 101–112 and bands 137–153, were removed. The spatial resolution is 1.5 m and spectral resolution is 10 nm. The ground truth of the image scene as shown in Fig. 2(b) provides the precise spatial locations of these 15 panels. Black pixels are panel center pixels, considered to be pure pixels, and the pixels in the white masks are panel boundary pixels mixed with background pixels, considered to be mixed pixels. Each element in this matrix is a square panel and denoted by p_{ij} with row indexed by $i = 1, \cdots, 5$ and column indexed by $j = a, b, c$. For each row $i = 1, \cdots, 5$, the three panels p_{ia}, p_{ib}, p_{ic} were made from the same material but have three different sizes. For each column $j = a, b, c$, the five panels p_{1j}, p_{2j}, p_{3j}, p_{4j}, p_{5j} have the same size but were made from five different materials. The sizes of the panels in the first, second, and third columns are 3 m $\times$ 3 m, 2 m $\times$ 2 m, and 1 m $\times$ 1 m, respectively. Five panel signatures were generated by averaging the black center pixels in each row, denoted by P1, P2, P3, P4, and P5. Their spectra are shown in Fig. 2(c). Table IV tabulates the spectral similarity values resulting from SAM and SID in the same manner as Table I. SID performed more effectively than SAM in terms of spectral discriminatory power. In order to see spectral discriminatory probabilities, the set $\Delta = \{\mathrm{P1}, \mathrm{P2}, \mathrm{P3}, \mathrm{P4}, \mathrm{P5}\}$ was used for the database, and a target pixel $\boldsymbol{t}$, was randomly extracted from the white mixed pixels of p_{5a} for identification. Table V shows the spectral discriminatory probabilities of Δ with respective to $\boldsymbol{t}$ using SAM and SID, where $p^{\mathrm{SAM}}_{\boldsymbol{t},\Delta}(\mathrm{P4}) : p^{\mathrm{SAM}}_{\boldsymbol{t},\Delta}(\mathrm{P5}) \approx 2.09$ and $p^{\mathrm{SID}}_{\boldsymbol{t},\Delta}(\mathrm{P4}) : p^{\mathrm{SID}}_{\boldsymbol{t},\Delta}(\mathrm{P5}) \approx 5.98$. SID was about three times more effective than SAM in identifying $\boldsymbol{t}$ as P5. Table V shows their respective spectral discriminatory entropies, where SID yielded the smallest entropy.

VI. CONCLUSION

In this correspondence, we present an information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. The Spectral Information Measure (SIM) is introduced to model a hyperspectral image pixel as a random variable to capture spectral variability in the pixel. Then the Spectral Information Divergence (SID) was further developed to measure the spectral similarity between two pixels. In order to compare two hyperspectral measures, a new definition of spectral discriminatory power was suggested for performance evaluation. Finally, for the purpose of material identification, the concept of spectral discriminatory probability vector was also developed based on a spectral database or library. The experiments demonstrate that the SIM-based criteria performed more effectively than the traditional spectral measure SAM, widely used in hyperspectral image analysis.

ACKNOWLEDGMENT

The author wishes to thank Dr. Harsanyi for providing the AVIRIS data, Dr. Q. Du and S.-S. Chiang for generating figures and tables, and Dr. M. L. G. Althouse for proofreading this paper.

REFERENCES

[1] G. Vane and A. F. H. Goetz, "Terrestrial imaging spectroscopy," *Remote Sensing Environ.*, vol. 24, pp. 1–29, 1988.
[2] C.-I Chang, T.-L. E. Sun, and M. L. G. Althouse, "An unsupervised interference rejection approach to target detection and classification for hyperspectral imagery," *Opt. Eng.*, vol. 37, no. 3, pp. 735–743, Mar. 1998.
[3] R. M. Fano, *Transmission of Information: A Statistical Theory of Communication*. New York: Wiley, 1961.
[4] S. Kullback, *Information Theory and Statistics*. New York: Wiley, 1959.
[5] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, 2nd ed. San Diego, CA: Academic, 1997.
[6] J. Harsanyi and C.-I Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 779–785, July 1994.
[7] C.-I Chang, X. Zhao, M. L. G. Althouse, and J. J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 898–912, May 1998.

