Band Selection for Hyperspectral Images: A Survey

Madhuri Kulkarni[1], Priyanka Kadam[2]

[1]Assistant Professor, Dept. of EXTC, SIES GST, Nerul, Maharastra, India
[2]Assistant Professor, Dept. of EXTC, SIES GST, Nerul, Maharastra, India

---***---

Abstract - *Remote sensing is the system which include monitoring, detecting and classifying objects without making physical contact. The data is captured by different remote sensing sensors using satellite communication such as telescopic sensors. The image is captured by this sensor based on reflected or emitted electromagnetic radiation from target on earth. So this type of image is called as Hyperspectral image which has more amount of details than the 2D images. Hyperspectral images having very narrow spectral band with small bandwidth which contains data in terms of light intensity. It is 3D image which gives information of coordinates of pixel in terms of x and y axis and z axis gives data regarding light intensity. As the image is having hundreds of sub bands from electromagnetic visible band, the processing of such large amount of data is difficult using any tool. So it is requiring to convert this infinite details into finite by selecting appropriate spectral bands without losing details of information. In this paper we are going to review different band selection methods for processing of hyperspectral images. Most of the bands are having same information. This redundant information is removed using band selection method. This is the first stage of hyperspectral image processing. The selection of appropriate spectral bands from hundreds of narrow bands can be done based on different statistical parameters, distance metric algorithm, clustering of bands, Spectral Angle mapper. Band selection algorithms are classified based on ranking, searching and clustering methods. These methods are broadly classify as supervisory and unsupervisory methods. Supervisory methods require prior knowledge of HIS data while doing band prioritization.*

Keywords: Hyperspectral image, Spectral bands, Band selection.

1. INTRODUCTION

The remote sensing is a system of data acquisition about an object without making a physical contact on the earth. The data is captured using satellite communication in terms of hyperspectral image in different remote sensing applications. Now a day's researchers are facing different challenges for acquiring details of information from hyperspectral image. As hyperspectral image having huge amount of details captured from electromagnet radiations. So processing of such huge data is very difficult by any tool. Band selection is effective way to reduce the size of hyperspectral image for object classification on earth. There are different band selection algorithms are used to reduce size of image.

Several studies have been proposed for band selection which is categorized as ranking based, searching based and clustering based methods.

Adjacent bands from hyperspectral image that are highly correlated are merged and select the bands that maximize the class separability using the Jeffries–Matusita distance. In the next stage bagger algorithm, SVM and KNN are used to classify the pixels. The Classification Errors Correction (CEC) algorithm is used for classification of objects. [6]

The original bands are grouped in clustered and representative band from each group is selected. Clustering-based band selection is done to contribute an optimal clustering framework (OCF). The rank on clusters strategy (RCS), is used to select bands on existing clustering structure. Also the of number of the required bands are selected using for getting better details from the distinctive information produced by certain number of bands. [2]

The unsupervised K-means Clustering method is used to select finite bands for the analysis of large dimensional hyperspectral image. The Yinyang K-means outperforms K-means algorithms by clustering the centers in the initial stage. The lower and upper bounds between each point leveraged between each point and cluster centre. The hyperspectral image analysis is done using graphics processing units (GPUs) algorithm. [5]

In this technique the hierarchical clustering structure is used to form groups of the bands. The aim of this technique is minimize the intracluster variance and maximize the inter cluster variance. The data redundancy is reduced using distances which are calculated based on mutual information or Kullback–Leibler divergence. This technique gives good accuracy for different dataset images. [7]

The Spectral Angle Mapper (SAM) is another algorithm used to identify the various classes in the image based on the calculation of the spectral angle. This angle is measured between the test vector which is built for each pixel and the reference vector built for each reference classes which is compared with determined threshold angle value. For SAM implementation all these results are gathered together to form three-dimensional Cube of images. [4]

Different methods are used for band selection of hyperspectral image. Every methods has its own advantage for selected spectral band. The band selection fusion (BSF) fuses all these set of selected bands which forms joint band subset (JBS).[1]

The redundant information is reduced using mutual information among spectral bands. But as mutual information does not take in account spatial redundancy of adjacent pixels, new method is used for band selection called as spatial mutual information. Weighting function is designed using difference of corresponding pixel intensity. This weighting function along with mutual information is used for calculation of spatial mutual information. [9]

In certain methods of band selection both supervised and unsupervised methods are combined to get more accurate and efficient result. Hyperspectral bands are ranked using unsupervised methods and accuracy of this result is increased by scoring those rankings. [8]

