We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

An Unsupervised Classification Method for Hyperspectral Remote Sensing Image Based on Spectral Data Mining

Xingping Wen and Xiaofeng Yang

Additional information is available at the end of the chapter

1. Introduction

Hyperspectral remote sensing is one of the most significant recent breakthroughs in remote sensing. It obtains image in a large number (usually more than 40), narrow (typically 10 to 20 nm spectral resolution) and contiguous spectral bands to enable the extraction of spectral information at a pixel scale, so it can produce data with sufficient spectral resolution for the direct recognition those materials with diagnostic spectral features [1]. Usually classification method of hyperspectral remote sensing data are divided into two categories [2]: using sub-pixel classification techniques [3] and spectral matching techniques [4]. In the former, the images should not need to atmospheric correction, however, due to higher dimension of hyperspectral image, it will lead to dimensionality disaster and Hughes phenomenon [5, 6] which refer to the fact that with the number of spectral bands increased the sample size required for training set grows exponentially. The solution methods usually are increasing sample size, thus this will cost a lot of human and material resources. Another simple but sometimes effective way to solve this problem is dimension reduction of hyperspectral data, but some useful information will be lost. Furthermore, it is hard to solve mixed pixels. In the latter, matched filtering method is successfully used in information extraction from hyperspectral remote sensing image. It classifies by computing the similarity of the pixel spectrum and the reference spectrum, and it needs no sample data but the image data should be atmospheric corrected beforehand. These methods based on the hypothesis that dark currents of the sensor and path radiation are removed and all spectra data have been calibrated to apparent reflectance. However, it is only the ideal condition for these effects are hard to be removed completely, so some mistakes will be caused due to atmospheric influence, especially for low reflectivity ground objects. This chapter proposed an unsupervised classification for hyperspectral remote sensing image. It can effectively extract

low reflectivity ground objects such as water or vegetation in shadowed area from hyperspectral remote sensing data using spectral data mining. Firstly, extracting more than 40 endmembers from the hyperspectral image using Pixel Purity Index (PPI) and calculating the spectral angle between the pixel spectra and each endmember spectra, the pixel was assigned to the endmember class with the smallest spectral angle. Then, endmember spectra were clustered based on K-mean algorithm. Finally, pixels in the same K-mean result class were combined to one class and the final classification outcome was projected and outputted. Comparing the classification result and field data, they are in accord with each other. This method can produce the objective result with no artificial interference. It can be an efficient information extraction method for hyperspectral remote sensing data.

