International Journal of Applied Earth Observation and Geoinformation 50 (2016) 141–149

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

A method to compute the n-dimensional solid spectral angle between vectors and its use for band selection in hyperspectral data

M. Tian [a], J. Feng [b], B. Rivard [b,*], C. Zhao [a]

[a] College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China
[b] Centre for Earth Observation Sciences, Department of Earth and Atmospheric Sciences, University of Alberta, 1-26 Earth Sciences Building, University of Alberta, Edmonton, Alberta T6G 2E3, Canada

ARTICLE INFO

Article history:
Received 17 November 2015
Received in revised form 16 March 2016
Accepted 26 March 2016
Available online 4 April 2016

Keywords:
Hyperspectral
Solid angle
Band selection
Collinearity
Unmixing

ABSTRACT

This study presents the calculation of spectral angle beyond two endmember vectors to the *n*-dimensional solid spectral angle (NSSA). The calculation of the NSSA is used to characterize the local spectral shape difference among a set of endmembers, leading to a methodology for band selection based on spectral shape variations of more than two spectra. Equidistributed sequences used in the quasi-Monte Carlo method (ESMC) for numerical simulations are shown to expedite the calculation of the NSSA. We develop a band selection method using the computation of NSSA(ϑ_n) in the context of a sliding window. By sliding the window over all bands available for varying band intervals, the calculated solid spectral angle values can capture the similarity of the endmembers over all spectral regions available and for spectral features of varying widths. By selecting a subset of spectral bands with largest solid spectral angles, a methodology can be developed to capture the most important spectral information for the separation or mapping of endmembers. We provide an example of the merits of the NSSA-ESMC method for band selection as applied to linear spectral unmixing. Specifically, we examine the endmember abundance errors resulting from the NSSA band selection method as opposed to using the full spectral dimensionality available.

1. Introduction

Hyperspectral data typically consists of hundreds of contiguous spectral bands and provides an ability to capture spectral features unique to varying image endmembers that can be used for mapping and detection. The spectral angle mapper algorithm (SAM) is one of the most commonly used mapping tools for analysis of hyperspectral data in the earth sciences. SAM measures the spectral similarity between two vectors, a reference spectrum and a target spectrum for a given number of bands (Kruse et al., 1993).

This first contribution of this study is to expand the calculation of spectral angle beyond two spectra to the *n*-dimensional solid spectral angle (NSSA) where *n* refers to the number of spectra. We therefore develop the method for calculating the NSSA. The method differs from the SAM method in that it is able to calculate the solid angle defined by a set of more than two endmember vectors.

As a second contribution, we then use the calculation of the NSSA to characterize the local spectral shape difference among a set of endmembers leading to a methodology for band selection. For many hyperspectral applications, spectral features are the basis for distinguishing and identifying materials (e.g. minerals and rocks, plant species or functional groups) and thus the use of band subsets or spectral regions can drive specific mapping goals. Determining a subset of bands pertinent to distinguish multiple classes or endmembers of interest becomes key to the success of spectral analysis for a given application. This relates to the collinearity or redundancy of adjacent bands that is a general preoccupation in hyperspectral analysis (Van der Meer and Jia, 2012). Reducing data redundancy while minimizing the impact on mapping accuracy is a challenge.

Band selection involves the selection of a subset of the original spectral bands using certain criteria. Several studies have focused on creating selection criteria based on information theory or spectral variance. Examples include using criteria based on information entropy (Bajcsy and Groves, 2004), information divergence (Ball et al., 2007), and mutual information (Guo et al., 2006; Kamandar and Ghassemian, 2011) to evaluate the information content of different bands. Other studies have addressed various issues of the band selection process. For example, Chang et al. (1999) used spectral variance to sort band importance; Sun et al. (2014) used a band quality index to address inter-band correlation and issues of signal-to-noise ratio in band selection; Qian et al. (2009) used affinity propagation to characterize representative bands; Tan et al. (2014) tried to integrate spatial and spectral information into the selection process; and Geng et al. (2014) calculated a volume-gradient to search for bands with large dissimilarity. Particle swarm optimization algorithms (e.g. Su et al., 2014; Yang et al., 2012), a

* Corresponding author.
E-mail address: benoit.rivard@ualberta.ca (B. Rivard).

http://dx.doi.org/10.1016/j.jag.2016.03.012
0303-2434/

simulated annealing approach (e.g. Chang et al., 2011), a nature-inspired framework (e.g. Nakamura et al., 2014), sparse theory (e.g. Sun et al., 2015; Zare and Gader, 2008), and forward and sequential search strategies (e.g. Du and Yang, 2008) were also explored for band selection. Jia et al. (2013) presented a detailed review of feature mining methods.

Most of these studies treat each band as an independent variable. However, it has long been recognized, certainly in mineral spectroscopy, that the shape of spectral features (position, width, asymmetry) from different scene materials should also be considered as variables. This is illustrated in several recent studies (e.g. Chang and Liu, 2014; Van der Meer, 2004; Wang and Chang, 2007). Thus band selection methods, that not only capture the presence of key absorption features but are also sensitive to the shape of spectral features, can offer advantages for specific domains including mineral detection. The band add-on (BAO) approach, proposed by Keshava (2004), is based on a measure of the spectral angle for an endmember pair. Key to the BAO method is the selection of a band set that maximizes the spectral angle between two endmembers. An impediment of this method is that spectral angle can only be calculated for 2 spectra/endmembers at a time thus a hierarchical processing architecture has to be defined to obtain a final result. Furthermore, BAO is a heuristic band selection method that relies on the selection of an initial band set, thus may not yield unique final band selection results and the initial band set impacts the final outcome.

