



Supervised and Unsupervised Clustering Based Dimensionality Reduction of Hyperspectral Data

B. A. Beirami*, M. Mokhtarzade

Department of Photogrammetry and Remote Sensing, Faculty of Geodesy and Geomatics, K. N. Toosi University of Technology, Tehran, Iran

PAPER INFO

Paper history:

Received 22 September 2020

Received in revised form 17 April 2021

Accepted 12 May 2021

Keywords:

Hyperspectral Image

Principal Component Analysis

K-means Clustering

Classification

Feature Extraction

Weighted Mean

ABSTRACT

Nowadays, hyperspectral images (HIs) are widely used for land cover land use (LCLU) mapping. Hyperspectral sensors collect spectral data in numerous adjacent spectral bands, which are usually redundant. Hyperspectral data processing comes with important challenges such as huge processing time, difficulties in transfer, and storage. In this study, two supervised and unsupervised dimensionality reduction methods are proposed for hyperspectral feature extraction based on the band clustering technique. In the first method, the unsupervised method, after the unsupervised band clustering stage with some statistical attributes, the principal component transform is used in each cluster, and the first PC component is considered an extracted feature. In the second method, the supervised method, bands are clustered based on training samples mean vectors of each class, and the weighted mean operator is used for feature extraction in each cluster. The experiment is conducted on the classification of real famous HI named Indian Pines. Comparing the obtained results and some other state of art methods proved the proposed method's efficiency.

doi: 10.5829/ije.2021.34.06c.03

1. INTRODUCTION

Hyperspectral sensors have high discrimination capabilities of ground surface materials due to recording hundreds of spectral portions of the electromagnetic wave. Classification is one of the most important methods for information extraction from HIs. Hughes phenomenon is the main problem in the supervised classification of HIs that degrades the classification accuracies. This phenomenon expresses that by increasing the number of features above some threshold, classification accuracy usually decreases [1, 2].

Dimensionality reduction is a common way to tackle this problem. Generally, feature selection and feature extraction are two main groups of dimensionality reduction methods. Feature selection methods try to find a lower-dimensional subset of the original feature so that essential discriminative information is preserved, while feature extraction methods try to find a transformation to map the features in lower-dimensional space. The main difference between these two methods lies in reduced

features' physical meaning, which is only preserved in feature selection methods. In this study, we focus on feature extraction methods.

Numerous studies are available in the literature for feature extraction from HIs. Principal components analysis (PCA) and minimum noise fraction (MNF) are two widely used unsupervised methods that map the original features in lower dimensional space so that the first few reduced features contain the most information [3, 4]. Another version of PCA named segmented principal component analysis (SPCA) is proposed, which enhanced the original PCA version in the computational time aspect [5, 6]. Also, the nonlinear version of PCA named Kernel Principal Component (KPCA) is used in several studies dimensionality reduction of hyperspectral data [7, 8]. The wavelet-based dimensionality reduction method is another unsupervised method [9, 10]. In this method, high frequency and low-frequency components of the spectral signature curve (SSC) are separated, and the smoother version of SSC is used as the reduced features. The rational function curve fitting method is

*Corresponding Author Institutional Email:
b_asghari@email.kntu.ac.ir (B. A. Beirami)

recently used for hyperspectral feature extraction [11, 12]. In this method, a specific rational function approximation is developed to fit each pixel's SSC, and coefficients of the numerator and denominator are considered the newly extracted features.

Another widely used method that uses labeled samples of classes is linear discriminant analysis (LDA) [13, 14]. This method tries to maximize the within-class similarity and between-class separability. A generalized version of LDA that uses the kernel function is proposed by Baudat and Anouar [15]. LDA has three important drawbacks; first, this method works well only if the distributions of classes are normal-like. Second, this method can only extract ($L-1$) features (L is the number of classes), and third, the singularity of the within-class scatter matrix often occurs in the face with hyperspectral data [16]. Another supervised method named decision boundary feature extraction (DBFE) is a method that is extracted discriminately feature from the decision boundary between classes [17]. Nonparametric weighted feature extraction (NWFE) is another important feature extraction method for hyperspectral data [18]. The main idea of NWFE is to put different weights on every sample, compute the weighted means, and then define nonparametric between-class, and within-class scatter matrices. As a drawback, this method takes enormous time for the data reduction process. The kernel version of NWFE named kernel NWFE (KNWFE) is proposed to extend the NWFE for a nonlinear situation [16].

