

## A Novel Method for Spectral Similarity Measure by Fusing Shape and Amplitude Features

J. G. Ding<sup>1,2,\*</sup>, X. B. Li<sup>1</sup> and L.Q. Huang<sup>3</sup>

<sup>1</sup> School of Resources and Safety Engineering, Central South University, Changsha 410083, China

<sup>2</sup> Changsha Engineering & Research Institute Ltd. of Nonferrous Metallurgy, Changsha, 410011, China

<sup>3</sup> Curtin University, Perth 6102, Australia

Received 18 June 2015; Accepted 19 December 2015

### Abstract

Spectral similarity measure is the basis of spectral information extraction. The description of spectral features is the key to spectral similarity measure. To express the spectral shape and amplitude features reasonably, this paper presents the definition of shape and amplitude feature vector, constructs the shape feature distance vector and amplitude feature distance vector, proposes the spectral similarity measure by fusing shape and amplitude features (SAF), and discloses the relationship of fusing SAF with Euclidean distance and spectral information divergence. Different measures were tested on the basis of United States Geological Survey (USGS) mineral\_beckman\_430. Generally, measures by integrating SAF achieve the highest accuracy, followed by measures based on shape features and measures based on amplitude features. In measures by integrating SAF, fusing SAF shows the highest accuracy. Fusing SAF expresses the measured results with the inner product of shape and amplitude feature distance vectors, which integrate spectral shape and amplitude features well. Fusing SAF is superior to other similarity measures that integrate SAF, such as spectral similarity scale, spectral pan-similarity measure, and normalized spectral similarity score(NS<sup>3</sup>).

*Keywords:* Spectral similarity measure, Shape features, Amplitude features, Shape features distance vector, Amplitude features distance vector

### 1. Introduction

A hyperspectrum can describe imaging targets accurately because of its approximately continuous spectral curve, wide wavebands range from dozens to thousands, and high resolution (10 nm). Surface features contain the specific reflection and radiation characteristics of electromagnetic waves. A hyperspectrum is the representation of the reflection or radiation characteristics of the electromagnetic waves of surfaces. Spectral similarity measure is the foundation of hyperspectral information acquisition and can be used to detect and recognize surface features effectively. Thus, spectral similarity measure should be extensively studied.

A spectral curve has shape and amplitude features (SAF). Spectral similarity measure methods based on shape features mainly include spectral angle mapper (SAM) (Kruse et al., [1]), spectral information divergence (SID) (Chang et al., [2]), spectral correlation coefficient (SCC) (Van de Meer et al., [3]), SID-SAM (Du et al., [4]), a hybrid measure using the spectral correlation angle (SCA) and SID, i.e., SID-SCA (Kumar et al.), and a hyperspectral matching by integrating the Jeffries-Matusita distance (JMD) and SAM, i.e., JMD-SAM (Padma et al., [6]). Similarity measures based on spectral shape features cannot reflect the spectral amplitude difference caused by spectral gain and are insensitive to the

spectral amplitude difference caused by offset. Furthermore, this type of similarity measure cannot comprehensively reflect the effect of spectral amplitude features on spectral similarity measure. Euclidean distance (ED) is the main spectral similarity measure method based on amplitude features. A similarity measure based on spectral amplitude features is over-sensitive to the spectral amplitude difference caused by gain and offset; thus, this measure is inadequate for comprehensively reflecting the effect of spectral shape features on spectral similarity. Research results demonstrate [7–9] that single spectral shape or amplitude features are inadequate in accurately measuring spectral similarity. Spectral similarity measures should consider both spectral shape and amplitude features.

Some scholars have studied spectral similarity measures that integrate shape and amplitude features. Granahan and Sweet [10] proposed a spectral similarity scale (SSS) by combining ED and SCC. Kong et al. [11] presented a spectral pan-similarity measure (SPM) by combining ED, SCC, and SID. Nidamanuri and Zbell [12] proposed a normalized spectral similarity score (NS<sup>3</sup>) method by combining normalized ED and SAM. The expressions of SSS, SPM, and NS<sup>3</sup> involve the sum of amplitude features and shape features. When these two features differ significantly, the smaller numerical value will be drowned out by the larger one. The features of large numerical values occupy significantly high proportions in the comprehensive measure, whereas the features of smaller numerical values will be neglected. This phenomenon influences the accuracy of spectral identification. By combining SCC and ED and by endowing and dynamically adjusting the weights of the

\* E-mail address: 17885988@163.com

shape index and amplitude index, Fang et al. [13] presented a spectral similarity measure method called dynamic weight adjustment method (DWAM). On the basis of the combination of SAM and ED, Zhang et al. [7] proposed a spectral similarity measure method called spectral changing-weight similarity measure (SCWSM) by endowing and dynamically adjusting the weights of the shape index and amplitude index. The weight adjustment criteria of DWAM and SCWSM are difficult to determine.

This paper provides the definitions of shape and amplitude features vectors, constructs the shape and amplitude features distance vectors, expresses spectral similarity measure with the inner product of shape and amplitude features distance vectors, and overcomes the shortages of existing measure methods that integrate the shape and amplitude features. Test data use United States Geological Survey (USGS) mineral\_beckman\_430. Results show that the proposed spectral similarity measure has significantly higher spectral identification accuracy than existing similarity measures.

