Anomaly Detection Rudiments for the Application of Hyperspectral Sensors in Aerospace Remote Sensing

Gerald Wong
Heriot-Watt University, Riccarton Campus, Edinburgh, EH14 4AS, United Kingdom

gw25@hw.ac.uk or gerald.wong@selexgalileo.com

Abstract. Hyperspectral imaging differs from conventional techniques by exploiting the spectral dimensionality of remote scenes. This additional information promotes discrimination of image elements, especially anomalies that are dissimilar with respect to global features. Algorithms for anomaly detection are designed to overcome the inherent difficulty of analysing hypercubes, which are the higher-dimensional analogues of conventional broadband images. Such algorithms are prolific in their variety and design, which could become an obstacle in choice or application for the neophyte researcher in this field. This paper seeks to consolidate this plethora of algorithms into succinct categories for clarity of rudimentary decision making.

1. Spectral and Sub-Pixel Mixing

The main benefit of hyperspectral imaging (HSI) with respect to conventional imaging is that otherwise identical image elements can be differentiated and even identified. Aerospace HSI is employed by the geophysical community for environmental monitoring, agricultural land surveys, in addition to natural mineral and resource scouting. Spectral reflectance is the parameter of interest, where the reflected to incident flux ratio is a function of the particular wavelength involved. Each type of terrain feature has a unique spectral signature, with a characteristic form and notable absorption bands thus allowing discrimination and identification.

For HSI systems with limited spatial resolution operating at high-altitude, the ground area subtended by a single detector pixel may exceed the size of details relevant to the application involved. In this case sub-pixel mixing can occur whereby the spectrum attached to a single pixel is actually composed of more than one distinct spectral signature. Spectral mixing is particular to the physical scene whereby actual mixing of materials occurs, such as desert sand lying upon vehicles – the net effect is identical. It can be stated using HSI nomenclature that the mixed composite spectrum is composed of individual endmember spectra in certain abundances, which each uniquely corresponding to a specific identifiable material or surface [1]. Figure 1 below illustrates a simplified composite spectrum that is composed of two distinct endmembers, while Figure 2 represents a linear mixing model for two wavebands and three endmembers.

Outwith intrinsic atmospheric effects, sub-pixel and spectral mixing are the most significant difficulties in HSI, and preoccupy the majority of computational efforts into spectral analysis. There are multitudes of novel and innovative methods for spectral unmixing, with direct implications for anomaly detection. These techniques form the subject of the next section on recognition algorithms, which begins with an overview of the basic RX algorithm before defining the major categories in use.
2. Recognition Algorithms

The use of HSI for the airborne detection of local anomalies requires algorithms that are able to identify regions possessing unusual spectral signatures, with respect to the overall global image [2,3]. There is generally no a priori information available, other than anomalies occurring with a low probability and with small relative sizes within a physical scene. The ground sampling distance (GSD) of an HSI detector will also normally exceed the anticipated anomalies size, causing sub-pixel mixing.

This is in addition to the spectral mixing that occurs because of real and physical contamination of an otherwise pure endmember, such as soil upon a land mine. For the remainder of this discussion, sub-pixel and spectral mixing are considered identical unless otherwise explicitly defined or mentioned. It is furthermore assumed that atmospheric effects, such as path radiance and absorption will have been separately accounted for with presented accepted methods [4].

Recognition algorithms deal with the distribution of hypercubes in spectral hyperspace, which exhibit clustered-type behaviour. Such behaviour is characterised by data clouds of hypercubes that can be potentially associated with a given background element. The extent and relative concentration of a particular data cloud is defined as the density spread, representing the net result of many effects.

These effects include aspects such as natural variability and illumination factors, which are beyond the scope of this paper. These density spreads are asserted as multivariate Gaussian distributions that
are mutually independent, which is an assumption believed to underpin the natural variability of spectra. Henceforth any small image elements with hypercubes lying far outwith the data clouds, which otherwise represent the majority of the image could potentially be anomalies.

The next sub-section covers the noteworthy initial developments into anomaly detection as a primer to the more recent advances discussed afterwards. The focus concentrates on the progressively inadequate K-Means clustering method and the ubiquitous RX algorithm, which still presently serves as an all-pervading comparison basis.

