

Progressive Band Selection

Kevin Fisher^a and Chein-I Chang^b

^a NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, Maryland USA 20771

^b University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland USA 21250

ABSTRACT

Progressive band selection (PBS) reduces spectral redundancy without significant loss of information, thereby reducing hyperspectral image data volume and processing time. Used onboard a spacecraft, it can also reduce image downlink time. PBS prioritizes an image's spectral bands according to priority scores that measure their significance to a specific application. Then it uses one of three methods to select an appropriate number of the most useful bands. Key challenges for PBS include selecting an appropriate criterion to generate band priority scores, and determining how many bands should be retained in the reduced image. The image's Virtual Dimensionality (VD), once computed, is a reasonable estimate of the latter. We describe the major design details of PBS and test PBS in a land classification experiment.

Keywords : Progressive band selection, dimensionality reduction, priority score.

1. INTRODUCTION

Hyperspectral images contain far more information than is necessary for most applications. The Hughes phenomenon [1], in fact, shows that trying to make use of all the data in an image yields less accurate results than if some information were intentionally thrown away. Discarding information therefore has two benefits: it reduces processing time and improves the quality of the application's output. Additionally, some hyperspectral imagers are installed on spacecraft with slow downlink transmitters, so removing information reduces the time needed to send the image to the ground.

Naturally, one must carefully choose which information to throw out. Often the information needed to distinguish two types of vegetation (in land classification scenarios) or detect a camouflaged tank (in target detection applications) lies in a few critical bands. Once these bands are identified, however, often the rest can be discarded safely.

We developed Progressive band selection (PBS) to detect the most critical image bands for a target application and build reduced images with only as many bands as are needed. The first step requires a *criterion* on which to judge the image bands and assign a numerical score. While PBS can use any real-valued function to generate these *priority scores*, most of the criteria studied here calculate a statistic (such as variance) for all pixels in the band.

After assigning scores to each band, PBS builds a reduced image containing only the highest-scoring bands. The user can either directly specify the number of bands to keep, or specify an acceptable level of application performance. In the latter case, PBS can use one of three methods to search for the proper number of bands to retain.

This section describes PBS and its benefits for hyperspectral image processing applications. Section 2 describes the criteria used to identify the most valuable bands in an image, and Section 3 describes how PBS finds the correct number of bands to include in its output. Section 4 describes an experiment we performed to test our band selection and output image sizing methodologies in a land classification application. Concluding remarks are found in Section 5.

2. BAND SELECTION CRITERIA

We use eight criteria in this experiment to rank image bands. Three compute central moments about the data, two more calculate more complex statistical measures, and three are used as experimental controls.

2.1. Central moments

These criteria treat the image band as a set of samples x_i from a random variable X , then compute its central moments. Although a data set with high variance is not automatically more useful than one with low variance, greater variance typically implies greater information content. Likewise, larger measures of other central moments generally indicate the image has features useful to the target application.

2.1.1. Variance (second central moment)

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (1)$$

where x_i is the value of pixel i , \bar{x} is the mean value of all pixels in the band, and N is the number of pixels.

2.1.2. Skewness (third central moment)

$$skewness = \frac{1}{N-1} \sum_{i=1}^N \frac{(x_i - \bar{x})^3}{\sigma^3} \quad (2)$$

where σ is the standard deviation of all pixels in the band.

2.1.3. Kurtosis (fourth central moment)

$$kurtosis = \frac{1}{N-1} \sum_{i=1}^N \frac{(x_i - \bar{x})^4}{\sigma^4} \quad (3)$$

2.2. Infinite-order statistics

2.2.1. Entropy

Entropy is the classic measure of information content in a data set, defined as $H(x)$:

$$H(x) = - \sum_{i=1}^N p(x_i) \log_2 p(x_i) \quad (4)$$

where $p(x)$ is the probability of pixel x within the image X (calculated by constructing a histogram of image pixels) and N is the number of pixels in the image. This equation calculates $H(x)$ in bits per pixel.

2.2.2. Information Divergence

Information Divergence [2] is an alternative measure of information content. It calculates the number of additional bits needed to represent a dataset with one distribution using a code designed for another distribution. For our purpose, the former (denoted p below) is the image band, and the latter (g) is a dataset with a Gaussian distribution with the same mean and variance as the image. The more the distribution of data in the image deviates from the Gaussian model, the higher its information divergence (D), and the more information it is presumed to contain.

$$D(p; g) = \sum_{i=1}^N p_i \log(p_i / g_i) + \sum_{i=1}^N g_i \log(g_i / p_i) \quad (5)$$

2.3. Control criteria

The remaining criteria can be thought of as naïve alternatives to computing statistical measures for each band. They are intended as control variables for the coming experiment.

