2

HYPERSPECTRAL MEASURES FOR SPECTRAL CHARACTERIZATION

A hyperspectral image can be considered as an image cube where the third dimension is represented by hundreds of contiguous spectral bands. As a result, a hyperspectral pixel is actually a column vector with dimensions equal to the number of spectral bands. Such between-band spectral information is very useful and can be used for spectral characterization. Many measures proposed in signal processing and pattern recognition can be used for this purpose. Nevertheless, most of them are spatial-based measures, and they are not particularly designed to measure spectral properties inherent in a single pixel vector. In this chapter, two new hyperspectral measures, spectral information measure (SIM) (Chang, 2000) and hidden Markov model (HMM)-based spectral measure (Du and Chang, 2001) are presented, both of which are derived from the Kullback-Leibler information distance to capture the spectral variability of a pixel vector. Additionally, spectral information divergence (SID), relative spectral discriminatory probability, relative spectral discriminatory power and relative spectral discriminatory entropy are also introduced to further account for spectral similarity and discrminability among pixel vectors.

2.1 MEASURES OF SPERCTRAL VARIABILITY

A hyperspectral image is generally acquired by hundreds of spectral channels. As a result, a scene pixel vector is usually represented by a column vector, in which each component contains specific spectral information provided by a particular channel. Therefore, a greater number of spectral channels translate more spectral information. This implies that a hyperspectral image pixel vector generally contains more spectral information than does a multispectral image pixel vector. In many situations, such spectral information is valuable and crucial in data analysis. In order to capture and characterize the spectral properties vector provided in a single pixel vector by hundreds bands, two statistical spectral measures are introduced in this section.

2.1.1 Spectral Information Measure (SIM)

The spectral information measure (SIM) is an information theoretic measure that models the spectral band-to-band variability as uncertainty resulting from randomness. It considers each pixel vector as a random variable with the probability distribution obtained by normalizing its spectral histogram to unity. With this interpretation, SIM can measure the spectral variability of a single hyperspectral pixel vector resulting from band-to-band correlation. It not only can describe the randomness of a pixel vector, but also can generate high-order statistics of each pixel vector based on its spectral histogram. So, SIM can be considered as a single pixel vector-based stochastic measure. It is particularly useful for hyperspectral image analysis. This is because the spectral information provided by each hyperspectral image pixel vector can improve material detection, discrimination, classification and identification. However, this advantage is also traded for extraction of many unknown interferers, which can be identified *a priori* (Chang et al., 1998). The effects caused by such interference can be only described by randomness but cannot be characterized deterministically. SIM is designed to meet this need. It can capture the uncertainty created by unknown signal sources in a stochastic manner. Therefore, the higher the data dimensionality, the more the randomness, which enhances the relative effectiveness of SIM. Because SIM is a statistical measure, it can generate higher order statistics that can further be used to characterize spectral variability such as variance (second order statistic used to measure for standard deviation), skewness (third order statistic used to measure for symmetry), kurtosis (fourth order statistic used to measure flatness). This advantage cannot be achieved by any deterministic measure such as Euclidean distance or spectral angle mapper (SAM) (Schowengerdt, 1997).

For a given hyperspectral pixel vector $\mathbf{x} = (x_1, \cdots, x_L)^T$, each component x_l represents a pixel in band image B_l which is acquired by a certain wavelength ω_l in a specific spectral range. Let $\mathbf{s} = (s_1, \cdots, s_L)^T$ be the corresponding spectral signature (i.e., spectrum) of $\mathbf{x}$ where s_l represents its spectral signature of x_l in the form of either radiance or reflectance values. Suppose that $\{\omega_l\}_{l=1}^{L}$ is a set of L wavelengths, each of which corresponds to a spectral band channel. Then $\mathbf{x}$ 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 this case, we let $\Omega = \{\omega_1, \omega_2, \cdots, \omega_{L-1}, \omega_L\}$ be the sample space and Σ be the power set of Ω, i.e., the set of all subsets of Ω and $\mathbf{x}(\omega_l) = s_l$. In order to define a legitimate probability measure P for $\mathbf{x}$, we first assume that all components s_l's associated with $\mathbf{x}$ are nonnegative. This is generally a valid assumption due to the nature of radiance or reflectance. With this assumption, we can normalize x_l's to the range of $[0, 1]$ as follows,

$$p_l = s_l / \sum_{k=1}^{L} s_k . \tag{2.1}$$

Using (2.1) we define a probability measure P for $\mathbf{x}$ by

$$P(\{\omega_l\}) = p_l . \tag{2.2}$$

The probability vector $\boldsymbol{p} = (p_1, p_2, \cdots, p_L)^T$ is the desired probability distribution of pixel vector $\mathbf{x}$. By means of this probability interpretation, any pixel vector $\mathbf{x} = (x_1, \cdots, x_L)^T$ can be viewed as a single information source with its statistics governed by $\boldsymbol{p} = (p_1, p_2, \cdots, p_L)^T$ defined by (2.1) and (2.2). As a result, these statistics can be used to describe the spectral variability of a pixel vector. For instance, we can define its statistics of different orders, such as mean $\mu(\mathbf{x}) = \sum_{l=1}^{L} p_l s_l$, variance $\sigma^2(\mathbf{x}) = \sum_{l=1}^{L} p_l \left(s_l - \mu(\mathbf{x})\right)^2$, third central moment $\kappa_3(\mathbf{x}) = \sum_{l=1}^{L} p_l \left(s_l - \mu(\mathbf{x})\right)^3$, fourth central moment, $\kappa_4(\mathbf{x}) = \sum_{l=1}^{L} p_l \left(s_l - \mu(\mathbf{x})\right)^4$, etc. Like 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 vector where all moments can be obtained via (2.1) and (2.2).

Since a hyperspectral image pixel vector $\mathbf{x}$ can be considered as an information source described by (2.1) and (2.2), from information theory (Fano, 1960) we can further define its self-information provided by band l by

$$I_l(x) = -\log p_l. \tag{2.3}$$

Using (2.3) the entropy of each hyperspectral image pixel vector $\mathbf{x}$, $H(\mathbf{x})$ can be obtained as

$$H(\mathbf{x}) = -\sum_{l=1}^{L} p_l \log p_l \tag{2.4}$$

which can be used to describe the uncertainty resulting from the pixel vector $\mathbf{x}$.

2.1.2 Hidden Markov Model (HMM)-Based Measure

Another statistical measure makes use of a hidden Markov model (HMM) to capture the unobserved and hidden spectral properties of a hyperspectral image pixel vector. HMM has been widely used in speech recognition (Rabiner and Juang, 1993) to model a speech signal as a doubly stochastic process with a hidden state process that can be only observed through a sequence of observations. Since the temporal variability of a speech signal is similar to the spectral variability of a hyperspectral image pixel vector, a similar idea can be applied to a hyperspectral spectral vector. In this case, a hidden Markov process is used to characterize spectral correlation as well as band-to-band variability with model parameters determined by the spectrum of the pixel vector that forms an observation sequence. In speech processing, the same word spoken in different times generally results in different speech spectra. This is also true for the case that the spectrum of the same material taken at different times varies.

In analogy with SID, we introduce a new HMM-based spectral measure, referred to as HMM information divergence (HMMID) that can be also derived from self-information specified by (2.3) and (2.4) (Du and Chang, 2001).

Let $\mathbf{o} = \left(o_1, o_2, \cdots, o_T\right)$ be an observation process with o_t being the observation taken place at time t and T is the number of observations made in the process. Assume

that there are N states denoted by $\{1,2,\cdots,N\}$ and the state at time t is denoted by q_t. Let $\mathbf{A} = \left\{a_{ij}\right\}_{1\le i,j\le N}$ be the state transition matrix with a_{ij} given by

$$a_{ij} = P(q_{t+1} = j | q_t = i) \tag{2.5}$$

and $\mathbf{B} = \left[b_j(o_t)\right]_{1\le j\le N, 1\le t\le T}$ be the observation probability density matrix where $b_j(o_t)$ is the probability density of the observation o_t at time t in the j-th state. We further assume that $\pi = \left(\pi_1, \pi_2, \cdots, \pi_N\right)$ is the initial state distribution with π_j given by

$$\pi_j = P(q_1 = j). \tag{2.6}$$

So, an HMM can be uniquely defined by a parameter triplet, denoted by $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ which can be estimated by the Baum-Welch algorithm using the maximum likelihood estimation as follows (Rabiner and Juang, 1993).

