SOME TECHNIQUES FOR ANOMALY DETECTION IN HYPERSPECTRAL IMAGERIES

José A. Malpica, Professor
María C. Alonso, Professor
Borja Rodríguez, Postgraduate Student
School of Geodesy and Cartography.
University of Alcala
28871 Madrid, Spain
josea.malpica@uah.es
mconcepcion.alonso@uah.es
borja.rodriguezc@uah.es

ABSTRACT

This research focuses on the very first step in the analysis of an image, the point at which one assumes no prior knowledge about the statistical characteristics of the pixels in the image and where little or nothing is known about the size and shape of the objects to be detected. Therefore, the only available option is to look for a point (or group of points) that deviates so much from other points as to arouse suspicion that it was generated by a different mechanism. This project does that by looking for one-dimensional projection (projection pursuit) optimizing some measurement of interest (index). Following work performed by the authors in visualization techniques for anomaly detection combining low components of PCA and RX (Alonso and Malpica, 2009), this work analyzes and compares index skewness and kurtosis with the popular RX index. The optimization for the plane projection is performed with a genetic algorithm. These indexes are tested in synthetic image and in AHS hyperspectral imagery. The current project shows how these indexes have their properties and characteristics and how they are superior to RX in many ways. Further, the authors proposed visualization techniques for anomaly detection combining the different indexes.

KEYWORDS: anomaly detection, genetic algorithm, hyperspectral imagery, projection pursuit, RX detector

INTRODUCTION

Remote sensing uses a variety of sensors onboard satellites. These sensors employ several bands and have been used in various fields including cartography, geology, hydrology, and urban planning. Some good references for information on remote sensing are Lillesand et al. (2007), Jensen (2006), and Campbell (2008), among others.

Hyperspectral imaging scans multiple regions of the electromagnetic spectrum to determine features of interest in a remote scene. Since hyperspectral images have a large number of bands, the feature space has a high dimensionality. Efficiently processing the huge amount of data collected by hyperspectral sensors is one of the challenges of working with this type of imagery.

Our work focuses on the detection of anomalies, which are considered targets (objects or materials), whose signatures are spectrally distinct from their background, with no a priori knowledge of the target's spectral signature. Anomalies are important features of special interest to image analysts in their daily routines. Methodologies and algorithms are needed to identify these atypical features, thereby allowing image analysts to decide whether to retain them as interesting information or whether to classify them as worthless information, probably noise, and remove them. We consider the detection of anomalies in hyperspectral images as equivalent to detecting outliers in the feature space. For the purposes of this paper, the terms *anomalies* and *outliers* are used synonymously, although usually the word *anomaly* belongs to the remote sensing domain while the word *outlier* belongs to the statistical domain.

MATERIAL

The experiment was performed on two types of imagery: a synthetic image and an image of a real scene. The synthetic image was made using two classes generated with a random Gaussian distribution using six bands and 250×250 pixels. The classes were relatively close to make it possible to see the method's potential for anomaly detection. One of the classes was small and hidden in the spectral bands (Figure 1).

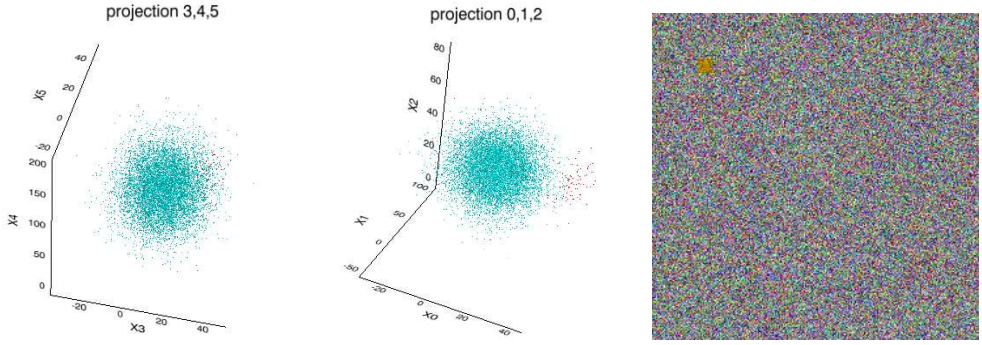

Figure 1. The synthetic image is created with six bands and two clusters generated by a Gaussian distribution. The two images on the left show two projections from a six-dimensional space into three-dimensional space: one for 3, 4, and 5 bands; the other for 0, 1, and 2 bands. The right graph shows the synthetic image where the 5, 1, and 3 bands represent the RGB.

