THE MODIFIED UNSUPERVISED SPECTRAL ANGLE CLASSIFICATION (MUSAC) OF HYPERION, HYPERION-FLASSH AND ETM+ DATA USING UNIT VECTOR

DAE SUNG KIM and YONG IL KIM

SPatial INformatics and Systems (SPINS) Laboratory, School of Civil, Urban, and Geosystem Engineering, Seoul National University, Seoul, Korea
mutul94@empal.com and yik@snu.ac.kr

ABSTRACT:

Unsupervised spectral angle classification (USAC) is the algorithm that can extract ground object information with the minimum "Spectral Angle" operation on behalf of "Spectral Euclidian Distance" in the clustering process. In this study, our algorithm uses the unit vector instead of the spectral distance to compute the mean of cluster in the unsupervised classification. The proposed algorithm (MUSAC) is applied to the Hyperion and ETM+ data and the results are compared with K-Means and former USAC algorithm (FUSAC). USAC is capable of clearly classifying water and dark forest area and produces more accurate results than K-Means. Atmospheric correction for more accurate results was adapted on the Hyperion data (Hyperion-FLAASH) but the results did not have any effect on the accuracy. Thus we anticipate that the "Spectral Angle" can be one of the most accurate classifiers of not only multispectral images but also hyperspectral images. Furthermore the cluster unit vector can be an efficient technique for determination of each cluster mean in the USAC.

KEY WORDS: Hyperspectral Hyperion Image, Unsupervised Spectral Angle Classification, Unit Vector, FLAASH

1. INTRODUCTION

The fundamental premise of the remote sensing of land cover/land use is that every surface object has its own unique pattern of reflected, emitted, and absorbed radiation across the spectral band and the same types of surface objects show similar spectral response patterns (James, 1996). Though hyperspectral data can derive complete reflectance spectrum, those with classical classification methods fails to provide reliable results due to the many bands, unstable training information and so on. Dimensionality reduction methods like Feature Extraction or Feature Selection have been used (Anil, 2003), and new algorithms for Hyperspectral data classification have been proposed to solve these problems.

Spectral angle is a new approach based on the fact that the spectra of the same type of surface objects in RS data are approximately linearly scaled variations of one another due to atmospheric and topographic effects (Youngsinn, 2002). SAC (Spectral Angle Classification) classifies the image pixels using the minimum "Angular Distance" rule and does not require the training data to be normally distributed. Furthermore, it is insensitive to the data variance and the size of the training data set. However, the researches on the SAC were mainly applied to the supervised classification (Youngsinn, 2002; Kruse et al, 1993b) and moreover the entire spectral angle concept could not be used in the unsupervised classification as the cluster mean had to be computed by distance concept (Kai-Yi Huang, 2002; Arel Weisberg, 1999).

In this paper, we applied the modified spectral angle technique to unsupervised classification based on previous research and obtained a more efficient algorithm by using the mean of the angle based on the unit vector (also called angle mean) for the calculation of a cluster mean. Our algorithm (MUSAC) was tested to the Hyperion, Hyperion-FLAASH and ETM+ data and compared the results with K-Means and former USAC algorithm (FUSAC).

2. UNSUPERVISED SPECTRAL ANGLE CLASSIFICATION (USAC)

The processing of USAC is as follows:

1. Select the seed points
2. Assessment the similarity through the angle distance between seed points and unknown pixels (sub-section 2.1)
3. Calculate each cluster mean (sub-section 2.2)
4. Repeat steps 2 and 3 until the iteration reaches the user specified threshold

2.1 Assessment of Similarity

Contrary to other conventional clustering algorithms, USAC computes the similarity of an unknown spectrum to a reference spectrum by using the spectral angle rule. The spectral angle between every pixel in the image and every cluster mean is found by using equation (1).

$$\theta_{i,c} = cos^{-1}\left[\frac{\sum_{k=1}^{m} x_{i,k}\mu_{c,k}}{\sqrt{\sum_{k=1}^{m} x_{i,k}^{2}\sum_{k=1}^{m}\mu_{c,k}^{2}}}\right] \qquad (1)$$

$$If \quad \theta_{i,cmin} = Min(\theta_{i,c}), \quad then \quad x_{i,c} \rightarrow x_{i,k,c} \qquad (2)$$

Small values of $\theta_{i,c}$ indicate that the two spectra are quantitatively similar (Arel et al, 1999, see the equation (2)), where, m is the number of band, and $x_{i,k}$ is one of the pixels that were randomly selected.