2. The study area

The study area is located at Heqing county in Yunnan province in southwest of China (25º38'N - 26º30'N; 99º58'E – 100º15'E) (figure 1). It covers mountainous terrain and has big altitude difference. It is situated in the Three Parallel Rivers of Yunnan Protected Areas which is a UNESCO World Heritage Site in China. The Three Parallel Rivers of Yunnan Protected Areas lies within the drainage basins of the upper reaches of three of Asia's great rivers run approximately parallel to one another though separated by higher mountain ranges with peaks over 6,000 meters. They are the Yangtze (Jinsha), Mekong (Lancang) and Salween (Nujiang) rivers, in the Yunnan section of the Hengduan Mountains. After this near confluence area, the rivers greatly diverge: the Nujiang River flows at Moulmein, Burma, into the Indian Ocean, the Mekong south of Ho Chi Minh City, Vietnam, into the South China Sea and the Yangtse impours into the East China Sea at Shanghai. It was awarded World Heritage Site status in 2003 for their richer biodiversity and spectacular topographical diversity [7].
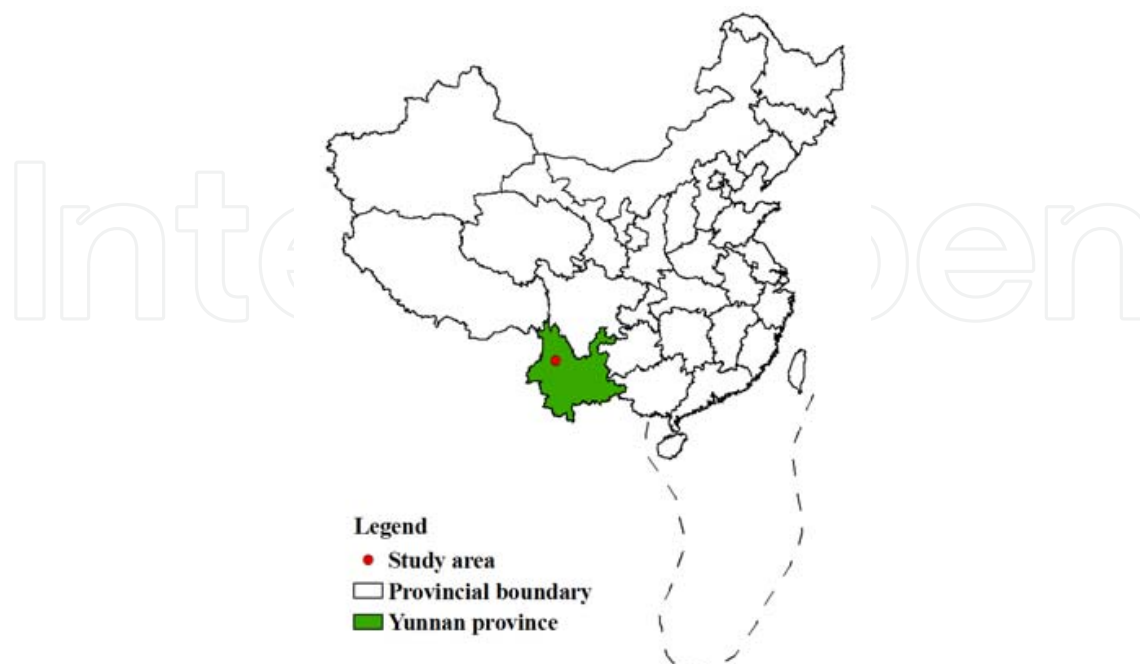

Figure 1. China provincial boundaries and the study area.

3. Remote sensing data

The image investigated in this chapter was obtained by Hyperion sensor boarded on EO-1 satellite in November 11, 2004, and it covers the 0.4 to 2.5 micrometer spectral range with 242 spectral bands at roughly 30m spatial resolution and 10nm spectral resolution over a 7.5 km wide swath from a 705 km orbit. The system has two grating spectrometers; one visible / near infrared (VNIR) spectrometer (approximately 0.4–1.0 micrometers) and one short-wave infrared (SWIR) spectrometer (approximately 0.9–2.5micrometers) (figure 2). Data are calibrated to radiance using both pre-mission and on-orbit measurements. Key Hyperion characteristics are discussed by Green et al. [8]. The image has a total of 242 bands but only 198 bands are calibrated. Because of an overlap between the VNIR and SWIR focal planes, there are only 196 unique channels [8, 9]. Due to water vapor absorption, some bands nearby 0.94, 1.38 and 1.87 micrometers also can not be available. The rest 163 bands can be used in research. Some pre-processing steps are necessary before using image. Firstly, some bad pixel value in original image were replaced by the means of two pixels value beside its two sides; then the image was radiometrically corrected using calibration coefficient; at last, the image was atmospheric corrected using FLAASH model [10]. Figure 3 and figure 4 shows different ground objects spectra before and after atmospheric correction. Shapes of different ground objects spectra after atmospheric correction are similar with shapes of standard laboratory spectra for the same ground object type.

