A NEW BAND SELECTION ALGORITHM FOR HYPERSPECTRAL DATA BASED ON FRACTAL DIMENSION

Hongjun Su [a, *], Yehua Sheng [a], Peijun Du [b]

[a] Key Laboratory of Virtual Geographic Environment (Ministry of Education), Nanjing Normal University, Nanjing, Jiangsu Province, 210046, P.R.China -hjsu1@163.com
[b] Department of Remote Sensing and Geographical Information Science, China University of Mining and Technology, Xuzhou, Jiangsu Province, 221008, P.R.China -dupjrs@hotmail.com

KEY WORDS: Hyper spectral; Imagery; Geometric information; Feature extraction; Algorithms

ABSTRACT:
Feature selection especially band selection plays important roles in hyperspectral remote sensed image processing. It is worth nothing that band selection approaches need to be combined with image spatial structure information so as to select valid bands and improve the performance. But all of the existing remote sensing data processing algorithms are used for the conventional broadband spectral data and can not process high dimensionality data effectively and accurately. According to the characteristic of HRS data, the algorithm which named optimal band index (OBI) based on fractal dimension was put forward in this paper. In OBI algorithm, firstly, the fractal dimension was used as the criterion to prune the bands which have noises, and the bands which have better spatial structure, quality and spectral feature were reserved. After that, the correlation coefficients and covariance among all bands were used to compute optimal band index, and then the optimum bands were selected. At last, in the experiment the proposed algorithm was compared with the other two algorithms (Adaptive Band Selection and Band Index), it proves that the OBI algorithm can work better on the band selection in hyperspectral remote sensing data processing than other algorithms.

1. INTRODUCTION

Hyperspectral remote sensing (HRS) technique is one of the breakthroughs in the earth observation in the 20th century 80s. It is well-known that hyperspectral has the characteristics of contiguous spectral range and narrow spectrum interval. Compared to the traditional broadband remote sensing, it can provide fine, detail and large volume spectral information, and could be used in target recognition, endmember extraction, state diagnosis and other sophisticated applications (R.L. Pu and P. Gong,2000; P.J. Du et al.,2007 ;H.J. Su et al.,2007). The more applications in many fields and the prosperous perspective of hyperspectral remote sensing call for advanced development of hyperspectral information processing technology, especially in how to deal with the high dimensionality and huge data.

The rapid development of the hyperspectral remote sensing is providing the useful data for the applications and demand to develop the corresponding data processing algorithm. But at present, hyperspectral remote sensing data processing technique has its insufficiency in development and cannot satisfy the requirement of hyperspectral applications. Specifically, because of high dimensionality of hyperspectral data, Hughes phenomenon i.e. the curse of dimensionality will be arises when using traditional algorithm to process hyperspectral data. Hughes phenomenon means that when the training sample number is a constant, the precision of classification will be decreased with the increasing of the Eigen-dimensionality. In order to process hyperspectral data effectively, it is necessarily to reduce the dimensionality of hyperspectral data. For hyperspectral remote sensing, dimensionality reduction references to compress the bands number without the decrease of the useful information. There are two ways to compress the bands number; one is feature selection that is to select some interesting bands or the bands with the more information and weak correlation relationships among all bands. The other is feature extraction, to compress all the bands using mathematic transformation.