Assume that the probability density $b_j(o_t)$ is a Gaussian mixture. It has been shown in Rabiner and Juang (1986) that it is equivalent to a multistage single Gaussian density given by

$$b_j(o_t) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left[-\frac{\left(o_t - \mu_j\right)^2}{2\sigma_j^2}\right] \tag{2.7}$$

where μ_j and σ_j^2 are the mean and variance of the observation o_t in the j-th state respectively.

Now, we define the forward probability $\alpha_j(t)$ as the joint probability of observing the first t observations $o_1, o_2, \cdots, o_t$ in the j-th state at time t. It can be solved by

$$\alpha_j(1) = \pi_j b_j(o_1) \tag{2.8}$$

$$\alpha_j(t) = \left[\sum_{i=1}^{N} \alpha_i(t-1) a_{ij}\right] b_j(o_t) \text{ for } 1 < t \le T. \tag{2.9}$$

Similarly, we define the backward probability $\beta_j(t)$ as the conditional probability of observing the observations $o_{t+1}, o_{t+2}, \cdots, o_T$ given that the state at time t is j. It can be solved by

$$\beta_j(T) = 1 \tag{2.10}$$

$$\beta_j(t) = \sum_{i=1}^{N} a_{ji} b_i(o_{t+1}) \beta_i(t+1) \text{ for } 1 < t \le T. \tag{2.11}$$

Let $\gamma_j(t)$ denote the conditional probability of $b_j(o_t)$ given that the observation $\mathbf{o} = \left(o_1, o_2, \cdots, o_T\right)$ and $\xi_{ij}(t)$ denote the conditional probability of a transition from state i to state j at time t+1 given the observation $\mathbf{o} = \left(o_1, o_2, \cdots, o_T\right)$. Then they can be calculated as follows

$$\gamma_j(t) = \frac{\alpha_j(t)\beta_j(t)}{\sum_{i=1}^{N}\alpha_i(t)\beta_i(t)} \tag{2.12}$$

$$\xi_{ij}(t) = \frac{\alpha_i(t)a_{ij}b_j(o_{t+1})\beta_j(t+1)}{\sum_{k=1}^{N}\sum_{l=1}^{N}\alpha_k(t)a_{kl}b_l(o_{t+1})\beta_l(t+1)} \ . \tag{2.13}$$

Using (8) and (9) the μ_j, σ_j^2, a_{ij} and π_j can be estimated by the following equations

$$\hat{\mu}_j(t) = \frac{\sum_{t=1}^{T}\left[\gamma_j(t)o_t\right]}{\sum_{t=1}^{T}\gamma_j(t)} \tag{2.14}$$

$$\hat{\sigma}_j^2(t) = \frac{\sum_{t=1}^{T}\left[\gamma_j(t)\left(o_t - \hat{\mu}_j(t)\right)^2\right]}{\sum_{t=1}^{T}\gamma_j(t)} \tag{2.15}$$

$$\hat{a}_{ij}(t) = \frac{\sum_{t=1}^{T-1}\xi_{ij}(t)}{\sum_{t=1}^{T-1}\gamma_i(t)} \tag{2.16}$$

$$\hat{\pi}_j = \gamma_j(1) \ . \tag{2.17}$$

As indicated previously, an endmember may be represented by variants of its true spectral signature from pixel vector to pixel vector because of unpredicted mixing occurring in a pixel vector. However, there must have some unobserved properties governed by this particular endmember that can distinguish itself from other endmembers. This spectral characterization is very similar to speech signals where one of their key features is pitches. If we assume the observation sequence $\mathbf{o} = \left(o_1, o_2, \cdots, o_T\right)$ is represented by the spectral signature **s** of a pixel vector **x** in a hyperspectral image, we can use the HMM to capture the unobserved and hidden spectral properties of **s**. Let λ_s be the parameter vector used to specify **s** and HMM(λ_s) be the HMM determined by λ_s. In analogy with (2.3) we can define the self-information of **s** provided by HMM(λ_s), denoted by $I_{\mathrm{HMM}(\lambda_s)}(\mathbf{s})$ as follows

$$I_{\mathrm{HMM}(\lambda_s)}(\mathbf{s}) = -1 / T\left[\log P(\mathbf{s}|\lambda_s)\right]. \tag{2.18}$$

2.2 SPECTRAL SIMILARITY MEASURES

In the previous section, two new hyperspectral measures were suggested to capture spectral variations of a single pixel vector. In this section, we consider the issue of how to measure the similarity between two pixel vectors. In the past, many pure pixel vector-based similarity measures have been used to evaluate the similarity between two pixel vectors. However, they may not be effective for hyperspectral pixel vectors since they do not take full advantage of band-to-band spectral correlation. In order to make comparison, we describe several commonly used measures which can be used for spectral similarity. First, we assume that there are two pixel vectors $\mathbf{r}_i = (r_{i1}, \cdots, r_{iL})^T$ and $\mathbf{r}_j = (r_{j1}, \cdots, r_{jL})^T$ with their respective spectral signatures given by $\mathbf{s}_i = (s_{i1}, \cdots, s_{iL})^T$ and $\mathbf{s}_j = (s_{j1}, \cdots, s_{jL})^T$.

2.2.1 Commonly Used Measures

Three different types of measures are of interest in pattern classification and can be briefly described as follows.

2.2.1.1 Distance-Based Measures

In multivariate analysis, communications and signal processing, distance-based measures are most commonly used metrics to measure the distance between sample data points. Three metrics that calculate the distance between the spectral signatures of two pixel vectors, $\mathbf{s}_i$ and $\mathbf{s}_j$ can be derived from l_1, l_2, l_∞-norms in real analysis.

- City block distance (CBD) corresponding to l_1-norm

$$\text{CBD}(\mathbf{s}_i, \mathbf{s}_j) = \sum_{l=1}^{L} |s_{il} - s_{jl}| \ . \tag{2.19}$$

- Euclidean distance (ED) corresponding to l_2-norm

$$\text{ED}(\mathbf{s}_i, \mathbf{s}_j) = \left\| \mathbf{s}_i - \mathbf{s}_j \right\| \equiv \left[\sum_{l=1}^{L} (s_{il} - s_{jl})^2 \right]^{1/2} \ . \tag{2.20}$$

- Tchebyshev distance (TD) or maximum distance corresponding to l_∞-norm

$$\text{TD}(\mathbf{s}_i, \mathbf{s}_j) = \max_{1 \le l \le L} \left\{ \left| s_{il} - s_{jl} \right| \right\} \ . \tag{2.21}$$

2.2.1.2 Orthogonal Projection-Based Measures

Two measures can be derived from orthogonal projection, SAM and orthogonal projection divergence (OPD) (Ren and Chang, 1998).

SAM is a widely used spectral similarity metric in remote sensing. It measures spectral similarity by finding the angle between the spectral signatures of two pixel vectors, $\mathbf{s}_i$ and $\mathbf{s}_j$.

$$\begin{aligned} \mathrm{SAM}(\mathbf{s}_i, \mathbf{s}_j) &\equiv \cos^{-1}\left(\mathbf{s}_i \cdot \mathbf{s}_j \,/\, \|\mathbf{s}_i\| \|\mathbf{s}_j\|\right) \\ &= \cos^{-1}\left(\sum_{l=1}^{L} s_{il} s_{jl} \,/\, \left[\sum_{l=1}^{L} s_{il}^2\right]^{1/2} \left[\sum_{l=1}^{L} s_{jl}^2\right]^{1/2}\right). \end{aligned} \tag{2.22}$$

The concept of the OPD is originated from the orthogonal subspace projection (OSP) (Harsanyi and Chang, 1994). It finds the residuals of orthogonal projections resulting from two pixel vectors, $\mathbf{s}_i$ and $\mathbf{s}_j$ given by

$$\mathrm{OPD}(\mathbf{s}_i, \mathbf{s}_j) = \left(\mathbf{s}_i^T P_{\mathbf{s}_j}^{\perp} \mathbf{s}_i + \mathbf{s}_j^T P_{\mathbf{s}_j}^{\perp} \mathbf{s}_j\right)^{1/2} \tag{2.23}$$

where $P_{\mathbf{s}_k}^{\perp} = \mathbf{I}_{L\times L} - \mathbf{s}_k(\mathbf{s}_k^T \mathbf{s}_k)^{-1} \mathbf{s}_k^T$ for $k = i, j$ and $\mathbf{I}_{L\times L}$ is the $L \times L$ identity matrix.