4. Methodology

4.1. Spectral angle mapper (SAM)

Spectral Angle Mapper (SAM) algorithm is successfully used in matched filtering based on hyperspectral remote sensing image [4, 11-18]. It computes the "spectral angle" between the pixel spectrum and the endmember spectrum. When used on calibrated data, this technique is comparatively insensitive to illumination and albedo effects. Smaller angles represent closer matches to the reference spectra. The result indicates the radian of the spectral angle computed using the following equation:

$$\alpha = \cos^{-1} \left(\frac{\sum_{i=1}^{m} t_i r_i}{\left(\sum_{i=1}^{m} t_i^2 \right)^{1/2} \left(\sum_{i=1}^{m} r_i^2 \right)^{1/2}} \right) \tag{1}$$

Where m=the number of bands.

t_i =pixel spectrum.

r_i =reference spectrum.

α = radian of the spectral angle

Figure 2. Dispaly bands as gray scale (40 = vnir, 93 = swir), Select three bands as RGB (29, 23, 16 = VNIR, 204,150, 93 = SWIR).
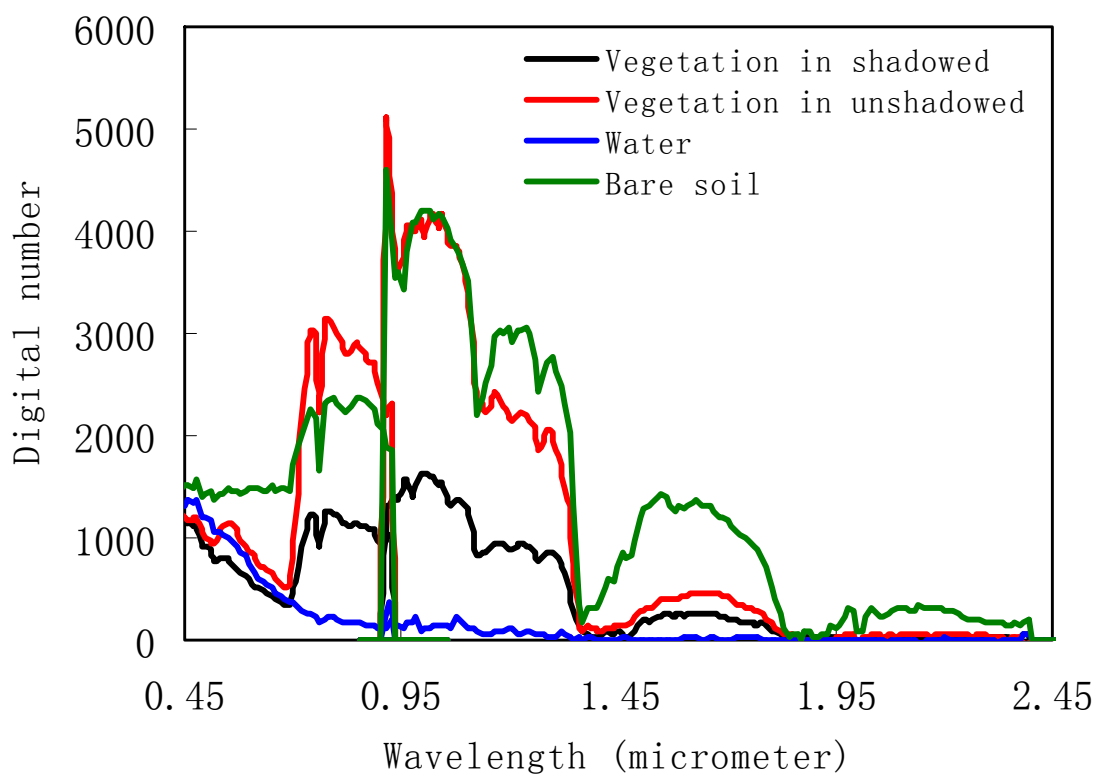
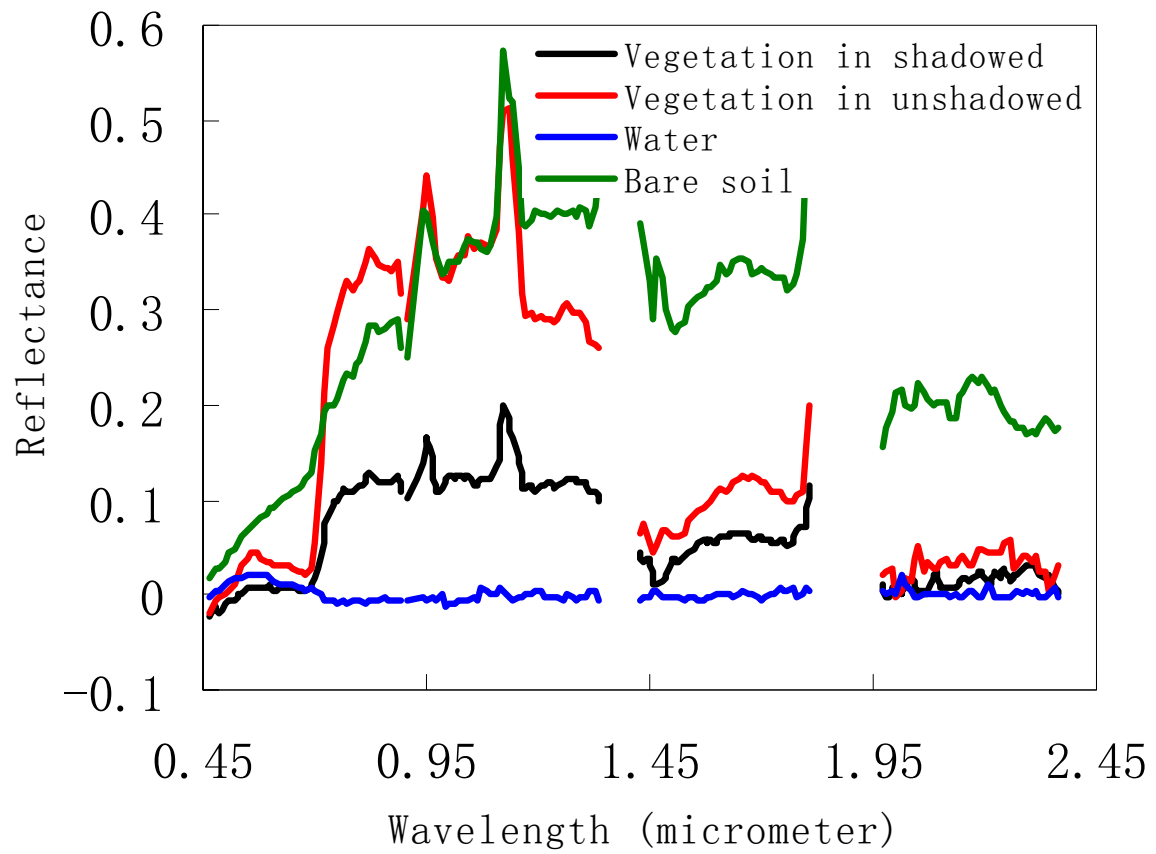**Figure 3.** Different ground objects curves before atmospheric correction