2.3.1. Random

The random classifier assigns a random score between 0 and 1 to each image band.

2.3.2. Uniform

The uniform classifier scores bands in a way that ensures the selected bands are distributed evenly across the original frequency range. For example, if we intend to select 100 of 200 bands from an image, the uniform classifier will assign a high score (between 2 and 3) to odd-numbered bands and a low score (between 0 and 1) to even-numbered bands. The image output by PBS will therefore contain every other band across the original range.

2.3.3. Sensor-based

The sensor-based classifier uses basic information about the image sensor to improve the output of the random classifier. This classifier assigns a zero score to bands known to be uncalibrated, masked, or of otherwise low quality, then gives random scores to the remaining bands. This optimization adds very little complexity to the random selection algorithm.

3. PROGRESSIVE BAND SELECTION

Progressive band selection is designed to tailor the number of bands in the reduced image for the target application. If the user does not directly specify the number of bands, PBS will search for the right number of bands to retain to meet an application-specific performance target. Figure 1 illustrates the three search strategies.

3.1. Progressive band expansion

Starting with an empty output image, Progressive band expansion (PBE) adds the highest-ranking image band to the output image and measures its performance in the target application. If higher performance is needed, PBE adds the next highest-ranking band and measures the performance again. This process continues until the application meets its performance target using the output image.

3.2. Progressive band reduction

The logical inverse of PBE, Progressive band reduction (PBR) starts with a copy of the input image as its output. It then removes the lowest-ranking band and re-measures application performance. The process continues until enough bands have been removed to reduce application performance to match (or fall just below) the target level.

3.3. Binary bisection band selection

PBR and PBE search iteratively for the right number of bands to retain. By contrast, Binary bisection band selection (BBBS) performs a binary search. BBBS begins with an output image containing the highest-ranking 50% of bands from the input image (the midpoint between an empty image and the full dataset). If application performance is above the threshold, the lowest-ranking half of these bands are removed (leaving the highest ranking 25%). However, if performance was inadequate, half of the bands that were excluded from the first dataset are added, such that the output image consists of the highest-ranking 75% of input bands. Performance is measured again, and if necessary the number of bands is adjusted by a factor of 1/8 (12.5%). The process continues until performance is within an acceptable range.

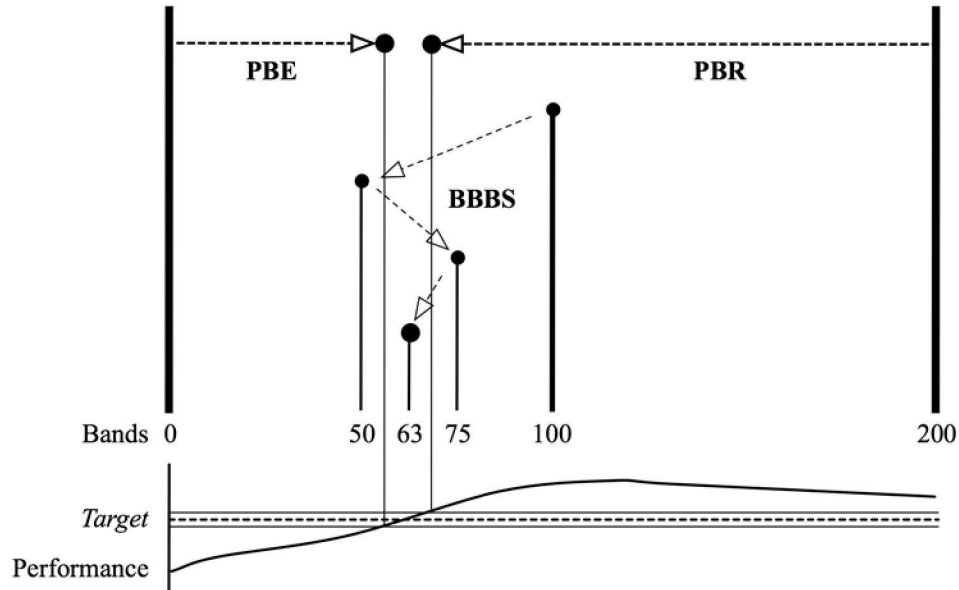


Figure 1. PBE and PBR search iteratively for the correct number of bands, while BBBS employs a binary search.