It is worth noting that if both $\mathbf{s}_i$ and $\mathbf{s}_j$ are normalized to unity, the relationship between $\mathrm{ED}(\mathbf{s}_i, \mathbf{s}_j)$ and $\mathrm{SAM}(\mathbf{s}_i, \mathbf{s}_j)$ can be established as follows.

$$\begin{aligned} \mathrm{ED}(\mathbf{s}_i, \mathbf{s}_j) &= \sqrt{2 - 2\langle \mathbf{s}_i, \mathbf{s}_j \rangle} = \sqrt{2\left(1 - \cos\left(\mathrm{SAM}(\mathbf{s}_i, \mathbf{s}_j)\right)\right)} \\ &= 2\sqrt{1 - \cos\left(\mathrm{SAM}(\mathbf{s}_i, \mathbf{s}_j)\right) / 2} = 2\sin\left(\mathrm{SAM}(\mathbf{s}_i, \mathbf{s}_j) / 2\right) \end{aligned} \tag{2.24}$$

where $\langle \mathbf{s}_i, \mathbf{s}_j \rangle = \sum_{l=1}^{L} s_{il} s_{jl}$.

When $\mathrm{SAM}(\mathbf{s}_i, \mathbf{s}_j)$ is small, $2\sin\left(\mathrm{SAM}(\mathbf{s}_i, \mathbf{s}_j) / 2\right) \approx \mathrm{SAM}(\mathbf{s}_i, \mathbf{s}_j)$, in which case $\mathrm{SAM}(\mathbf{s}_i, \mathbf{s}_j)$ is nearly the same as $\mathrm{ED}(\mathbf{s}_i, \mathbf{s}_j)$. Fig. 2.1 shows the geometric interpretations of ED and SAM, and their relationship.

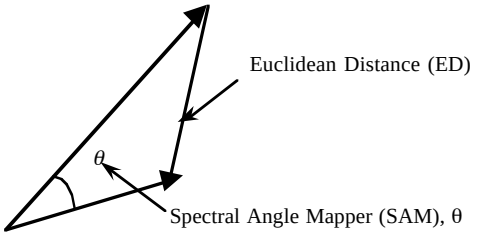

Figure 2.1. Geometric relationship between ED and SAM

2.2.2 Spectral Information Divergence (SID)

A new criterion to measure spectral similarity, called spectral information divergence (SID) is presented in this section (Chang, 2000). It originates from the concept of divergence in information theory and measures the discrepancy of probabilistic behaviors between the spectral signatures of two pixel vectors. In other words, the spectral similarity between two pixel vectors is measured by SID based on the discrepancy between their corresponding spectral signature-derived probability distributions. The idea

of using divergence is not new and has been also found in pattern recognition (Tou and Gonzalez, 1974) and band selection (Mausel et al., 1990, Conese and Maselli, 1993, Stearns et al. 1993, Jensen, 1996) such as Jeffries-Matusita measure. But what is new for SID is that it is designed from SIM for spectral similarity. Comparing to SAM and ED that extract geometric features, i.e. angle and spatial distance between two pixel vectors, SID measures the distance between the probability distributions produced by the spectral signatures of two pixel vectors. Accordingly, SID may be more effective than ED and SAM in capturing spectral variability.

Now, following (2.1) we calculate the probability vectors $\boldsymbol{p} = (p_1, p_2, \cdots, p_L)^T$ and $\boldsymbol{q} = (q_1, q_2, \cdots, q_L)^T$ for the spectral signatures of two pixel vectors, $\mathbf{s}_i$ and $\mathbf{s}_j$ where $p_k = s_{ik} / \sum_{l=1}^{L} s_{il}$ and $q_k = s_{jk} / \sum_{l=1}^{L} s_{jl}$. So, the self-information provided by $\mathbf{r}_j$ for band l is defined by (2.3) and given by

$$I_l(\mathbf{r}_j) = -\log q_l. \tag{2.25}$$

Using (2.3) and (2.25) we can further define $D_l(\mathbf{r}_i||\mathbf{r}_j)$(Cover and Thomas, 1993), the discrepancy in the self-information of band l in $\mathbf{r}_j$ relative to the self-information of band l in $\mathbf{r}_i$ by

$$\begin{aligned} D_l(\mathbf{r}_i||\mathbf{r}_j) &= I_l(\mathbf{r}_j) - I_l(\mathbf{r}_i) = \left(-\log q_l\right) - \left(-\log p_l\right) \\ &= \log\left(p_l / q_l\right) \end{aligned}. \tag{2.26}$$

Averaging $D_l(\mathbf{r}_i||\mathbf{r}_j)$ in (2.16) over all bands $1 \le l \le L$ with respect to $\mathbf{r}_i$ results in

$$\begin{aligned} D(\mathbf{r}_i||\mathbf{r}_j) &= \sum_{l=1}^{L} p_l D_l(\mathbf{r}_i||\mathbf{r}_j) = \sum_{l=1}^{L} p_l \left(I_l(\mathbf{r}_j) - I_l(\mathbf{r}_i)\right) \\ &= \sum_{l=1}^{L} p_l \log\left(p_l / q_l\right) \end{aligned} \tag{2.27}$$

where $D(\mathbf{r}_i||\mathbf{r}_j)$ is the average discrepancy in the self-information of $\mathbf{r}_j$ relative to the self-information of $\mathbf{r}_i$. In context of information theory $D(\mathbf{r}_i||\mathbf{r}_j)$ in (2.27) is called the relative entropy of $\mathbf{r}_j$ with respect to $\mathbf{r}_i$ which is also known as Kullback-Leibler information measure, directed divergence or cross entropy (Kullback, 1968). Similarly, we can also define the average discrepancy in the self-information of $\mathbf{r}_j$ relative to the self-information of $\mathbf{r}_i$ by

$$\begin{aligned} D(\mathbf{r}_j||\mathbf{r}_i) &= \sum_{l=1}^{L} q_l D_l(\mathbf{r}_j||\mathbf{r}_i) = \sum_{l=1}^{L} q_l \left(I_l(\mathbf{r}_i) - I_l(\mathbf{r}_j)\right) \\ &= \sum_{l=1}^{L} q_l \log\left(q_l / p_l\right) \end{aligned}. \tag{2.28}$$

Summing (2.27) and (2.28) yields spectral information divergence (SID) defined by

$$\mathrm{SID}(\mathbf{r}_i, \mathbf{r}_j) = D(\mathbf{r}_i || \mathbf{r}_j) + D(\mathbf{r}_j || \mathbf{r}_i), \tag{2.29}$$

which can be used to measure the spectral similarity between two pixel vectors $\mathbf{r}_i$ and $\mathbf{r}_j$. It should be noted that while $\mathrm{SID}(\mathbf{r}_i, \mathbf{r}_j)$ is symmetric, $D(\mathbf{r}_i || \mathbf{r}_j)$ is not. This is because $\mathrm{SID}(\mathbf{r}_i, \mathbf{r}_j) = \mathrm{SID}(\mathbf{r}_j, \mathbf{r}_i)$, $D(\mathbf{r}_i || \mathbf{r}_j) \neq D(\mathbf{r}_j || \mathbf{r}_i)$. Compared to other popular similarity measures, SID offers a new look of spectral similarity by taking advantage of relative entropy to account for the spectral information provided by each pixel vector.