Ideally a band selection method would be able to select an optimal band subset that (a) captures the dominant spectral differences for all classes/endmembers, (b) is sensitive to spectral shape differences rather than amplitude difference, the latter being dramatically affected by non-compositional factors such as illumination and imaging geometry, and (c) improves the overall accuracy of mapping or classification results. The proposed NSSA based band selection method aims to capture the overall spectral shape difference amongst all endmembers investigated independent of the overall spectral amplitude. By selecting a subset of spectral bands with largest NSSA, a methodology can be developed to capture the most important spectral information for the separation or mapping of a suite of endmembers.

Following an explanation of the calculation of the NSSA and its use for band selection we provide an example applied to linear spectral unmixing. Van der Meer and Jia (2012) recently addressed the challenges imposed by the collinearity of endmembers in linear spectral unmixing that can impact the reliability of fractional abundance estimates. We therefore examine endmember abundance errors resulting from the NSSA band selection method as opposed to using the full spectral dimensionality available.

2. Background

2.1. Spectral angle mapper algorithm (SAM)

The SAM is commonly used to measure the spectral similarity between two vectors, a reference spectrum and a target spectrum for a given number of bands (Kruse et al., 1993). The angle is computed as:

$$SAM(\boldsymbol{x},\boldsymbol{y}) = \cos^{-1}\frac{<\boldsymbol{x},\boldsymbol{y}>}{||\boldsymbol{x}||||\boldsymbol{y}||} \tag{1}$$

where **x** and **y** represent the reference and target spectrum. A smaller angle represents a higher similarity. A benefit of the SAM algorithm is that it is insensitive to multiplicative gain factors, which impact the vector length rather then the vector direction. Thus, spectral variation introduced by changes in illumination due to topography, or target particle size in the case of mineral substrates, do not influence SAM results. SAM is therefore used to quantitatively highlight differences in spectral shape (Crosta et al., 1998; Sohn and Rebello, 2002; Van der Meer et al., 1997) but it can only calculate an angle between two vectors.

2.2. Mathematical framework for the estimation of the n-dimensional solid angle

The solid angle between three or more vectors can be described mathematically. The positive linear combinations of an independent vector set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \cdots, \boldsymbol{x}_n\}$ in n-dimensional Euclidean space constitute a polygonal cone C_n (Rockafellar, 1970),

$$C_n = \{\boldsymbol{x} : \boldsymbol{x} = \sum_{i=1}^{n} \varpi_i \boldsymbol{x}_i,\ \varpi_i \geq 0\} \tag{2}$$

where $\{\varpi_i\}$ is a positive weighted coefficient set. The solid angle is defined as the measure of the intersection of a polygonal cone C_n and the corresponding unit sphere surface S_{n-1} (Hajja and Walker, 2002). Fig. 1 illustrates the solid angle in 2, 3 and n (n is an integer and larger than 3) dimensions. In this paper, each vector has a dimension (number of bands) and the letter n refers to the dimensionality or number of spectral vectors (e.g. endmembers). As shown in Fig. 1, for 2-dimensional Euclidean space, the solid angle ϑ_2 (in radians) between two vectors is numerically equal to the intersection of cone C_2 and unit circle S_1, which is an arc. This is the angle measured by the SAM algorithm. For three vectors, the solid angle ϑ_3 (Fig. 1) basically corresponds to a section of spherical surface on unit sphere S_2. When $n > 3$ simple visual descriptors fail because the intersection is on a hypersphere. However, the calculation of a solid angle ϑ_n constituted by n independent vectors in n-dimensional Euclidean space was solved mathematically (Hajja and Walker, 2002). Thus far, no research has explored the application of this mathematical tool in analysis of hyperspectral data.

3. N-dimensional solid spectral angle (NSSA)

To obtain the solid spectral angle for n endmember spectra, the spectra are represented by an endmember vector matrix $\boldsymbol{E} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \ldots, \boldsymbol{e}_n\}$ in which every column endmember vector is normalized (length equals 1.0). The n-dimensional solid spectral angle (NSSA) for the n endmembers is defined as:

$$\mathrm{NSSA}(\vartheta_n) = |\det(\boldsymbol{E})| \int_s ||\mathbf{EV}||^{-n} ds \tag{3}$$

where $\mathrm{NSSA}(\vartheta_n)$ is the n-dimensional solid spectral angle (in radians) and $||\bullet||$ is the Euclidean norm. The determinant is necessary for the normalization, a procedure essential to ensure that the calculation of the $\mathrm{NSSA}(\vartheta_n)$ occurs on the related unit sphere S_{n-1}. $\boldsymbol{V} = [v_1, v_2, v_3, \cdots, v_i, \cdots, v_n]^{\mathrm{T}}, v_i > 0$ is the spherical vector of a unit sphere for n dimensions expressed in polar Cartesian coordinates. Specifically,

$$\begin{cases} v_1 = \cos\theta_1, & 0 \leq \theta_1 \leq \pi/2 \\ v_2 = \sin\theta_1 \cos\theta_2, & 0 \leq \theta_2 \leq \pi/2 \\ v_3 = \sin\theta_1 \sin\theta_2 \cos\theta_3, & 0 \leq \theta_3 \leq \pi/2 \\ \ldots\ldots \\ v_{n-1} = \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{n-2} \cos\theta_{n-1}, & 0 \leq \theta_{n-1} \leq \pi/2 \\ v_n = \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{n-2} \sin\theta_{n-1}, & 0 \leq \theta_{n-1} \leq \pi/2 \end{cases} \tag{4}$$

The differential element ds is the surface area of a n-1 dimensional sphere S_{n-1} and is expressed as:

$$ds = \sin^{n-2}(\theta_1)\sin^{n-3}(\theta_2), \cdots, \sin(\theta_{n-2}) d\theta_1 d\theta_2 \cdots d\theta_{n-1} \tag{5}$$