Review

Systematic Review of Anomaly Detection in Hyperspectral Remote Sensing Applications

Ivan Racetin [1] and Andrija Krtalić [2,*]

[1] Faculty of Civil Engineering, Architecture and Geodesy, University of Split, 21000 Split, Croatia; racetin@gradst.hr

[2] Faculty of Geodesy, University of Zagreb, 10000 Zagreb, Croatia

[*] Correspondence: andrija.krtalic@geof.unizg.hr

Abstract: Hyperspectral sensors are passive instruments that record reflected electromagnetic radiation in tens or hundreds of narrow and consecutive spectral bands. In the last two decades, the availability of hyperspectral data has sharply increased, propelling the development of a plethora of hyperspectral classification and target detection algorithms. Anomaly detection methods in hyperspectral images refer to a class of target detection methods that do not require any a-priori knowledge about a hyperspectral scene or target spectrum. They are unsupervised learning techniques that automatically discover rare features on hyperspectral images. This review paper is organized into two parts: part A provides a bibliographic analysis of hyperspectral image processing for anomaly detection in remote sensing applications. Development of the subject field is discussed, and key authors and journals are highlighted. In part B an overview of the topic is presented, starting from the mathematical framework for anomaly detection. The anomaly detection methods were generally categorized as techniques that implement structured or unstructured background models and then organized into appropriate sub-categories. Specific anomaly detection methods are presented with corresponding detection statistics, and their properties are discussed. This paper represents the first review regarding hyperspectral image processing for anomaly detection in remote sensing applications.

Keywords: target detection; Reed-Xiaoli algorithm; background models; kernel-based methods; representation models

1. Introduction

Hyperspectral imaging (HSI) is an established and recognized technique in numerous applications, such as agriculture and forestry [1], the food industry [2], humanitarian demining [3,4], medicine [5], search and rescue missions [6,7], and water resources [8]. Hyperspectral sensors are passive instruments that record reflected electromagnetic radiation in tens or hundreds of narrow and consecutive spectral bands. In remote sensing applications, the primary radiation source in visible, near-infrared, and shortwave infrared electromagnetic spectrum regions is the Sun. In contrast, in other applications (i.e., food science or medicine), artificial radiation sources can be used. HSI relies on the fact that every material possesses a unique spectral signature. A spectral signature or spectral reflectance refers to unique reflectance (ratio of reflected and incident electromagnetic radiation) variation as a function of wavelength.

The ability to detect or identify materials based on their spectral signatures resulted in the development of two leading processing chains of hyperspectral images: classification and target detection (TD). Classification arranges pixels in spectrally disjoint thematic classes. TD aims to find predefined objects or materials in the image or solve the binary problem of whether the pixel under test (PUT) belongs to some general pattern of the

image or deviates from it. In other words, TD needs to determine whether the PUT belongs to the background or is a target. Even though TD can be treated as a binary classification problem, at least on the theoretical scale, there are compelling reasons why these two processing chains should be separated. For something to be considered a target, it should cover only a negligible image area. An image would then consist of a background class that would contain all or almost all image pixels and a scarcely populated or empty target class. An optimization criterion that minimizes classification error would lead to classifying all pixels as a background and result in missed detections, simultaneously achieving very high classification accuracy. More detailed discussion on this topic can be found in [9].

When TD algorithms are used to discover predetermined materials or objects, they are called spectral matching detection algorithms [10] or spectral signature-based target detectors [11]. These algorithms can be regarded as examples of supervised learning or supervised TD, as they require a-priori knowledge of target spectral signature. Spectral matching detection algorithms assess the similarity between the PUT and the known spectral signature. Required spectral signatures can be obtained from a spectral library, by a field measurement with a spectroradiometer, or by identification of the target pixel in the hyperspectral image. If spectral signatures are acquired from a spectral library or by a spectroradiometer, they are usually expressed in reflectance units. Although TD in hyperspectral images relies on spectral signatures that consist of dimensionless reflectance values, hyperspectral sensors originally measure spectral radiance instead of reflectance. Spectral radiance is a radiant flux in a given direction per unit projected area per unit solid angle as a function of wavelength [12]. Illumination conditions, sensor characteristics, and atmospheric transmission are the leading causes of the differences between reflectance and radiance spectrum. Conversion of at-sensor radiance to ground pixel reflectance spectrum requires radiometric calibration that consists of sensor calibration and atmospheric, solar, and topographic correction [13]. This step is essential, as it dramatically influences detection results. Radiometric calibration is unnecessary if a target pixel can be identified in the hyperspectral image, but this case is usually unfeasible in real-life applications.

Anomaly detection methods in hyperspectral images refer to a TD class that does not require any a-priori knowledge about a hyperspectral scene or target spectrum and, therefore, can be regarded as unsupervised learning techniques. Based on statistical techniques, models, or assumptions, unsupervised learning methods aim to find regularities in the input data [14]. A structure in the input space is presumed such that some patterns happen more than others, while some occur rarely. The goal of unsupervised learning is to discover and analyze these patterns. Hence, anomaly detection methods can be performed in reflectance, radiance, or any other units. Basically, anomaly detection methods try to model the image background, and pixels with a spectrum that does not fit the defined background model are then proclaimed as anomalies [9–11,15]. Consequently, anomaly detectors are generally unable to distinguish detected targets between each other, nor can they judge whether the detected anomaly is just a rare pixel or a target of interest. However specific techniques, such as described in [16], have been developed to discriminate anomalies between each other.