For many years, the design of efficient and robust feature extraction and feature selection especially band selection algorithms has been the most important issue addressed by remote sensed image users. Strong effort has been devoted to elaborate new band selection algorithms and improve techniques used to reduce dimensionality. Many band selection algorithms in feature selection for hyperspectral remote sensing were proposed in the past years. Martínez-Usó A et al. developed a novel technique which using mutual information-based clustering to deal with multispectral images; in this algorithm, a distance based on mutual information is used to construct a hierarchical clustering structure with the multispectral bands. Moreover, a criterion function is used to choose automatically the number of final clusters (Martínez-Usó A et al., 2006), Also band selection method based on mutual information (MI) was developed (B.F. Guo et al., 2006). Wang et al. proposed a novel entropy-based band selection method which used spectral gradient based on entropy to define the average normalized information carried by each set of selected bands, and developed a fast derivative descend algorithm to search for the optimal solution by maximizing the average normalized information (Wang et al.,2004). In 2005, a statistical procedure to provide optimal sensor settings for a classification task at hand was developed. The procedure selects wavelength band settings which optimize the separation between the different spectral classes. The method is applicable as a band reduction technique, but it can as well serve the purpose of data interpretation or sensor design (De Backer S. et al., 2005). The band selection algorithm based on constrained energy minimization (CEM) was explored, the algorithm

* Corresponding author. Email: hjsu1@163.com, hjsurs@hotmail.com, Tel: +86-13851706937

referred to as constrained band selection (CBS) for hyperspectral imagery. It interprets a band image as a desired target signature vector while considering other band images as unknown signature vectors (Chein-I Chang et al., 2006). Alam and Ochilov also presented a new technique for adaptive band selection from hyperspectral image cubes for detecting small targets using an anomaly detector (Alam and Ochilov, 2005). Enve in 2003, a band selection method based on Independent Component Analysis (ICA) was presented and as a significant benefit, the ICA-based band selection retains most physical features of the spectral profiles given only the observations of hyperspectral images (H.T. Du et al., 2003).

But all of the existed remote sensed data processing algorithms are used for the conventional broadband spectral data and can not process high dimensionality data effectively and accurately. It is urgent and necessarily to develop the new algorithms which can be used in high dimensional data. And recent works on spectral band selection involves two separate tasks: feature bands selection and redundancy reduction. However, a better subset of bands can be determined if these two tasks are considered together instead of separately.

In this paper, our research aims at finding an effective band selection algorithm for it. According to the characteristic of HRS data, the algorithm which named optimal band index (OBI) based on fractal dimension was put forward. This algorithm used the fractal dimension which describing the quality of remote sensing images as the weighed index, and then used the OBI as band selection criterion.

2. DATASETS

The dataset used in our experiments is consists of some spectral curves from OMIS aerial hyperspectral remote sensed imagey. The data was obtained by Operational Modular Imaging Spectrometer (OMIS) in Beijing. It has 64 bands and image size is 536×512. The spectral range is from 460 to 1100nm, and the wavelength interval is 10nm. It is located by geo-coordinates 116.258722E and 40.262377N at upper left corner. Figure 1 has shown the cube image of hyperspectral image about study area (false color synthesized image by channel 12, 24, 20).

Figure 1. Hyperspectral image of study area

3. METHODOLOGIES

3.1 Fractal geometry in remote sensing

Fractal theory is a useful tool for science research, and fractal research is a fairly new field of interest. The fractal objects have two fundamental characteristics. One is it has a fine structure at arbitrarily small scales. The underlying principle of fractals is that a simple process that goes through infinitely much iteration becomes a very complex process. Fractals attempt to model the complex process by searching for the simple process underneath. The other is it is self-similar at least approximately and almost all fractals are at least partially self-similar. This means that a part of the fractal is identical to the entire fractal itself except smaller. Fractal theory is mainly study the disorder and self-similar systems in the nature and social activities.

Because it is practical and valid to describe the internal regulations of irregular phenomenon in the nature, Fractal dimension has been used in many fields. It is no doubt that remote sensing information is the reflection of the objects in nature, so it also has the characteristics of spatial self-similarity and time serial fractal dimension. In addition, remote sensing image also has fractal characteristics (D. Jiang et al., 2000). In fact, the remote sensed imagery is the projection of 3D objects on a 2D surface and is the fractal Brown-surface, and the fractal dimension of the image is equal to the fractal normal of spatial 3D objects. So the fractal dimension of remote sensed imagery can express the fractal characteristics of true ground surface.