4. EXPERIMENT

4.1. About the test image

The test image is a small subset (shown in Figure 2) of an image taken by the Hyperion instrument onboard the NASA EO-1 spacecraft. The image depicts a suburban and mountainous forest area near Tuscon, Arizona on June 17, 2003 at 10-10:15am local time. [4] The original image is 5585 pixels long, however for this experiment we use a 512 pixel subset of the image, centered over the suburban area. Both images have 256 spatial columns, and each pixel covers a 30m x 30m area. There are 242 spectral bands in the test image, covering wavelengths between 400 and 2500 nm.

Atmospheric correction of the image is not necessary since all the training data needed for both classification algorithms are taken from the image itself. The same atmospheric effects affect all training and test pixels.

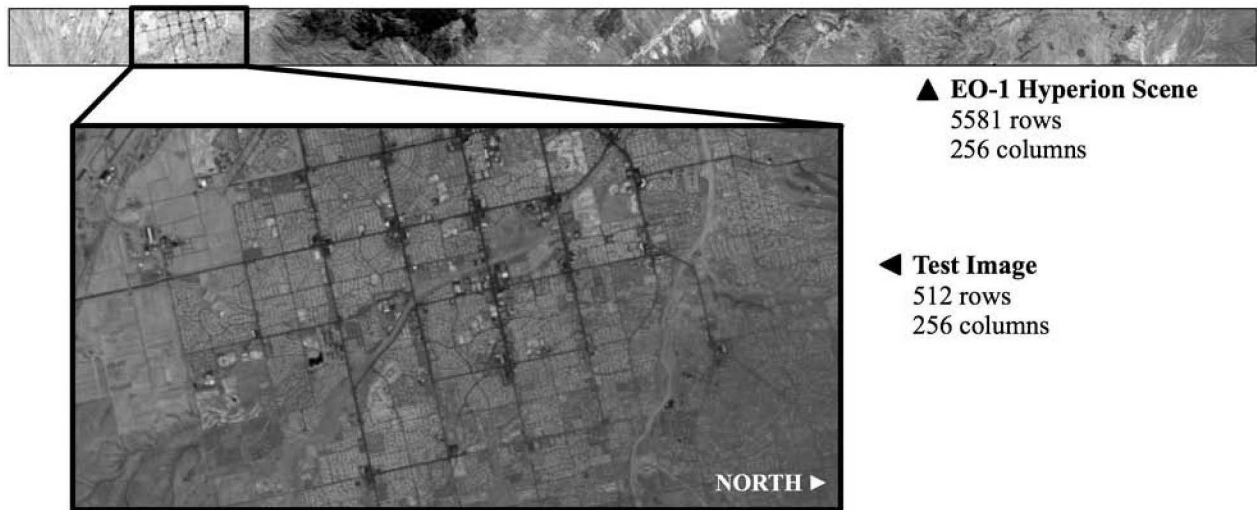


Figure 2. The test image is a small subset of a scene generated by EO-1 Hyperion.

Figure 3 shows the major features in the dataset. A suburb covers the center of the image; a grid of paved streets encircles residential areas and has shopping malls and parking lots at its intersections. The residential areas consist of grassy lawns along narrow side streets. The largest expanse of paved ground is the runway at the airport, south of the suburbs, and the largest expanses of grass are found at the golf course just to its north. A forest fire rages in the mountains to the north, and while the fire itself is not included in the test image, the upper-right corner is shrouded by smoke from the blaze. The only detectable pools of water in this parched suburb are a lake and a small reservoir.