The actual scene is one of airborne hyperspectral imagery acquired with the sensor Airborne Hyperspectral Scanner (AHS). The AHS, an airborne scanner of the whiskbroom type with 80 bands in the electromagnetic spectrum, was acquired by the Spanish National Institute for Aerospace Technology, INTA (Instituto Nacional de Técnica Aeroespacial). The AHS image used here was obtained in May 2006. This dataset was taken at a height of approximately 1300 m, with 2700 m cross-track and 14 km along-track, and a resolution of approximately 3.5 m. The experiment was performed on the original radiance data, and no correction was performed. The image in Figure 2 was made using bands 2, 18, and 44 for the red, green, and blue colors, respectively. The size of the image is 262 × 249 pixels.

The imagery used in our experiment is of a rural site, El Escorial, a location near Madrid. In our analysis we consider only two classes, background and certain man-made targets. Our analysis aims to distinguish the latter from many sources of background variation, such as forest, bushes, roads, bare soil, rocks, and other natural objects. Before the flight took place, several plots of land at El Escorial were prepared with different types of healthy and diseased vegetation. The algorithms developed in this study use raw data. When the flight was planned, the party of scientists and technicians were not aware of any buildings in the surrounding area, as these did not appear in the cartographic map used. These constructions were discovered only after the flight took place and after we studied the images. Other objects were present in the survey area at the time of the flight, such as the two cars that took the researchers to the field, some mats of synthetic fabric, and two panels and a crane used to measure canopy reflectance taken as targets by the detection algorithm, since these objects are man-made objects in a natural background. The scene is ideal for the study of anomaly detection; we have a hyperspectral image with known man-made objects, but it is not possible to identify these by merely looking at the image.

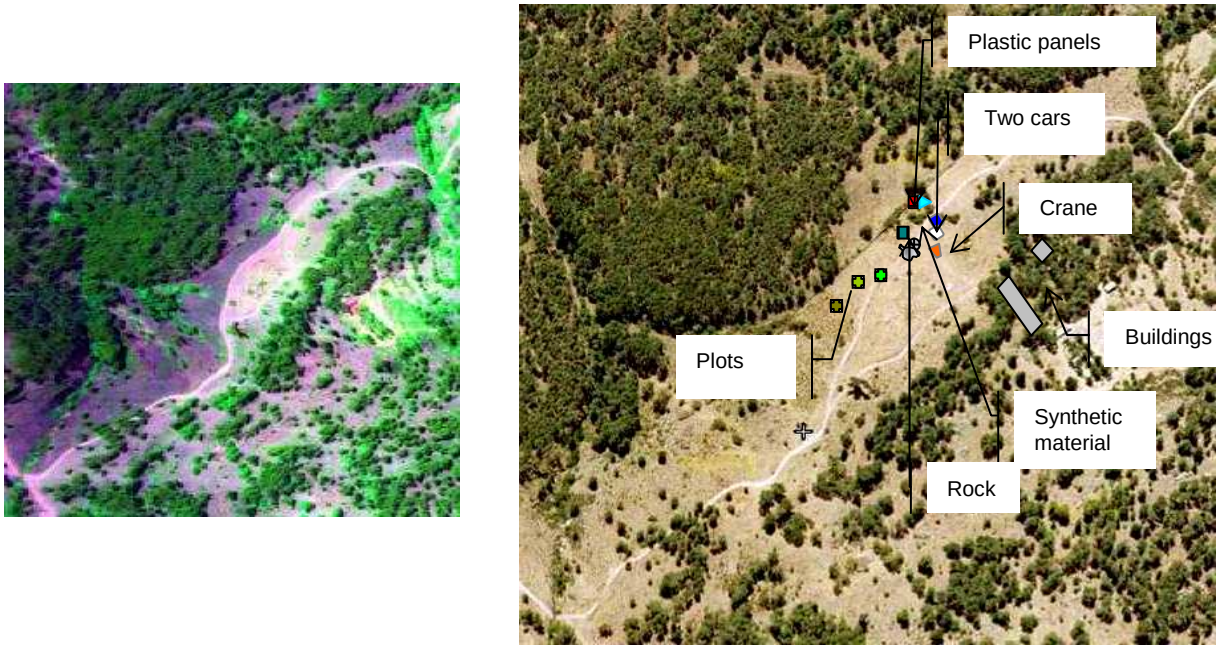

Figure 2. The AHS imagery of the El Escorial site. The RGB image on the left was obtained from AHS imagery with bands 2, 18, and 44, respectively. The image at the right corresponds to an aerial image of the same area taken at another time, with the targets drawn in. These targets are difficult to make out in the AHS image at the top, especially the cars, crane, and metallic panel.