Band selection is done by observing adjacent bands. The highly correlated bands are merged and select the bands that maximize the class separability using the Jeffries–Matusita distance.[6]

In some algorithms correlation between different sub bands are used to form clustering. In that AP algorithm is used along with wavelet decomposition to processes HSI data. As the image are having high frequency and low frequency components. High frequency component gives details of sharpness of image and noise information and low frequencies are deals with image smoothing information. For clustering this high frequency component is used which gives correlation in each band and signal to noise ratio.[10]

2. THE NECESSITY OF HYPERSPECTRAL BAND SELECTION

In the remote sensing techniques, the data from earth field is captured using telescopic sensors in terms of hyperspectral images without physical contact. This data having huge amount of details related to the object to be detected. As we know that this data is captures using electromagnetic radiation in terms of light intensity. The hyperspectral images are differ from normal 2D images in terms of dimensions. These images having large no. of spectral bands. So each pixel represent hundreds of values in different sub bands. As the processing of such large amount of data is very difficult using any tool. So it is required to reduce the number of spectral bands from image with preserving details of the object. Now a days this is recent trends for researcher to restrict the no. of bands from image so that further processing is possible and gives efficient result. The main objective of this band selection technique is to remove redundant data from the huge amount of data. This will preserve processing time for further analysis.

3. REVIEW OF BAND SELECTION METHOD

Band selection methods can be classified as:

- Ranking-based methods
- Searching-based methods
- Clustering-based methods

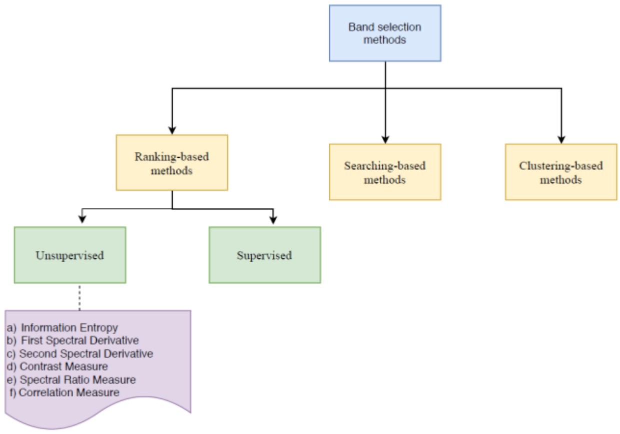

Figure 1: Classification of band selection methods

The band selection methods are classified based on different criteria. Appropriate method is selected based on application of hyperspectral images. In ranking based methods, ranks are given to different sub-bands based on different parameters shown in figure 1.

The best band is selected using optimization problem of a given criterion function in searching based method. In clustering based method, the sub-bands are grouped with the help of different parameters.

A. Ranking-based methods

In ranking-based methods the all spectral bands are sorted according to predefined criteria asper importance of each band. Ranking-based methods aim at designing a criterion to evaluate the importance of each band, and use the top-rank bands to constitute the band subset. The ranking-based methods are again classified into two types as unsupervised and supervised.

1) Unsupervised

Unsupervised methods work on generic information evaluation approaches gives fast and efficient result. Very less or no any preprocessing is needed for this type methods to work on hyperspectral images.

a) Information Entropy

In this method the band selection is done by calculating information entropy of each band using following formula. The mathematical description of this method is shown below,

$$H(\lambda) = \sum_{i=1}^{m} P_i \, In \, P_i$$

H is the entropy measure, *P* is the probability density function of reflectance values in a hyperspectral band and *m* is the number of distinct reflectance values.

A histogram of reflectance values are used to calculate the probabilities. As we know that if amount of information is more the entropy is high. So all the bands are arranged in ascending order with respect to their entropy so band who has high entropy will come first in prioritization.

b) First Spectral Derivative

In this method the adjacent bands who are not differ much are reduced with the help of band similarity using ***First Spectral Derivative***. If two adjacent bands do not differ much then that part of image from geo-spatial property is described using only one band. The mathematical description of this method is shown below,

$$D_1(\lambda_i) = \sum_x \|I(x, \lambda_i) - I(x, \lambda_{i+1})\|$$

Where *I* represents the hyperspectral value, *x* indicate a spatial location and λ is the central wavelength. Thus, if D_1 is equal to zero then one of the band is redundant.

c) Second Spectral Derivative

As we know that the information is more in that subset if derivative of it is large. So another method can be used to sort spectral bands from HSI data is second spectral derivative as a function of added information. With the help of linear interpolation, the middle band is predicted using adjacent two bands which will be considered as redundant band. The high value of second spectral derivative D_2 indicate large amount of information is preserved in selected band.

