Exemplar Component Analysis: A Fast Band Selection Method for Hyperspectral Imagery

Kang Sun, Xiurui Geng, and Luyan Ji

***Abstract*—How to find the representative bands is a key issue in band selection for hyperspectral data. Very often, unsupervised band selection is associated with data clustering, and the cluster centers (or exemplars) are considered ideal representatives. However, partitioning the bands into clusters may be very time-consuming and affected by the distribution of the data points. In this letter, we propose a new band selection method, i.e., exemplar component analysis (ECA), aiming at selecting the exemplars of bands. Interestingly, ECA does not involve actual clustering. Instead, it prioritizes the bands according to their exemplar score, which is an easy-to-compute indicator defined in this letter measuring the possibility of bands to be exemplars. As a result, ECA is of high efficiency and immune to distribution structures of the data. The experiments on real hyperspectral data set demonstrate that ECA is an effective and efficient band selection method.**

***Index Terms*—Band selection (BS), cluster centers, clustering, hyperspectral data.**

I. Introduction

THE hyperspectral sensors can now simultaneously capture hundreds of image bands with wavelength ranging from the visible spectrum to the infrared region, due to the great development in recent decades [1]. Although the large number of hyperspectral bands provide abundant information to differentiate materials, they bring about also some problems, e.g., Hughes phenomena [2]. Moreover, the process of the mass data also demands considerable computation power. As a result, dimensionality reduction has been one of the most important preprocess steps in hyperspectral data analysis to address these issues.

Band selection (BS) is an effective approach for hyperspectral dimensionality reduction, which has been paid increasing attention in recent years. The existing BS techniques can be grouped into two broad types, namely supervised and unsupervised. The supervised methods require training samples that may be practically not available [3]. Thus, this letter mainly focuses on unsupervised BS techniques. The unsupervised techniques can be understood as the process of selecting a qualified subset from a larger set of bands, without any prior knowledge. Many of the unsupervised BS methods are based on information evaluation means. They try to selected the subset with large information [4]–[6], low similarity [3], [7], and some other information measures [8]–[10]. However, the bands obtained by these methods are generally extreme ones and even outliers.

In order to select representative instead of extreme bands, recently, the researchers interpret BS as a clustering problem, i.e., a process of partitioning the bands into groups of similar clusters. In these circumstances, the cluster centers are generally considered preferable to marginal (often extreme) bands. For instance, Martínez–Usó proposed to group the bands using hierarchical clustering method to minimize the intracluster variance and maximize the intercluster variance [11]. A BS method based on affinity propagation clustering [12] was proposed in [13] and [14], which aims to select the centers of the band clusters. In [15], Hedjam presented a new BS algorithm that is accomplished by performing Markov clustering to the band adjacency graph. Based on different distance measures, such as interquartile range, correlation coefficient, and covariance, Ahmad proposed to cluster and therefore select the bands by corresponding different k-means versions [16].

These clustering-based methods can be understood as the direct application of the clustering methods to hyperspectral bands. Although the centers of the clusters seem to be a better choice, these methods suffer from large computation complexity. Moreover, these methods may be greatly influenced by the number of clusters. Last but not least, these clustering methods generally require spherical distribution of the data since a data point is always assigned to the nearest center. However, as for the bands of hyperspectral data, the spherical distribution requisite may not hold.

In fact, there is no need to perform cluster analysis to hyperspectral bands since what we need is the cluster centers (or exemplars) only. In this letter, we select the cluster centers without actual clustering. Specifically, for each band, we use an indicator termed exemplar score (ES) to measure the possibility of a band to be an exemplar. The ES utilizes two reasonable assumptions [17], namely, the exemplars have the highest local density and they are at a relatively large distance from points of higher density. Based on ES, we present a fast BS method, i.e., exemplar component analysis (ECA), which aims to select the bands with high probability to be exemplars (or high ES). Interestingly, ECA does not involve actual clustering; instead, it prioritizes the bands according to their ESs. ECA has quite low computation complexity since it is actually a band-ranking method, avoiding the subset search step. In addition, as will be verified in the experiment, ECA is able to detect nonspherical clusters. In fact, ECA has no distribution requirements of the data points.

II. ECA

BS has its basis in the fact that hyperspectral bands have high correlation. For a highly correlated subset of bands, we

Manuscript received August 3, 2014; revised September 15, 2014 and October 10, 2014; accepted October 30, 2014.