In the picture on the right of Figure 2, the objects have been drawn on an aerial image from another flight. In the AHS image, at the left of Figure 2, these objects cannot be recognized because they are too small for the resolution of the image.

METHODS

RX

Several methods have been used for anomaly detection, although the first one to be developed and of particular interest to us is that of Reed and Yu (1990), referred to as the RX detector, after the initials of its proponents, Reed and Xiaoli Yu. It is a widely used anomaly detector for multispectral and hyperspectral images (Yu et al., 1997). The RX algorithm is a likelihood ratio detector based on Mahalanobis distance. It is the normalized difference between the value of the test pixel and the mean of the background

$$\partial_{RX}\left(\vec{x}\right) = \left(\vec{x} - \vec{\mu}\right)^{t} \sum\nolimits_{MxM}^{-1} \left(\vec{x} - \vec{\mu}\right)$$

where $\vec{x}$ is the pixel spectral vector, μ is the mean spectral vector for the area of interest, M the number of spectral bands, and Σ is the covariance matrix. The Mahalanobis distance is used to calculate how far each point is from the center of the cloud formed by the rest of the pixels, and the shape of the cloud is taken into account through Σ. The mean is obtained by considering only a window around the pixel being examined, not the whole image, by sliding this window over every spectral pixel and calculating the mean of the spectral pixels falling within the window. The algorithm returns a matrix of confidence for each pixel. If confidence is greater than a defined threshold, then the pixel $\vec{x}$ is taken as an anomaly.

Projection pursuit. *Projection pursuit* refers to a method that tries to maximize some measure of "interestingness"' called an index, and for this it uses one or more linear combinations of the original features. The geometric intuition tells us that the n-dimensional space of features can be rotated and that the first few dimensions of the rotated space are then retained.

In our work the measure of interestingness will be given for a structure where it is easy to make out outliers. We

have used projection pursuit to project from the multidimensional space of feature space to one dimension, one straight line, such as the skewness (positive or negative) or kurtosis is maximized. The values of this projection can be represented as a gray level image.

We will first review skewness and kurtosis as indicated by the following equation:

$$\text{Skewness} = \frac{1}{N}\sum_{j=0}^{N-1}\left(\frac{x_i - \text{mean}}{\sqrt{\text{variance}}}\right)^3 \quad (1)$$

Skewness measures the asymmetry of a distribution; if the skewness is different from zero, then that distribution is asymmetrical, whereas normal distributions are perfectly symmetrical, and its skewness is zero. A negative skew indicates that the tail on the left side of the probability density function is longer than the tail on the right side, and the bulk of the values (including the median) lies to the right of the mean. When the tail on the right side is longer than that of the left side and the bulk of the values lies to the left of the mean, a positive skew occurs. When the skew is zero, the values are relatively evenly distributed on both sides of the mean, but this is not necessarily implying a symmetric distribution.

The following graphs reveal three distributions with different skewness (Figure 3).

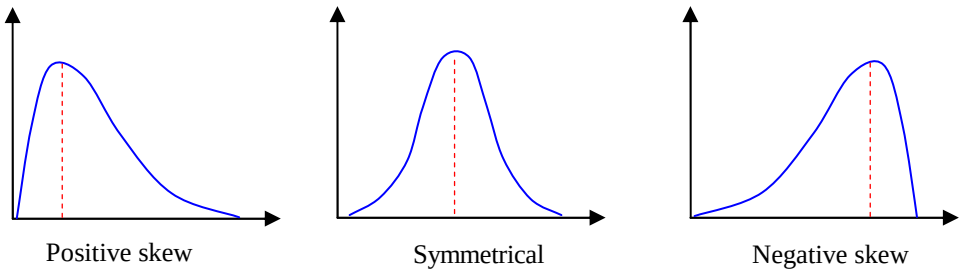

Figure 3. Skewness for different distributions.