## 2. Methodology

### 2.1 Shape and amplitude features vectors

Spectral amplitude features vector expresses band-wise brightness and is defined by spectral vector  $V$ , which is expressed in Equation (1), where  $n$  is the total waveband number of the hyperspectrum and  $v_i$  is the spectral radiance or reflectance amplitude of the waveband  $b_i$ . Spectral shape features can be expressive of the relative value of band-wise brightness, and a spectral shape feature vector is defined by the unit vector  $E$  of  $V$ .  $E$  is expressed as Equation (2). Spectral linear transformation, i.e., gain and offset, influences vector  $V$  and  $E$  (Equations (3) and (4)). In Equations (3) and (4),  $k$  is the gain factor and  $b$  is the offset factor. Equation (3) shows that spectral gain transformation changes the band-wise brightness but does not change  $E$ . Thus, the shape features vector expresses the direction but not the amplitude of the spectral vector  $V$ . Equation (4) reflects that spectral offset transformation changes both the band-wise brightness and  $E$ . According to the applied spectral features vector, spectral similarity measure methods can be divided into shape feature-based measure (SFM), amplitude feature-based measure (AFM), and measure by integrating SAF.

$$V = (v_1, v_2, \dots, v_i, \dots, v_n)^T \quad (1)$$

$$E = \frac{V}{\|V\|} \quad (2)$$

$$V_g = kV, E_g = \frac{kV}{\|kV\|} = \frac{V}{\|V\|} = E \quad (3)$$

$$V_o = V - b, E_o = \frac{V - b}{\|V - b\|} = \frac{V_o}{\|V_o\|} \quad (4)$$

### 2.2 Spectral similarity measure methods based on shape features

Spectral similarity measure methods based on shape features include SAM [1, 14, 15], SID [2, 16], SCC [3], SID-SAM [4], SID-SCA [5], and JMD-SAM [6]. The metrics of SAM, SID, SID-SAM, and JMD-SAM are not affected by spectral gain transformation and are affected by but not sensitive to

spectral offset transformation. The metrics of SCC are not affected by both spectral gain and offset transformation. The metrics of SID-SCA, i.e., combined indices of SID and SCC, are not affected by spectral gain transformation and are affected by but not sensitive to spectral offset transformation. Other measure methods such as spectral binary coding, spectral absorption features, and spectral gradient characteristics, are not influenced by spectral gain factor and belong to spectral similarity measures based on shape features.

#### 2.2.1 SAM

SAM is the generalized angle  $\theta$  between the test spectral vector  $T$  and the reference spectral vector  $R$ . SAM can be expressed as follow:

$$\theta(T, R) = \cos^{-1} \frac{\sum_{i=1}^n t_i \cdot r_i}{\sqrt{\sum_{i=1}^n t_i^2} \sqrt{\sum_{i=1}^n r_i^2}}, \theta \in [0, \pi/2] \quad (4)$$

where  $t_i$  is the  $i^{\text{th}}$  component of  $T$ ,  $r_i$  is the  $i^{\text{th}}$  component of  $R$ , and  $n$  is the dimensions of the vector. Let  $e^t$  be the unit vector of  $T$  and  $e^r$  the unit vector of  $R$ . The generalized angle can then be expressed as follow:

$$\theta(T, R) = \cos^{-1}(e^t \cdot e^r) \quad (5)$$

From Equations (2) and (6), SAM can be expressed by unit vectors  $e^t$  and  $e^r$  of vectors  $T$  and  $R$  directly; hence, SAM is an SFM. SAM is not affected by spectral gain transformation and is affected by but not sensitive to spectral offset transformation.

#### 2.2.2 SID

SID is the application of information theory in the spectral classification field and belongs to the spectral information measure. It calculates the information entropy contained in each waveband amplitude of the spectral vector and represents the spectral shape similarity by the sum of information entropy. SID is expressed as follows:

$$D(T \| R) = \sum_{i=1}^n e_i^t \log\left(\frac{e_i^t}{e_i^r}\right) \quad (6)$$

$$D(T \| R) = \sum_{i=1}^n e_i^t \log\left(\frac{e_i^t}{e_i^r}\right) \quad (7)$$

$$D(R \| T) = \sum_{i=1}^n e_i^r \log\left(\frac{e_i^r}{e_i^t}\right) \quad (8)$$

From Equations (2), (7), (8), and (9), SID can be expressed by unit vectors  $e^t$  and  $e^r$  of vectors  $T$  and  $R$  directly; hence, SID is an SFM. SID is not affected by spectral gain transformation and is affected by but not sensitive to spectral offset transformation.

#### 2.2.3 SCC

The spectral correlation coefficient of spectrum, i.e., the Pearson's correlation coefficient, is actually the cosine of the generalized angle of vectors after the spectral curves minus their means. It can be expressed as follows:

$$r(T, R) = \frac{\sum_{i=1}^n (t_i - \bar{t})(r_i - \bar{r})}{\sqrt{\sum_{i=1}^n (t_i - \bar{t})^2} \sqrt{\sum_{i=1}^n (r_i - \bar{r})^2}} \quad (9)$$

$$r(T, R) = \frac{\sum_{i=1}^n t_i' \cdot r_i'}{\sqrt{\sum_{i=1}^n t_i'^2} \sqrt{\sum_{i=1}^n r_i'^2}} = e^t \cdot e^r \quad (10)$$