Figure 4. Different ground objects spectra after atmospheric correction.

Usually, a constant threshold is assigned firstly. When α is lower than the constant threshold, that means the pixel spectrum and the reference spectrum are similar with each other, and then assigned the pixel into the reference spectral class.

4.2. Extracting endmembers

The reference spectra can be selected from Spectral libraries, acquired by the handhold spectroradiometer, or extracted from the image itself. The commonly used technique is to extract the reference spectra from the image, for this method has the advantage that endmembers were collected under similar atmospheric conditions and pixel scale. A variety of methods have been used to find endmembers in multispectral and hyperspectral images. Iterated Constrained Endmembers (ICE) is an automated statistical method to extract endmembers from hyperspectral images [19]. [20] found a unique set of purest pixels based upon the geometry of convex sets. Probably Pixel Purity Index (PPI) is the most widely used algorithm [21]. In this chapter, PPI was used to find the most spectrally pure pixels in hyperspectral images as reference spectra. Firstly, the image was applied to a dimensionality analysis and noise whitening using the Minimum Noise Fraction (MNF) transform process [22, 23]. Then, the data are projected onto random unit vectors repeatedly and the total number of each pixel marked as an extreme pixel is noted. At last, the purest pixels in the scene are rapidly identified. In this chapter, 48 endmember spectra were extracted from hyperspectral image.