K. Sun is with the Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China.

X. Geng is with the Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China (e-mail: gengxr@sina.com.cn).

L. Ji is with Centre for Earth System Science, Tsinghua University, Beijing 100084, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2014.2372071

may need only one of them, i.e., the most representative one. As a result, BS is naturally associated with data clustering; the cluster centers can be regarded as the most representative bands. Then, how to locate the cluster centers (exemplars) accurately and quickly has been the kernel of BS. It is worth noting that the "clustering" here is not applied to pixels (or ground objects) but to bands. Therefore, in this letter, each band of hyperspectral data are regarded as a data instance (point) in high-dimensional space. The main purpose of this letter is to present an algorithm that can find the exemplars of these points accurately and efficiently.

Rodriguez *et al.* recently proposed a new clustering algorithm and claimed that clusters are recognized regardless of their shape and of the dimensionality of the space in which they are embedded [17]. This method utilizes two reasonable assumptions: One is that the cluster centers must have the highest local density (ρ_i), and the other is that they have relatively larger distance (δ_i) to the points of higher density. In the following, we first illustrate the computation of ρ_i and δ_i in detail.

Assume the hyperspectral data set is $\mathbf{X}_{N\times L} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L]$ where $\mathbf{x}_i = [x_{1i}, x_{2i}, \ldots, x_{Ni}]^T$ is the vector constructed by the ith band, N is the number of pixels, and L is the number of total bands. First the distance matrix of bands needs to be computed. Let d_{ij} denote the Euclidean distance between the ith band $\mathbf{x}_i$ and jth band $\mathbf{x}_j$, then

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2 \tag{1}$$

where $\|\bullet\|$ is the 2-norm operator. In [17], ρ_i is defined as the number of points that are adjacent to point i, which can be understood as using a hard threshold to define neighbors. In this letter, we employ a soft threshold, i.e., a Gaussian heat kernel function, to define ρ_i as follows:

$$\rho_i = \sum_{j=1}^{L} \exp\left(-\frac{d_{ij}}{2\sigma^2}\right) \tag{2}$$

where σ is a adjusting parameter, controlling the weight degradation rate as the function of distances. Therefore, σ has a similar function of determining the neighborhood range of the points. The computation of δ_i is quite simple, as its definition is the nearest distance to the points of higher density from point i, as follows:

$$\delta_i = \min_{j:\rho_j>\rho_i} (d_{ij}). \tag{3}$$

For the point with the highest local density, we simply let $\delta_i = \max_j(d_{ij})$. We can see from the definition of δ_i that only the points with local or global maxima density that have a relatively large δ_i. In fact, δ_i plays an important role in the suppression of highly correlated bands.

These two indicators can properly characterize the location of the points (bands) in data clouds and play a vital role in the presentation of our method. More specifically, when ρ_i is large and δ_i is small, it implies that the ith band is close to but not the cluster center since there is a point that is in the same cluster (since δ_i is small) but has a larger ρ_i. When ρ_i and δ_i are both small, this point lies around the edge of the clusters. When ρ_i is small and δ_i is large, on the other hand, the point is away from the entire data cloud, indicating that the point is probably an outlier. It is only when both ρ_i and δ_i are relatively large that the point can be exemplars.

Based on this, we use the product of ρ_i and δ_i [17] to measure the probability of the points (bands) to be exemplars. In this letter, the indicator is termed ES as follows:

$$\mathrm{ES}_i = \rho_i * \delta_i. \tag{4}$$

We can see that ES imposes punishment on small values of ρ_i and δ_i. Only when ρ_i and δ_i are both of large values will the i-th band have large ES_i, indicating a high chance to be exemplars.

Once the ES for the bands has been calculated, the BS step for ECA is quite simple, ranking the bands according to their ESs from high to low and selecting the top ones. Therefore, ECA is essentially a band prioritization method.