Generally speaking, there are two aspects of applications about the fractal. One is to simulate the objects and terrain in nature, by add more details about existed data using interpolation methods. The other is to represent the complexity of the spatial information about objects. In this paper, our research in calculating the fractal dimension of remote sensed imagery belongs to the latter one.

The changes of object details can be described by a fractal dimension *D*. Fractal dimension is the parameters of quantitative characterization of fractal, which is the measurement for the fractal object internal uneven, the quantitative characteristics of overall level of structural, the complexity of fractal and the rough and chronology of the objects. The fractal dimension *D* can be calculated by

$$D = \ln n / \ln(1/s) \tag{1}$$

Where, n is the number of subpart in the whole object and s is the scaling factor. Fractal dimension reflects the spatial structure of geo-objects, and it is important for the applications of remote sensing. The information extracted from images is varying with the different spatial structures of images. So it is possible to select band by band selection index based on fractal dimension in hyperspectral remote sensing.

3.2 Fractal dimension for remote sensed imagery band selection

Some fractal dimension calculation methods can be used in remote sensing, such as line divider method, triangular prism method and double blanket method. In this paper, we used the Double Blanket Method (DBM) (H.G. Zhang et al., 2005).

According to the fractal theory, the surface of the remote sensed imagery has the characteristic of fractal in some scaling extent.

In order to calculate the fractal dimension of the remote sensed imagery, two “blankets” are designed near the top and bottom of the image surface in some distance, and the blankets enveloped the image surface. Then, the volume between the two blankets can be obtained and the surface area of the two blankets can be calculated. It is should be promised that the surface area of the two blankets are varying with the diverse scales.

The idea of this algorithm is, supposing that the gray values of pixel(X, Y) in the images are a whole surface, two blankets are put near the top and bottom of remote sensed imagery surface in some distance respectively and enveloped the surface. Then, the volume between the two blankets can be get and the surface area of the two blankets can be derived from the volume.

In the initial step, the two blankets are superposed and became the same layer. That is,

$$f(i,j) = u_0(i,j) = d_0(i,j) \tag{2}$$

Where $u_0(i,j)$ and $d_0(i,j)$ are the value of top-blanket and bottom-blanket in the place (i,j) of image when the scale is 0. The value of top-blanket and bottom-blanket can be defined by formula (3) and formula (4) when the scale $\varepsilon = 1, 2, \ldots$:

$$u_\varepsilon(i,j) = \max\{u_{\varepsilon-1}(i,j)+1, \max[u_{\varepsilon-1}(m,n)]\} \tag{3}$$

$$d_\varepsilon(i,j) = \min\{d_{\varepsilon-1}(i,j)-1, \min[d_{\varepsilon-1}(m,n)]\} \tag{4}$$

Where, (m,n) represents the four neighbour points of (i,j). The volume between two blankets can be gotten by:

$$V_\varepsilon = \sum_{i,j}[u_\varepsilon(i,j) - d_\varepsilon(i,j)] \tag{5}$$

And the surface area can be calculated by:

$$A(\varepsilon) = V_\varepsilon / (2*\varepsilon) \tag{6}$$

After the above steps, a serial list of “scale-surface area” can be achieved, and the slope *K* of the regression line can be derived by logarithm regress for the scale and surface area. The fractal dimension of the remote sensed imagery can be gotten by *D=2.0-K*.

4. OPTIMIZE BAND INDEX BASED ON FRACTAL DIMENSION

The band selection algorithms used widely can be categorized into two classes, one is band selection based on the information, such as entropy and combined entropy, optimize index factor (OIF), covariance matrix of combined bands; and the other is band selection based on separability of the same classes, such as dispersibility, standard distance between mean values, B distance, separability of the same classes and so on (C.H.Zhao et al., 2004; J.P. Liu et al., 2001).

