Unsupervised Hyperspectral Band Selection Using Parallel Processing

He Yang, Qian Du‡
Department of Electrical and Computer Engineering
Mississippi State University, USA
‡corresponding and presenting author, du@ece.msstate.edu

A hyperspectral image cube contains hundreds of spectral bands with very fine spectral resolution. It contains abundant spectral information that permits precise object identification. But its vast data volume brings about problems in data transmission, storage, and analysis. As a general approach, band selection can extract several most important bands and discard the rest to achieve dimension reduction by taking advantage of high spectral correlation in hyperspectral imagery. In most application areas, the goal of hyperspectral image analysis is to detect or classify objects. Driven by detection or classification accuracy, one would expect that using a subset of original bands the accuracy is unchanged or tolerably degraded while computational burden is significantly relaxed.

In terms of object information availability, band selection techniques can be divided into two categories: supervised and unsupervised. Supervised methods are to preserve object information, which is known *a prior*, while unsupervised methods do not assume any object information. Although supervised techniques clearly aim at selecting bands that include important object information and the selected bands generally provide better detection or classification than unsupervised techniques, the required prior knowledge may be unavailable in most of practical applications. So it is a need to develop unsupervised band selection methods that can generally provide good performance regardless of object types.

The objective of unsupervised band selection techniques is to select the most informative and distinctive bands. In our research, we use signal-to-noise ratio (SNR) to measure the information contained in each band, and a spectral similarity metric for band distinctiveness. In particular, the approaches that are proposed to search for distinctive spectral signatures as endmembers were applied to find the most distinctive bands sequentially in [1]. An effective method was also proposed to find the suboptimal initial band pair for the band selection algorithm. This unsupervised band selection technique outperformed other algorithms published in the literature.

To further reduce the computational complexity, we will implement this algorithm in parallel. Let L denote the number of bands in a hyperspectral image. The parallel algorithm is implemented in Open MPI and C++. Its major steps are:

1. Prepare data:
   1.1 Distribute all the bands to each processor. Each processor has the data from all the bands to avoid communication burden.
   1.2 Assign each processor the index of target bands for processing.
2. Select starting band pairs:
   2.1 For each processor, start with initial band 1 and add band 1 to the searched band pool.

2.2 For each processor, compute the similarity between bands in the searched band pool and each band in the target band pool with the method in [1].
2.3 Use all-to-one reduction to get the least similar band globally.
2.4 Use the found band as newly initial band, use one-to-all broadcast to send the band index to all the processors, and update the searched band pool with the newly found initial band's index. Store previously searched bands as the previous initial bands.
2.5 Loop Step 2.2-2.4 until it achieves convergence, i.e. the newly found initial band is the previously initial band. Then store the newly found band to the searched band pool.
2.6 Broadcast the initial band pair to all the processors.

3. Adjust loading balancing:
   Reassign the (L-2) target band indices to each processor.
4. Search the bands to be selected:
   4.1 For each processor, select the next band with the approach in [1] using the global searched band pool and the local target band pool.
   4.2 Use all-to-one reduction to find the next global selected band and add it to the global searched band pool.
   4.3 Redo Steps 4.1-4.2 until the number of bands added to the searched band pool is enough. Adjust the loading balancing in Step 3 as needed.
5. Output the final searched band pool as the parallel selected bands.

Preliminary results show that the parallel algorithm provides very similar selected bands. In the final paper, we will present the detailed performance discussion of the proposed parallel band selection algorithm.

References:

[1] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 564-568, Oct. 2008.