The kurtosis is given by the equation:

$$\text{Kurtosis} = \frac{1}{N}\sum_{j=0}^{N-1}\left(\frac{x_i - \text{mean}}{\sqrt{\text{variance}}}\right)^4 - 3 \quad (2)$$

Kurtosis is a measure of a distribution's peak. If the kurtosis clearly differs from zero, then the distribution is either flatter or more peaked than normal; the kurtosis of the normal distribution is zero, and it is called *mesokurtic*. Kurtosis is a measure of how extreme observations are in a data set. The greater the kurtosis coefficient, the more peaked the distribution around the mean.

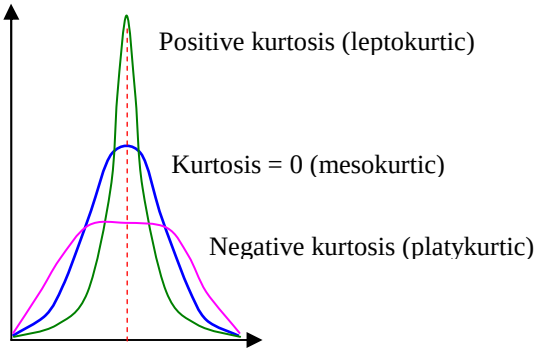

Figure 4. Kurtosis for different distributions.

Genetic algorithm. Now the problem is to find the maximum or minimum of the index, since there is no known analytical procedure to do the optimization; therefore, we use a Genetic Algorithm (GA) to optimize them. Stochastic global search methods, GAs generate solutions to optimization problems using techniques inspired by evolution, such as

survival of the fittest, using operators borrowed from natural genetics, such as mutation and crossover. GAs differ from traditional search and optimization methods in that they do not require derivative information, use probabilistic transition rules, and work on encoding the parameter set rather than in the parameter set itself. More detail about genetic algorithms can be found in Golberg (1989) and many other sources.

We have optimized the skewness (Equation 1) with the positive sign (that we will call SP) and with the negative sign (SN). We have also minimized Kurtosis (KN) for Equation 2, or what equals minimizing:

$$\sum_{j=0}^{N-1}\left(\frac{x_i - \text{mean}}{\sqrt{\text{variance}}}\right)^4$$

This signifies obtaining the most platykurtic distribution possible or getting the most rounded peak curve possible with skinny tails (Figure 4).

Because the GA is a stochastic search method, it is difficult to formally specify convergence and termination criteria. For a number of iterations the optimal solution may remain static before a superior optimal solution is found. We have found that within 100 iterations the algorithm finds acceptable solutions.

RESULTS WITH SINTECTIC IMAGE

The experiment was performed on the synthetic image of Figure 1 by applying RX, SP, SN, and KN methods.

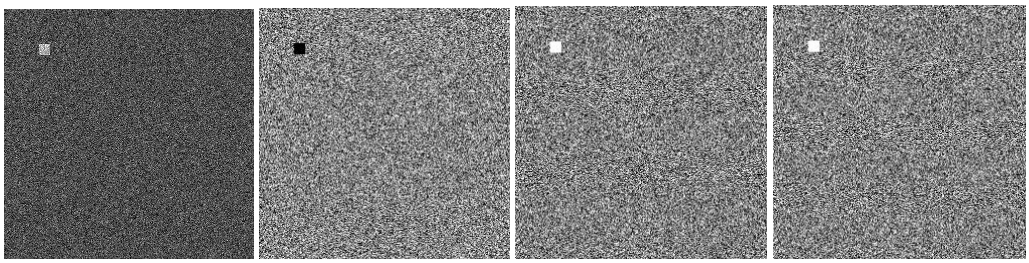

Figure 5. Results of applying RX (a), SP (b), SN (c), and KN (d) to the synthetic image of Figure 1.

Figure 5(a) shows the results of applying RX to the synthetic image of Figure 1. The degree of detection is represented by the brightness value; although in the case of Figure 5(b), the darker pixels represent the detection by minimizing the positive skewness. Minimizing SP and SN gave opposite values since one is the opposite of the other (as explained above). The kurtosis also gave good results as can be seen in Figure 5.

Using the naked eye, one cannot determine which algorithm has performed the best detection; the only clear result is that RX has performed less satisfactorily than the other three. To do this we have utilized receiver operating characteristics (ROC) analysis. Such ROC graphs are useful for organizing classifiers and visualizing their performance. The ROC provides a probability of detection versus a probability of false alarm curve (Fawcett, 2006).