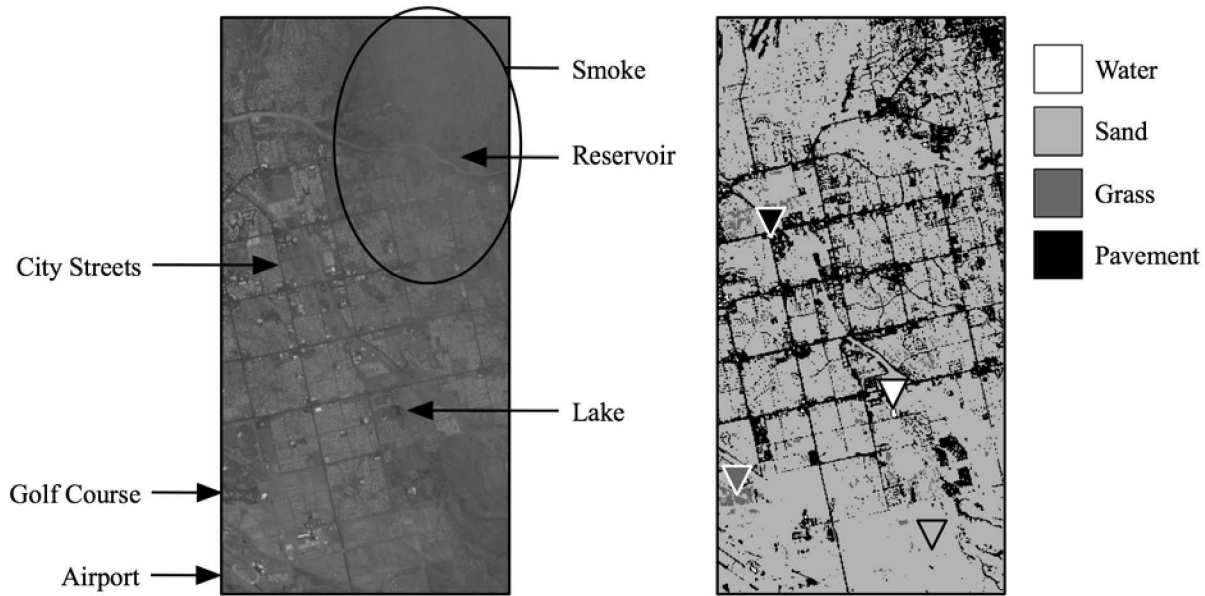


Figure 3. Left: Major features in the test image. Right: map of materials present in the test image. Triangles indicate where spectral samples of each material were taken.

4.2. Experiment

Typically, PBS uses one of the three methods described above to determine an acceptable number of bands to include in the output image. However, one aim of this experiment is to test Virtual Dimensionality (VD) as an initial estimate of the output band count. VD is a measure of the number of spectrally distinct signals that combine to produce the spectra in the output image. [3] PBS will prepare output images with band counts equal to several multiples of VD: [0.25, 0.5, 1, 1.5, 2, 2.5]. Since the Virtual Dimensionality of the test image is 35, we configured PBS to prepare output with 9, 18, 35, 53, 70 and 88 bands images using each of the eight criteria. The reduced images PBS produces will serve as input to two land classification algorithms.

PBS requires a certain amount of processing time. Ideally, the time saved by processing the reduced image in place of the full-size image will more than offset the cost of using PBS. Figure 4 shows that the time needed to process the test image with PBS varies with the criterion used. The skewness and kurtosis criteria are the most computationally expensive, followed closely by information divergence. All three use expensive mathematical operations to generate priority scores, extensive processing that may not be worth the effort.

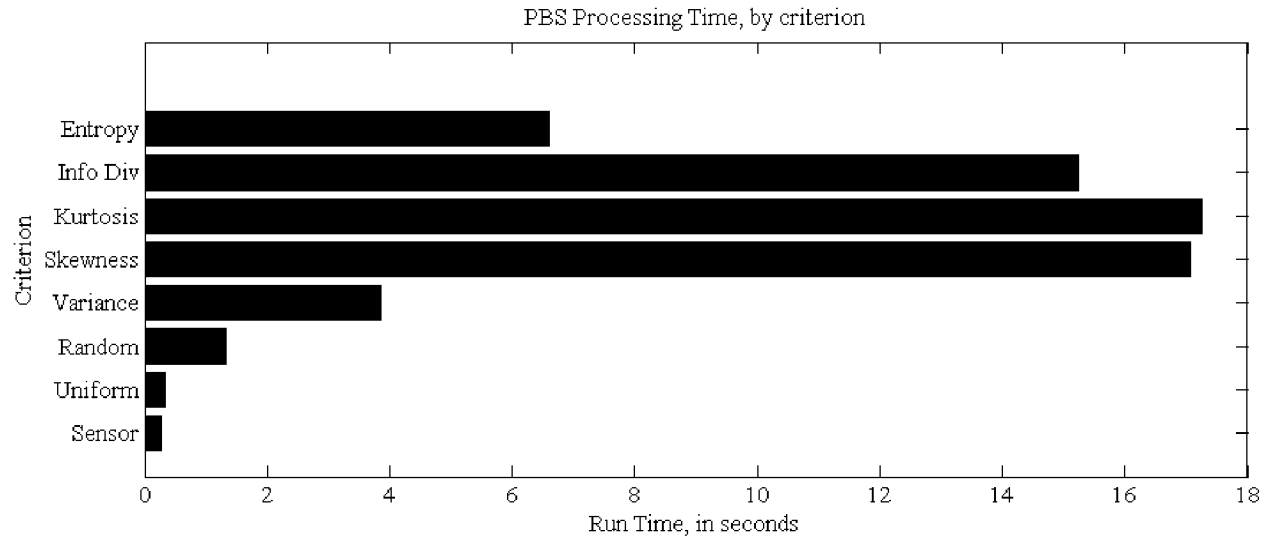


Figure 4. Computationally complex criteria exact a steep performance penalty.