Some remarks for ECA are noteworthy. To our knowledge, the existing band-prioritization-based BS methods, such as [5], [8], generally suffer from the high correlation among selected bands since the similar bands usually have similar scores. Interestingly, ECA gives full consideration of the correlation between bands, although it is also a band-prioritization-based method. More specifically, suppose the order of ES is $\mathrm{ES}_{m_1} > \mathrm{ES}_{m_2} > \cdots > \mathrm{ES}_{m_L}$, $1 \le m_1, m_2, m_L \le L$. The first band, numbered m_1, corresponds to the center of the largest cluster since it has the highest ES. Then, band m_2, which has the second largest ES, must be at a large distance from band m_1. Otherwise, if band m_2 is close to band m_1, it will lead to a small δ_{m_2} and therefore a small ES_{m_2}. Similarly, for any band m_i, it must have a relatively large distance to the bands of larger ES. In fact, ES_{m_i} indicates the possibility of band m_i to be an exemplar in the premise that bands $m_1, m_2, \ldots, m_{i-1}$ have been already identified as exemplars. Since ECA simultaneously considers the local density and correlation between bands, it is expected to select more reasonable subset of bands.

From the previous analysis, we can see that ECA has the following advantages.

1) *Low computation complexity*. The computation of the distance matrix spends most of the computation time, involving $L^2N/2$ float operations (flops). The computation complexity of ES can be omitted compared with the distance matrix. Thus, the flops needed for ECA is about $L^2N/2$.
2) Be effective to nonspherically distributed data. ECA can find the cluster centers from not only spherical data but also from the data with other distributions. This is because ECA is based on two assumptions shared by many distribution structures.
3) Does not need subset search. As a prioritization-based method, ECA can be very fast and avoid the local optimum caused by subset search.
4) Has the potentiality to automatically determine the number of bands. As will be shown in the experiment, ES has a distinct inflection point that can be used to determine the number of selected bands.

III. EXPERIMENT

Here, we evaluate the performance of ECA on two real hyperspectral data sets (both available at www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes). For comparison, we choose four different types of unsupervised BS method,

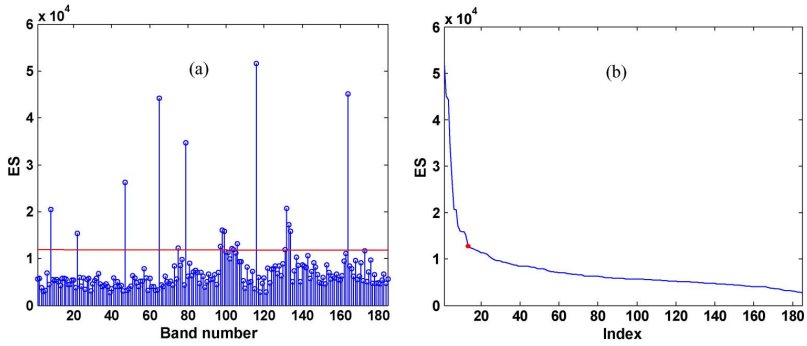

Fig. 1. (a) ES for each band. (b) Sorting the ES from high to low. The red line in (a) and red point in (b) denote the inflection point.

namely, information divergence BS (IDBS) [8], maximum ellipsoid volume (MEV) [4], similarity-based unsupervised BS (SUBS) [3] and affinity propagation based BS (APBS) [13]. To evaluate their performance, both the accuracy and computing time are compared. In addition, as for the parameter σ in (2), we assigned it with a self-adjusting value as follows:

$$\sigma = \frac{1}{30}\text{mean}(d_{ij}) \tag{5}$$

where mean(d_{ij}) is the mean value of all the distances between pair bands.

A. Indian Pines Data set

The data set used in this experiment is the famous Indian Pines data set that has been widely used in remote sensing experiments. This data set was acquired by Airborne Visible Infrared Imaging Spectrometer (AVIRIS) and has spatial size of 145 $\times$ 145 pixels, containing 220 bands with a range from 400 to 2500 nm. Due to the heavy water vapor absorption and low SNR, the noise bands numbered 1–3, 103–112, 148–165, and 217–220 are manually removed; thus a total of 185 valid bands participate in the following experiment.

The ES for each band was computed according to (4), as shown in Fig. 1(a). Then, the ES was sorted in decreasing order [shown in Fig. 1(b)]. From Fig. 1(a), we can see that the bands with relatively high ESs distributed uniformly among the full set of bands, demonstrating that ECA discourages the selection of highly correlated bands. Moreover, from Fig. 1(b), it can be seen that, ES has a notable inflection point, denoted by the red dot. ES declines rapidly before the inflection point and has a relatively slow declining rate after the inflection point. Therefore, ES can be considered an indicator to measure the importance of bands. In ECA, the most important band will be the first to be extracted, followed by less significant bands. For this data set, the number of bands with ES larger than the inflection point is 15, which is consistent with [11] and [14]. We selected the 15 bands with ES higher than that of the inflection point [above the red line in Fig. 1(a)].