A measure similar to SID, called Jeffries-Matusita measure that has been used for band selection (Jensen, 1996) can be also defined as Jeffries-Matusita distance (JMD) by

$$\mathrm{JMD}(\mathbf{r}_i, \mathbf{r}_j) = \sqrt{\sum_{l=1}^{L} \left[\sqrt{p_l} - \sqrt{q_l} \right]^2}. \tag{2.30}$$

2.2.3 Hidden Markov Model-Based Information Divergence (HMMID)

Following an approach similar to SID, we can also define an HMM-based information divergence for two pixel vectors. Suppose that a hyperspectral pixel vector $\mathbf{s}_k$ specified by $\mathrm{HMM}(\lambda_{\mathbf{s}_k})$ can be also modeled by another HMM(λ). In this case, the information discrepancy of $\mathbf{s}_k$ provided by $\mathrm{HMM}(\lambda_{\mathbf{s}_k})$ and HMM(λ) is the entropy of HMM(λ) relative to $\mathrm{HMM}(\lambda_{\mathbf{s}_k})$ and is given by

$$\begin{aligned} J_{\mathrm{HMM}}(\lambda_{\mathbf{s}_k}; \lambda) &= I_{\mathrm{HMM}(\lambda_{\mathbf{s}_k})}(\mathbf{s}_k) - I_{\mathrm{HMM}(\lambda)}(\mathbf{s}_k) \\ &= 1/T \left[\log P(\mathbf{s}_k | \lambda) - \log P(\mathbf{s}_k | \lambda_{\mathbf{s}_k}) \right]. \end{aligned} \tag{2.31}$$

Using (2.31) we can further define an HMM-based information distance between two hyperspectral pixel vectors $\mathbf{r}_i = (r_{i1}, \cdots, r_{iL})^T$ and $\mathbf{r}_j = (r_{j1}, \cdots, r_{jL})^T$. Let $\mathbf{s}_i = (s_{i1}, \cdots, s_{iL})^T$ and $\mathbf{s}_j = (s_{j1}, \cdots, s_{jL})^T$ be their respective spectral signatures specified by their associated hidden Markov models, $\mathrm{HMM}(\lambda_{\mathbf{s}_i})$ and $\mathrm{HMM}(\lambda_{\mathbf{s}_j})$. An HMM information divergence (HMMID) between $\mathbf{r}_i$ and $\mathbf{r}_j$ is defined by Du and Chang (2001)

$$\mathrm{HMMID}(\mathbf{r}_i; \mathbf{r}_j) = J_{\mathrm{HMM}}(\lambda_{\mathbf{s}_i}; \lambda_{\mathbf{s}_j}) + J_{\mathrm{HMM}}(\lambda_{\mathbf{s}_j}; \lambda_{\mathbf{s}_i}). \tag{2.32}$$

2.3 MEASURES OF SPECTRAL DISCRIMINABILITY

In Section 2.2, spectral similarity measures between two pixel vectors were developed. When there are more than two pixel vectors, how do we discriminate one from another?

In this section, we study three measures that can be used for spectral discriminability among a set of pixel vectors (Chang, 2000).

2.3.1 Relative Spectral Discriminatory ProBability (RSDPB)

In many remote sensing applications it is often the case that we are required to identify a pixel vector of interest using an existing spectral library or data base. Under this circumstance, we are interested in how much likelihood of a particular spectral signature **t** will be identified by a selective set of spectral signatures, Δ, which serves as a spectral database or spectral library. To meet this need, a new criterion, called relative spectral discriminatory probability (RSDPB) is proposed. It calculates the discriminatory probabilities of all the spectral signatures in a spectral library or database relative to the spectral signature of a pixel vector to be identified. The resulting probabilities show the likelihood of the pixel vector to be identified by the materials in the library or database. Although SAM has been also widely used for this purpose, it is shown in experiments that SID seems to have advantages over SAM in characterizing spectral similarity and variability. As explained previously, this may be due to the fact that SAM only preserves the angle between two pixel vectors rather than their random behaviors in nature. Since remotely sensed image pixel vectors are generally corrupted by noise and other unknown factors during acquisition, using information theoretic-based measures may be a better alternative to traditional measures, which allow each pixel vector to have a certain degree of spectral variations.

Let $\{\mathbf{s}_k\}_{k=1}^{K}$ be K spectral signatures in the set Δ which can be considered as a database and **t** be any specific target spectral signature to be identified using Δ. We define the spectral discriminatory probabilities of all $\mathbf{s}_k$'s in Δ relative to **t** as follows.

$$p_{\mathbf{t},\Delta}(k) = m(\mathbf{t},\mathbf{s}_k) / \textstyle\sum_{j=1}^{L} m(\mathbf{t},\mathbf{s}_j) \text{ for } k = 1,\cdots,K \tag{2.33}$$

where $\sum_{j=1}^{L} m(\mathbf{t},\mathbf{s}_j)$ is a normalization constant determined by **t** and Δ. The resulting probability vector $\mathbf{p}_{\mathbf{t},\Delta} = (p_{\mathbf{t},\Delta}(1), p_{\mathbf{t},\Delta}(2),\cdots, p_{\mathbf{t},\Delta}(K))^T$ is called relative spectral discriminatory probability (RSDPB) of Δ with respect to **t** or spectral discriminatory probability vector of Δ relative to **t**. Then, using (2.33) we can identify **t** via Δ by selecting the one with the smallest relative spectral discriminability probability. If there is a tie, either one can be used to identify **t**. Fig. 2.2 shows a graphical representation of RSDPB of Δ with respect to **t**.

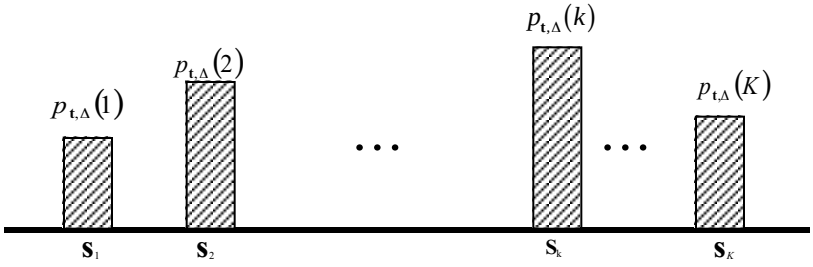

Figure 2.2. Graphical representation of RSDPB, $\mathbf{p}_{\mathbf{t},\Delta}$ of Δ with respect to **t**

2.3.2 Relative Spectral Discriminatory PoWer (RSDPW)

In the previous section, we focused on the similarity between the spectral signatures of two pixel vectors. However, if we are given two spectral similarity measures, how do

we evaluate which one is more effective than the other? In order to cope with this issue, a new criterion, called relative spectral discriminatory power (RSDPW) is suggested in this section. It is designed based on the power of discriminating one pixel vector from another relative to a reference pixel vector.

Assume that $m(\cdot,\cdot)$ is any given hyperspectral measure. Let $\mathbf{d}$ be the spectral signature of a reference pixel vector and $\mathbf{s}_i$, $\mathbf{s}_j$ be the spectral signatures of any pair of two pixel vectors. The RSDPW of $m(\cdot,\cdot)$, denoted by $\text{RSDPW}_m(\mathbf{s}_i,\mathbf{s}_j;\mathbf{d})$ is defined by

$$\text{RSDPW}_m(\mathbf{s}_i,\mathbf{s}_j;\mathbf{d}) = \max\left\{m(\mathbf{s}_i,\mathbf{d}) / m(\mathbf{s}_j,\mathbf{d}), m(\mathbf{s}_j,\mathbf{d}) / m(\mathbf{s}_i,\mathbf{d})\right\}. \quad (2.34)$$

More precisely, $\text{RSDPW}_m(\mathbf{s}_i,\mathbf{s}_j;\mathbf{d})$ selects as the discriminatory power of $m(\cdot,\cdot)$ the maximum of two ratios, ratio of $m(\mathbf{s}_i,\mathbf{d})$ to $m(\mathbf{s}_j,\mathbf{d})$ and ratio of $m(\mathbf{s}_j,\mathbf{d})$ to $m(\mathbf{s}_i,\mathbf{d})$. The $\text{RSDPW}_m(\mathbf{s}_i,\mathbf{s}_j;\mathbf{d})$ defined by (2.34) provides a quantitative index of spectral discrimination capability of a specific hyperspectral measure $m(\cdot,\cdot)$ between two spectral signatures $\mathbf{s}_i$, $\mathbf{s}_j$ relative to $\mathbf{d}$. Obviously, the higher the $\text{RSDPW}_m(\mathbf{s}_i,\mathbf{s}_j;\mathbf{d})$ is, the better discriminatory power the $m(\cdot,\cdot)$ is. In addition, $\text{RSDPW}_m(\mathbf{s}_i,\mathbf{s}_j;\mathbf{d})$ is symmetric and bounded below by one, i.e., $\text{RSDPW}_m(\mathbf{s}_i,\mathbf{s}_j;\mathbf{d}) \geq 1$ with equality if and only if $\mathbf{s}_i = \mathbf{s}_j$.