4.2.1. Spectral angle mapping (SAM)

Spectral angle mapping (SAM) computes the similarity of two image pixels by treating them as vectors in a high-dimensional space and measuring the angle between them. This technique is useful for satellite images because it ignores differences in the intensity of sunlight in each pixel, focusing instead on the differences in the at-sensor spectral radiance of the two materials. SAM makes material classification maps by comparing each pixel to a set of representative material spectra and coloring the map according to which material spectra makes the smallest angle with each pixel.

The four dominant materials present in the test image are grass, pavement, soil and water. The triangles in Figure 3 show where pixels representing these materials were selected from the original image. The materials in these pixels were verified using Google Earth imagery of the same area. Figure 3 also shows a material map created using the full test image. In this experiment, output images from PBS are processed by SAM in order to attempt to generate the same map.

Figure 7 shows the maps generated by SAM from each of the reduced images. The information divergence, kurtosis and skewness criteria failed to make reduced images with fewer than 53 bands that SAM could process. This is because the criteria selected mostly dark and uncalibrated bands, due to the fact that these bands have data that are heavily skewed from the expected Gaussian distribution. These criteria consider them valuable, when in fact they are just noisy. Larger datasets produced by these criteria (which are guaranteed to contain at least some calibrated bands) still produced poor material maps: kurtosis caused misclassification of the smoky area in the upper-right corner, and skewness caused most of the central area to be classified as pavement.

Other criteria fared better: with as few as 18 bands (one half of the image's Virtual Dimensionality), entropy, variance and the three control criteria (random, uniform and sensor-based) produced images that yielded nearly correct SAM output. All five appeared to perform almost equally, indicating that 18 bands ($1/2 \times VD$) is enough to perform the classification and still allow for slight variance in the set of bands chosen. Nine bands ($1/4 \times VD$), however, appears to be a bridge too far: only the uniform classifier produced reliable output.

Smaller images can be processed faster than larger ones, but Figure 5 shows that processing an 18-band version of the test image saves only three seconds over processing an 88-band version. Since all five of the non-control criteria take longer than three seconds to execute, PBS pre-processing is uneconomical at best.

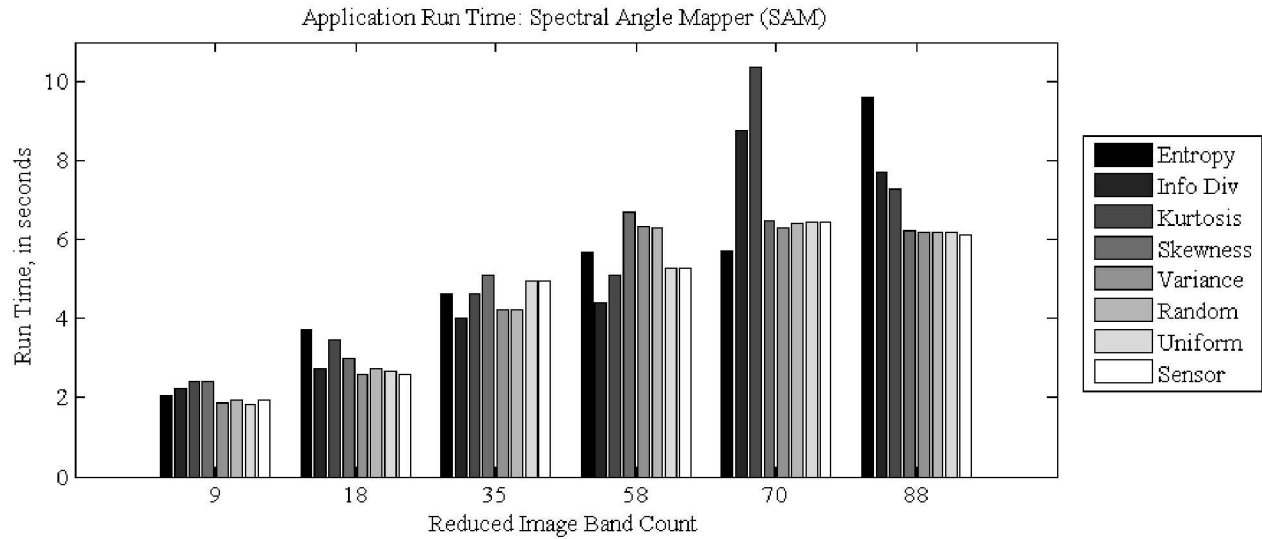


Figure 5. Smaller input images require less processing time, meanwhile some PBS criteria take longer to evaluate the image than others.