$$r(T, k \cdot R) = \frac{\sum_{i=1}^n (t_i - \bar{t})(k \cdot r_i - k \cdot \bar{r})}{\sqrt{\sum_{i=1}^n (t_i - \bar{t})^2} \sqrt{\sum_{i=1}^n (k \cdot r_i - k \cdot \bar{r})^2}} = r(T, R) \quad (12)$$

$$r(T, R + b) = \frac{\sum_{i=1}^n (t_i - \bar{t})((r_i + b) - (\bar{r} + b))}{\sqrt{\sum_{i=1}^n (t_i - \bar{t})^2} \sqrt{\sum_{i=1}^n ((r_i + b) - (\bar{r} + b))^2}} = r(T, R) \quad (13)$$

where  $r$  is the Pearson's correlation coefficient of  $T$  and  $R$ ,  $r \in [-1,1]$ ;  $\bar{t}$  is the mean of the test spectral vector components;  $\bar{r}$  is the mean of the reference spectral vector components. When  $r > 0$ ,  $T$  and  $R$  are positively correlated. When  $r < 0$ ,  $T$  and  $R$  are negatively correlated. To represent whether the correlation degree is positive or negative, the spectral correlation will be expressed with  $r^2$ , i.e., determination coefficient  $R^2$ ,  $R^2 \in [0,1]$ . With increasing shape similarity between  $T$  and  $R$ ,  $R^2$  gradually approaches one. To ensure the consistency of the SCC result with the SID and SAM results, this study used  $1-R^2$  to express the measured result of SCC. From Equations (10) and (11), although unit vectors  $e^t$  and  $e^r$  of  $T$  and  $R$  are changed by the spectral offset transformation first, SCC is an SFM because it can be expressed by unit vectors  $e^t$  and  $e^r$  directly. Equations (12) and (13) reveal that SCC is affected by neither reference spectral vector  $R$  gain nor offset transformation.

### 2.2.4 SID-SAM

The SID-SAM method is formed by combining SID and SAM. SID-SAM has two expressions:

$$SID - SAM_{TAN} = SID(T, R) \cdot TAN(\theta(T, R)) \quad (14)$$

$$SID - SAM_{SIN} = SID(T, R) \cdot SIN(\theta(T, R)) \quad (15)$$

When two curves are similar in shape, Equations (14) and (15) can obtain similar results.

### 2.2.5 SID-SCA

The SID-SCA method is formed by combining SID and SCA. SCA is expressed as follows:

$$SCA(T, R) = \cos^{-1}\left(\frac{r(T, R) + 1}{2}\right) \quad (16)$$

where  $r(T, R)$  is Pearson's correlation coefficient of  $T$  and  $R$ ,  $SCA \in [0, \pi/2]$ . When  $T$  and  $R$  are similar in shape, SCA approaches zero. SID-SCA has two expressions:

$$SID - SCA_{TAN} = SID(T, R) \cdot TAN(SCA(T, R)) \quad (17)$$

$$SID - SCA_{SIN} = SID(T, R) \cdot SIN(SCA(T, R)) \quad (18)$$

### 2.2.6 JMD-SAM

The JMD-SAM method is formed by combing JMD and SAM. JMD is the Jeffried-Matusita distance [17, 18]. The calculation formula of JMD is expressed as follows:

$$JMD(T, R) = \sqrt{\sum_{i=1}^n (\sqrt{e_i^t} - \sqrt{e_i^r})^2} \quad (19)$$

where  $e_i^t$  is the  $i^{\text{th}}$  component of unit vectors of  $T$ ;  $e_i^r$  is the  $i^{\text{th}}$  component of the unit vector of  $R$ . SID-SCA has two expressions:

$$JMD - SAM_{TAN} = JMD(T, R) \cdot TAN(SAM(T, R)) \quad (20)$$

$$JMD - SAM_{SIN} = JMD(T, R) \cdot SIN(SAM(T, R)) \quad (21)$$

### 2.2.7 Calculation example

This calculation example elaborates that the measure result based on spectral shape features is only related to shape features and is insensitive to the spectral amplitude difference caused by gain or offset. Thus, the effect of spectral amplitude features on spectral similarity is not reflected comprehensively. Data were collected from USGS mineral\_beckman\_430. Kaolinite CM3 was used as the test spectrum, and Kaolinite CM9, Illite IL101, and Hydrogrossular NMNH120555 were used as reference spectra (Fig. 1). SID was used for SFM, and ED was used for the AFM. The calculated results are shown in Table 1.

Although the reference spectrum and the test spectra show close curve shapes, they have significant amplitude differences (Fig. 1). The reference spectral curve can be viewed approximately as the gain or offset of the test spectral curve. In Table 1, the SID of the test spectrum and the reference spectrum is small (0.0070–0.0079) and  $SID_{\max}/SID_{\min} = 1.1$ , thus indicating the small variation of SID. However, their ED is relatively high (1.4560–4.6367) and  $ED_{\max}/ED_{\min} = 3.2$ , thus indicating a significant ED difference. The information entropy of SID measure and ED measure are 1.5834 and 1.4154, respectively, thus indicating the greater divergence of the ED measure. This example reflects that measures based on shape features cannot reflect the amplitude difference characteristics. Kaolinite CM3 and Kaolinite CM9 pertain to the same type of spectrum; however, measures based on shape features only cannot classify them accurately.