2.3.3 Relative Spectral Discriminatory Entropy (RSDE)

Since $\mathbf{p}_{\mathbf{t},\Delta} = (p_{\mathbf{t},\Delta}(1), p_{\mathbf{t},\Delta}(2),\cdots, p_{\mathbf{t},\Delta}(K))^T$ given by (2.33) is the relative spectral discriminability probability vector of $\mathbf{t}$ using a selective set of spectral signatures, $\Delta = \{\mathbf{s}_k\}_{k=1}^K$, we can further define the relative spectral discriminatory entropy (RSDE) of the spectral signature $\mathbf{t}$ with respect to the set Δ, denoted by $H_{\text{RSDE}}(\mathbf{t};\Delta)$ by

$$H_{\text{RSDE}}(\mathbf{t};\Delta) = -\sum_{k=1}^K p_{\mathbf{t},\Delta}(k)\log p_{\mathbf{t},\Delta}(k). \quad (2.35)$$

Equation (2.35) provides an uncertainty measure of identifying $\mathbf{t}$ resulting from using $\Delta = \{\mathbf{s}_k\}_{k=1}^K$. A higher $H_{\text{RSDE}}(\mathbf{t};\Delta)$ may have a less chance to identify $\mathbf{t}$. For example, if $\mathbf{p}_{\mathbf{t},\Delta}$ is uniformly distributed, i.e., $p_{\mathbf{t},\Delta}(k) = 1/K$ for $k = 1,\cdots,K$, then $H_{\text{RSDE}}(\mathbf{t};\Delta)$ achieves its maximum, $\log K$, namely, $H_{\text{RSDE}}(\mathbf{t};\Delta) \leq \log K$. This is the worst scenario of material identification. In other words, when the probability of using any $\mathbf{s}_k$ in Δ to identify $\mathbf{t}$ is equally likely, there is no way for us to know which signature is more likely to be $\mathbf{t}$. In this case, any signature can be used to identify $\mathbf{t}$. However, if $H_{\text{RSDE}}(\mathbf{t};\Delta)$ is small, it only provides information that some candidates in Δ are more likely to be used to identify $\mathbf{t}$, but it does not necessarily mean that their relative discriminability rates are high. As an illustrative example, we consider a hypothetical case that the relative spectral discriminability probabilities for two signatures $\mathbf{u}$, $\mathbf{v}$ in Δ are very small with equal probability 2^{-4}, while others are all zeros except for one signature $\mathbf{s}_k \in \Delta$ with

$p_{\mathbf{t},\Delta}(k) = 14/16$. The RSDE $H_{\text{RSDE}}(\mathbf{t};\Delta) = 1/2 + 14/16\left(4 - \log_2 14\right)$ bits, which is 0.6686, bits less than one bit. So, the chance to identify **t** must be good. However, this is only partly true in the sense that we may conclude that **t** must be either **u** or **v** but there is no way for us to know which one would be since their relative spectral discriminability probabilities are equal to 2^{-4}. Under this circumstance, one of them can be used to identify **t**.

2.4 EXPERIMENTS

Two sets of hyperspectral image data, AVIRIS and HYDICE will be used for experiments. Since the performance of CBD, TD, OPD and JMD is very similar to that of ED and SAM according to experiments conducted based on these two data sets, only four representative spectral measures, ED, SAM, SID and HMMID, one from each category described in Section 2.2 are selected for performance evaluation.

2.4.1 AVIRIS Data

The data set to be used in this section is Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) reflectance data shown in Fig. 1.5. For spectral variability, Tables 2.1 tabulates the statistics generated by SIM for these five signatures up to four moments (not central moments) using (2.1) and (2.2) where μ_i is the i-th moment, and Fig.2.3 plots values in Table 2.1 for graphical interpretation.

Table 2.1. Four moments generated by SIM for the five signatures in Fig. 1.5

	μ_1	μ_3	μ_3	μ_4
blackbrush	0.2670	0.0786	0.0245	0.0079
creosote leaves	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

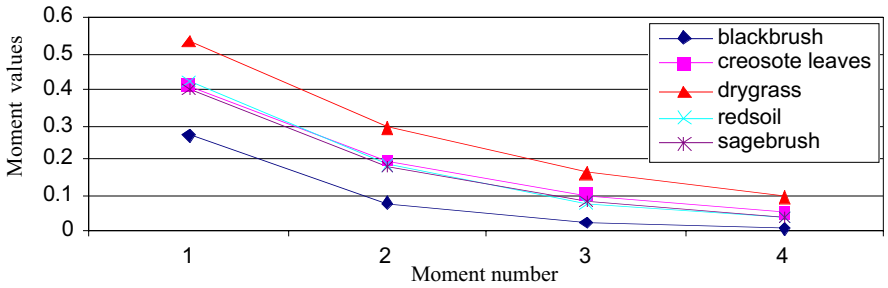

Figure 2.3. Plots of Table 2.1

Table 2.2 also tabulates the self-information of these five signatures using SID specified by (2.3) and HMM specified by (2.18) respectively and Figure 2.4 is their graphical plots.

Table 2.2. Self-information of the five signatures in Fig. 1.5 using SIM and HMM

	blackbrush	creosote leaves	drygrass	redsoil	sagebrush
SID	7.1814	7.0564	7.2740	7.2356	7.1709
HMM	3.5639	3.3771	3.9115	3.6489	3.5959

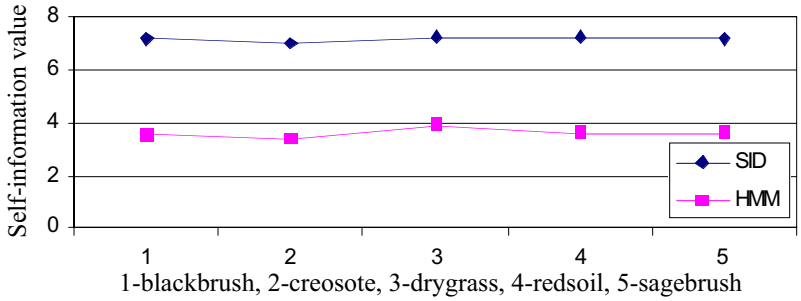

Figure 2.4. Plots of Table 2.2

As we can see in Tables 2.1 and 2.2, the HMM-based measure offers more valuable information than SID does. For example, SID produces very close values of self-information for all the five signatures. On the contrary, HMM categorizes the values of self-information of these five signatures into three classes, one for three signatures: blackbrush, creosote leaves and sagebrush, one for red soil and another for dry grass. From Fig. 1.5 we can see that blackbrush (open circle) creosote leaves (asterisked line) and sagebrush (diamonded line) have very similar spectra while red soil and dray grass are very distinct. This fact is also reflected in Table 2.1 where all their four moments generated by SIM are very close. However, it is very difficult to determine which signature is more close to another by visual inspection from the figure.