4.2.2. Support vector machine (SVM)

We used libsvm 2.89 [5] to create a support vector machine (SVM)-based classifier. The classifier uses C-support vector classification (C-SVC), the radial basis function (RBF) kernel, and the one-against-one approach to multi-class classification. Using a gamma value of 2 gives the most accurate output in this configuration.

We chose forty-eight pixels from the test image to serve as the training set, twelve for each material. The pixels were gathered from the regions marked with triangles in Figure 3. Training the SVM took an insignificant amount of time.

SVM output maps, shown in Figure 8, look different than their SAM counterparts. SAM tends to find a "default" material choice; all pixels tend to be classified as this material (sand, in the case of the test image) unless their spectra are strikingly similar to one of the other materials. In SVM output maps, more pixels are classified as pavement or grass, even when they contain a substantial proportion of sand.

As noted previously, the information divergence, skewness and kurtosis criteria failed to provide reduced images with fewer than 53 bands that could be processed by SVM. Even when allowed to select 70 or 88 bands, these three criteria produced images that made it easy for SVM to misclassify pixels in areas obscured by smoke.

Results for other criteria also mirrored the results obtained using SAM. Entropy, variance and the control criteria produced images with as few as 18 bands ($1/2 * VD$) that yielded output maps that were virtually identical to those made with reduced images with 88 bands or the full image. Given only nine bands, SVM appears to outperform SAM: maps produced with the variance and sensor-based criteria are remarkably similar to those made with the full-size image. However, the performance of the other criteria given so few bands indicates that this output is unreliable.

Figure 6 shows that SVM is a more efficient algorithm than SAM, making similar maps in roughly half the time. However, as with SAM the time saved is far less than the time PBS spends to process the images.

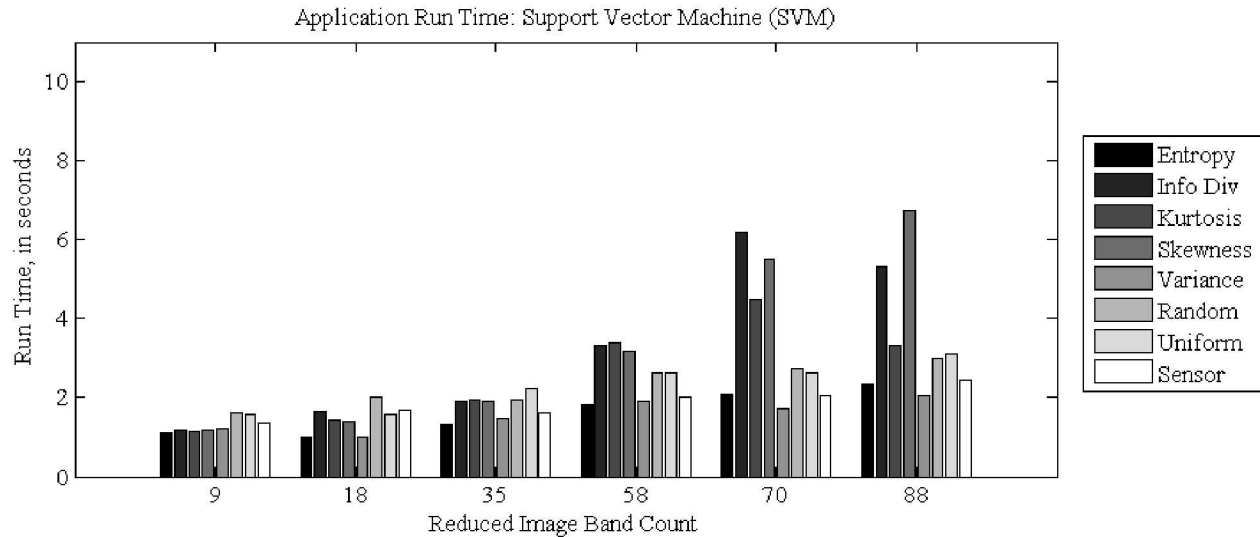


Figure 6. SVM is more efficient than SAM, but still benefits from smaller input images.