Recently, clustering-based methods are used for feature extraction [19, 20]. A supervised method named PSBS has been proposed based on k-means clustering of bands for band selection of HIs [21]. In a detailed paper, previous authors proposed the PSFE method based on fuzzy c-means clustering and feature weighting by class membership values [22]. Clustering-based feature extraction (CBFE) is proposed by Imani and Ghassemian [23]. This method considers a vector of mean values of training samples of all classes in each band and then a clustering algorithm such as k-means, groups these vectors in some clusters, and subsequently, the mean of bands whose associated vectors are located in a cluster is considered as an extracted feature. Usually, clustering-based methods work well even in small sample size situations (SSS situation) and do not have the other problems of some conventional supervised feature extraction methods such as enormous computation times, limitation of $L-1$ features (in LDA), and problems in estimating covariance matrix. An unsupervised version of band selection based on clustering of some bands statistics such as variance, mean absolute deviation (MAD), and standard deviation are for endmember extraction and classification of HIs [24-28].

This study introduced two unsupervised and supervised methods for dimensionality reduction of hyperspectral data based on bands clustering. As a result

of the literature, when PCA transforms modified so that transformation is carried out by avoiding the low correlations between highly correlated blocks, PCA's efficiency improved [29]. This study's first proposed method is the unsupervised method named unsupervised clustering based principal components analysis (UCPCA). In this method, in the first stage, bands are clustered based on some statistical attributes. We expect adjacent and non-adjacent correlated bands to concentrate in the same cluster. In the second stage in each cluster, we carried out the PCA transform, and the first principal component (PC) is considered reduced features in each cluster. Unsupervised band clustering is the main novelty of this method, making UCPCA, like the original PCA, an unsupervised method. In the second proposed method, the supervised method named weighted mean clustering-based feature extraction (WMCFE), bands are clustered based on the mean values of training samples in each class, and the weighted mean of bands is based on MAD considered as an extracted feature in each cluster. The major differences between UCPCA and WMCFE can be summarized as follows:

1. Each cluster of bands in the UCPCA is formed based on some statistical attributes of bands in an unsupervised manner without using any training samples, but bands clustering space in the WMCFE is formed based on the mean value of training samples in each class.
2. The UCPCA method used PCA transform for extracting the informative feature of each cluster of bands, but the WMCFE used a weighted mean for feature extraction in each cluster. Based on this explanation UCPCA method is grouped in the information-based feature extraction method, but WMCFE is the statistical feature extraction method.

In the next part, the methodology of the study that consists of two proposed methods is presented. In the second part, after introducing the data set results and related analysis of experiments are obtained. Finally, in the last part conclusion is presented.

2. METHODOLOGY

2. 1. UCPCA

Unsupervised clustering-based principal components analysis (UCPCA) is an unsupervised method the same as PCA. This method has two stages. In the first stage, hyperspectral bands are grouped by K-means clustering based on some statistical attributes. Each band of HI can be represented by three statistical measures that are shown in Table 1.

In the above formulas, B_i is the value of i^{th} pixel, M , and N are image dimensions, $\text{mean}(b)$ is the mean of all pixel values in band b . In other words, each band can be shown as a point in 3-dimensional statistical space, and the K-means clustering algorithm tries to group each

TABLE 1. Statistical attributes

Statistical attribute	Formula
Variance (Var)	$SM1_b = \frac{1}{M,N} \sum_{i=1}^{M,N} (B_i - mean(b))^2$
Mean	$SM2_b = \frac{\sum_{i=1}^{M,N} B_i}{M,N}$
Kurtosis	$SM3_b = \frac{\frac{1}{M,N} \sum_{i=1}^{M,N} (B_i - mean(b))^4}{(\frac{1}{M,N} \sum_{i=1}^{M,N} (B_i - mean(b))^2)^2}$

band of HI so that the bands with similar information lie in the same cluster. K-means is one of the simplest unsupervised learning algorithms and consisted of four stages [30]:

- 1) Place K points into the attribute space represented by the statistics of bands. K is the number of desired reduced features.
- 2) Assign each band to the group that has the closest centroid based on some distance metrics such as square Euclidean
- 3) recalculate the positions of the K centroids by taking the mean of all bands assigned to that centroid's cluster
- 4) Repeat steps 2 and 3 until stopping criteria met (such as no centroid move, the sum of the distances is minimized, or user-defined maximum number of iterations)

Because of the algorithm's sensitivity to the initial randomly selected cluster centers and the probability of local optimum situation of results, the k-means algorithm may run multiple times.