The mathematical description of this method is shown below,

$$D_2(\lambda_i) = \sum_x \|I(x, \lambda_{i-1}) - 2I(x, \lambda_i) + I(x, \lambda_{i+1})\|$$

Where D_2 indicates second spectral derivative, I is a hyperspectral value, x is a spatial location and λ is the central wavelength.

d) Contrast Measure

In this method each band from HSI data is used for classification. A classification error of only one particular band is measured to find out usefulness of the band. The desirable band is selected based on highest amplitude discrimination (image contrast) to minimize this classification error.

The sum of all contrast values along the boundaries are used to calculate this measure, if class boundaries were known in advance.

In general, bands are sorted as per their *ContrastM value.* Higher values of *ContrastM* bands are ranked first than lower values of *ContrastM*. In discrete case the *ContrastM* is calculated with the help of following formula,

$$ContrastM(\lambda) = \sum_{i=1}^{m} \|f_i - E(f)\| * f_i$$

f is the histogram of all contrast values computed across one band by using Sobel edge detector, *E(f)* is the sample mean of the histogram *f* and λ is the central wavelength. *m* is the number of distinct contrast values in a discrete case.

e) Spectral Ratio Measure

This method explores the band ratio quotients for ranking bands and identifies bands that differ just by a scaling factor. The larger the deviation from the average of ratios *E(ratio)* over the entire image, the higher the *RatioM* value of the band.

The mathematical description of this method is shown below,

$$RatioM(\lambda_i) = \sum_x \frac{I(x, \lambda_i)}{I(x, \lambda_{i+1})} - E\left(\frac{I(x, \lambda_i)}{I(x, \lambda_{i+1})}\right)$$

Where *RatioM* represents the measure, *I* is a hyperspectral value, *x* is a spatial location and λ is the central wavelength.

f) Correlation Measure

If a signal-to-noise ratio is large enough of received signal, the normalized correlation metric gives efficient result. In this measure correlation is calculated just like the spatial autocorrelation method for ranking of all adjacent bands. This method is less sensitive to local mismatches. The normalized correlation measure is calculated as shown below,

$$CorM(\lambda_i) = \frac{E(I(\lambda_i) * I(\lambda_{i+1})) - E(I(\lambda_i)) * E(I(\lambda_{i+1}))}{\sigma(I(\lambda_i)) * \sigma(I(\lambda_{i+1}))}$$

Where *CorM* represents the measure, *I* is a hyperspectral value, *x* is a spatial location and λ is the central wavelength. *E* denotes an expected value and σ is a standard deviation. After selecting the first least correlated band based on all adjacent bands, the subsequent bands are chosen as the least correlated bands with the previously selected bands.

2) Supervised Rank Based Method

The supervised methods require training data to develop an internal predictive model. A training data set is obtained via registration of calibrated hyperspectral imagery with ground measurements. Number of training sets are used to get better accurate result as compare to no. of bands to be selected.

Supervised methods provide more accurate results than unsupervised methods [8]

B. Searching Based Method

In this method band selection is converted into an optimization problem of a given criterion function and the best bands is searched to get desired solution. Two key issues exist in searching-based methods:

a. The criterion function and

b. The searching strategy.

The criterion function can be similarity-based measurements which can be calculated using Euclidean distance (ED), Bhattacharyya distance, Jeffries–Matusita (JM) distance, spectral angle mapping (SAM), spectral information divergence (SID), transformed divergence, MI, and spatial entropy based MI. The searching strategy determines the best way to find an optimal or suboptimal solution.

1) Spectral Angular Mapper Algorithm

The Spectral Angle Mapper (SAM) algorithm is based on an ideal assumption that a single pixel of remote sensing images represents one certain ground cover material, and can be uniquely assigned to only one ground cover class.

This algorithm works simply on the principle of measurement of spectral similarity between two spectra. The spectral similarity is obtained by considering each spectral band as a vector of q number of spectral bands.

The spectral similarity is measured using angle made by two spectra treating them as vectors in a space with dimensionality equal to the number of bands.