5. CONCLUSION

PBS attempts to reduce processing time by removing spectral redundancy from hyperspectral images. Our experiment confirms that in land classification scenarios smaller datasets are processed in less time than the full image yet produce similar output. However, the savings does not offset the overhead of running PBS with any of the five non-control criteria. It is worth pointing out, though, that the control criteria, meant to represent naïve methods of band selection, produce reduced images with equal or greater quality than the non-control criteria, and run quickly enough to make PBS an economical way of improving processing efficiency. Future research will search for band selection criteria that can outperform naïve approaches.

6. REFERENCES

- [1] G. Hughes. "On the mean accuracy of statistical pattern recognizers", IEEE Trans. Information Theory, (1968).
- [2] S. Kullback. "The Kullback-Leibler distance", The American Statistician, 41:340–341 (1987).
- [3] J. Harsanyi, W. Farrand, and C.-I Chang, "Determining the number and identity of spectral endmembers: An integrated approach using Neyman–Pearson eigenthresholding and iterative constrained RMS error minimization", Proc. 9th Thematic Conference on Geologic Remote Sensing, (1993).
- [4] EO-1 Hyperion image EO1H0360372003184110KW. Data available from the U.S. Geological Survey.
- [5] C.-C. Chang and C.-J. Lin, "LIBSVM : a library for support vector machines", (2001). Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

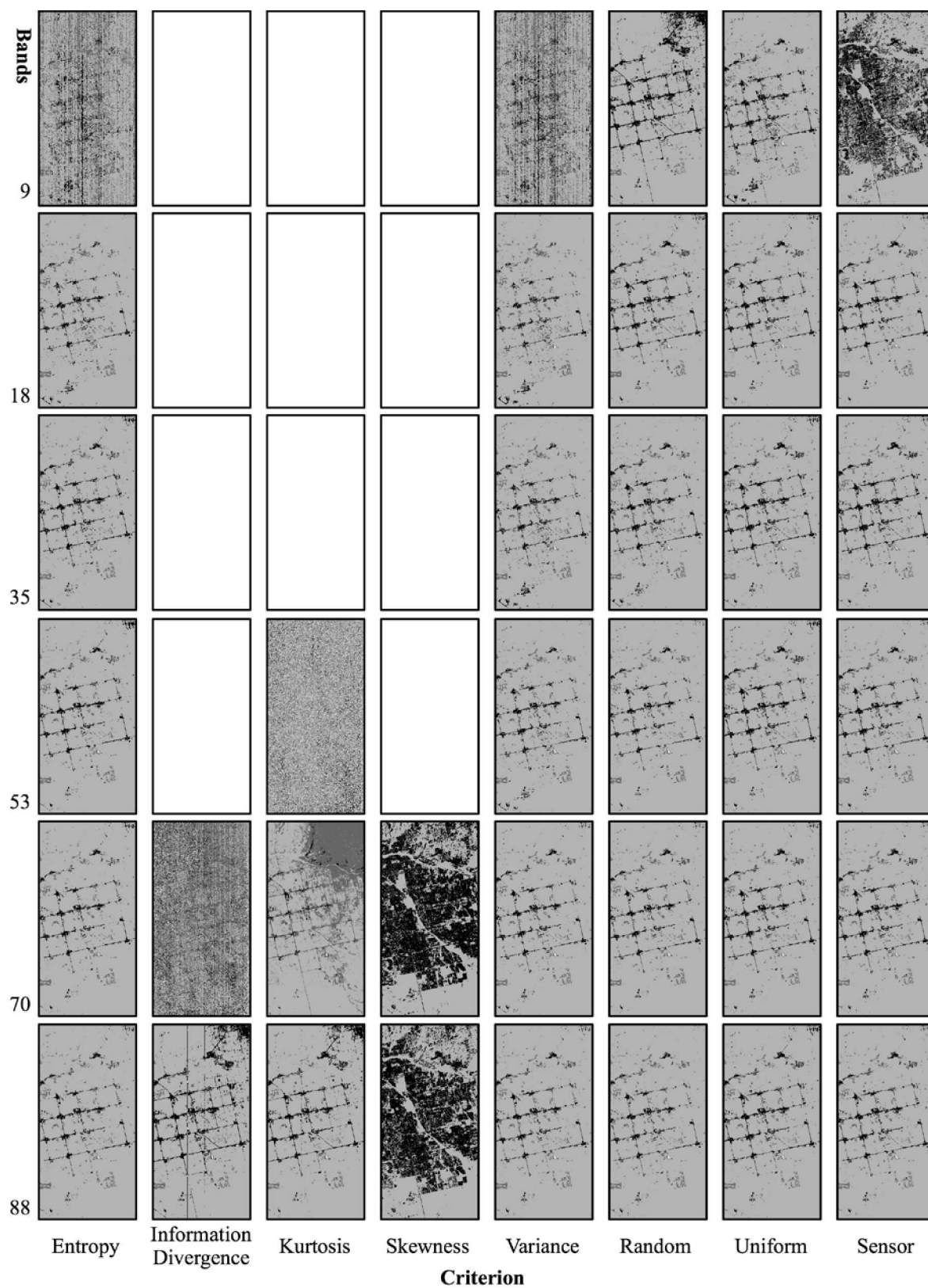


Figure 7. Output of spectral angle mapping (SAM) classifier given reduced images.