SAM was used to match each pixel spectrum to 48 endmembers. Figure 5 (b) is the classification result using the constant threshold. Different color represents different classes and black refer to unclassified classes. Comparing figure 5(a) with figure 5(b), it is shown that most vegetation in shadowed region and water are classified into unclassified classes, so some vegetation and water information are lost. One of reason is that the radiance of low reflectivity ground object such as water and vegetation in shadowed area are severely weakened by atmosphere influence when they arrive at the satellite. As is shown from figure 4, the digital numbers value of the vegetation in unshadowed area are 2 times higher than theirs in shadowed region. After atmospheric correction, the reflectance of the vegetation in shadowed area are obviously lower than theirs in unshadowed region, and reflectance in some spectral range (0.46-0.68 and 1.98-2.37 micrometers) are near zero. However, comparing the shape of vegetation spectra in shadowed and unshadowed area, they are similar with each other, so it is possible to identify the low reflectance ground object using SAM algorithm for it is relatively insensitive to illumination and albedo effects, but the constant threshold is not suitable. In this chapter, 48 endmember spectra were used as reference spectra, so the hypothesis that 48 endmember spectra include all land cover type is reasonable. The spectral angle of every pixel was calculated using with all endmembers, and the pixel belongs to the class which has the smallest spectral angle. Figure 5(c) is the processed classification result using adjustable threshold. Comparing with figure 5(b) and figure 5(c), they are the same except unclassified pixel in figure 5(b) which belongs to the certain land cover type in figure 5(c). This method improved the constant threshold SAM classification result, so it is more effective than using constant threshold.
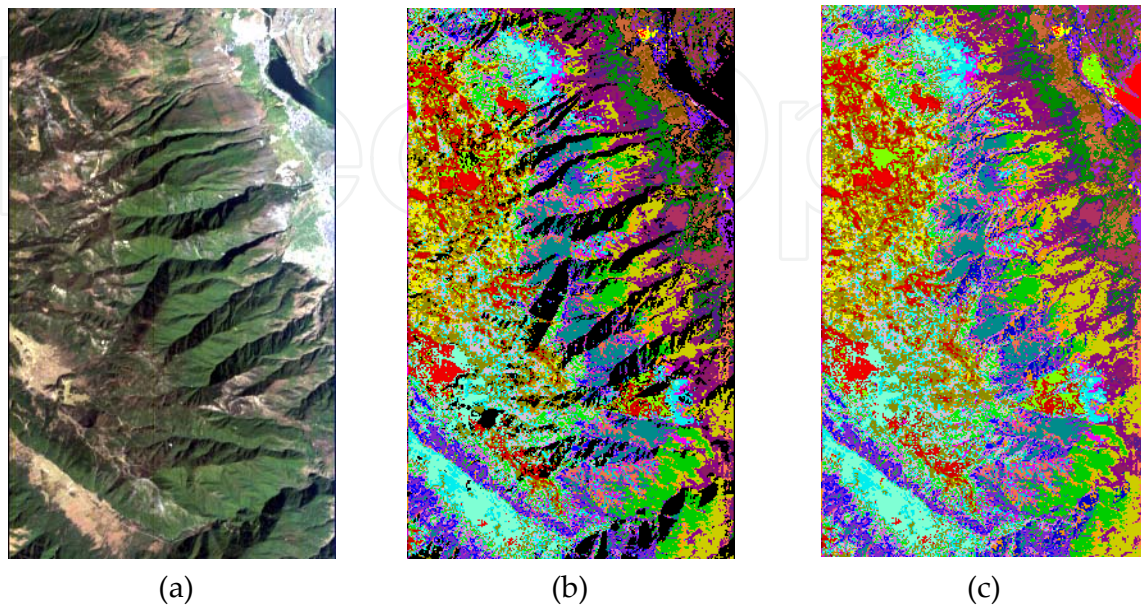
Figure 5. (a). The hyperspectral image after atmospheric correction. (b). the classification result of SAM using constant threshold. (c). the classification result of SAM using adjustable threshold