The SAM algorithm generalizes this geometric interpretation to n-dimensional space. SAM determines the similarity by applying the following equation:

$$a = \cos^{-1}\left(\frac{\sum_{i=1}^{nb} t_i r_i}{\sqrt{\sum_{i=1}^{nb} t_i^2}\sqrt{\sum_{i=1}^{nb} r_i^2}}\right)$$

Where nb is the number of bands in the image, t is the pixel spectrum, r is the reference spectrum and a is spectral angle [4].

2) Jeffries–Matusita (JM) distance

The Jeffries– Matusita (JM) distance to measure the class separability which is widely used for feature selection.

Consider two classes *i* and *j*, the JM distance between the class *i* and the class *j* is given by the following Eq.

$$J_{ij} = \sqrt{2(1 - e^{B_{ij}})}$$

Where B_{ij} is the Bhattacharyya distance defined as

$$B_{ij} = \frac{1}{8}(m_i - m_j)^T\left(\frac{\Sigma_i + \Sigma_j}{2}\right)(m_i - m_j) + \frac{1}{2}In\left[\frac{|(\Sigma_i + \Sigma_j)/2|}{|\Sigma_i|^{1/2}|\Sigma_j|^{1/2}}\right]$$

The Jij distance is applied for the feature selection in a binary classification problem. To proceed to a feature selection in a multi classification problem, we must to find the features that give the largest average JM distance [6].

The average distance is given by Eq.

$$D_b = \sum_{i=1}^{c}\sum_{j=1}^{c} p(\omega_i)p(\omega_k)J_{ij}$$

C. Clustering based method

In this method, whole spectral bands are divided into different clusters and one of the band from each cluster acts as representative band for further processing. This representative band from each cluster is selected using information measures such as mutual information or Kullback–Leibler divergence [2].

a) Mutual-Information based approach

To remove redundant information, the correlation between the sub bands is used. The dissimilarity approach gives this correlation and that can be measured using mutual information.

The Shannon entropy of a random variable X is given as

$$H(x) = -\int p(x) logp(x)\,dx$$

Where p(x) is probability density function.

In the case of a discrete random variable X, entropy H(X) is expressed as

$$H(x) = -\sum_{x\epsilon\Omega} p(x) log p(x)$$

Where p(x) represents the probability of an event x ∈Ω from a finite set of possible values.

The average amount of *information* of a random variable is called as Entropy. The mutual information quantifies the statistical dependence of random variables. It gives us effectiveness of prediction of one variable on other.

Consider a set of *L* discrete random variables having their corresponding bands $X_1,\ldots X_L$ from a hyperspectral image. $I(X_i, X_j)$ is defined as

$$I(x_i, x_j) = \sum_{x_i\epsilon\Omega} \sum_{x_i\epsilon\Omega} p(x_i, x_j) log \frac{p(x_i, x_j)}{p(x_i)p(x_j)}$$

I is always a nonnegative quantity for two random variables. The value of I is zero indicates the given two random variables are statistically independent.

The variables are mostly dependent on each other if value of I is large.

Mutual information *I* can be calculated with the help of entropy as shown below,

$$I(x_i, x_j) = H(x_i) + H(x_j) - H(x_i, x_j)$$

Where $H(x_i, x_j)$ is the joint entropy calculated using joint probability distribution $p(x_i, x_j)$.

b) Divergence-Based Criterion:

Another approach to find out dissimilarity measure between two image bands is with the help of by their corresponding probability distributions.

The dissimilarity distance between two probability distributions is measured using the Kullback–Leibler divergence. The symmetric version of the Kullback–Leibler divergence is often used which gives the real distance.

Let us consider x_i and x_j two random variables that are defined in Ω space, representing the i^{th} and j^{th} bands of a hyperspectral image. Let us assume that $P_i(x)$ and $P_j(x)$ are the probability distributions of these random variables [7]. Thus, the symmetric Kullback–Leibler divergence can be expressed in the discrete domain as follows:

$$D_{KL}(x_i, y_j) = \sum_{x\epsilon\Omega} P_i(x) log \frac{P_i(x)}{P_j(x)} + \sum_{x\epsilon\Omega} P_j(x) log \frac{P_j(x)}{P_i(x)}$$

The Kullback–Leibler divergence is always nonnegative, when pi(x) and pj(x) are the same probability, the divergence is being zero.

c) The K-means clustering algorithm

K-means is one of the easiest unsupervised learning algorithms and most widely used method to group data in a specified number of clusters.