Table 2.3. Similarity values produced by ED among the five signatures in Fig. 1.5

	blackbrush	creosote leaves	drygrass	redsoil	sagebrush
blackbrush	0	0.1765	0.2568	0.4031	0.0681
Creosote leaves		0	0.4182	0.5637	0.1288
drygrass			0	0.2175	0.2957
redsoil				0	0.4477
sagebrush					0

Table 2.4. Similarity values produced by SAM among the five signatures in Fig. 1.5

	blackbrush	Creosote leaves	drygrass	redsoil	sagebrush
blackbrush	0	0.1767	0.2575	0.4058	0.0681
Creosote leaves		0	0.4213	0.5714	0.1289
drygrass			0	0.2179	0.2968
redsoil				0	0.4515
sagebrussh					0

Table 2.5. Similarity values produced by SID among the five signatures in Fig. 1.5

	blackbrush	creosote leaves	drygrass	redsoil	sagebrush
blackbrush	0	0.0497	0.0766	0.1861	0.0063
creosote leaves		0	0.2298	0.4154	0.0303
drygrass			0	0.0640	0.0973
redsoil				0	0.2340
sagebrush					0

Table 2.6. Similarity values produced by HMMID among the five signatures in Fig. 1.5

	blackbrush	creosote leaves	drygrass	redsoil	sagebrush
blackbrush	0	1.5390	3.6717	3.9182	0.2263
creosote leaves		0	6.2549	6.7333	1.4102
drygrass			0	1.1302	4.0073
redsoil				0	3.4166
sagebrush					0

Tables 2.3-2.6 are produced for spectral similarity among these five signatures by ED, SAM, SID and HMMID using (2.20), (2.22), (2.29) and (2.32) respectively. According to Tables 2.3-2.6, the smaller value between two signatures is, the more similar the two signatures are. The results produced by ED and SAM were nearly the same because of (2.24). Moreover, as we can see, blackbrush is closest to sagebrush, while creosote leaves is closest to sagebrush. If we examine the last column under sagebrush in Tables 2.3-2.6, it found that sagebrush is more closer to creosote leaves than to blackbrush. The similarity values produced by ED and SAM between blackbrush and sagebrush was about twice between creosote leaves. The similarity values produced by SID between blackbrush and sagebrush was about five times between creosote leaves. The similarity values produced by HMMID between blackbrush and sagebrush was about six times between creosote leaves. For signatures whose spectra are dissimilar, HMMID produced even much greater values than do other three measures. For example, the similarity values between red soil and blackbrush, between red soil and creosote leaves, between red soil and sagebrush were 3.9182, 6.7333 and 3.4166 respectively which are much greater values compared to their counterparts produced by ED, SAM and SID. An assessment based on Tables 2.3-2.6 may be subjective and difficult to compare the discriminatory power of the four measures. In order to better illustrate the results in Tables 2.3-2.6, we also plot their values in Fig. 2.5 where B,C,D,R,S represent blackbrush, creosote leaves, dray grass, red soil and sagebrush respectively where the plots produced by ED and SAM were nearly identical.

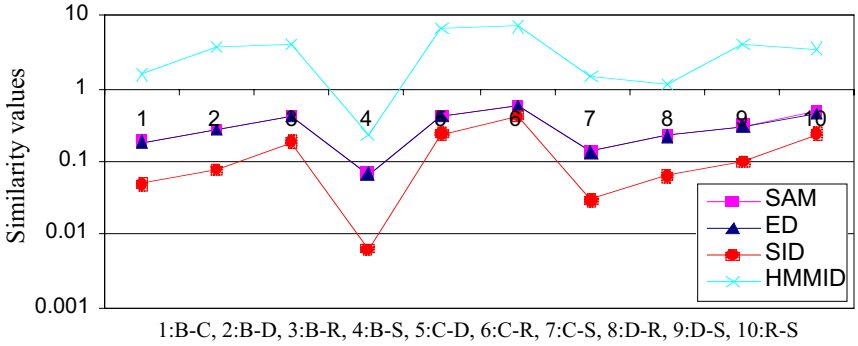

Figure 2.5. Plots of Tables 2.3-2.6

However, it should be noted that the magnitude of the similarity values in Fig. 2.5 does not imply the discriminatory power of each measure. In this case, we calculate RSDPW for ED, SAM, SID and HMMID to evaluate their spectral discriminatory power relative to a reference signature $\mathbf{d}$, which is chosen to be the red soil. As we know, the signature of dry grass is very close to that of red soil. So, the power of HMMID discriminating $\mathbf{s}_i$ = dry grass from the other three signatures $\mathbf{s}_j$ = blackbrush, creosote

leaves, sagebrush, were given by 3.4665, 5.9571, 3.0227 respectively. The values of RSDPW produced by HMMID are much greater than those produced by ED, SAM and SID. This implies that the signature of dry grass is very different from those of blackbrush, creosote leaves and sagebrush.

Comparing $\text{RSDPW}_{\text{HMMID}}(\mathbf{s}_i = \text{dry grass}, \mathbf{s}_j = \text{blackbrush}; \mathbf{d} = \text{red soil}) \approx 63$ to $\text{RSDPW}_{\text{HMMID}}(\mathbf{s}_i = \text{dry grass}, \mathbf{s}_j = \text{sagebrush}; \mathbf{d} = \text{red soil}) \approx 69$ suggests that both blackbrush and sagebrush must have very similar signatures, which is indeed the case. This evidence is also provided by finding RSDPWs of blackbrush and sagebrush with respect to the reference signature **d** specified by red soil.

$$\text{RSDPW}_{\text{SAM}}(\mathbf{s}_i = \text{blackbrush}, \mathbf{s}_j = \text{sagebrush}; \mathbf{d} = \text{red soil}) = 1.1126 \quad (2.36)$$

$$\text{RSDPW}_{\text{ED}}(\mathbf{s}_i = \text{blackbrush}, \mathbf{s}_j = \text{sagebrush}; \mathbf{d} = \text{red soil}) = 1.1106 \quad (2.37)$$

$$\text{RSDPW}_{\text{SID}}(\mathbf{s}_i = \text{blackbrush}, \mathbf{s}_j = \text{sagebrush}; \mathbf{d} = \text{red soil}) = 1.2576 \quad (2.38)$$

$$\text{RSDPW}_{\text{HMMID}}(\mathbf{s}_i = \text{blackbrush}, \mathbf{s}_j = \text{sagebrush}; \mathbf{d} = \text{red soil}) = 1.1468 \quad (2.39)$$

which also show how close the signatures of blackbrush and sagebrush are and is smaller than 1.1106, 1.1126 and 1.2576 produced by ED, SAM and SID respectively.

Tables 2.7-2.10 tabulate the RSDPWs of ED, SAM, SID and HMMID respectively with red soil used as a reference signature **d** where Fig. 2.6 plots their corresponding values.

Table 2.7. RSDPW of ED using red soil as a reference signature **d**

	blackbrush	creosote leaves	drygrass	sagebrush
blackbrush	1	1.3984	1.8533	1.1106
creosote leaves		1	2.5917	1.2591
drygrass			1	2.0584
sagebrush				1

Table 2.8. RSDPW of SAM using red soil as a reference signature **d**

	Blackbrush	creosote leaves	drygrass	sagebrush
blackbrush	1	1.4081	1.8623	1.1126
creosote leaves		1	2.6223	1.2656
drygrass			1	2.0721
sagebrus				1

Table 2.9. RSDPW of SID using red soil as a reference signature **d**

	blackbrush	creosote leaves	drygrass	sagebrush
blackbrush	1	2.2321	2.9078	1.2574
creosote leaves		1	6.4906	1.7752
drygrass			1	3.6562
sagebrush				1

Table 2.10. RSDPW of HMMID using red soil as a reference signature **d**

	blackbrush	creosote leaves	drygrass	sagebrush
blackbrush	1	1.7185	3.4665	1.1468
creosote leaves		1	5.9571	1.9708
drygrass			1	3.0227
sagebrush				1