We mapped the 185 bands into 2-D plane according to their Euclidean distance by multidimensionality scaling (MDS), as shown in Fig. 2. The 15 bands selected by ECA are marked by color dots in the figure. It can be seen that the distribution structure of the bands is irregular and the bands selected by ECA lie closely to the center of the clusters. This demonstrates that ECA has the ability to detect exemplars for nonspherical

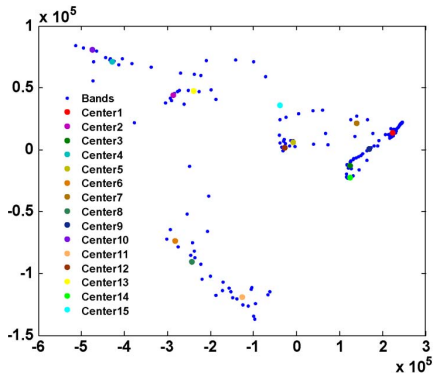

Fig. 2. Bands are mapped into 2-D space by MDS according to their Euclidean distance. The selected bands spread roughly in the centers of the clusters.

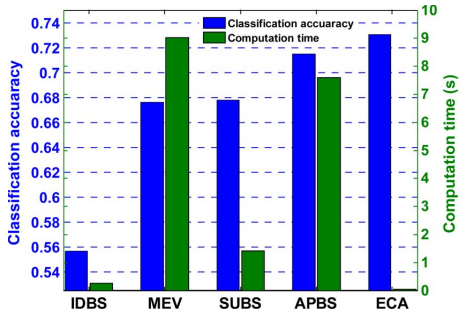

Fig. 3. SVM Classification accuracy and computation time for different methods.

clusters. The point with the highest ES (center 1 in the figure) is located in the center of the largest cluster. The size of the cluster represented by the exemplar is roughly declining with the decrease of ES. This experiment demonstrates that the use of ES as the measurement of possibility to be exemplars is reasonable.

In order to evaluate the performance of the subset selected by ECA, we used the other four methods, namely, IDBS, MEV, SUBS, and APBS to select the same number (15) of bands. Then, the selected subsets are used for support vector machine (SVM) classification, with 20% randomly chosen training samples. The kernel mapping function of SVM is the radial basis function, and a one-against-one strategy is adopted for classification of multiclass. The obtained overall classification accuracy and computation time are shown in Fig. 3.

It can be seen from Fig. 3 that, the clustering-based BS methods ECA and APBS are superior to the other methods in terms

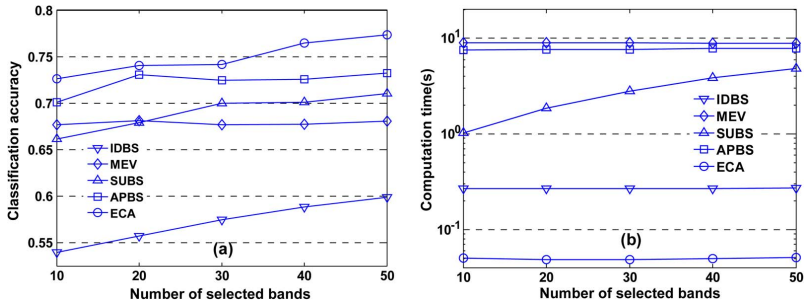

Fig. 4. (a) SVM classification accuracy using bands selected by different methods with different numbers. (b) Computation time (in seconds) for different BS methods.

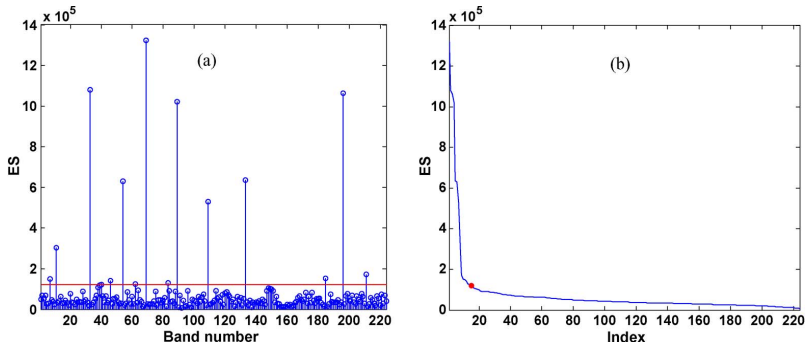

Fig. 5. (a) ES values. (b) Sorting the ES from high to low. The red line in (a) and red point in (b) denote the inflection point.