Generally speaking, there are three principles for band selection. Firstly, the information in the selected band should be optimized and achieved the max value; secondly, the relativity between the selected band and other bands should be minimized; thirdly, the spectral response characteristics of the objects in the study area can distinguish the interesting objects easily (J.P. Liu et al., 2001; X.G. Jiang et al., 2002). Those bands that have the abundance information, minimized relativity, more discrepancy and separability are the best bands for band selection algorithm. It is should point out that an excellent band selection algorithm should take information in each band, relativity among all bands and standard deviation coefficient of band gray into account.

There are many band selection algorithms in feature selection for hyperspectral remote sensing. Take band index (BI) which was put forward based on the relationship coefficient matrix by X. G. Jiang (2002) for an example, it has considered the information content and relativity of different bands, but the index was designed for the characteristics of block-structure in HRS image data. So when we divided all bands into several subspaces, the dimension of these subspaces was disparate, and because of the result of BI was influenced by the factitious factors. What’s more, if we do not take care of the spatial structure information or have not pre-process noise bands for the image; perhaps it would badly impact on the results (H.J. Su et al., 2007).

Supposing there are *N* spectral bands in the whole hyperspectral space and are divided into k groups. The numbers of bands in each group are *n1*, *n2*, *n3*,...,*nk*, the BI can be stated as:

$$P_i = \frac{\sigma_i}{R_i} \tag{7}$$

$$R_i = R_w + R_a \tag{8}$$

Where σ_i is the RMS (error of mean square) of the i^{th} band, R_w is the mean of the sum of the absolute value of relationship coefficients between the i^{th} band and other bands which in the same group. R_a is the sum of the absolute value of relationship coefficients between the i^{th} band and other bands which in the different groups.

Based on the above analysis and considering the correlativity of spatial and spectra information adequately, a new algorithm, named Optimal Band Index (OBI) was put forward in this paper, which was used as the criterion to select characteristic bands. In OBI algorithm, firstly, the fractal dimension was used as the criterion to prune the bands which have noises, and the bands which have better spatial structure, quality and spectral feature were reserved. After that, the correlation coefficients and covariance among all bands were used to compute optimal band index, and then the optimum bands were selected. The detailed steps were as follows: Firstly, the DBM method was used to compute the fractal dimension of HRS image, and the whole bands can be divided into several subspaces according to the decrease trend of the OBI index. Finally, the OBI is used for characteristic features band selection.

After the whole space of the image is divided into some subspace by fractal dimension of the image, and then the optimize band index are used for band selection:

$$OBI_i = \frac{\sigma_i}{R_{s,i}} \tag{9}$$

Where OBI_i is the optimize band index of the i^{th} band, σ_i is the covariance the i^{th} band, $R_{s,i}$ is the sum of the relationship between the i^{th} band and the other bands. The value of optimize index of each band is calculated and sorted by descending.

5. EXPERIMENTS AND DISCUSSIONS

5.1 Experiments

In order to validate the algorithm we proposed in the paper, some experiments are carried out. For the comparison, some other traditional algorithms which mostly based on the information statistics parameters such as mean, standard deviation coefficient and so on are also used in the experiments.

At last, the proposed algorithm was compared with the other two algorithms, Adaptive Band Selection (ABS) (C.H. Liu et al., 2005) and Band Index (BI), the results are list in table 1(only top ten bands).