In the second stage of the proposed algorithm, when the clusters formed, principal components analysis is applied in each cluster (for more information about the principal component analysis, referred to literature [31]), and the first PC of each cluster that contains the most information is considered as a new extracted feature.

2. 2. Weighted Mean Clustering-based Feature Extraction

Weighted mean clustering-based feature extraction (WMCFE) is a supervised method based on the bands' K-means clustering and weighed mean operator. This method has three stages. In the first stage, based on training samples, we calculate the mean vector of training samples in each class for each band of HI. This vector is as follows:

$$b_i = [m_{i1} \ m_{i2} \ m_{i3} \ \dots \ m_{ic}]^T \quad (1)$$

In the above vector b_i is i^{th} band of HI and m_{ic} is mean of class c in i^{th} band. In other words, if we want to describe this method as the same as the UCPCA method, in the WMCFE method, the means of classes are attributes of each band. In the second stage, k-means clustering is introduced to group the bands the same as the UCPCA. The number of clusters in k-means

clustering is user-defined and is equal to the desired numbers of reduced extracted features. In the last stage, the weighted mean of bands in each cluster is considered as extracted features. In this study, the weighted mean operator in each cluster is formulated as follows:

$$\text{weighted mean}(\mathbf{B}) = \frac{\sum_{i=1}^k w_i b_i}{\sum_{i=1}^k w_i} \quad (2)$$

In the above, b_i is the i^{th} band in a cluster, w_i is its corresponding weight and k is the number of bands in each cluster. Weights in this study are defined by MAD statistical measure:

$$\frac{1}{M,N} \sum_{i=1}^{M,N} |B_i - mean(b)| \quad (3)$$

Bands with higher values of MAD get larger weights in each cluster because these bands contain more information, and so have more contribution to the extracted feature of each cluster [27].

3. EXPERIMENTAL RESULTS

3. 1. Data Set

The airborne visible/ Infrared Imaging Spectrometer (AVIRIS) sensor acquired the Indian Pines scene from a mixed forest/agricultural from the Indian Pines Site in the USA. This data set consists of 145×145 pixels and 224 spectral bands. This data set's spatial resolution is 20 m, and band wavelengths are from 400 to 2500 nm with 10 nm resolution. After removing water absorption and noisy bands, the remaining 200 bands are used in this study. This scene contains 16 different land covers. Figure 1 illustrates the false-color composition of the original image and the ground-truth map .

After discarding four classes named alfalfa, grass-pasture-mowed, oats, and stone-steel-towers with a small number of samples in the ground truth map, the other 12 classes are used in this study.

3. 2. Results

For evaluating proposed dimensionality reduction methods, a conventional maximum likelihood classifier is used to classify extracted features. The classification results for proposed methods are compared with four other dimensionality reduction methods named PCA, CBFM and NWFE. In this study, average and overall accuracies are two approaches used accuracy metrics [32].

Experiments are done in a few training-size situations with 60 training samples per class randomly chosen from the ground truth map. This training set is used for training classifiers and feature extraction. Half of these training samples are used in supervised feature extraction methods. The obtained overall and average accuracies values in the mentioned experiment for 60 training sample sizes are shown in Figures 2(a) and 2(b).