4.3. Clustering using K-mean algorithm

Figure 5(c) contains so many classes, and some classes may belong to the same class. In this chapter, these classes were clustered using K-mean algorithm [24], which is a straightforward and effective algorithm for finding clusters in data. It classifies pixels based on features into k centroids of group, one for each cluster. These centroids shoud be placed as much as possible far away from each other, then take each point belonging to a given data set which associate to the nearest centroid. The algorithm proceeds as follows. Firstly, the number of classes which the image should be partitioned into is inputted, and k records are randomly assigned to be the initial cluster center. Then, for each record, find the nearest cluster center. For each of the k clusters, find the cluster centroid, and update the location of each cluster center to the new value of the centroid. Repeat steps until convergence or termination. In this chapter, endmember spectra were clustered using K-mean algorithm and final 5 spectral classed were outputted. Then, classification result using adjustable threshold were merged according the K-mean algorithm result. Final classification result is shown in figure 6. Comparing the classification result and field data, they are in accord with each other.
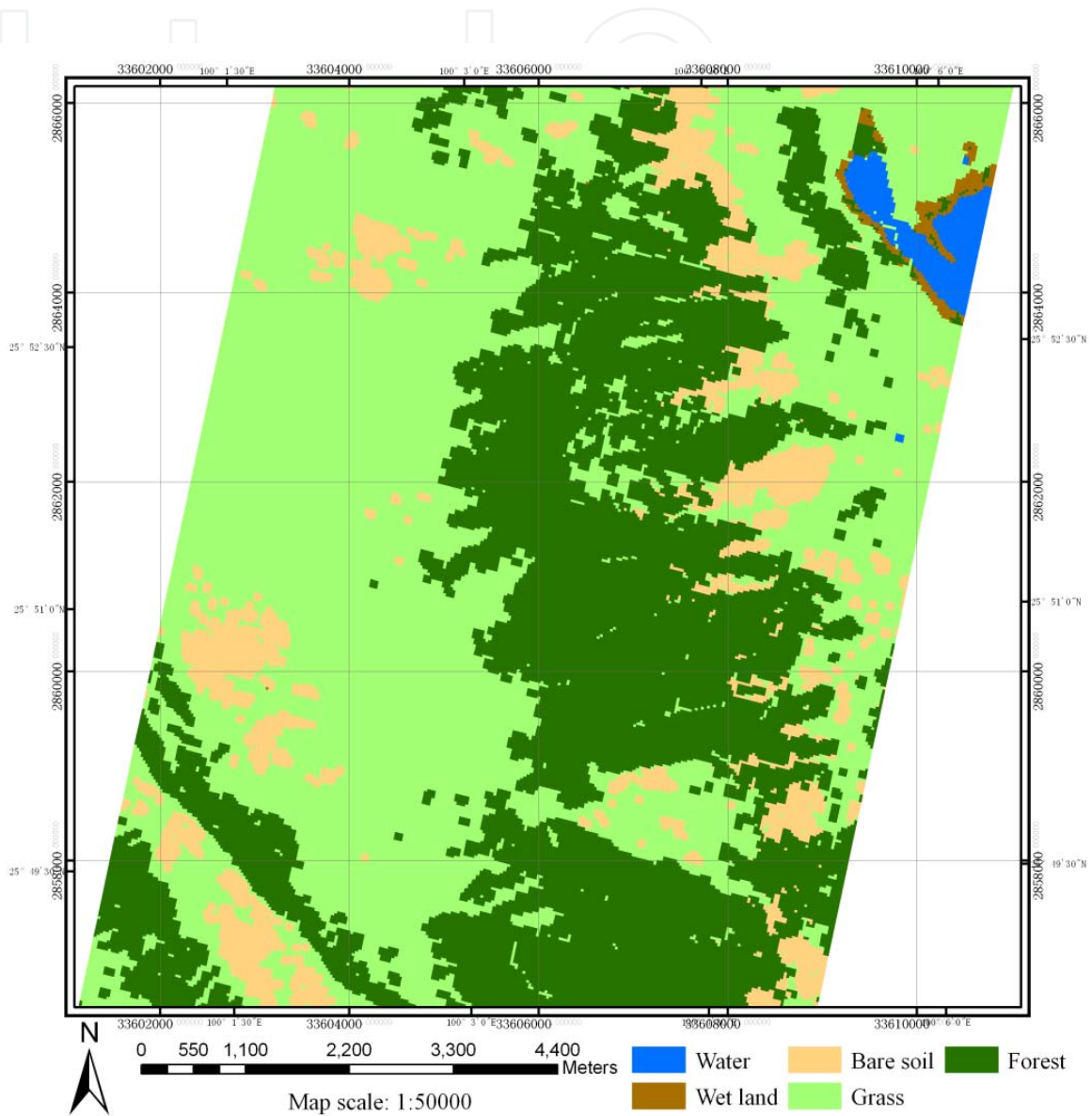
Map Projection: Gauss Kruger 3 Degree, Zone 33, False Easting: 500 kilometer, Central Meridian: 99º.