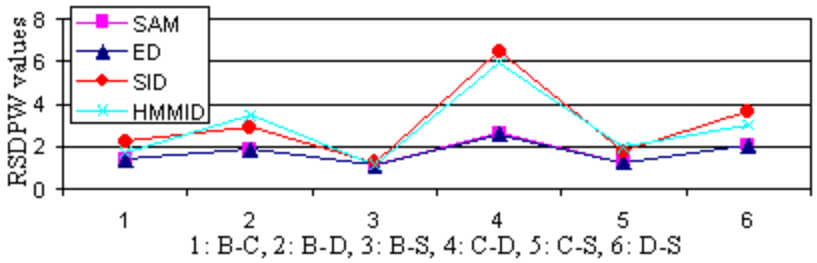

Figure 2.6. Plots of Tables 2.7-2.10

From Fig. 2.6 it is obvious that the stochastic information measures, HMMID and SID were more effective than deterministic measures ED and SAM. In order to evaluate which measure is more effective in terms of spectral discriminability, a mixed spectral signature was randomly generated to be used as a target signature, **t** for identification. It was composed of 0.1055 blackbrush, 0.0292 creosote leaves, 0.0272 dry grass, 0.7588 red soil and 0.0974 sagebrush. It should be noted that the **t** was generated randomly, not intended for a particular preference. From Tables 2.1-2.4, the spectrum of red soil is very similar to that of dry grass. Using (2.33) Table 2.11 tabulates the RSDPB of ED, SAM, SID and HMMID where Fig. 2.7 also plots the values of Table 2.11.

Table 2.11. RSDPB produced by ED, SAM, SID and HMMID with **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 leaves	drygrass	redsoil	sagebrush
ED	0.2215	0.3417	0.1049	0.0773	0.2547
SAM	0.2212	0.3430	0.1044	0.0769	0.2546
SID	0.1897	0.4933	0.0588	0.0112	0.2500
HMMID	0.2395	0.4911	0.0511	0.0036	0.2147

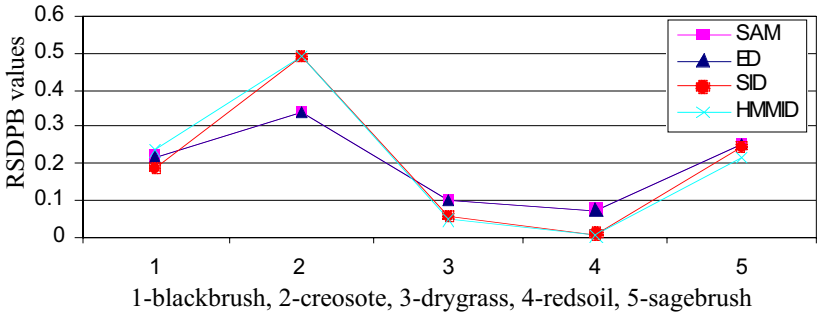

Figure 2.7. Plot of Table 2.11

According to Table 2.11, the ratio of using ED and SAM to identify **t** as red soil to as dry grass was $0.1049:0.0773 \approx 0.1044:0.0769 \approx 1.36$. Compared to ED and SAM, SID and HMMID yielded $0.0588:0.0112 \approx 5.25$ and $0.0511:0.0036 \approx 14.19$ respectively. A similar observation from Fig. 2.7 also shows that SID and HMMID were more effective than other three measures in identifying **t** as red soil with nearly zero value of spectral discriminatory probability at red soil. Table 2.12 calculated RSDE of ED, SAM, SID and HMMID using Table 2.11 where HMMID produced the least entropy with Table 2.13 produced by the RSDE results of HMMID using different numbers of states.

Table 2.12. RSDE of Table 2.11 produced by ED, SAM, SID and HMMID

	ED	SAM	SID	HMMID
RSDE	1.4835	1.4822	1.2274	1.1940

Table 2.13. RSDEs resulting from different number of states

# of states	3	4	5	6	7	8	9	10
RSDE	1.1940	1.1939	1.1938	1.1936	1.1932	1.1943	1.1941	1.1941

It was found that RSDE decreased slightly as the number of states increased. However, a high number of states generally increases computational complexity of parameter estimation process significantly. So, according to our experiments, a reasonable number of states ranges from 4 to 6 that allows us to achieve good compromise.

2.4.2 HYDICE Data

Unlike the AVIRIS data studied in the previous section, the HYDICE data used in the following experiments were directly extracted from a HYDICE image scene of size 64×64 shown in Fig. 1.6(a). The spectra of P1, P2, P3, P4 and P5 in Fig. 1.7(c) form a data set $\Delta = \{P1, P2, P3, P4, P5\}$ that would be used for panel identification. Tables 2.14-2.17 tabulate the similarity values resulting from ED, SAM, SID and HMMID.

Table 2.14. Similarity values produced by ED among the five panel signatures in Fig. 2.5

	P1	P2	P3	P4	P5
P1	0	1301.6	2033.3	4107.3	4831.6
P2		0	1340.4	5064.1	5733.0
P3			0	5434.1	5968.7
P4				0	1125.4
P5					0

Table 2.15. Similarity values produced by SAM among the five panel signatures in Fig. 2.5

	P1	P2	P3	P4	P5
P1	0	0.0435	0.0673	0.1144	0.1240
P2		0	0.0430	0.1479	0.1567
P3			0	0.1652	0.1710
P4				0	0.0248
P5					0

Table 2.16. Similarity values produced by SID among the five panel signatures in Fig. 2.5

	P1	P2	P3	P4	P5
P1	0	0.0039	0.0086	0.0233	0.0313
P2		0	0.0033	0.0385	0.0484
P3			0	0.0476	0.0570
P4				0	0.0025
P5					0

Table 2.17. Similarity values produced by HMMID among the five panel signatures in Fig. 2.5

	P1	P2	P3	P4	P5
P1	0	0.0255	0.0291	0.2935	0.4798
P2		0	0.0215	0.2891	0.4483
P3			0	0.3590	0.5483
P4				0	0.0186
P5					0

It is interesting to note that unlike Tables 2.3-2.4 which show that ED and SAM generated very close values, the values generated by ED and SAM in Tables 2.14-2.15 are very different. However, if the five panel signatures were normalized, the results produced by ED and SAM turned out to be very close. In analogy with Fig. 2.5, Fig. 2.8 also plots similarity values of the four measures in Tables 2.14-2.17 for comparison.

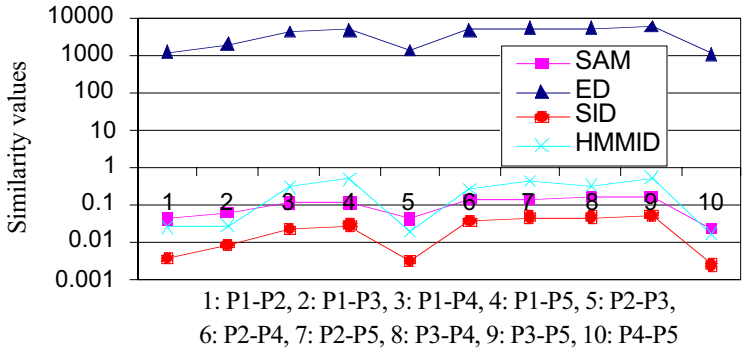

Figure 2.8. Plots of Tables 2.14-2.17

The above HYDICE experiments also showed that ED and SAM performed very similarly because of (2.24). Tables 2.14-2.17 and Fig. 2.8 show that the signatures of P1, P2 and P3 are very close. Similarly, both P4 and P5 have very close signatures, but are very distinct from those of P1, P2 and P3.

Tables 2.18-2.21 tabulate RSDPW values for ED, SAM, SID and HMMID where P2 was used as a reference signature **d**. These tables also demonstrate that SID and HMMID outperformed than ED and SAM in terms of RSDPW and HMMID performed better than SID.