2.2 Determination of Cluster Means

After finding the similar elements of the clusters, the means of elements were generally recalculated by using "Distance Mean" in the former researches. In this study, we introduce a more efficient algorithm at the spectral angle hypothesis with "Unit Vector" (or Angle Mean, equation (3)). The related equation is as follows:

$$unitvector = (p^{2}_{i,1} + \Lambda + p^{2}_{i,m}) = \sqrt{\sum_{k=1}^{m} p^{2}_{i,k}} = 1 \qquad (3)$$

$$\theta_{i,k} = cos^{-1}(x_{i,k} / \sqrt{\sum_{k=1}^{m} x^{2}_{i,k}}) \qquad (4)$$

$$\mu_{\theta_{i,k}} = \frac{1}{N_c}\sum_{i=1}^{N_c}\theta_{i,k} \qquad (5)$$

$$\mu_{c,k} = unitvector \times cos(\mu_{\theta_{i,k}}) = cos(\mu_{\theta_{i,k}}) \qquad (6)$$

The angle $\theta_{i,k}$ between seed points and each band is calculated by equation (4), and $\mu_{\theta_{i,k}}$ is corresponded to the mean of angle. Here, N_c is the number of elements of any cluster. The mean of cluster ($\mu_{c,k}$) is computed by multiplying the unit vector by the cosine angle of the mean of angle (equation (6)). In consequence, the spectral angle is calculated again through the equation (1).

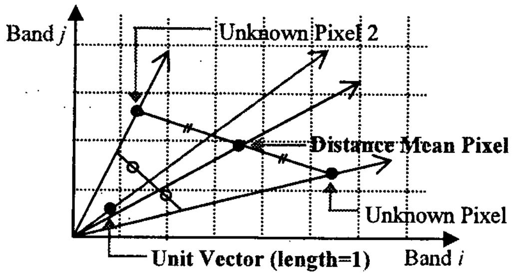


Figure 1. Comparison of Mean Pixel and Unit Vector

As shown in Figure 1, if unknown pixels 1 and 2 are included in the same cluster, FUSAC algorithm calculates the cluster mean with certain distance at each band but is not suitable for some cases such as where the pixel exists near 0. We believe that the use of the unit vector is a more appropriate technique to be applied to remotely sensed data because RS data generally exist in the middle of the space plane, and the unit vector is insensitive to the outlier.

Finally, the above steps are repeated until the iterations reach the user specified convergence threshold. In this study, 10 seed points for clustering were selected by equal division of the range of DN each band, and the iterative calculations just stopped when the number of elements changed by less than 1%.

3. IMPLEMENTATION

3.1 Data and Preprocessing

The USAC algorithm was applied to two different sets of data, the Hyperion data which is loaded on the EO-1 satellite and ETM+ on Landsat-7. The Hyperion data provides a high spectral resolution hyperspectral imager capable of resolving 242 spectral bands (from 0.4 to 2.5 μm) with a 30-meter resolution (same as ETM+ spatial resolution). Each image was scanned on the same day, i.e., April 3, 2002, and the test was conducted in the southern part of Seoul, Korea.

As each image must be of the same spot, ETM+ was registered on the basis of the Hyperion image and each image made by re-sampling has the same size (200 pixels by 500 pixels). Ninety-three bands having the same band width as the ETM+ wavelength were selected, and each band corresponded to the following: band 1 (ETM+): band 10-17 (Hyperion), band 2: band 18-26, band 3: 27-34, band 4: band 39-55, band 5: band 140-161, and band 7: band 192-220. The false color composite Hyperion images, which were preprocessed for implementation, are shown in Figure 2.

FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes), which is normally used as atmospheric correction, was developed with the support of the U. S. Air Force Research Lab as the first-principles atmospheric correction modelling tool for retrieving spectral reflectance from hyperspectral radiance images (A. Berk et al, 1989). Unlike many other atmospheric correction techniques, FLAASH incorporates the MODTRAN4 (Matthew et al, 2000) radiation transfer code.

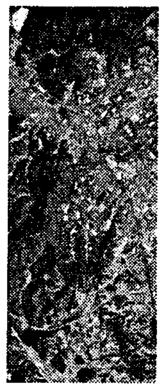


Figure 2. Hyperion Data

3.2 Validation

Three data (Hyperion, Hyperion-FLAASH, and ETM+) were passed through the process with selected ten seed points. On the basis of the three data, two techniques, spectral K-Means and FUSAC, were applied to similarity calculation. The results were comprised of six products. Finally, three data out of the six results from the MUSAC were tested for recalculating the cluster means. Therefore, finally nine result images were produced and assessed. The indices, ⓐ to ⓘ, is given in Figure 3. The nine results, which were classified through each clustering algorithm, are presented in the Figure 4 (You can find the results at the end of this paper).

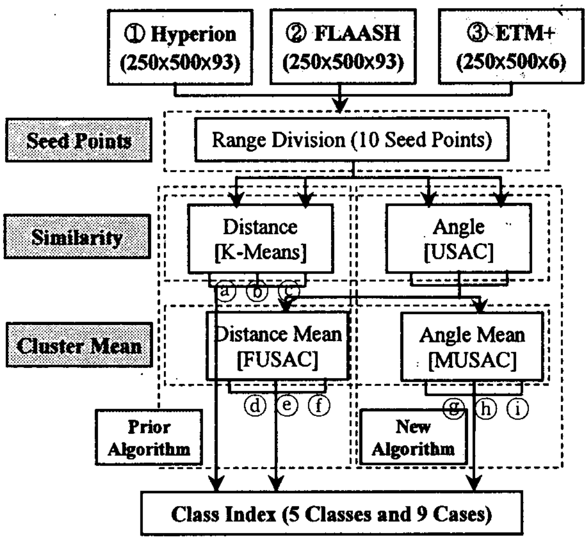


Figure 3. Flow Chart

4. RESULT

Thematic reference map was used for quantitative test of clustering results. MLC (Maximum Likelihood Classifier) and the supervised SAM (Spectral Angle Mapper), MDC (Minimum Distance Classifier), and ECHO (Extraction and Classification of Homogeneous Objects) classification techniques were applied and the pixels included in the same area of each result were selected for organizing the error matrix. Five classes, namely Water, Forest, Soil, Urban and Grass, were trained as a test set, and the training set were processed with the Multispec program.

The accuracy of USAC was higher than the results of K-Means. In this perspective, the spectral angle is an effective method for the hyperspectral data in addition to multispectral data. This result was already demonstrated by Youngsinn Sohn et al (Youngsinn, 2002), and we confirmed this result by visual and quantitative tests. In the visual estimation, the result of K-Means clustering can not distinguish the dark forest from water but USAC (especially MUSAC) can (see the Figure 5).

Proving that the atmospheric correction method needs to be applied to the unsupervised classification algorithm to obtain a better product, the atmospheric corrected Hyperion data (ⓑ, ⓔ and ⓗ) and not (ⓐ, ⓓ and ⓖ) were compared. The results did not show any other difference through visual approach and also the classification accuracy of Hyperion-FLAASH data showed a similar or somewhat lower result in comparison with Hyperion. In the assessment of time and stability of astringency, you can find that the result of Hyperion shows better than Hyperion-FLAASH (Figure 6).

From Figure 7, we can see the particular result in the case of ⓐ. That is, the five clusters only appeared, and the rest was not classified. This is due to the fact that K-Means is insensitive to the outlier. Figure 7 shows the particular result with a binary image and 6th-10th images look black.