Figure 6. The classification map of hyperspectral remote sensing image

5. Results and discussions

Matching filter was maturely used in spectral classification in hyperspectral remote sensing image, however, due to atmospheric effect, it is hard to extract low reflectivity ground object. This chapter proposed an unsupervised classification method. Firstly, the hyperspectral remote sensing image was atmospherically corrected. Accuracy atmospheric correction is the key to the classification. Then, endmember spectra were extracted using PPI algorithm, and the image was classified using SAM. Traditionally SAM algorithm used constant threshold. This chapter improved and used adjustable threshold, and the pixel belong to class which has the smallest spectral angle. Finally, the endmember spectra were clustered based on K-mean algorithm and classes were combined according to the K-mean algorithm result. The final classification map was projected and outputted. It is an effective classification method especially for hyperspectral remote sensing image. Users also can adjust the endmember and classes number according to their applications.

Author details

Xingping Wen
Faculty of Land Resource Engineering,
Kunming University of Science and Technology, Kunming, China

Xiaofeng Yang
Research Center for Analysis and Measurement,
Kunming University of Science and Technology, Kunming, China

Acknowledgement

6. References

[1]   A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock (1985) Imaging Spectrometry for Earth Remote Sensing, Science, vol. 228, pp. 1147-1153.

[2]   F. van der Meer (2006) The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery, International Journal of Applied Earth Observation and Geoinformation, vol. 8, pp. 3-17.

[3]   J. J. Settle and N. Drake (1993) Linear mixing and the estimation of ground cover proportions, International Journal of Remote Sensing, vol. 14, pp. 1159-1177.

[4]   F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz (1993) The Spectral Image-Processing System (SIPS) -

Interactive Visualization and Analysis of Imaging Spectrometer Data, Remote Sensing of Environment, vol. 44, pp. 145-163.

[5] K. Fukunaga and R. R. Hayes (1989) Effects of sample size in classifier design, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, pp. 873-885.

[6] G. Hughes (1968) On the mean accuracy of statistical pattern recognizers, Information Theory, IEEE Transactions on, vol. 14, pp. 55-63.

[7] Available: http://en.wikipedia.org/wiki/Three_Parallel_Rivers_of_Yunnan_Protected_Areas.

[8] R. O. Green, B. E. Pavri, and T. G. Chrien (2003) On-orbit radiometric and spectral calibration characteristics of EO-1 Hyperion derived with an underflight of AVIRIS and in situ measurements at Salar de Arizaro, Argentina, Geoscience and Remote Sensing, IEEE Transactions on, vol. 41, pp. 1194-1203.