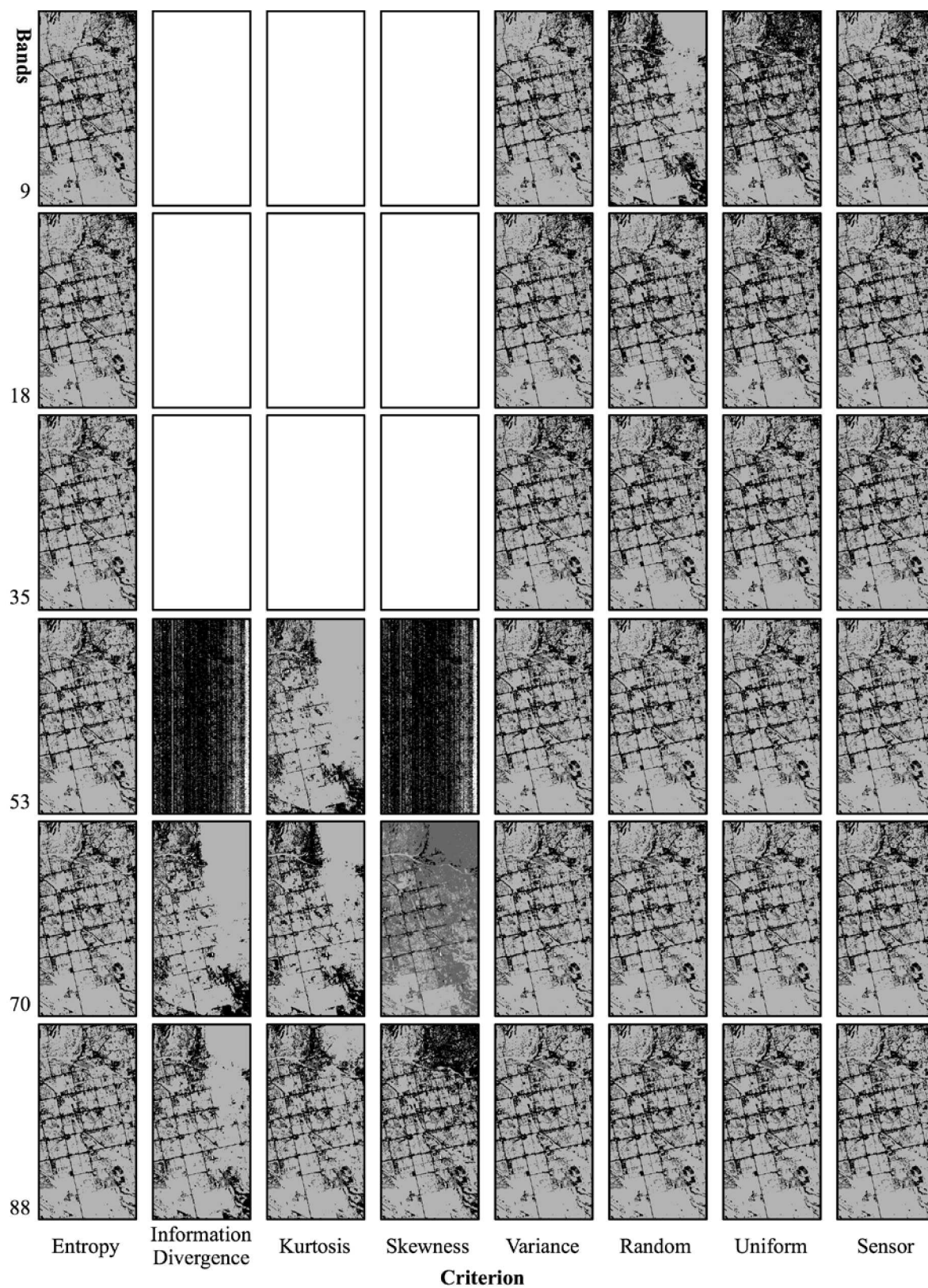


Figure 8. Output of support vector machine (SVM) classifier given reduced images.