As can be observed in Figure 6 the three—SN, SP, and KN—perform better than RX for a small range (between 0.0 and 0.2) of false alarms, and SP, SN, and KN work similarly. Some authors take the area under the ROC curve (AUROC) as a measurement of the model's accuracy (Hosmer and Lemeshow, 2000). The area is a measure of discrimination, that is, the ability of the technique to classify those pixels correctly with and without some probability of belonging to the target. This threshold-independent measure of discrimination between both classes takes values between 0.5 (no discrimination) and 1 (perfect discrimination). Therefore, the closer the ROC plot is to the upper left corner, the higher the overall accuracy of the test. An area of 1 represents a perfect test and an area of 0.5 represents a worthless test. A rough guide for knowing the accuracy of a classifier is 0.5–0.6 for a fail, 0.6–0.7 for poor, 0.7–0.8 for fair, 0.8–0.9 for good, and 0.9–1 for excellent. The area corresponding to the RX algorithm is 0.98695; while for both the SN and the SP it is 0.99990, and for the KN it is almost the same at 0.99985. This indicates a very high predictive capacity for the three projection pursuit models for this synthetic image. Table 1 allows us to see the difference between the methods from the point at which the ROC curves are constructed (Figure 6). The ROC curve for the SN and SP are quite similar and better than the ROC curve for the KN, while the RX has the poorest ROC curve.

Table 1. Quantitative details of probability of detection (first column) versus probability of false alarm (second column) for each detector

RX			SP			SN			KN	
0,000000	0,000000		0,000000	0,056766		0,000000	0,053752		0,000000	0,078777
0,000000	0,000000		0,000000	0,130925		0,000000	0,124882		0,000000	0,167556
0,000000	0,000000		0,000000	0,248593		0,000000	0,239135		0,000000	0,302730
0,000000	0,000000		0,000000	0,406611		0,000000	0,395774		0,000000	0,471697
0,000000	0,000000		0,000000	0,584540		0,000000	0,573446		0,000000	0,646452
0,000000	0,000000		0,000000	0,744048		0,000000	0,735039		0,000000	0,793777
0,000000	0,000000		0,000000	0,864939		0,000000	0,858622		0,000000	0,898267
0,000000	0,000000		0,000000	0,939034		0,000000	0,935667		0,000000	0,956780
0,000000	0,000000		0,000000	0,977092		0,000000	0,975200		0,000000	0,984754
0,000000	0,000000		0,000000	0,992946		0,000000	0,992321		0,008264	0,995912
0,000000	0,149137		0,008264	0,998269		0,008264	0,998044		0,016529	0,998990
0,000000	0,756344		0,049587	0,999663		0,041322	0,999695		0,066116	0,999840
0,049587	0,960227		0,115702	0,999936		0,107438	0,999936		0,198347	0,999984
0,280992	0,995479		0,314050	0,999984		0,297521	0,999984		0,388430	0,999984
0,611570	0,999423		0,537190	0,999984		0,495868	0,999984		0,619835	0,999984
0,776860	0,999952		0,685950	1,000000		0,685950	1,000000		0,760331	1,000000
0,900826	1,000000		0,842975	1,000000		0,842975	1,000000		0,884298	1,000000
0,942149	1,000000		0,933884	1,000000		0,933884	1,000000		0,942149	1,000000
0,983471	1,000000		0,975207	1,000000		0,966942	1,000000		0,975207	1,000000
0,991736	1,000000		0,991736	1,000000		0,991736	1,000000		0,991736	1,000000