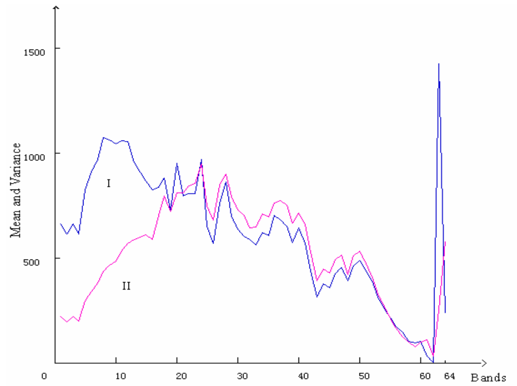

Figure 2. The mean and variance of each band

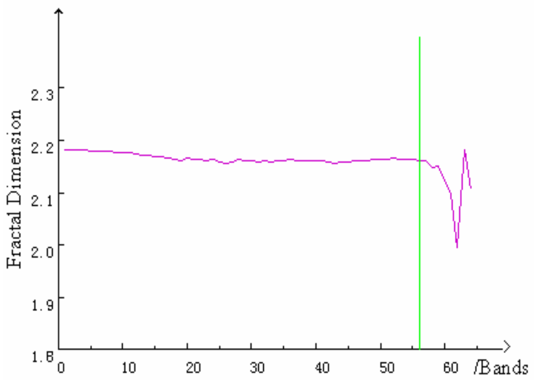

Figure 3. The fractal dimension of each band of HRS image

It is known that the mean value reflects the lightness information of image band and standard deviation coefficient represents the information underlying in the band, so the mean values and standard deviation coefficient (variance) of each band are also derived. Shown in Figure 2, where curve I is the mean of the spectral lightness and curve II is the standard deviation coefficient of the spectral lightness. Figure 3 has shown the results of fractal dimension curve calculated by DBM method.

5.2 Discussions

The fractal dimension reflects the characteristics of the objects surface, and for the remote sensed imagery, it interprets the spatial complexity. The experiments results for fractal dimension are mainly between 2 and 3 except one value is 1.9964. From the curve in the figure4, we can know that the curve is changed placidly from band 1 to band 57; while from band 58 to band 64, the curve is waved acutely. It is known that the changes of fractal dimension indicate the change directions for spatial structure, quality and spectral characteristics of remote sensed imagery, so the images quality from band 1 to band 57 are better for those from band 58 to band 64.

The result in Figure 3 are inosculating with the results in figure 2, it is because the larger standard deviation coefficient, the more dispersed degree and information underlying in bands. At the same time, the minimized absolute values for the relationship among all bands illustrate the more independency and redundancy in data and information. So the OBI can reflect information content and relationships among all bands synthetically, and is a useful tool for band selection.

Sort order	ABS	BI	OBI
1	24	21	24
2	28	20	28
3	23	18	27
4	27	19	23
5	22	17	18
6	20	15	22
7	21	14	20
8	18	16	21
9	29	13	29
10	37	24	37

Table 1. The band results using different band selection algorithms

As shown in table 1, the results derived from ABS and OBI is almost equally, for example the top 10 bands are equal completely except the sorting order. For BI algorithm, there are the same four bands compare to the ABS and OBI algorithms. What's more, among the top five bands from the table 1 and figure 3, the max standard deviation coefficient and min standard deviation coefficient are similar for different algorithms, for example for ABS algorithm they are 947.066 and 843.136, for OBI algorithm are 947.066 and 794.514,for BI algorithm are 812.619 and 703.881. It proves that the OBI algorithm can work better on the band selection in hyperspectral remote sensing data processing than other algorithms. The algorithm proposed in this paper is effective tool for band selection because it takes spatial correlation and spectral correlation among all bands in the remote sensed imagery.

6. CONCLUSIONS

The fractal dimension can quantificational interprets spatial structure and spatial complexity of remote sensed imagery. And the fractal characteristics on spatial and energy of sensors and observation objects of remote sensing results the comprehensive applications trends. In order to process hyperspectral remote sensing data quickly, a new algorithm, named OBI was put forward in this paper based on the traditional algorithm and fractal dimension. In OBI algorithm, firstly, the fractal dimension was used as the criterion to prune the bands which have noises, and the bands which have better spatial structure, quality and spectral feature were reserved. After that, the correlation coefficients and covariance among all bands were used to compute optimal band index, and then the optimum bands were selected. At last, in the experiment the proposed algorithm was compared with the other two algorithms (ABS and BI), it proves that the OBI algorithm can work better on the band selection in hyperspectral remote sensing data processing than other algorithms.