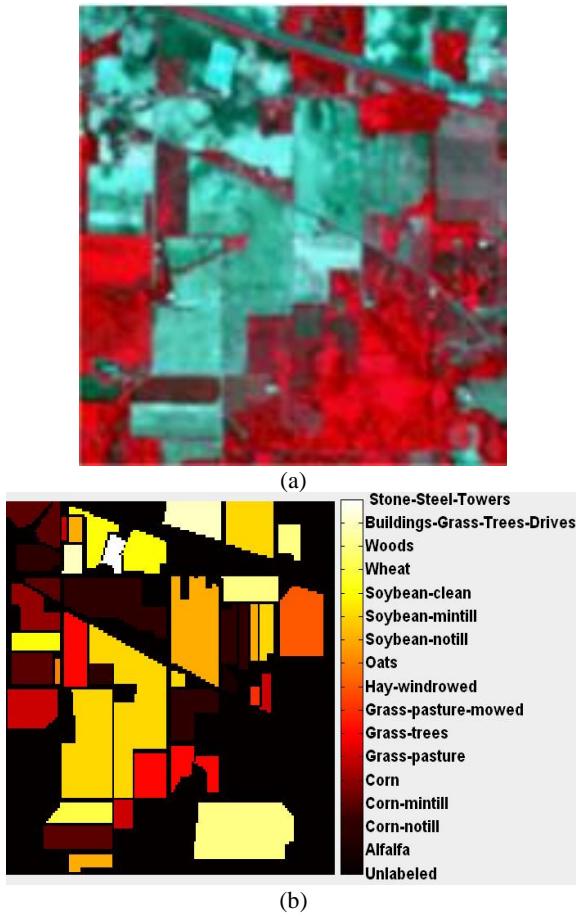
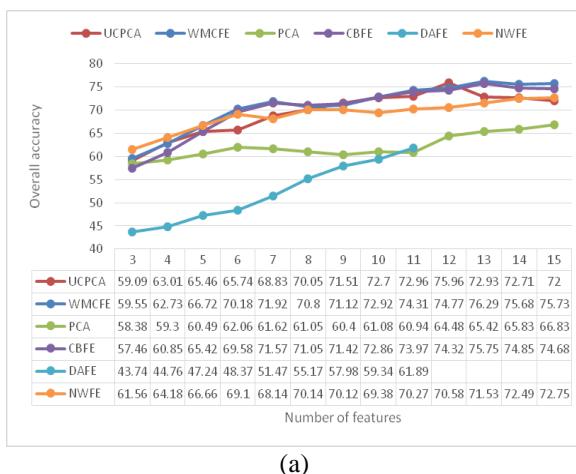
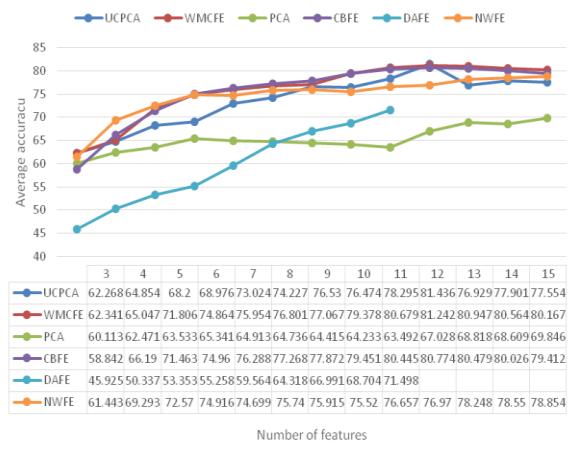


Figure 1. Indian pines data set - a) color composite image - b) ground-truth map

The highest ML achieved classification overall accuracies of extracted features from different methods



(a)



(b)

Figure 2. Classification accuracies for different feature extraction methods in different number of features- a) overall accuracy - b) average accuracy

that are shown in Table 2 demonstrate that both proposed methods (UCPCA and WMCFE) perform better than the other feature extraction methods and at last WMCFE is superior. The main reason for the superiority of the UCPCA against the PCA is the bands clustering technique, which leads to the more homogenous band groups, and as a result, improves the performance of the PCA. In the comparison to CBFM, WMCFE uses the weighted mean operator that informative band has more contribution to the extracted feature of each cluster. Compared to some classical methods such as the DAFE and the NWFE, whose performance is dependent on a large number of the training samples (for accurate parameter estimation of models), the proposed method (UCPCA and WMCFE) works well even when few training samples are available.

The ground truth map of the Indian data set and the classified maps of different methods in the situation of Table 2 are shown in Figure 3. Based on Figure 3, one can understand that the proposed methods produced more smooth classification results.