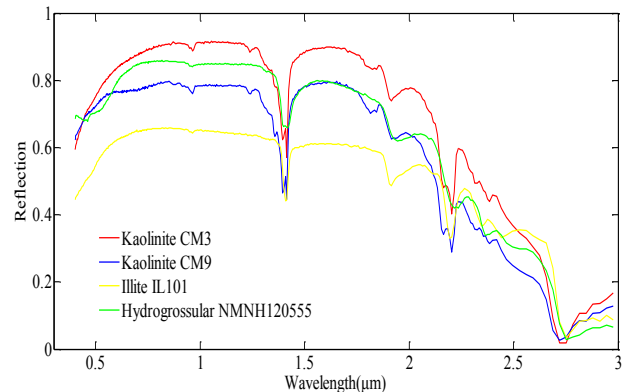


Fig.1. Five spectral curves of Kaolinite CM3 and others.

**Table 1.** ED and SID of Kaolinite CM3, Kaolinite CM9, and others

Measure	Test spectrum	Reference spectra			Entropy
		Kaolinite CM9	Illite IL101	Hydrogrossular MNH120555	
SID	Kaolinite CM3	0.0070	0.0075	0.0079	1.5834
ED	Kaolinite CM3	2.1656	4.6367	1.4560	1.4154

**2.3 Spectral similarity measure methods based on amplitude features**

Spectral similarity measure methods based on amplitude features include ED, Manhattan distance (MD), and Chebyshev distance (CD). ED is the most common one. Furthermore, a support vector machine [19] that directly uses spectral amplitude is a spectral similarity measure method based on amplitude features.

**2.3.1 ED**

ED has two expressions:

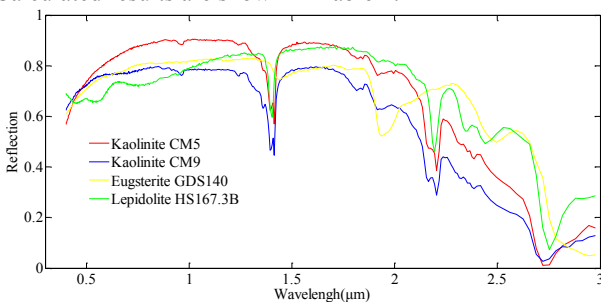
$$ED(T, R) = \sqrt{\sum_{i=1}^n (t_i - r_i)^2} \tag{22}$$

$$ED(T, R) = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_i - r_i)^2} \tag{23}$$

The result of Equation (22) is influenced by the component number of the spectral vector; however, such effect is eliminated in Equation (23). ED in Tables 1 and 2 is calculated by Equation (22).

**2.3.2 Calculation example**

This calculation example elaborates that the measure result based on spectral amplitude features is only related to amplitude features. This example also confirms that the measure result based on spectral amplitude features cannot reflect the differences in spectral shape features and the effect of spectral shape features on spectral similarity comprehensively. Data were collected from USGS mineral\_beckman\_430. Kaolinite CM5 was used as the test spectrum. Kaolinite CM9, Eugsterite GDS140, and Lepidolite HS167.3B were used as reference spectra (Fig. 2). SID was used for SFM and ED was used for AFM. Calculated results are shown in Table 2.



**Fig.2.** Five spectral curves of Kaolinite CM5 and others.

**Table 2.** ED and SID of Kaolinite CM5, Kaolinite CM9, and others

Measure	Test spectrum	Reference spectra			Entropy
		Kaolinite CM9	Eugsterite GDS140	Lepidolite HS167.3B	
ED	Kaolinite CM5	2.009	2.0479	2.0544	1.5849
SID	Kaolinite CM5	0.0084	0.0351	0.0287	1.3964

Fig. 2 reflects that the reference spectrum is close to the test spectrum in the overall distance; however, evident differences in shape features are observed between them. In Table 2, ED of the test spectrum and the reference spectrum varies within a small range (2.009–2.0544,  $ED_{max}/ED_{min} = 1.02$ ). However, their SID changes within a relatively wide range (0.008–0.0351,  $SID_{max}/SID_{min} = 4.2$ ), thus indicating the significant differences in shape features between them. The information entropies of the ED and SID measures are 1.5849 and 1.3964, respectively. This results shows that the ED measured result remains basically the same, whereas the SID measured result has high divergence. This example confirms that the spectral similarity measure based on amplitude features only cannot reflect the difference in shape features. The measure method involving amplitude features alone cannot accurately classify the spectrum.

**2.4 The spectral similarity measure by integrating SAF**

**2.4.1 SSS**

SSS [10] combines ED and SCC. ED represents the amplitude features, and SCC represents the shape features. The calculation formula of SSS is expressed as follow:

$$SSS(T, R) = \sqrt{ED(T, R)^2 + \hat{R}^2(T, R)^2} \tag{24}$$

where  $ED(T, R)$  is calculated by Equation (23),  $\hat{R}^2(T, R) = 1 - R^2$ , and  $R^2$  is the determination coefficient. SSS integrates shape and amplitude features and has high spectral identification accuracy. However, one index in SSS is easy to be drowned out by another one. When calculating the gain or offset of almost the same spectral curve, the SSS result only reflects the effect of  $ED(T, R)^2$ . This result is almost equal to the similarity measure based on amplitude features.

**2.4.2 SPM**

SPM [11] integrates ED, SCC, and SID. ED is the amplitude features, whereas SCC and SID represent the shape features. The calculation formula of SPM is expressed as follows:

$$SPM(T, R) = SID(T, R) \cdot TAN(\sqrt{ED(T, R)^2 + SSD(T, R)^2}) \tag{25}$$

$$SSD(T, R) = \left(\frac{1-r}{2}\right)^2 \tag{26}$$

where  $ED(T, R)$  is calculated by Equation (23) and  $r$  is calculated by Equation (10). SPM still cannot avoid SCC being drowned out by ED when calculating gain or offset of almost the same spectral curves.