REFERENCES

Alam M.S. and Ochilov S, 2005. Adaptive hyperspectral band selection. *SPIE-The International Society for Optical Engineering, Vol5908*.

Baofeng Guo, Steve Gunn, Bob Damper et al, 2006. Band Selection for Hyperspectral Image Classification Using Mutual Information.*IEEE Geoscience and Remote Sensing Letters*,3(4), pp. 522-526.

Chein-I Chang and Su Wang, 2006. Constrained Band Selection for Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44(6), pp.1575-1585.

Chunhong Liu, Chunhui Zhao and Lingyan Zhang, 2005. A New Method of Hyperspectral Remote Sensing Image Dimensional Reduction(in Chinese). Journal of Image and Graphics, 10(2):218-222.

Chunhui Zhao and Chunhong Liu, 2004. Research and Analysis of Hyperspectral Remote Sensing Image Dimensional Reduction (in Chinese).*Chinese space science and technology*, vol(5):28-36.

De Backer S., Kempeneers P., Debruyn W., et al, 2005. Band selection for hyperspectral remote sensing. *In: IEEE geoscience and remote sensing letters*,2:(3), pp.319-323.

Dong Jiang and Jianhua Wang, 2000. The fractal thought in remote sensing information science(in Chinese).*Journal of Gansu Sciences*,12(1):53-57.

Hongjun Su,Yehua Sheng, Peijun Du,2007. Study on Auto-Subspace Partition for Band Selection of Hyperspectral Image (in Chinese). *Geo-Information Science*, 9(4):123-128.

Hongjun Su, Yehua Sheng, Yongning Wen, et al, 2007. Study on Data Mining Technology in Hyperspectral Remote Sensing. *in:Geoinfomatics2007: Remotely sensed Data and Information, Proceedings of SPIE*, Vol.6752, 67520R.

Hongtao Du, Hairong Qi, Xiaoling Wang, et al, 2003. Band Selection Using Independent Component Analysis for Hyperspectral Image Processing. *32nd Applied Imagery Pattern Recognition Workshop (AIPR2003)*.pp.93.

Huaguo Zhang, Weigen Huang,Changbao Zhou, et al, 2005. Fractal Characterization of IKONOS Imagery (in Chinese). *Bulletin of Surveying and Mapping*, 5:15-18.

Jianping Liu, Yingshi Zhao and Shuling Sun,2001. Exper imen tal Studies aboutMethods on Optimal Bands Selection of Hyperspectral Remote Sensing Datasets(in Chinese). *Remote Sensing Technology and Application*, 16(1):7-13.

Martínez-Usó A, Pla F, Sotoca JM, García-Sevilla P, et al, 2006. Automatic Band Selection in Multispectral images using Mutual Information-based Clustering, *CIARP'2006*, Cancun, Mexico, pp.644-654.

Peijun Du, Hongjun Su and Wei Zhang, 2007. Spectral Features Recognition Based on Data Mining Algorithms. *in:Geoinfomatics2007: Remotely sensed Data and Information, Proceedings of SPIE*, Vol.6752, 675204.

Rui Liang Pu and Peng Gong,2000. Hyperspectral Remote Sensing and its Application (in Chinese). *Beijing: Higher Education Press*.

Wang Hongzhi and Angelopoulou, *2004*. Elli. Average Normalized Information: a Novel Entropy-Based Band Selection Method. *Stevens Institute of Technology Technical Report, CS, 13*.

Xiaoguang Jiang, Lingli Tang and Changyao Wang, 2002. Spectral Characteristics and Feature Sel ection of Hyperspectral Remote Sensing Data--Takjng Shunyi Region of Beijing as a Study Area(in Chinese).*Remote Sensing Technology and Application*, 17(2):59-65.

ACKNOWLEDGEMENTS

We thank the National Natural Science Foundation of China (No. 40401038) and the Major Program of Basic Research of University of Jiangsu Province (No.07KJA42005) for providing the funds for this study.