[9] R. Beck (2003) EO-1 User Guide v. 2.3: http://eo1.usgs.gov & http://eo1.gsfc.nasa.gov.

[10] T. Cooley, G. P. Anderson, G. W. Felde, M. L. Hoke, A. J. Ratkowski, J. H. Chetwynd, J. A. Gardner, S. M. Adler-Golden, M. W. Matthew, A. Berk, L. S. Bernstein, P. K. Acharya, D. Miller, and P. Lewis (2002) FLAASH, a MODTRAN4-based atmospheric correction algorithm, its application and validation, in Geoscience and Remote Sensing Symposium, pp. 1414-1418.

[11] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman (1992) Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm, in Summaries of the Third Annual JPL Airborne Geoscience Workshop, Pasadena CA, pp. 147-149.

[12] W. M. Baugh, F. A. Kruse, and W. W. Atkinson (1998) Quantitative geochemical mapping of ammonium minerals in the southern Cedar Mountains, Nevada, using the airborne visible/infrared imaging spectrometer (AVIRIS), Remote Sensing of Environment, vol. 65, pp. 292-308.

[13] F. van der Meer (2006) Indicator kriging applied to absorption band analysis in hyperspectral imagery: A case study from the Rodalquilar epithermal gold mining area, SE Spain, International Journal of Applied Earth Observation and Geoinformation, vol. 8, pp. 61-72.

[14] F. Van der Meer, M. Vasquez-Torres, and P. M. Van Dijk (1997) Spectral characterization of ophiolite lithologies in the Troodos Ophiolite complex of Cyprus and its potential in prospecting for massive sulphide deposits, International Journal of Remote Sensing, vol. 18, pp. 1245-1257.

[15] J. Schwarz and K. Staenz (2001) Adaptive Threshold for Spectral Matching of Hyperspectral Data, Canadian Journal of Remote Sensing, vol. 27, pp. 216-224.

[16] E. L. Hunter and C. H. Power (2002) An assessment of two classification methods for mapping Thames Estuary intertidal habitats using CASI data, International Journal of Remote Sensing, vol. 23, pp. 2989-3008.

[17] X. Wen, G. Hu, and X. Yang (2007) Combining the Three Matched Filtering Methods in Mineral Information Extraction from Hyperspectral Data, Journal of China University of Geosciences, vol. 18, pp. 294-296.

[18] X. Wen, G. Hu, and X. Yang (2007) A Simplified Method for Extracting Mineral Information From Hyperspectral Remote Sensing Image Using SAM Algorithm, in 12th Conference of International Association for Mathematical Geology, Geomathematics and GIS Analysis of Resources, Environment and Hazards, Beijing, China, pp. 526-529.

[19] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington (2004) ICE: a statistical approach to identifying endmembers in hyperspectral images, Geoscience and Remote Sensing, IEEE Transactions on, vol. 42, pp. 2085-2095.

[20] M. E. Winter (1999) Fast autonomous spectral endmember determination in hyperspectral data, in Proceedings of the Thirteenth International Conference on Applied Geologic Remote Sensing, Vancouver, British Columbia, Canada, pp. 337-344.

[21] J. W. Boardman, F. A. Kruse, and R. O. Green (1995) Mapping Target Signatures Via Partial Unmixing of Aviris Data, in Summaries of the Fifth Annual JPL Airborne Earth Science Workshop, Washington, D. C, pp. 23-26.

[22] A. A. Green, M. Berman, P. Switzer, and M. D. Craig (1988) A transformation for ordering multispectral data in terms of imagequality with implications for noise removal, Geoscience and Remote Sensing, IEEE Transactions on, vol. 26, pp. 65-74.

[23] J. W. Boardman (1993) Automated spectral unmixing of AVIRIS data using convex geometry concepts: in Summaries, in Fourth JPL Airborne Geoscience Workshop, Arlington, Virginia, pp. 11–14.

[24] J. MacQueen (1967) Some methods for classification and analysis of multivariate observations, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281-297.