**2.4.3 NS3**

NS3 [12] combines ED and SAM. ED represents the amplitude features, whereas SAM represents the shape features. The calculation formula is:

$$NS3(T, R) = \sqrt{\hat{A}(T, R)^2 + (1 - \cos(\theta(T, R)))^2} \tag{27}$$

where  $\hat{A}(T, R)$  is the normalized ED. ED is calculated by Equation (23). As a general rule,  $(1 - \cos(SAM(T, R)))^2$  is far smaller than  $\hat{A}(T, R)$ , thus indicating that shape features will be drowned out by the amplitude features.

**2.4.4 The proposed method for spectral similarity measure by fusing SAF**

To measure the difference of shape features between  $T$  and  $R$  accurately, spectral shape distance vector ( $S$ ) is constructed. The definition of  $S$  can be expressed by the following:

$$S1 = ((e_1^t - e_1^r) \log(\frac{e_1^t}{e_1^r}), \dots, (e_i^t - e_i^r) \log(\frac{e_i^t}{e_i^r}), \dots, (e_n^t - e_n^r) \log(\frac{e_n^t}{e_n^r}))^T \tag{28}$$

$$S2 = (\sqrt{(e_1^t - e_1^r) \log(\frac{e_1^t}{e_1^r})}, \dots, \sqrt{(e_i^t - e_i^r) \log(\frac{e_i^t}{e_i^r})}, \dots, \sqrt{(e_n^t - e_n^r) \log(\frac{e_n^t}{e_n^r})})^T \tag{29}$$

where  $S$  depicts every component difference of the unit vectors of  $T$  and  $R$ . Thus,  $S$  can specifically reflect the shape feature differences between spectra. SAM, SID, SCC, SID-SAM, SID-SCA, and JMD-SAM only represent the overall shape feature differences but cannot reveal the difference between spectral components.

To measure the amplitude features difference between  $T$  and  $R$  accurately, spectral amplitude distance vector  $A$  is constructed. Its definition has two expressions:

$$A1 = ((t_1 - r_1)^2, \dots, (t_i - r_i)^2, \dots, (t_n - r_n)^2) \tag{30}$$

$$A2 = (|t_1 - r_1|, \dots, |t_i - r_i|, \dots, |t_n - r_n|) \tag{31}$$

The proposed fusing SAF is expressed as follow:

$$SAF_F(T, R) = S \cdot A \tag{32}$$

where  $SAF_F$  is the proposed method for spectral similarity measure by fusing SAF. Fusing SAF can be understood from the following three aspects:

(1) Relationship between fusing SAF and ED  
When  $A$  is expressed by Equation (31) and ED is expressed by Equation (22), Equation (32) expresses the following:

$$SAF_F(T, R) = S \cdot A = |S| \cdot |A| \cdot \cos \phi = |S| \cdot \cos \phi \cdot ED(T, R) \tag{33}$$

where  $\phi$  is the generalized angle of  $S$  and  $A$ . From Equation (33), the value of  $SAF_F$  is equal to the ED corrected by shape features. The correction factor is  $|S| \cdot \cos \phi$ .

(2) Relationship between fusing SAF and SID  
When  $S$  is expressed by Equation (29), Equation (32) concludes that:

$$SAF_F(T, R)^2 = (S \cdot A)^2 = (|S| \cdot |A| \cdot \cos \phi)^2 = (|A| \cdot \cos \phi)^2 \cdot SID(T, R) \tag{34}$$

From Equation (34), the value of  $SAF_F(T, R)^2$  is equal to SID corrected by amplitude features. The correction factor is  $(|A| \cdot \cos \phi)^2$ .

(3) Relationship of fusing SAF with ED and SID

When  $A$  is expressed by Equation (31),  $S$  is expressed by Equation (29), and ED is expressed by Equation (22), the following can be concluded from Equation (32):

$$SAF_F(T, R) = S \cdot A = |S| \cdot |A| \cdot \cos \phi = \sqrt{SID(T, R)} \cdot ED(T, R) \cdot \cos \phi \tag{35}$$

In Equation (35), the value of  $SAF_F(T, R)$  is not the simple product of  $\sqrt{SID}$  representing the shape features and ED that representing the amplitude features. SID and ED express the overall features of vectors, but cannot reflect relations between wavebands of  $T$  and  $R$ .  $\cos \phi$  reflecting band-wise relations between  $T$  and  $R$ .

According to Equations (33), (34), and (35), fusing SAF is different from SID and ED but also has a certain relationship with them. Compared with SID, ED, SSS, and SPM, fusing SAF can accurately reflect the relations between wavebands of  $T$  and  $R$  and has higher spectral identification accuracy.

The spectral similarity measure function  $SAF_F$  meets the following three properties:

- (1) Non-negativity:  $SAF_F(T, R) \geq 0$ ;
- (2) Reflexivity:  $SAF_F(T, R) = 0$  only when  $T = R$ ;
- (3) Symmetry:  $SAF_F(T, R) = FSAF(R, T)$ .