TABLE 2. highest ML classification overall accuracies

Methods

	UCPCA	WMCFE	PCA	CBFM	DAFE	NWFE
OA	75.96	76.29	66.83	75.75	61.89	72.75
# of features	12	13	15	13	11	15

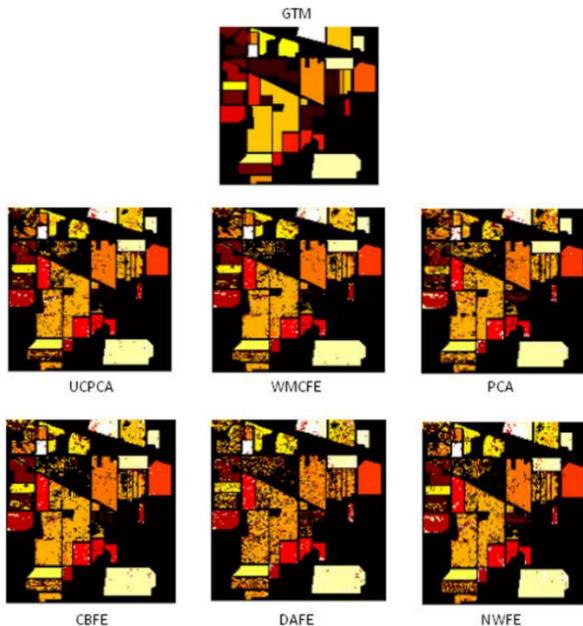


Figure 3. Ground truth map (GTM) and classified maps of different feature extraction methods

4. CONCLUSION

In this study, two unsupervised and supervised feature extraction methods based on bands clustering have been introduced for dimensionality reduction of hyperspectral data. The unsupervised method, named UCPCA, is based on unsupervised bands clustering and principal components analysis feature extraction. The supervised method, named WMCFE, is based on supervised bands clustering based on samples mean of classes and weighed mean operator based on the MAD measure. The experimental results showed better performance of proposed methods compared with some conventional feature extraction methods such as PCA, CBFE and NWFE.