K-means is most widely used algorithm for clustering in different areas. In this method we have to convert n data points into k clusters based on similarities between data points. The centroid of the cluster is calculated using simple averaging. Based on the similaries, data points are merged in that cluster whose centroid is close towards that data point. The centroid of that cluster is again redefined by calculating average of it. This process is continue unless all data points will cover and merged in clusters. [5]

d) Affinity Propagation Clustering (AP)

This algorithm is stable and gives effective result for multi-class and large-scale data. Affinity Propagation is the exemplar based clustering method in which cluster centers are indicated by real data points called as exemplar. All the data points from HSI data are considered as initial exemplar. In AP clustering algorithm similarities between data points are considered to form clustering. The similarity is nothing but how well the data point is suited to be the exemplar. These similarities may be positive, negative or non-symmetric in AP method[11]. To find out theses similarity most probably negative squared Euclidean distance is used using following formula as,

$$S(x_i, x_j) = -\|x_i - x_j\|^2$$

4. DISCUSSION

The techniques used for band selection should remove the redundant information while keep significant information for further analysis. Ranking based methods are less complex as compared to other methods but it gives less accurate result.

Searching based and clustering based band selection methods have more consistent behavior and steady outputs on different databases.

5. CONCLUSION AND FUTURE SCOPE

In this paper we have introduced classification of band selection techniques for hyperspectral images in remote sensing applications. In recent trends variety of band selection techniques are proposed for selection of proper bands and reducing redundant band from HSI data. These all techniques are broadly classified into three categories such as ranking based, searching based and clustering based. Ranking based methods gives us statistical analysis such as information entropy, first spectral derivative, contrast measure, spectral ratio measure of HSI data and according to that the spectral narrow bands are sorted in ascending order so select appropriate band for further analysis.

In searching based methods work on criterion function and the searching strategy to select appropriate band from HSI data. The criterion function gives measure of Euclidean distance (ED), Bhattacharyya distance, Jeffries–Matusita (JM) distance, spectral angle mapping (SAM). Using searching strategy desired best way is find out to get optimal solution.

In clustering based methods the entire spectral bands from HSI data are divided into different clusters and one band from each cluster is selected to form final band subset. To remove the redundant bands, representative bands are selected using information measurements or Kullback–Leibler divergence. Some clustering algorithms are k-means, affinity propagation (AP), and graph clustering.

As for our future work, we are interested to analyze performance of all techniques on same database to compare results for best band selection. Also we can work on number of bands used for further processing. The no. of bands selected after applying band selection method should not be too small or too large.

REFERENCES

[1] Y. Wang, L. Wang, H. Xie and C. I. Chang, "Fusion of various band selection methods for hyperspectral imagery," *Remote Sensing,* vol. 11, no. 18, pp. 1-19, 2019.

[2] Q. Wang, F. Zhang and X. Li, "Optimal Clustering Framework for Hyperspectral Band Selection," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 56, no. 10, pp. 5910-5922, 2018.

[3] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geoscience and Remote Sensing Magazine,* vol. 7, no. 2, pp. 118-139, 2019.

[4] S. Rashmi, S. Addamani, Venkat and S. Ravikiran, "Spectal Angle Mapper Algorithm for remote Sensing Image Classification," *International Journal of Innovative Science, Engineering & Technology,* vol. 1, no. 4, pp. 201-205, 2014.

[5] M. Paoletti, J. Haut, J. Plaza and A. Plaza, "Yinyang K-means clustering for hyperspectral image analysis," Proceedings of the 17th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2017, pp. 1625-1636, 2017.

[6] S. A. Medjahed, T. Ait Saadi, A. Benyettou and M. Ouali, "A new post-classification and band selection frameworks for hyperspectral image classification," *Egyptian Journal of Remote Sensing and Space Science,* vol. 19, no. 2, pp. 163-173, 2016.

[7] A. Martínez-Usó, F. Pla, J. M. Sotoca and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 45, no. 12, pp. 4158-4171, 2007.

[8] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selection," *Photogrammetric Engineering and Remote Sensing,* vol. 70, no. 7, pp. 793-802, 2004.

[9] A. Amankwah, "Spatial mutual information based hyperspectral band selection for classification," *Scientific World Journal,* vol. 2015, 2015.

[10] Z. Ren and L. Wu, "Affinity Propagation for Hyperspectral Band Selection," *IOP Conference Series: Earth and Environmental Science,* vol. 170, no. 2, 2018.

[11] L. Jiao, J. Feng, F. Liu, T. Sun and X. Zhang, "Semisupervised Affinity Propagation Based on Normalized Trivariable Mutual Information for Hyperspectral Band Selection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 8, no. 6, pp. 2760-2773, 2015.