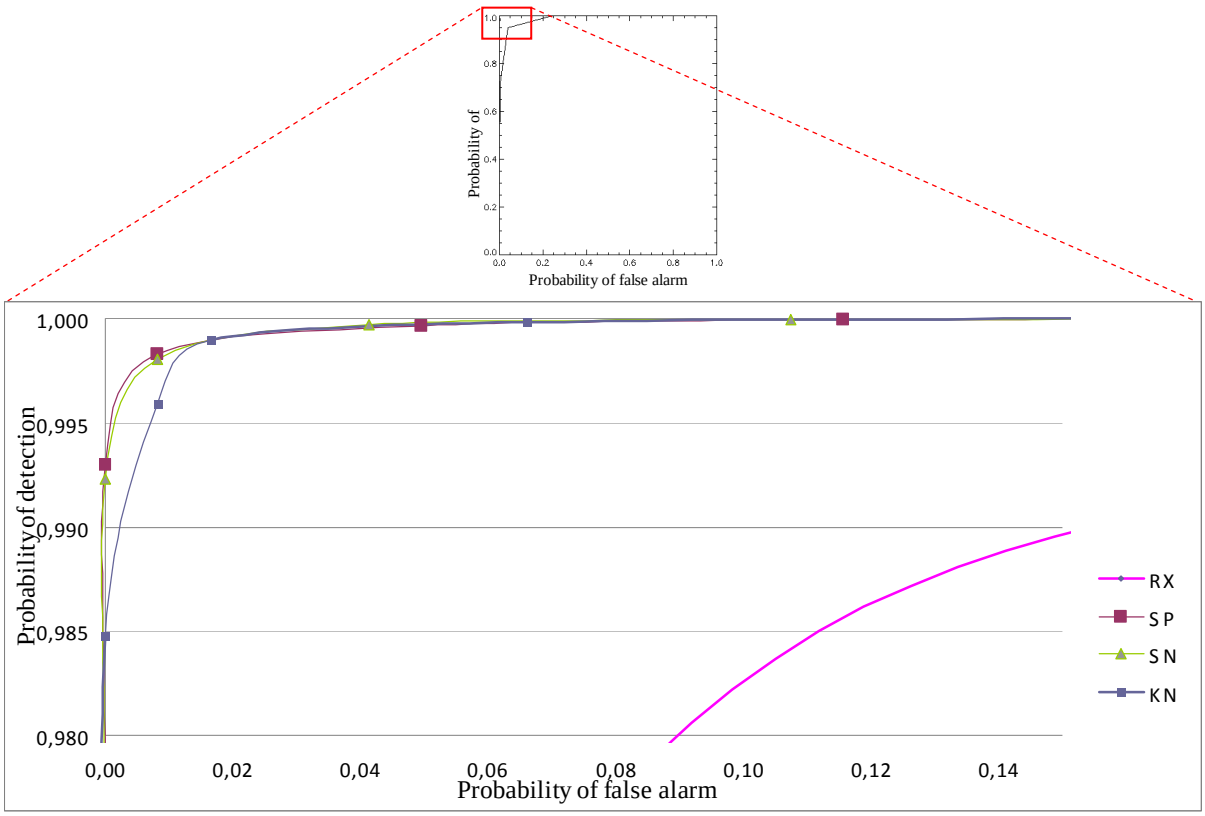

Figure 6. The ROC curves for the RX algorithms with the synthetic image; the graph represents the probability of detection versus the probability of false alarms. The curve for the SP, SN, and RP cannot be seen because it is confounded with the frame of the graph (above). Below is the zoom graph, where it can be distinguished.

RESULTS WITH THE REAL IMAGE

Some authors have shown (Eisman et al. 1995; Malpica et al. 2008) that hyperspectral imagery could aid in discriminating between natural and man-made objects; the spectral characteristics of one class differ in significant ways from the spectral characteristics of the other.

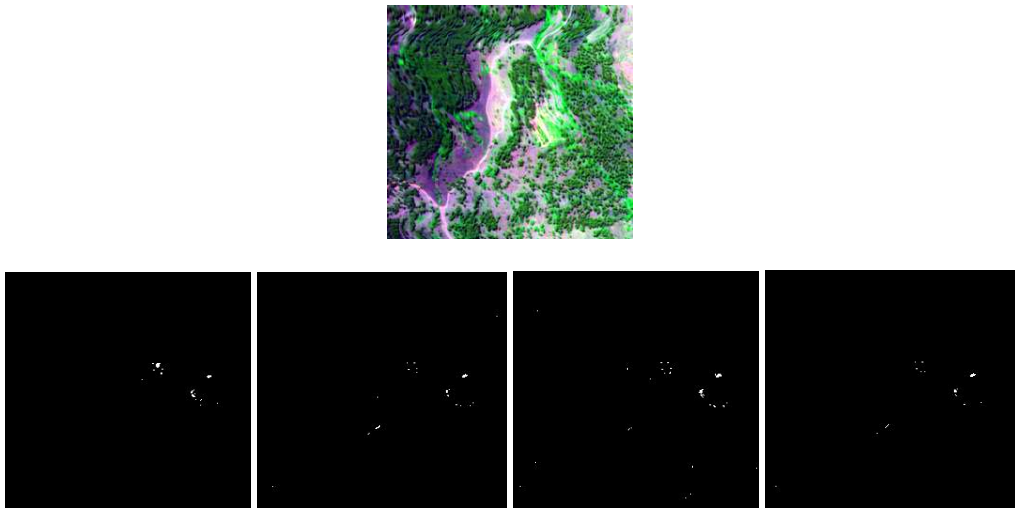

Figure 7**.** Original image and detection results with RX, SP, SN, KN, respectively.