**3. Test and Result Analysis**

**3.1 Test method**

**3.1.1 Evaluation methods of RSDPB and RSDE**

Given a known spectral library  $\square$ , spectra  $R$  in  $\square$  is used as the reference spectra. By using the spectral measure method  $M$ , the category or name of the test spectrum  $T$  can be determined. To evaluate  $M$ , Ref. [17] provided two methods: RSDPB and RSDE.

(1) RSDPB

RSDPB is used to calculate the possibility for determining the test spectrum  $T$ :

$$p_{T,\Delta}^M(j) = \frac{M(T, R_j)}{\sum_{i=1}^n M(T, R_i)} \quad \text{for } j = 1, 2, \dots, n \tag{36}$$

where  $M(\cdot, \cdot)$  is the similarity measure function;  $T$  is the test spectrum;  $R$  is the reference spectra;  $n$  is the total number of spectra in library  $\square$ ;  $p_{T,\Delta}^M(j)$  is the possibility for  $M$  to determine  $T$  as a spectrum in the library  $\square$ .

(2) RSDE

On the basis of the definition of entropy in information theory, the relative spectral identification entropy RSDE can be derived according to RSDPB:

$$H_{RSDE}^M(T, \Delta) = -\sum_{j=1}^n p_{T,\Delta}^M(j) \log_2 p_{T,\Delta}^M(j) \tag{37}$$

Where  $H_{RSDE}^M(T, \Delta)$  is the information entropy of  $T$  confirmed by the spectra in  $\square$ . A smaller information entropy leads to a more concentrated possibility distribution for  $T$  to be confirmed by the spectra in  $\square$ . Thus,  $T$  can be easily determined. Otherwise, such determination possibility is more scattered and  $T$  is more difficult to be determined. When  $H_{RSDE}^M(T, \Delta)$  is close to  $\log_2 n$ , the determination possibility of  $T$  by the spectra in  $\square$  is almost equal,  $T$  cannot be determined.

RSDPB and RSDE are widely used. They provide references for evaluating spectral similarity measure methods. However, they assume that the measure method has good monotony and rationality to spectral similarity measure. Monotony means that with the increase in similarity, similarity measure decreases. Rationality means that measured result is fair to similarity. They can only evaluate discrimination of spectral identification of the method from numerical value distribution of measured results. Without judgment on whether the spectral matching results are correct, they cannot evaluate the spectral identification accuracy of the method.

**3.1.2 Kappa factor**

Cohen presented the Kappa factor in 1960 for evaluating classification result of hyperspectral images. The Kappa factor can establish an error matrix by statistics on classification image and actual image pixels, thus enabling the accurate verification of the classification of hyperspectral images. Its calculation formula is expressed as follows:

$$k = \frac{P_0 - P_e}{(1 - P_e)} \tag{38}$$

$$P_0 = \frac{1}{N} \sum_{i=1}^n P_i \tag{39}$$

where  $P_0$  is the overall classification accuracy;  $P_e$  is the probability of obtaining the correct classification caused by accident;  $n$  is the number of types;  $N$  is the sample size;  $P_i$  is the sample size of the  $i^{th}$  type that are classified accurately. When the classification image tends to be consistent with the actual image, the Kappa factor approaches one. The Kappa factor is relatively reliable because it is based on actual image classification. However, it has to conduct field surveys on actual surface types of hyperspectral images.

**3.1.3 Test method based on USGS mineral\_beckman\_430**

**3.1.3.1 Reference spectral library**

A spectral library is a set of spectral data of various surface features reflection (or radiation) collected by hyperspectral imaging spectrometer under certain conditions. With the growing studies on development and application of hyperspectral technology, higher requirements on spectral library are proposed. Representative surface feature spectral libraries in the world mainly include USGS, JPL, JHU, IGCP264, ASTER, etc. In this paper, spectral similarity measure method was tested by using the USGS mineral\_beckman\_430, which covers 430 spectral curves of 207 categories of minerals. The distribution of intra-category spectrum quantity is displayed in Table 3. The table shows a total of 300 spectra with intra-category quantity  $\geq 2$  and 148 spectra with intra-category quantity  $\geq 5$ . All spectral data were measured on a calibrated Beckman5270 spectrometer.

The spectral range of this Beckman5270 spectrometer is 0.2 - 3.0  $\mu\text{m}$ . All spectral reflectance data were corrected to absolute reflectance.

Mineral category of spectra is introduced in the mineral\_beckman\_430, which provides references for judging spectral classification accuracy. This overcomes the shortage of RSDPB and RSDE that cannot evaluate spectral identification accuracy. Therefore, 207 categories of minerals in the mineral\_beckman\_430 are almost covering all common minerals, which reflect the high reliability of the evaluation.

**Table 3.** Spectrum quantity of intra-category spectrum quantity in mineral beckman 430

intra-category spectrum quantity	1	2	3	4	5	6	7	8	9
In-library spectrum quantity	128	74	30	48	35	30	7	16	18
intra-category spectrum quantity	10	11	12	13	14	15	16	17	18
In-library spectrum quantity	0	11	0	13	0	0	0	0	18

**3.1.3.2 Test spectra**

Test spectra were chosen from the mineral\_beckman\_430 according to the following steps:

- (1) Set the mineral\_beckman\_430 as the spectra set S0. SID of spectra in S0 is calculated through cross-footing. Spectral curves with  $SID(S_0, S_0) \leq 0.01$  are chosen to form the spectra set S1;
- (2) Choose T from S1. T shall meet intra-category spectrum ( $SID(T, S_0) \leq 0.01$ ) quantity  $\geq 2$  and spectrum ( $SID(T, S_0) \leq 0.01$ ) quantity in S0  $\geq 10$ .