2.1. K-Means Clustering

One of the first algorithms for anomaly detection was the K-Means clustering method, which aims to group together hyperpixels into partitions that could be allocated to image elements. The name derives from its origin in expectation-maximization theory, whose objective is to determine the k means of Gaussian distributed data [5]. The mathematical premise of K-Means clustering is to minimize the intra-cluster variance \( V \), which known as the squared error function. This is given within equation (1) below for the intra-variance \( V \) of k clusters \( S_i \) where \( i=1…k \) with \( \mu_i \) centroids

\[
V = \sum_{i=1}^{k} \sum_{x \in S_i} \left| x - \mu_i \right|^2
\]  

(1)

The first step requires dividing up the hyperspace data cloud into a certain number of partitions, to be decided upon by the application or particular situation involved. This initial division can be carried out at random or with any \textit{a priori} knowledge available. The centroid of each partition is calculated and new partitions are substituted, which each hyperpixel is reassigned to. The iterative process of centroid calculation and hyperpixel reassignment continues until a stable situation is achieved. This stability is characterized by fixed centroids and hence no more reassignments of hyperpixels. Within the field of cluster analysis, K-Means for reference lies between K-Medoids and Fuzzy C-Means in efficacy [5].

Although relatively fast at processing hyperspectral imagery, K-Means yields inferior results [6] with respect to nearly all other approaches to be discussed in following sections. The natural variability of spectral signatures does not exhibit the spherical Gaussian distributions assumed by K-Means clustering. Combined with even minor mixing of endmembers, the net effect is significantly incorrect identification of image elements. Figure 3 below illustrates the inherent difficulties of K-Means clustering – note especially the spurious roughness of the identified image elements. It has also been demonstrated that K-Means clustering is analogous to principle components analysis and is hence now relegated to the status of a pre-processing algorithm at best [7,8].

![Figure 3: K-Means classification of a test scene, where the crosses depict a wrongly labelled class [9]](image)
2.2. RX Algorithm

Within HSI anomaly detection, the RX algorithm is the industry benchmark that forms the basis of comparison for more recent developments. Numerous derivatives of the RX algorithm cover countless subtle variations, but all still incorporate the basic premise that follows [14,15]. The RX algorithm identifies anomalies with respect to the spectral background, which is taken to be an annular region around the given test pixel. The Mahalanobis distance is the primary metric employed to determine the spectral similarity between image pixels [16]. In order to further characterize the local background area, a multivariate Gaussian distribution can also be modeled to this annular region. If the test pixel does not show similar behavior with respect to neighbors, then it is designated an anomaly.

The RX algorithm and its variants (such as GRX) suffer from difficulties, with the foremost being the nature of the background region involved. This annular area may encompass more than one type of material surface, thus rendering a single Gaussian distribution as inadequate. The background region could be reduced in size to assist in isolating one particular surface, but this solution is in conflict with a requirement for sufficient pixel numbers with which to generate accurate statistics. In varied physical environments, this condition may not be possible to achieve in any measure at all. This can thus cause the RX algorithm to be reduced down to edge-detection, creating large false alarm rates.

Due to the limited annular region commonly employed, an additional problem is that a local anomaly may not be a global anomaly and will be incorrectly reported. A prime example of such false alarms could be evergreen trees in an otherwise sparse frozen environment. Despite its failings, the RX algorithm is more resource efficient than most recent methods and is capable of yielding reasonable first-order results. It is also used as a common denominator for facilitating comparison between disparate algorithms.

In terms of how imagery is actually analysed presently, two broad categories currently exist for anomaly recognition algorithms – geometric endmember determination and statistical independent component analysis. Both of these modern techniques primarily rely upon spectral unmixing (SU), in order to classify image elements and thus highlight anomalies. Countless algorithms can be assigned to each category with innumerable variations, which sometimes merely duplicate previous techniques but with simply a different name. It is far beyond the scope of this report to elucidate every such algorithm, but the following sections will briefly explain the two main categories. Reference [10] is an excellent publication for further information on spectral unmixing.

3. Geometric Endmember Determination

The first category is geometric endmember determination (GED) and involves the actual layout of hypercubes in spectral hyperspace [6,9]. Recalling the illustration in Figure 2, the principle assumption behind GED is that endmembers must necessarily reside at the cusps of the mixing space enclosing the hypercubes. Within N-dimensional hyperspace, this enclosing region is known as the simplex and possesses a dimensionality of (N+1). Referring back to the example of Figure 2, the simplex for the two-waveband system has three vertices and is hence triangular in 2D appearance.