The degree of detection is represented by the gray level of the result layer of the corresponding algorithm. The pixels with the highest value occur in the patch formed by the two cars. If the first few pixels with higher, brighter values are taken, then the cars, the metallic panel and some of the buildings are detected. Figure 7 shows the results of algorithms RX, SP, SN, and KN, applying a threshold adjustments or brightness tuning, this threshold shows the outliers that better represent each algorithm (Figure 7). As can be observed, all the algorithms have detected man-made objects.

CONCLUSIONS

This paper has presented new indexes for projection pursuit for anomaly detection in hyperspectral imagery. The main objective was to seek a projection that diversified from the main body of the distribution using the tails of the distribution for detecting outliers.

The RX detector is prone to high false alarms because the local Gaussian assumption is largely inaccurate in real data. However, the indexes proposed and used in this study—based in statistical moments—do not suffer from this problem; therefore they find outliers not found by RX. In that sense they are superior to RX.

The behavior of the different indexes is coherent with the theoretical interpretation for each of them, i.e., skewness with its asymmetry provides the outlier to the left or the right of the distribution, and the kurtosis, when it is rolled down by the optimization effect of the genetic algorithm exposes the outliers.

We can say that no unique index in projection pursuit serves best for all cases; each one has its advantages and disadvantages depending on the data and the type of outliers to be detected. Therefore, the authors recommend using a combination of all possible indexes and procedures, including RX, in the search for outliers.

ACKNOWLEDGMENTS

The authors would like to thank the Instituto Nacional de Técnica Aeroespacial (INTA) for providing the AHS data for this research. The authors wish also to thank the Spanish MICINN for financial support, Project no. CGL2010-15357.

REFERENCES

Alonso, M. C., and J. A. Malpica, 2009. The combination of three statistical methods for visual inspection of anomalies in hyperspectral imageries. *Seventh International Conference on Advances in Pattern Recognition*, ICAPR, pp.377-380, 2009.

Campbell, J.B., 2008. *Introduction to Remote Sensing*. The Guilford Press (fourth edition) pp 626.

Eismann, M. T., J. H. Seldin, C. R. Schwartz, J. R. Maxwell, K. K. Ellis, J. N. Cederquist, A. D. Stocker, A. Oshagan, R. O. Johnson,W. A. Shaffer, M. R. Surette, M. J. McHugh, A. P. Schaum, and L. B. Stotts, 1995. Target detection in desert backgrounds: Infrared hyperspectral measurements and analysis, *SPIE: Signal and Data Processing of Small Targets*, Vol.2561, pp. 80–97.

Fawcett, T, 2006. An introduction to ROC analysis, *Pattern Recognition Letters*, 27: 861–874.

Goldberg, D. E., 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley Publishing Company.

Hosmer, D. W. and S. Lemeshow, 2000. *Applied Logistic Regression*, 2nd edition, Wiley, New York.

Lillesand, T. M., R. W. Kiefer, and J. Chipman, 2007. *Remote Sensing and Image Interpretation*, Wiley (sixth edition).

Jensen, J. R., 2006. *Remote Sensing of the Environment: An Earth Resource Perspective*. Prentice Hall; (2nd edition).

Malpica, J. A., J. G. Rejas, M. C. Alonso. 2008. A projection pursuit algorithm for anomaly detection in hyperspectral imagery. *Pattern Recognition*, 41(11): 3313-3327.

Reed, I. S., and X. Yu, 1990. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(10):1760–1770.

Yu, X., L. E. Hoff, I. S. Reed, A. M. Chen, and L. B. Stotts, 1997. Automatic targe detection and recognition in multiband imagery: A unified ML detection and estimation approach, *IEEE Transactions Image Processing*, 6:143–156.