5. REFERENCES

- Hughes, G. "On the mean accuracy of statistical pattern recognizers", *IEEE Transactions on Information Theory*, Vol. 14, 1, (1968), 55-63. DOI: 10.1109/TIT.1968.1054102
- Landgrebe, D., "Hyperspectral image data analysis", *IEEE Signal Processing Magazine*, Vol. 19, 1, (2002), 17-28. DOI: 10.1109/79.974718
- Lim, S., K.H. Sohn, and C. Lee. "Principal component analysis for compression of hyperspectral images", In IEEE International Geoscience and Remote Sensing Symposium., IGARSS S'01. (2001). DOI: 10.1109/IGARSS.2001.976068
- Harris, J. R., Rogge, D., Hitchcock, R., Ijewliw, O., Wright, D. "Mapping lithology in Canada's Arctic: application of hyperspectral data using the minimum noise fraction transformation and matched filtering", *Canadian Journal of Earth Sciences*, Vol. 42, 12, (2005), 2173-2193. DOI: 10.1139/e05-064
- Jia, X. and J.A. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification", *IEEE Transactions on Geoscience and Remote Sensing*, 1999. Vol. 37, 1, (1999), 538-542. DOI: 10.1109/36.739109
- Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., & Chanussot, J., "Hyperspectral remote sensing data analysis and future challenges", *IEEE Geoscience and Remote Sensing Magazine*, Vol. 1, 2, (2013), 6-36. DOI: 10.1109/MGRS.2013.2244672
- Datta, A., S. Ghosh, and A. Ghosh, "Unsupervised band extraction for hyperspectral images using clustering and kernel principal component analysis", *International Journal of Remote Sensing*, Vol. 38, 3, (2017), 850-873. DOI: 10.1080/01431161.2016.1271470
- Fauvel, M., J. Chanussot, and J.A. Benediktsson. "Kernel principal component analysis for feature reduction in hyperspectral images analysis". IEEE Proceedings of the 7th Nordic Signal Processing Symposium (2006). DOI: 10.1109/NORSIG.2006.275232
- Kaewpijit, S., J. Le Moigne, and T. El-Ghazawi, "Automatic reduction of hyperspectral imagery using wavelet spectral analysis", *IEEE Transactions on Geoscience and Remote Sensing*, 2003. Vol. 41, 4, (2003), 863-871. DOI: 10.1109/TGRS.2003.810712
- Prabhu, N., M.K. Arora, and R. Balasubramanian, "Wavelet Based Feature Extraction Techniques of Hyperspectral Data", *Journal of the Indian Society of Remote Sensing*, Vol. 44, 3, (2016), 373-384. DOI: 10.1007/s12524-015-0506-9
- Hosseini, S.A. and H. Ghassemian. "A novel approach to hyperspectral data feature extraction using rational function curve fitting". IEEE International Conference on Signal and Image Processing Applications (ICSIPIA), (2015). DOI: 10.1109/ICSIPIA.2015.7412241
- Hosseini, S.A. and H. Ghassemian, "Rational function approximation for feature reduction in hyperspectral data", *Remote Sensing Letters*, Vol. 7, No. 2, (2016), 101-110. DOI: 10.1080/2150704X.2015.1101180
- Fukunaga, K., Introduction to statistical pattern recognition. 2013: Academic Press. Hardcover ISBN: 9780122698514
- David, L., "Hyperspectral image data analysis as a high dimensional signal processing problem", *IEEE Signal Processing Magazin*, Vol. 19, No. 1, (2002), 17-28. <https://engineering.purdue.edu/~landreb/SigProcPaper.pdf>
- Baudat, G. and F. Anouar, "Generalized discriminant analysis using a kernel approach", *Neural Computation*, Vol. 12, No. 10, (2000) 2385-2404. DOI: 10.1162/089976600300014980
- Kuo, B.-C., C.-H. Li, and J.-M. Yang, "Kernel nonparametric weighted feature extraction for hyperspectral image classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47, No. 4, (2009), 1139-1155. DOI: 10.1109/TGRS.2008.2008308
- Lee, C. and D.A. Landgrebe, Feature extraction based on decision boundaries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 4, (1993), 388-400. DOI: 10.1109/34.206958
- Kuo, B.-C. and D.A. Landgrebe, "Nonparametric weighted feature extraction for classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, No. 5, (2004), 1096-1105. DOI: 10.1109/TGRS.2004.825578
- Asghari Beirami, B., and M. Mokhtarzade. "Introducing an unsupervised method for feature extraction of hyperspectral images using bands clustering in the prototype space." *Journal of Geomatics Science and Technology*, Vol. 9, No. 2 (2019), 195-207. <http://jgst.issge.ir/article-1-907-en.pdf>