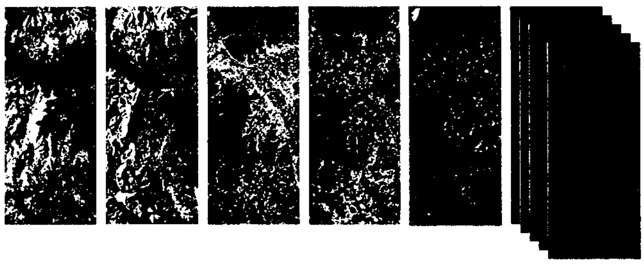


Figure 7. Particular Result in the Case of ⓐ

5. CONCLUSION

In this paper, we examined one of the classification algorithms, MUSAC, and also discussed the test performed on the Hyperion and ETM+ data. Atmospheric correction, FLAASH, was used for the pre-processing of the Hyperion data. Contrary to the FUSAM algorithm, the angle mean for the cluster center was used in this study and compared with results of other algorithms.

Some essential conclusions from the result using nine cases which were branched off from three data can be summarized as follows:

1. Atmospheric correction has little influences on the result of the clustering algorithms.
2. The spectral angle can be one of the most accurate classifiers and a valuable tool for clustering.
3. The unit vector can be an efficient technique for determination of each cluster mean in the USAC.

USAC is insensitive to near dark point because it uses only the direction of the spectra and not the length. Based on the understanding of the limitation of the USAC algorithm, we plan to conduct further studies on the following topics:

1. Calculate of the cluster mean by applying the split and merge technique
2. Study the combination of the angle and distance concept as a solution of the above second limitation
3. Research on the SAC value files for the weighted solution

REFERENCES

A. Berk, L. S. Bernstein and D. C. Robertson, 1989. MODTRAN: a Moderate Resolution Model for LOWTRAN7, MA, Air Force Geophys. Lab., Hanscom AFB, pp. 38.

Anil Cheriyadat, 2003. Limitations of Principal Component Analysis for Dimensionality-Reduction for Classification of Hyperspectral Data, MA, Mississippi State University, pp. 8-24.

Arel Weisberg, Michelle Najarian, Brett Borowski, Jim Lisowski, Bill Miller, 1999. Spectral Angle Automatic cLuster rouTine (SAALT): An Unsupervised Multispectral Clustering Algorithm, *Aerospace Conference, IEEE*, Vol. 4, pp. 307-317.

James B. Campbell, 1996. *Introduction to Remote Sensing*, Second Edition, The Guilford Press, New York, pp. 5.

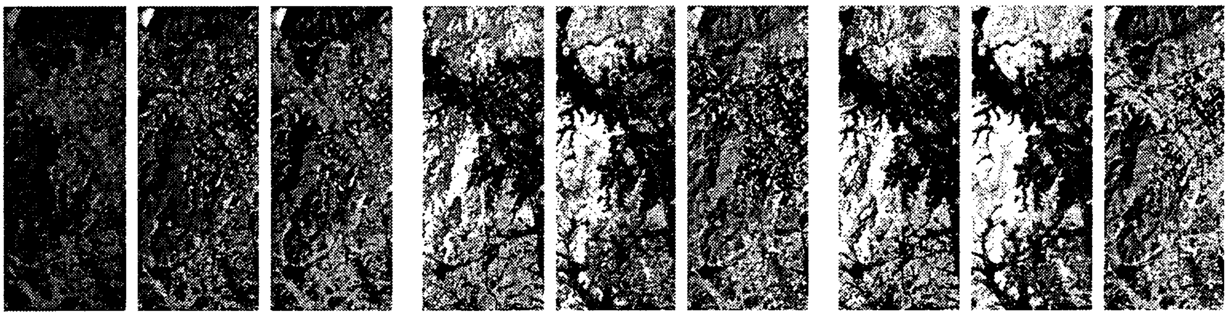
Kai-Yi Huang, 2002. A Synergistic Automatic Clustering Technique (SYNERACT) for Multispectral Image Analysis, *PE & RS, ASPRS*, Vol. 68, No. 1, pp. 33-40.

Kruse FA, Lefkof AB, Boardman JB, Heidebrecht KB, Shapiro AT, Barloon PJ, Goetz AFH, 1993b. The Spectral Image Processing System (SPIS)-Interactive Visualization and Analysis of Imaging Spectrometer Data, *Remote Sensing of the Environment*, No. 44, pp 309-336.