A total of 38 spectral curves were chosen from S0, which form the test spectrum set T (Table 4).

**Table 4.** Test spectra

No.	Name	category
8	Albite GDS30	Albite
9	Albite HS143.3B	Albite
10	Albite HS324.3B	Albite
11	Albite HS66.3B	Albite
70	Calcite CO2004	Calcite
71	Calcite HS48.3B	Calcite
72	Calcite WS272	Calcite
154	Grossular WS483	Grossular
155	Grossular WS484	Grossular
156	Grossular WS485	Grossular
215	Kaolinite CM3	Kaolinite
216	Kaolinite CM5	Kaolinite
261	Montmorillonite CM26	Montmorillonite
262	Montmorillonite CM27	Montmorillonite
266	Montmorillonite STx-1	Montmorillonite
267	Montmorillonite SWy-1	Montmorillonite
274	Muscovite GDS114	Muscovite
276	Muscovite GDS117	Muscovite
277	Muscovite GDS118	Muscovite
278	Muscovite GDS119	Muscovite
279	Muscovite GDS120	Muscovite
388	Talc TL2702	Talc
393	Topaz Cameron Cone #42	Topaz
394	Topaz Crystal Park #2	Topaz
395	Topaz Glen Cove #6	Topaz
396	Topaz Glen Cove #8	Topaz
398	Topaz Harris Park #17	Topaz
399	Topaz Harris Park #3	Topaz

400	Topaz Harris Park #9	Topaz
401	Topaz Jos #22	Topaz
402	Topaz Little 3 Mine #41	Topaz
404	Topaz Tarryalls #4	Topaz
405	Topaz Wigwam Area 2 #12	Topaz
406	Topaz Wigwam Area 3 #13	Topaz
407	Topaz Wigwam Area 4 #14	Topaz
408	Topaz Wigwam Area 5 #15	Topaz
409	Topaz Wigwam Area 6 #16	Topaz
410	Topaz Wigwam Area A #10	Topaz

**3.2 Test results**

In this paper, SAM, SID, SCC, SID-SAM, SID-SCA, JMD-SAM, ED, SSS, SPM, NS<sup>3</sup>, and fusing SAF were evaluated on the basis of the mineral\_beckman\_430. Among them, shape features (S) in fusing SAF were calculated through Equation (28) and amplitude features (A) in fusing SAF were calculated through Equations (30) and (31). Test results are shown in Table 5.

**Table 5.** Test results of similarity measure methods

Similarity measure methods		Identification of test spectra											Identified quantity	Identification rate	Mean identification rate
		10	11	70	71	72	155	261	262	267	274	393			
SFM	SAM	√	×	√	√	√	×	×	√	×	√	√	30	79%	79%
	SID	×	×	√	√	√	√	×	√	×	√	√	30	79%	
	SCC	√	√	×	√	√	√	×	√	×	√	√	31	82%	
	SID TAN SAM	×	×	√	√	√	×	×	√	×	√	√	29	76%	
	SID SIN SAM	×	×	√	√	√	×	×	√	×	√	√	29	76%	
	SID TAN SCA	√	×	√	√	√	√	×	√	×	√	√	31	82%	
	SID SIN SCA	√	×	√	√	√	√	×	√	×	√	√	31	82%	
	JMD TAN SAM	×	×	√	√	√	×	×	√	×	√	√	29	76%	
JMD SIN SAM	×	×	√	√	√	×	×	√	×	√	√	29	76%		
AFM	ED	√	×	×	×	×	×	√	×	√	×	×	26	68%	68%
Integrating SAF	SSS	√	×	×	√	√	√	√	√	√	√	√	32	84%	82%
	SPM	×	×	√	√	√	×	√	√	√	√	√	31	82%	
	NS <sup>3</sup>	√	×	×	×	×	×	√	×	√	×	×	26	68%	
	Fusing SAF (S1, A1)	√	×	√	√	√	√	√	√	√	√	√	33	87%	
	Fusing SAF (S1, A2)	√	×	√	√	√	√	√	√	√	√	√	33	87%	

For convenience, similar results of measure methods are omitted from Table 5. Five test spectra (No.8, 154, 156, and 266) cannot be identified by all methods, and 23 test spectra (No.9, 215, 216, 276-388, and 394-410) that can be identified by all methods. Test results demonstrate that:

- (1) Spectral similarity measure based on shape features identifies 29–31 spectra, showing an identification rate of 76%–82%, 79% in average. The overall identification accuracy is higher than the spectral similarity measure based on amplitude features but lower than that based on comprehensive shape and amplitude features.
- (2) Spectral similarity measure based on amplitude features identifies 26 spectra, showing an identification rate of 68%. Its overall identification accuracy is the lowest, thus indicating that amplitude feature-based identification is inferior to shape feature-based identification.
- (3) Spectral similarity measure based on comprehensive shape and amplitude features identifies 26–33 spectra, showing an identification rate of 68%–87%, 82% in average. The overall identification accuracy is higher than that based on shape features and amplitude features.
- (4) The proposed fusing SAF identifies 33 spectra, achieving an identification rate of 87%. It is confirmed having significantly higher spectral identification accuracy than in other methods.