of classification accuracy. In particular, the subset selected by ECA has the highest classification accuracy (73.1%). ECA outperforms the other clustering-based method APBS mainly because of that APBS is suitable for spherical clusters, whereas ECA can deal with nonspherical data points. From Fig. 2 we can see that the distribution of the bands in Euclidean space is irregular. Therefore, ECA can find the exemplars more accurately. Moreover, MEV, which aims to select most informative bands has a very close performance with the similarity-based method SUBS. Both the two methods prefer to the extreme bands, and they have lower accuracy than APBS and ECA. This demonstrates that, compared with outlier bands, the exemplars may be a better choice for BS in the context of classification. In addition, the computation time (right y-axis) of these methods show that ECA is very fast and spends the least time (about 0.05 s). The other clustering-based BS method, i.e., APBS, is time-consuming, spending about 7.9 s for this data set.

To further quantitatively compare the performance of these five methods, we use them to select different numbers of bands ranging from 10 to 50 with steps of 10. Then, the selected subsets are used for SVM classification. The overall SVM classification accuracy and the corresponding computation time are shown in Fig. 4(a) and (b), respectively.

As shown in Fig. 4(a), the subsets selected by ECA can always obtain the highest classification accuracy, regardless of the number of selected bands. This indicates that the subset selected by ECA is very effective in terms of classification. For the other clustering-based method APBS, although it is not as accurate as ECA, it outperforms all the other methods. Therefore, we may draw the conclusion that the exemplars are more suitable for classification than the extreme bands (MEV and SUBS). In addition, the band-prioritization-based BS method IDBS always has the lowest accuracy. This can be attributed to the fact that IDBS does not take the correlation between bands into consideration. Another possible reason for the poor performance of IDBS is that the non-Gaussian criterion is unsuitable in performing BS for classification [11].

From the computation time in Fig. 4(b), we can see that the computation time for ECA (about 0.05 s) is much less than that of the other methods and is independent of the number of selected bands. In fact, the computation time of all the band-prioritization-based methods is independent of the number of selected bands (for instance, IDBS spent about 0.3 s) since all the bands need to be ranked regardless of the number of the selected bands. Moreover, the computation time for APBS is also independent of the number of bands (about 7.9 s). The reason is that APBS is essentially a cluster algorithm, for which the number of clusters is determined by parameters and thus has little influence to the computation time. The computation time for MEV is declining slightly with the increase in the number of selected bands since the sequential backward selection is adopted as the subset search strategy. On the contrary, SUBS adopts a sequential forward selection strategy; thus, the computation time is approximately linearly increasing as the size of selected subset grows.

B. Salinas Data Set

To further evaluate the performance of ECA, we use another well-known data set acquired by AVIRIS, i.e., Salinas. This radiance data set consists of 224 bands with spatial size of 512×217 pixels. Generally, the noise bands, numbered 1–3, 106–114, 150–167, and 221–224, should be manually removed,

TABLE I
BANDS SELECTED BY DIFFERENT METHODS FROM SALINAS DATA

Methods	Selected band indices (15 from 224)
IDBS	25, 26, 27, 28, 29, 30, 109, 154, 158, 159, 160, 164, 165, 166, 224
MEV	1, 2, 3, 4, 5, 9, 19, 32, 34, 37, 40, 42, 54, 67, 183
SUBS	1, 2, 3, 4, 5, 12, 19, 32, 38, 41, 45, 55, 67, 126, 171
APBS	14, 68, 103, 108, 111, 113, 152, 154, 156, 158, 161, 163, 164, 166, 168
ECA	7, 11, 33, 40, 46, 54, 62, 69, 83, 89, 109, 133, 185, 196, 211

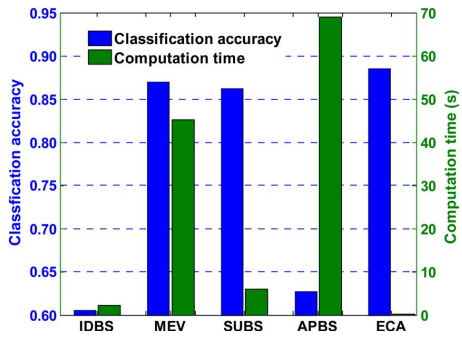

Fig. 6. Salinas data set: SVM classification accuracy and computation time (in seconds) for different BS methods.

but in this experiment, these bands are purposefully kept as natural test for the robustness of these methods to noise.