M. W. Matthew, S. M. Adler-Golden, A. Berk, S. C. Richtsmeire, 2000. Status of Atmospheric Correction Using a MODTRAN4-based Algorithm, *SPIE Proceeding, Algorithm for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, 4049, pp. 199-207.

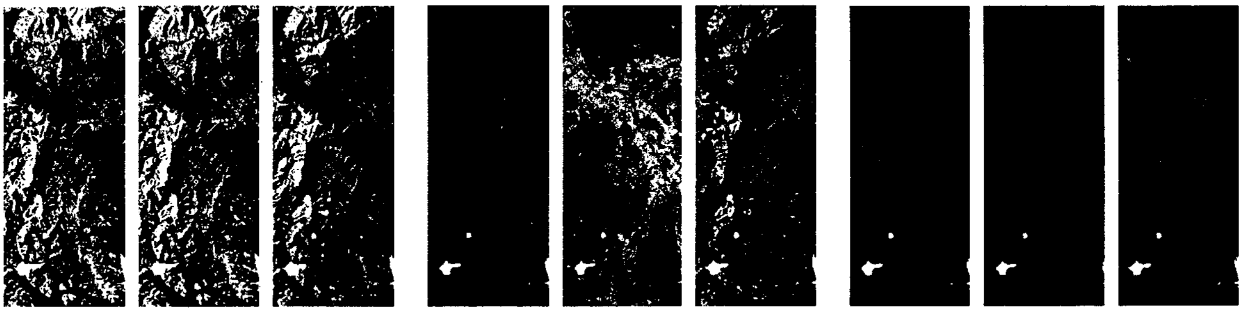
Richard J., 1999. *Remote Sensing Digital Image Analysis (An Introduction)-Third Edition*, Springer-Verlag, Germany, No. 4, pp. 230-236

Youngsinn Sohn and N. Sanjay Rebello, 2002. Supervised and Unsupervised Spectral Angle Classifiers, *PE & RS, ASPRS*, Vol. 68, No. 12, pp. 1271-1280.



(a) K-Means (ⓐ, ⓑ, and ⓒ) (b) FUSAC (ⓓ, ⓔ, and ⓕ) (c) MUSAC (ⓖ, ⓗ, and ⓘ)

Figure 4. Nine Results with each Clustering Algorithm



(a) K-Means (ⓐ, ⓑ, and ⓒ) (b) FUSAC (ⓓ, ⓔ, and ⓕ) (c) MUSAC (ⓖ, ⓗ, and ⓘ)

Figure 5. Critical Difference of Water Class in each Case (Note: The white color corresponds to water class.)

Table 1. Accuracy and Iteration Number of each Classification Result

Cases	ⓐ	ⓑ	ⓒ	ⓓ	ⓔ	ⓕ	ⓖ	ⓗ	ⓘ
Accuracy	69.31%	75.98%	78.01%	71.32%	60.58%	73.64%	77.33%	74.55%	70.88%
Iteration No.	35	16	20	30	13	37	24	11	29

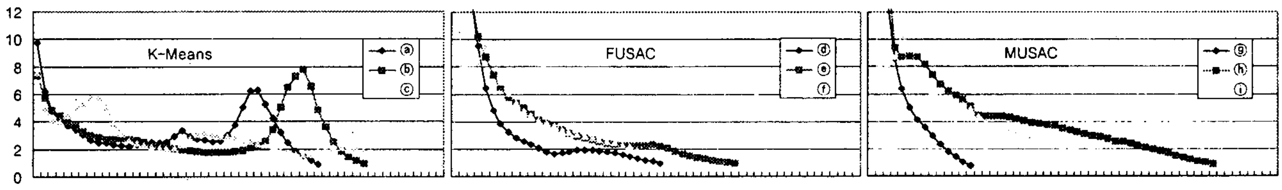


Figure 6. Assessment of Astringency for Iterative Calculation (x-axis: Iteration No. , y-axis: Changed Pixel (%))