Table 2.18. RSDPW of ED using P2 as a reference signature **d**

	P1	P3	P4	P5
P1	1	1.0298	3.8907	4.4046
P3		1	3.7781	4.2771
P4			1	1.1321
P5				1

Table 2.19. RSDPW of SAM using P2 as a reference signature **d**

	P1	P3	P4	P5
P1	1	1.0116	3.4000	3.6023
P3		1	3.4395	3.6442
P4			1	1.0595
P5				1

Table 2.20. RSDPW of SID using P2 as a reference signature **d**

	P1	P3	P4	P5
P1	1	1.1818	9.8718	12.4103
P3		1	11.6667	14.6667
P4			1	1.2571
P5				1

Table 2.21. RSDPW of HMMID using P2 as a reference signature **d**

	P1	P3	P4	P5
P1	1	1.1860	11.3373	17.5804
P3		1	13.4465	20.8512
P4			1	1.5507
P5				1

From Tables 2.18-2.21 we also plot their RSDPW values in Fig. 2.9. Like the AVIRIS experiments, the stochastic information measures, HMMID and SID performed more effectively than deterministic measures ED and SAM.

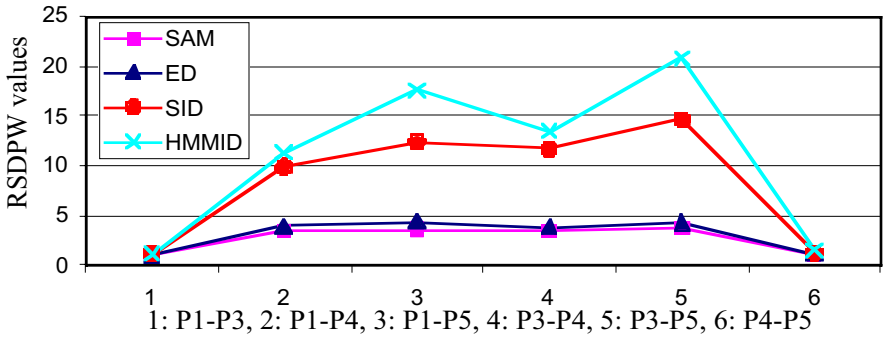

Figure 2.9. Plots of Tables18-2.21

In order to evaluate the RSDPB, a target pixel vector randomly extracted from the white mask of p_{23} was chosen as **t** for identification. It was based on the fact that P2 is very close to both P1 and P3, but more closer to P3. This selection enables us to evaluate the effectiveness of RSDPB. The pixel vector **t** was a panel edge pixel vector mixed with the background grass signature. Table 2.22 tabulates its discriminatory probabilities against five panel signatures using ED, SAM, SID and HMMID where Fig. 2.10 plots their RSDPB values in Table 2.22.

Table 2.22. RSDPB values produced by ED, SAM, SID and HMMID with **t** chosen from a pixel vector in the white mask of p_{23}

	P1	P2	P3	P4	P5
ED	0.1530	0.1339	0.1578	0.2631	0.2922
SAM	0.1544	0.1108	0.1482	0.2837	0.3028
SID	0.1029	0.0520	0.0813	0.3419	0.4218
HMMID	0.0994	0.0419	0.0939	0.3680	0.3968

As we can see from Table 2.22, the RSDPBs among P1, P2 and P3 using ED and SAM were very close, but both SID and HMMID were clearly better than ED and SAM. This fact can be better demonstrated in Fig. 2.10, where SID and HMMID produced lowest discriminatory probability of **t** against P2.

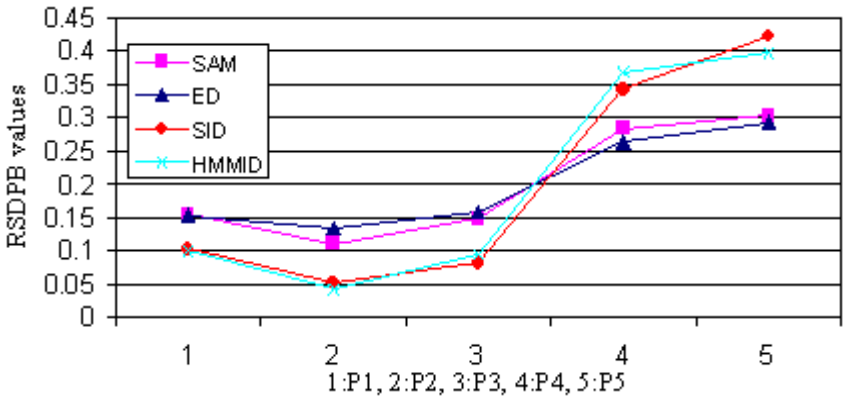

Figure 2.10. Plot of Table 2.22

If we further calculate the ratio of the second smallest RSDPB to the smallest RSDPB for ED, SAM, SID and HMMID respectively,

$$p_{\mathbf{t},\Delta}^{\mathrm{ED}}(\mathrm{P3}):p_{\mathbf{t},\Delta}^{\mathrm{ED}}(\mathrm{P2}) = 0.1530:0.1339 \approx 1.14, \tag{2.32}$$

$$p_{\mathbf{t},\Delta}^{\mathrm{SAM}}(\mathrm{P3}):p_{\mathbf{t},\Delta}^{\mathrm{SAM}}(\mathrm{P2}) = 0.1482:0.1108 \approx 1.34, \tag{2.33}$$

$$p_{\mathbf{t},\Delta}^{\mathrm{SID}}(\mathrm{P3}):p_{\mathbf{t},\Delta}^{\mathrm{SID}}(\mathrm{P2}) = 0.0813:0.0520 \approx 1.56 \tag{2.34}$$

$$p_{\mathbf{t},\Delta}^{\mathrm{HMMID}}(\mathrm{P3}):p_{\mathbf{t},\Delta}^{\mathrm{HMMID}}(\mathrm{P2}) = 0.0939:0.0419 \approx 2.24 \tag{2.35}$$

HMMID was about twice as effective as ED, SAM, and SID to identify **t** as P2.

Table 2.23. RSDE of Table 2.21 produced by ED, SAM, SID and HMMID

ED	SAM	SID	HMMID
1.5586	1.5344	1.3230	1.3190

Table 2.23 tabulates their respective RSDEs. Once again, HMMID produced the least entropy. More experiments can be found in Chang (2000), Du and Chang (2001) and Du (2000).

2.5 CONCLUSIONS

This chapter presents two new information theoretic hyperspectral measures, spectral information measure (SIM) and hidden Markov model (HMM)-based measure for spectral characterization. Both measures use self-information to characterize spectral variability, similarity, and discrimination for hyperspectral image analysis. SIM considers a hyperspectral image pixel vector as a random variable so that the spectral variability of the pixel vector can be more effectively described by randomness in nature. The Hidden Markov Model (HMM) is introduced to model the unobserved and hidden spectral properties of a hyperspectral image pixel vector as a Markov random process. With these interpretations, a SIM-derived Spectral Information Divergence (SID) and an HMM-derived Information Divergence (HMMID) are further developed to measure the spectral similarity between two pixel vectors. In order to evaluate the effectiveness of a

hyperspectral measure, a new definition of spectral discriminatory power is suggested for performance analysis. Finally, the concept of spectral discriminatory probability vector is also developed for the purpose of material identification. Experiments have demonstrated that SID and HMMID capture spectral characteristics more effectively than do commonly used spectral similarity measures, Euclidean distance (ED), spectral angle mapper (SAM). This is because both SID and HMMID are statistical measures as opposed to ED, SAM, which are considered to be deterministic measures. In particular, the HMM can be used to specify complicated hidden spectral properties that cannot be observed in the spectrum of a hyperspectral pixel vector. As the number of spectral channels grows, the uncertainty and unpredictability in spectral analysis increase. Under this circumstance, SID and HMMID have better ability than do ED, SAM in characterizing spectral properties. This has been demonstrated in the experiments. The disadvantage of HMMID is the complexity in the implementation of HMM where the parameter vector $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ used in HMM must be estimated before it is used. As a final note, a discrete HMM-based spectral measure was also developed in (Du and Chang, 1999; Du, 2000) which was derived from the continuous HMM described in Section 2.1.2 by uniformly quantizing the spectrum of each pixel vector into a finite number of discrete values. The quantization used in the discrete HMM can simply analysis, but may lead to loss of information since an unsupervised clustering technique such as C-means (or K-means) method must be used to minimize the information loss and capture the major features of a pixel vector.