20. Esfandian, N., and K. Hosseinpourb. "A Clustering-Based Approach for Features Extraction in Spectro-Temporal Domain Using Artificial Neural Network." *International Journal of Engineering, Transactions B: Applications*, Vol. 34, No. 2 (2021): 452-457. DOI: 10.5829/IJE.2021.34.02B.17
21. Mojariadi, B., Emami, H., Varshosaz, M., & Jamali, S. "A novel band selection method for hyperspectral data analysis", Int Arch Photogramm Remote Sens Spat Inf Sci, (2008).
22. Mojariadi, B., Abrishami-Moghaddam, H., Zanj, M. J. V., & Duin, R. P. "Dimensionality reduction of hyperspectral data via spectral feature extraction", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47, No. 7, (2009), 2091-2105. DOI: 10.1109/TGRS.2008.2010346
23. Imani, M. and H. Ghassemian, "Band clustering-based feature extraction for classification of hyperspectral images using limited training samples", *IEEE Geoscience and Remote Sensing Letters*, Vol. 11, No. 8, (2014), 1325-1329. DOI: 10.1109/LGRS.2013.2292892
24. Ahmad, M., Haq, D. I. U., Mushtaq, Q., & Sohaib, M. "A new statistical approach for band clustering and band selection using K-means clustering", *IACSIT International Journal of Engineering and Technology*, Vol. 3, No.6, (2011), 606-614.
25. Sohaib, M., Ihsan-Ul-Haq and Q. Mushtaq, "Dimensional Reduction of Hyperspectral Image DataUsing Band Clustering and Selection Based on Statistical Characteristics of Band Images", *International Journal of Computer and Communication Engineering*, Vol. 2, No. 2, (2013), 101. <http://www.ijcce.org/papers/148-B091.pdf>
26. Datta, A., S. Ghosh, and A. Ghosh, "Combination of clustering and ranking techniques for unsupervised band selection of hyperspectral images", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 8, No. 6, (2015), 2814-2823. DOI: 10.1109/JSTARS.2015.2428276
27. I ul Haq, I. and X. Xu. "A new approach to band clustering and selection for hyperspectral imagery". IEEE 9th International Conference on Signal Processing (ICSP), (2008). DOI: 10.1109/ICOSP.2008.4697345
28. Saqui, D., Saito, J. H., Campos, J. R., & Jorge, L. A. D. C. "Approach Based on Fuzzy C-Means for Band Selection in Hyperspectral Images", *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, Vol. 10, No.5, (2016), 889-895.<https://publications.waset.org/10004410/approach-based-on-fuzzy-c-means-for-band-selection-in-hyperspectral-images>
29. Naik, G.R., Advances in Principal Component Analysis: Research and Development. 2017, Springer. Hardcover ISBN 978-981-10-6703-7. DOI:10.1007/978-981-10-6704-4
30. MacQueen, J. "Some methods for classification and analysis of multivariate observations". In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. (1967). Oakland, CA, USA.
31. Richards, J.A. and J. Richards, Remote sensing digital image analysis. Vol. 3. 1999: Springer. Hardcover ISBN 978-3-642-30061-5. DOI: 10.1007/978-3-642-30062-2
32. Beirami, B.A. and Mokhtarzade, M, "SVM classification of hyperspectral images using the combination of spectral bands and Moran's I features". IEEE 10th Iranian Conference on Machine Vision and Image Processing (MVIP), (2017). DOI: 10.1109/IranianMVIP.2017.8342334
33. Beirami, Behnam Asghari, and Mehdi Mokhtarzade. "Band Grouping SuperPCA for Feature Extraction and Extended Morphological Profile Production from Hyperspectral Images." *IEEE Geoscience and Remote Sensing Letters*, Vol. 17, No. 11, (2020), 1953-1957. DOI: 10.1109/LGRS.2019.2958833
34. Beirami, Behnam Asghari, and Mehdi Mokhtarzade. "Superpixel-Based Minimum Noise Fraction Feature Extraction for Classification of Hyperspectral Images." *Traitement du Signal*, Vol. 37, No. 5, (2020). DOI: <https://doi.org/10.18280/ts.370514>

Persian Abstract

چکیده

در سال‌های اخیر از تصاویر ابرطیفی به طور گسترده‌ای برای تهیه تقشه از کاربری اراضی تحت پوشش زمین استفاده می‌شود. سنسورهای ابرطیفی باندهای طیفی را در نواحی مجاور متعددی که معمولاً اضافی هستند جمع آوری می‌کنند، از این رو پردازش داده‌های طیفی فوق با چالش‌های مهمی مانند زمان پردازش زیاد، مشکلات در انتقال و ذخیره‌سازی داده روپرتو است. در این مطالعه، دو روش کاهش ابعاد تحت نظارت و بدون نظارت برای استخراج ویژگی‌های طیفی بر اساس روش خوشبندی باند پیشنهاد شده است. در روش اول، که روش بدون نظارت است، پس از مرحله خوشبندی بدون نظارت باندها با استفاده از مشخصات آماری، در هر خوشه، از تبدیل مولفه اصلی استفاده می‌شود و اولین مولفه آن در هر خوشه به عنوان ویژگی استخراج شده در نظر گرفته می‌شود. در روش دوم که روش نظارت شده است، باندها بر اساس میانگین بردارهای نمونه‌های آموزشی هر کلاس خوشبندی شده و از عملکر میانگین وزن‌دار برای استخراج ویژگی در هر خوشه استفاده می‌شود. آزمایش‌های این مقاله بر روی تصویر ابرطیفی مشهور واقعی این‌دین پاین انجام شده است. مقایسه بین نتایج بدست آمده و برخی دیگر از روش‌های مرسوم بیانگر کارآیی بیشتر روش پیشنهادی است.