For HSI systems with several hundred wavebands, the GED technique may become computationally demanding and time consuming. Principle component analysis (PCA) is commonly used to reduce the hyperspace dimensionality to a reasonable level without significant loss of spectral contrast. It can be observed that many bands in a hyperspectral image do not contain significant energy or brightness with respect to others. In order to reduce the computational burden, these low-value bands can hence be either disregarded entirely or folded into adjacent bands. In mathematical terms, the premise of PCA involves determining the optimum number of eigenvalues that contain the bulk of the cumulative normalized energy.

After dimensional reduction has been carried out, a simplex is superimposed onto the hyperspace and iteratively reduced in size until the minimum is achieved. The resultant simplex at this point should only have hypercubes residing on or within its surfaces. Endmembers can then be extracted.
from the vertices and SU can be subsequently performed using linear regression. After image elements have then been classified, those with the lowest abundances should be indicative of local anomalies.

Although the strength of GED is its capability of locating low-probability anomalies, there are a number of problems associated with the approach. They are namely that reflectances of determined endmembers may have non-physical values of either greater than unity or less than zero. The reason stems from pixel clusters with significant variability and either very high or low brightness values. The resultant simplex may require vertices to be extended beyond physical meaning, in order that all hypercubes are properly enclosed. An additional drawback is that numerous high purity endmember pixels must be present in the image for reasonable overall results. This may not always be possible to achieve within noisy environments, thus producing spurious outcomes.

Figure 4 illustrates the classification capability of GED, through a derivative variant known as simplex volume maximisation (SVM). Note in particular the ability to handle complicated histograms that exhibit clear non-Gaussian behaviour, without causing spurious classifications or false alarms.

![Figure 4: SVM element classification and abundance histogram of alfalfa (top) and wheat crop (bottom) in an airborne image of an agricultural test scene (Barrax, Spain) [9]](image)

4. Independent Component Analysis

The second SU category of independent component analysis (ICA) broadly covers those methods that involve mathematical abstraction by considering the statistical spread of data clouds [9]. ICA assumes that the fundamental endmembers are linearly mixed and statistically independent, rather just merely uncorrelated [11,12]. The basic premise behind ICA is that the endmembers can be determined by maximising their statistical independence and non-Gaussian nature. As with the non-statistical GED approach, PCA is commonly employed as pre-processing to reduce the overhead before ICA proper.

Each hypercube in the image is assumed as being a linearly mixed combination of the underlying endmembers, which is a critical HSI assumption. These are mapped to hypercubes through an initially unknown mixing matrix that represents the relative abundances of each endmember within each particular hypercube. The endmembers can therefore be recovered from the source data through appropriate matrix techniques that seek to determine the mixing matrix. Since the presence of unmixed
pixels is not explicitly required, ICA henceforth presents an advantage over GED in moderately mixed images. Figure 5 below illustrates the ability of ICA in classifying images with multiple elements and associated endmembers — contrast with the better class definition within Figure 4.

A significant difficulty of ICA and other derived approaches, such as independent factor analysis (IFA), is that they are rendered ineffective when the source data exhibits non-Gaussian complicated multi-structures. The resultant effect is that misclassification of image elements can occur, which can mask the presence of anomalies — such as soil being mixed with wheat as in Figure 5.

The assumption of statistically independent endmembers is also not strictly accurate, since all abundances must sum to unity and hence a degree of interdependency exists. Four well-known ICA based algorithms (including IFA) have been tested in recent times and it was discovered that endmembers are always incorrectly unmixed, despite using artificial images to prevent it [13]. Analyzing images that contain numerous endmembers and significant signature variability aids in improving unmixing results. The increased spectral diversity has the effect of suppressing any statistical dependency of endmembers and assists in satisfying the key assumptions of ICA [12].

Along with the ubiquitous RX algorithm, the two general categories of GED and ICA underpin the recognized algorithms in current usage for anomaly detection. There are many emerging developments that do not clearly fit into either such group, which employ both inventive and novel techniques. These advances such as adaptive neural networks or sonic transformation show potential and opportunity for HSI, but reside outwith the present discussion scope due to their disparate nature.

The following reference is an excellent starting point for the study of neural networks and their application to HSI anomaly detection [17]. The novel process of sonic transformation in converting spectra into audio waveforms to synergize with human resources is covered within reference [18]. There have also been recent advances into the underlying theory that underpins most algorithm assumptions, suggesting that elliptically contoured (EC) distributions like Gaussians are not suitable. The evidence for “fat-tailed” distributions is discussed within reference [19] and may require a critical reassessment of the assumptions supporting most algorithms used at present for anomaly detection.