**4. Conclusions**

According to definitions of shape features vector and amplitude features vector in this paper, spectral similarity measure methods are divided into SFM, AFM, and integrating SAF. The test results of spectral similarity measure methods based on the USGS mineral\_beckman\_430 reveal that integrating SAF shows the highest mean identification accuracy, followed by SFM and AFM successively.

The discussed spectral similarity measure methods integrating SAF such as SSS, SPM, and NS<sup>3</sup>, neglect the relation between shape feature distance vector and amplitude feature distance vector. They only combine results of shape features and amplitude features; thus, significantly influenced by weights of shape and amplitude features. Any improper weight selection will lead to single shape measure or amplitude measure. For example, NS<sup>3</sup> will obtain a result equal to the AFM under certain conditions.

Fusing SAF not only combines spectral shape and amplitude features, but also considers the relationship between shape feature distance vector and amplitude feature distance vector. Therefore, it achieves higher identification accuracy than SSS, SPM, and NS<sup>3</sup>, which is superior to a certain extent.

---

## References

1. Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., and Goetz, A. F. H., "The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data", *Remote sensing of environment*, 44(2), 1993, pp. 145-163.
2. Chang, C. I., "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis", *Information theory, IEEE transactions on*, 46(5), 2000, pp. 1927-1932.
3. van der Meer, F., and Bakker, W., "Cross correlogram spectral matching: application to surface mineralogical mapping by using AVIRIS data from Cuprite, Nevada", *Remote sensing of environment*, 61(3), 1997, pp. 371-382.
4. Du, Y., Chang, C. I., Ren, H., Chang, C. C., Jensen, J. O., and D'Amico, F. M., "New hyperspectral discrimination measure for spectral characterization", *Optical engineering*, 43(8), 2004, pp. 1777-1786.
5. Naresh Kumar, M., Seshasai, M. V. R., Vara Prasad, K. S., Kamala, V., Ramana, K. V., Dwivedi, R. S., and Roy, P. S., "A new hybrid spectral similarity measure for discrimination among Vigna species", *International journal of remote sensing*, 32(14), 2011, pp. 4041-4053.
6. Padma, S., and Sanjeevi, S., "Jeffries Matusita based mixed-measure for improved spectral matching in hyperspectral image analysis", *International journal of applied earth observation and geoinformation*, 32, 2014, pp. 138-151.
7. Zhang, J., Zhu, W., and Dong, Y., "A spectral similarity measure based on changing-weight combination method", *Acta Geodaetica et Cartographica Sinica*, 42(3), 2013, pp. 418-424.
8. Lhermitte, S., Verbesselt, J., Verstraeten, W. W., and Coppin, P., "A comparison of time series similarity measures for classification and change detection of ecosystem dynamics", *Remote sensing of environment*, 115(12), 2011, pp. 3129-3152.
9. Vishnu, S., Nidamanuri, R. R., and Bremananth, R., "Spectral material mapping using hyperspectral imagery: a review of spectral matching and library search methods", *Geocarto international*, 28(2), 2013, pp. 171-190.
10. Granahan, J. C., and Sweet, J. N., "An evaluation of atmospheric correction techniques using the spectral similarity scale", *In geoscience and remote sensing symposium, IGARSS'01, IEEE*, 5, 2001, pp. 2022-2024.
11. Kong, X., Shu, N., Tao, J., and Gong, Y., "A new spectral similarity measure based on multiple features integration", *Spectroscopy and spectral analysis*, 31(8), 2011, pp. 2166-70.
12. Nidamanuri, R. R., and Zbell, B., "Normalized spectral similarity score (NS<sup>3</sup>) as an efficient spectral library searching method for hyperspectral image classification", *Selected topics in applied earth observations and remote sensing, IEEE*, 4(1), (2011), pp. 226-240.
13. Fang, S., and Gong, H., "Spectral similarity scale based on dynamic weighting adjustment method", *Geomatics and information science of Wuhan university*, 31, 2006, pp. 1044-1046.
14. Singh, K. D., Ramakrishnan, D., and Mansinha, L., "Relevance of transformation techniques in rapid endmember identification and spectral unmixing: A hyperspectral remote sensing perspective", *In geoscience and remote sensing symposium (IGARSS)*, 2012, pp. 4066-4069.
15. Abdi, M., and Karimpour, M. H., "Application of Spectral Angle Mapper Classification to Discriminate Hydrothermal Alteration in Southwest Birjand, Iran, Using Advanced Spaceborne Thermal Emission and Reflection Radiometer Image Processing", *Acta geologica sinica (english edition)*, 86(5), 2012, pp. 1289-1296.
16. Zhang, E., Zhang, X., Yang, S., and Wang, S., "Improving hyperspectral image classification using spectral information divergence", *Geoscience and Remote Sensing Letters, IEEE*, 11(1), 2014, pp. 249-253.
17. Chang, C. I., "Hyperspectral imaging: techniques for spectral detection and classification", *Kluwer Academic/ Plenum Publishers*, 2003.
18. Ghiyamat, A., Shafri, H. Z. M., Mahdiraji, G. A., Shariff, A. R. M., and Mansor, S., "Hyperspectral discrimination of tree species with different classifications using single-and multiple-endmember", *International journal of applied earth observation and geoinformation*, 23, 2013, pp. 177-191.
19. Bahria, S., Essoussi, N., and Limam, M., "Hyperspectral data classification using geostatistics and support vector machines", *Remote Sensing Letters*, 2(2), 2011, pp. 99-106.