Hyperspectral band selection via clustering algorithms

Barat Mojaradi
Hamid Abrishami Moghadam
Mohammad Javad Valdan Zoeje

Hyperspectral technology, compared to the multi-spectral, is capable of reconstructing spectral signatures of phenomena, as well as producing a spectral library for earth observation. Determination of informative bands is a challenge for hyperspectral data analysis. In particular spectral based band selection (SBBS) with limited training data or based on a spectral library is demanded for remote sensing applications. In general, supervised feature selection methods, regardless of their search strategy stage, can be classified into two approaches based on their predefined criteria. The first one, called the filter approach, is based on optimising a discrimination measure, such as the Mahalanobis distance, Bhattacharya distance, etc. The filter approach operates independently from any classification algorithm, so undesirable features are omitted before the classification process begins. The second one, called the wrapper approach, tries to optimize the classification accuracy of the desired classifier by selecting feature subsets. Feature shaving is an example of this category. Recently also unsupervised feature selection procedures based on feature similarity and mutual information have been proposed. Each band selection category has its special merits and shortcomings based on the optimisation and computation costs. The band selection methods result in suboptimal solutions. This is because they depend on the search strategy with some of them trying to optimize predefined pair-wise class discriminant measures. They suffer from some shortcomings, such as high correlation between neighbouring bands (when computing a separability measure) and classification with limited training samples in high dimensional space at the beginning of backward search or at the final steps of forward search algorithms.

This paper proposes an innovative SBBS method, so called prototype space band selection (PSBS), which uses only the class spectra. The main novelties of PSBS lie in: band representation in new space (called prototype space), where bands are characterized in terms of class reflectivity to pose reflectance properties of classes to bands; and using uncertainty and angle measures to distinguish highly correlated and informative bands. In general, feature vectors consist of a set of elements that describe objects. From a pattern recognition point of view, a space should possess some particular properties so that a finite representation of objects can be characterized for the learning process. For hyperspectral images, a feature vector is defined in terms of spectral response of pixel $\mathbf{x}=[x_1,x_2,...,x_n]^T$, where n is the number of bands; hence, the pixels of an image are represented in the band space (axes of this space are made by bands of hyperspectral data) to perform different types of analysis, such as clustering and classification. The representation of pixels in feature space is thereby appropriate for image classification tasks. The band selection methods try to study and analyse similar bands to represent and condense pixels spectra with low dimensions. Intuitively, to study and find relevant bands it is reasonable to express bands in terms of their properties. Accordingly, we propose to represent bands required for this type of analysis in the prototype space. In this space a band is characterized in terms of the spectral response of different classes to pose reflect properties of classes to bands. Indeed feature vectors in this space describe the band behaviour in terms of their reflectance in dealing with imaging scene prototype.

Let us assume that the pixels in an image belong to L classes, the spectra are given by n bands, and the classes can be represented by a single prototype spectrum like the class mean. We can, then, denote the characteristic vector of band i $\mathbf{h_i}=[m_{1i},m_{2i},\ldots, m_{ji},...,m_{Li}]^T$ in prototype space, where m_{ji} is the refers to the class j in band i and L is the number of classes. Therefore, the prototype space has L dimensions. We will use this space to study and cluster bands based on the similarity of their behaviour.

Having clustered the bands by clustering algorithms like K-means and Mode seeking in prototype space, highly correlated bands and isolated bands are separated by entropy measure. Consequently, PSBS is built by choosing nearest band to the cluster centers as proper representative of bands that fall in each cluster. Furthermore, we will enrich PSBS with informative channels which are outlier channels identified through their angle with respect to diagonal centers in the prototype space. Compared to previous band selection methods like Sequential Floating Forward Selection (SFFS) and Sequential Floating Backward Selection (SFBS), in PSBS, the search strategies are substituted with clustering algorithms to find relevant bands. Moreover, instead of optimizing separability criteria, the overall accuracy of classification over validation data set is used to decide which bands yield maximum accuracy. Experimental results demonstrated higher performance of the PSBS method with respect to its conventional counterparts in terms of overall accuracy with limited sample sizes.