Figure 5: ICA element classification and abundance histogram of alfalfa (top) and wheat crop (bottom) in an airborne image of an agricultural test scene (Barrax, Spain) [9]
5. Conclusion

An interesting adjunct to this discovery of non-EC characteristics is that angular density variations in hyperspace may hold the key for improved sub-pixel anomaly detection. Non-EC studies have indicated that certain directions have clusters of large-radii hypercubes, which can appear incorrectly anomalous to present methods [19]. Early research into anomaly detection using angular rarity or spectral direction exhibits potential for synergy with conventional techniques. Future results will demonstrate if there is prospect for incorporating this new approach into reappraised algorithms using non-EC assumptions. Assigning due diligence to this non-EC evidence, it can be concluded with care that GED may hence be more efficacious than ICA, due to not assuming Gaussian-type distributions.

References

[1] Randall B Smith, “Introduction to Hyperspectral Imaging” MicroImages Inc (July 2006)
[2] C-I Chang and S-S Chiang, “Anomaly Detection and Classification for Hyperspectral Imagery” IEEE Transactions on Geoscience and Remote Sensing, Volume 40, Number 6 (June 2002)
[3] Nirmal Keshava, “Angle-Based Band Selection for Material Identification in Hyperspectral Processing” Proceedings of SPIE, Volume 5093 (April 2003)
[4] Frederick Smith (Editor), “Atmospheric Propagation of Radiation” IR/EO Systems Handbook: Volume 3, SPIE Optical Engineering Press & Infrared Information Analysis Centre (1996)
[5] Douglas Steinley, “K-Means Clustering: Half-Century Synthesis” British Journal of Mathematical and Statistical Psychology, Volume 59, Part One, pages 1-34 (May 2006)
[6] Nirmal Keshava, “A Survey of Spectral Unmixing Algorithms” Lincoln Laboratory Journal, Volume 14, Number 1, page 55 (2003)
[7] Hongyuan Zha et al, “Spectral Relaxation for K-Means Clustering” Neural Information Processing Systems, Volume 14, pages 1057-1064 (December 2001)
[8] Chris Ding and Xiaofeng He, “K-Means Clustering via Principal Component Analysis” Proceedings of the 21st Int'l Conference on Machine Learning , pages 225-232 (July 2004)
[9] Peter WT Yuen et al, “Geometric and Statistical Spectral Unmixing for Subpixel Target Detection” Technical conference paper, EMRS DTC Proceedings, B17 (2005)
[10] Chein-I Chang, “Hyperspectral Imaging: Techniques for Spectral Detection and Classification” Kluwer Academic and Springer, ISBN 0306474832 (July 2003)
[11] Aapo Hyvärinen and Erkki Oja, “Independent Component Analysis: Algorithms and Applications” Neural Networks, Volume 13, pages 411-430 (2000)
[12] Erkki Oja et al, “Independent Component Analysis” John Wiley & Sons Inc, Chapter One, ISBN 047140540X (June 2001)
[13] José MP Nascimento and José M Bioucas Dias, “Does Independent Component Analysis Play a Role in Unmixing Hyperspectral Data?” IEEE Transactions on Geoscience and Remote Sensing, Volume 43, Number 1 (January 2005)
[14] CJ Willis “Anomaly Detection in Hyperspectral Imagery using Statistical Mixture Models” Technical conference paper, EMRS DTC Proceedings, B16 (2005)
[15] Heesung Kwon and Nasser M Nasrabadi, “Kernel RX-Algorithm: A Nonlinear Anomaly Detector for Hyperspectral Imagery” IEEE Transactions on Geoscience and Remote Sensing, Volume 43, Number 2 (February 2005)
[16] Prasanta Chandra Mahalanobis, “On the Generalized Distance in Statistics” Proceedings of the National Institute of Science of India, Vol 12, pages 49-55 (1936)
[17] Owen Thomas, “Target Confirmation in LWIR Hyperspectral Data using Neural Networks” Technical conference paper, EMRS DTC Proceedings, B19 (2005)
[18] Mark Bernhardt et al, “Sonification of Hyperspectral Image Data” Technical conference paper, EMRS DTC Proceedings, B3 (2007)
[19] Mark Bernhardt et al, “Hyperspectral Anomaly Detection and Un-Mixing in Fat-Tailed Clutter” Technical conference paper, EMRS DTC Proceedings, B21 (2006)