Generalized Matched Filters and Univariate Neyman–Pearson Detectors for Image Target Detection

Robert S. Caprari

***Abstract*—I derive two-stage, statistically suboptimal target detectors for images. The first, or transformation, stage is a "generalized matched filter" (GMF) that linearly transforms the input image. I propose three rational signal-to-noise-ratio criteria whose maximization yields the three GMFs. The second, or detection, stage is a univariate "Neyman–Pearson detector" (NPD), which executes a pointwise likelihood ratio test on GMF transformed images. Experiments on infrared and synthetic-aperture radar imagery compare GMF/NPDs with several established detectors.**

***Index Terms*—Image processing, linear discriminators, pattern recognition, statistical signal detection.**

I. INTRODUCTION

INVESTIGATE statistical target detection in images by methods composed of two suboptimal stages—"transformation" followed by "detection." Transformation is accomplished by operating on the input image with a linear filter that is designed to maximize one of three rational "signal-to-noise-ratio" (SNR) functions. Detection is accomplished by operating on the transformed image with the statistically optimal univariate, or point, detector. This detector correctly accounts for the empirical target and clutter statistical distributions, rather than being based on idealized distribution (e.g., normal) assumptions. The theoretical development is supported by experimental results on infrared (IR) and synthetic-aperture radar (SAR) images, including effectiveness and complexity comparisons with several established image target detectors.

A note on the terminology that is used here is worthwhile. I distinguish between three grades of detector optimality: optimal detectors maximize, without constraint, a direct measure of detector effectiveness (e.g., detection probability corresponding to prescribed false-alarm probability); suboptimal detectors also maximize some objective function, but either they only achieve a constrained maximization, or the objective function is only an indirect and imperfect measure of true detector effectiveness; and nonoptimal detectors do not maximize any objective function whatsoever. The transformation filters proposed here are suboptimal in two respects: they are constrained to be linear; and they maximize SNR functions that have a strong, but nevertheless only indirect, influence on detector effectiveness. The detectors proposed here are suboptimal in the sense that they are constrained to be univariate rather than free to be multivariate. The classification scheme used here is not universal, with authors often omitting the suboptimal category, and only distinguishing between optimal and nonoptimal categories. In any credible two-category scheme, both the filters and detectors proposed here would emphatically fit in the optimal category—in no sense are these filters and detectors nonoptimal. Although truly optimal target detectors are theoretically known, for realistic image statistics and manageable image ensemble sizes evaluation of these optimal detectors is an intractable problem, which leaves suboptimal detectors as the most objective of tractable detectors.

II. IMAGE TARGET DETECTION

The target detector will be designed to distinguish between the two hypotheses

Clutter Hypothesis: the image contains pure clutter;

Target Hypothesis: the image contains a correctly centered target embedded in clutter.

In essence, the target detector is required to distinguish members of a *pure-clutter* random field $P(i)$ (clutter hypothesis) from members of a *target-in-clutter* random field $Q(i)$ (target hypothesis). $P(i)$, with mean vector $\boldsymbol{\mu}_P$ and covariance matrix C_P, is assumed to be strict-sense stationary, which is reasonable for distortionless narrow field imaging, although not necessarily so for wide-field imaging, with its attendant distortion aberration [1]. The strict-sense stationarity of $P(i)$ ensures that a single target detector designed according to the prescription of Section IV will be suboptimal at every point in the image. If $P(i)$ were only wide-sense stationary, in general different points would have distinct suboptimal detectors, because the detector at point i depends on the probability distribution of $P(i)$ in a neighborhood of point i. However, wide-sense stationarity of $P(i)$ is sufficient to render a single transformation filter designed according to one of the prescriptions of Section III suboptimal at every point in the image. $Q(i)$ is a nonstationary random field with mean vector $\boldsymbol{\mu}_Q$ and covariance matrix C_Q.

III. GENERALIZED MATCHED FILTERS

The first stage of the target detector being proposed here is to transform the input image by a *generalized matched filter* (GMF), which is a linear filter whose filter function $\boldsymbol{f}$ (i.e., impulse response or point-

Manuscript received August 2, 1999; revised March 17, 2000.
The author is with the Defence Science and Technology Organisation, Salisbury SA 5108, Australia (e-mail: robert.caprari@dsto.defence.gov.au).
Communicated by J. M. F. Moura, Guest Editor.
Publisher Item Identifier S 0018-9448(00)05776-X.