We first computed the ES for each band [Fig. 5(a)] and then sorted the ES from high to low [see Fig. 5(b)]. It can be seen from Fig. 5(b) that ES has a clear inflection point (denoted by the red dot). There are 15 bands that have larger ES than the inflection point; thus, the number of selected bands is set as 15. The bands selected by different methods are listed in Table I.

From the BS results in Table I, we can see that ECA can keep clear of the noisy (water vapor absorption) bands successfully. The other compared methods, particularly IDBS and APBS, are severely affected by the noisy bands.

In order to quantitatively evaluate the performance of these techniques, we performed SVM classification with randomly 10% training samples based on the selected subsets. The overall classification accuracy and the computation time of the methods are shown in Fig. 6.

From the classification accuracy (left axis), we can see that the subsets selected by ECA always have the highest classification accuracy (about 88.5%). On the other hand, for this data set, APBS has a much worse performance than that for the Indian Pines data set. This is because the noise bands have not been manually removed, implying that APBS is sensitive to noise (also verified in Table I). In addition, MEV and SUBS have quite similar performances; both are worse than ECA. The computation times (right axis) demonstrate that ECA takes the least time (about 0.25 s), whereas APBS takes the most (about 69 s).

IV. CONCLUSION

Compared with the extreme bands (for instance with largest information or least similarity), the exemplars are more desirable since they can represent the similar bands better. The results of the experiments show that the exemplars do have higher classification accuracy. However, the determination of the exemplars is very difficult, generally involving bands clustering that is time-consuming and tends to be influenced by the distribution of the data. In this letter, we have presented an efficient unsupervised BS ECA, which selects the exemplars without actual clustering. The experiments verified the validity and efficiency of ECA. Moreover, we can see that the ES curve has a notable inflection point that may be used to determine the number of selected bands. Thus, how to establish the inflection point of ES is our future research interest.

REFERENCES

[1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
[2] B. Kim and D. A. Landgrebe, "Hierarchical classifier design in high-dimensional numerous class cases," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 4, pp. 518–528, Jul. 1991.
[3] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.
[4] C. Sheffield, "Selecting band combinations from multispectral data," *Photogramm. Eng. Remote Sens.*, vol. 51, no. 6, pp. 681–687, 1985.
[5] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.
[6] X. Geng, K. Sun, L. Ji, and L. Ji, "A fast volume-gradient based band selection method for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7111–7119, Nov. 2014.
[7] P. Mitra, C. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 301–312, Mar. 2002.
[8] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.
[9] M. Ghamary Asl, M. R. Mobasheri, and B. Mojaradi, "Unsupervised feature selection using geometrical measures in prototype space for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 3774–3787, Jul. 2014.
[10] K. Sun, X. Geng, L. Ji, and Y. Lu, "A new band selection method for hyperspectral image based on data quality," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2697–2703, Jun. 2014.
[11] A. Martínez–Usó, F. Pla, J. M. Sotoca, and P. Garcia-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.
[12] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.
[13] Y. Qian, F. Yao, and S. Jia, "Band selection for hyperspectral imagery using affinity propagation," *IET Comput. Vis.*, vol. 3, no. 4, pp. 213–222, Dec. 2009.
[14] S. Jia, Z. Ji, Y. Qian, and L. Shen, "Unsupervised band selection for hyperspectral imagery classification without manual band removal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 531–543, Apr. 2012.
[15] R. Hedjam and M. Cheriet, "Hyperspectral band selection based on graph clustering," in *Proc. 11th Int. Conf. Inf. Sci., Signal Process. Appl.*, 2012, pp. 813–817.
[16] M. Ahmad, D. I. U. Haq, Q. Mushtaq, and M. Sohaib, "A new statistical approach for band clustering and band selection using K-means clustering," *IACSIT Int. J. Eng. Technol.*, vol. 3, no. 6, pp. 606–614, Dec. 2011.
[17] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.