Desired characteristics of anomaly detectors can be identified: high probability of detection and low false alarm rate, repeatability of detection performance in different test scenarios and scenes with different levels of complexity, automatic determination of detection (test) statistic threshold, low computational complexity, ability to perform in a real-time, constant false alarm rate (CFAR) operation under the selected statistical model, and robustness of detection performance in regard to a parameter selection (detection performance is not significantly influenced by small changes in its parameters).

A variety of approaches in background modeling have been developed on the basis of a general background model, a spatial model (global, local, or quasi-local), the ability to detect resolved (full pixel) or unresolved (sub-pixel) anomalies, detection statistic selection, and threshold determination.

To the best of our knowledge, to date no review papers have been published on hyperspectral image processing for anomaly detection in remote sensing applications. However, there are papers that are formally classified as research papers but provide a good overview of the subject area, including Manolakis [10], Matteoli, Diani, and Corsini [11], Matteoli, Diani, and Theiler [17], and Nasrabadi [18]. Nevertheless, the most recent of these was written almost 10 years ago, and in the interim, some new approaches in hyperspectral anomaly detection have been developed. This paper is the first review dealing with hyperspectral image processing for anomaly detection in remote sensing applications.

The paper is organized into two parts, A and B. The first part deals with the bibliometric analysis, and the second part, with a methodological overview of the field. The research in bibliometric analysis was conducted to determine the trends and future research challenges in this field. Part B contains the mathematical framework of anomaly detection methods used on hyperspectral images along with a presentation of relevant anomaly detection techniques.

2. Part A: Bibliometric Analysis

The main goal of the bibliometric analysis performed in this research was to answer the following research questions (RQ):

- **RQ 1**. What is the trend among scientific publications on hyperspectral image processing for anomaly detection in remote sensing applications?

- **RQ 2**. What are future research directions in this scientific field?

Bibliographic databases were analyzed in this research, resulting in the selection of the Web of Science (Web of Knowledge, https://apps.webofknowledge.com) and Scopus (https://www.scopus.com/) databases [19,20] as the most relevant for this research topic. Using the same search strategy and criteria, Web of Science found 725 documents, while the Scopus database provided 973 documents. Hence, the results of the Scopus bibliographic database were used for subsequent analysis. According to the Scopus Content Coverage Guide [21] data from January 2020, it covers more than 5000 publishers and contains over 25,000 active titles with more than 77 million publications. The Scopus records date back to 1788, with over 6.6 million records published prior to 1970 [21]. The main focus of the SCOPUS database is classified in four subject clusters [21]: health sciences (30.4%), physical sciences (28.0%), social sciences (26.2%) and life sciences (15.4%).

To select relevant publications on hyperspectral image processing for anomaly detection in remote sensing applications, the PRISMA methodology [22] for systematic reviews was followed. The search strategy was conceived around three keywords: hyperspectral, anomaly, and detection, which were interconnected with the conjunction AND. The chosen terms were searched in the publication titles, abstracts, or authors' keywords to compile an extensive document list. As the research was conducted in January 2021, it encompassed publications published up to the end of 2020. No exclusion criterion (EC) was defined for the research time span, as the goal of RQ 1 was to identify the research trend over the years. To emphasize, the manuscript selection was initiated with two inclusion criteria (IC):

- **IC 1**. The search string (TITLE-ABS-KEY (hyperspectral AND anomaly AND detection))
- **IC 2**. The publications are written in English.

The exclusion criteria (EC) define what publications should be discarded from the research collection. In this work, two EC were defined:

- **EC 1**. Reviews and conference reviews, books and book chapters, letters and notes.
- **EC 2**. Publications with less than three citations per year.

The first EC resulted in the selection of articles and conference papers that could be seen as the primary sources of scientific contributions, whereas the excluded document

types might not have provided sufficiently novel material on the research subject. The second EC served as the relevance criterion, such that only highly-cited publications were selected for final bibliometric analysis. The overall procedure is depicted in Figure 1.

The bibliometric analysis was performed in the programming language R, using the bibliometrix [23] package. Bibliometrix is an open-source R-package that enables comprehensive science mapping analysis [23].
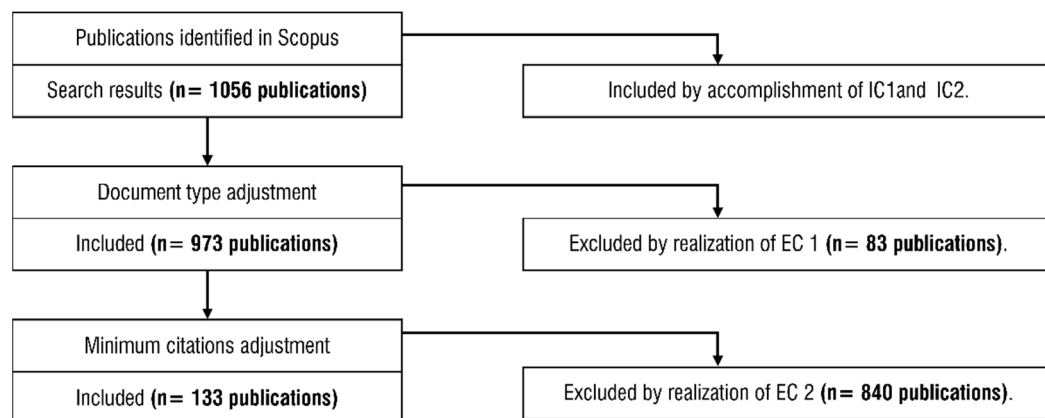

Figure 1. PRISMA flow chart of the procedure for selection of relevant publications.

2.1. Descriptive Bibliometric Analysis

The overview of the descriptive bibliometric statistics for the dataset (Figure 1) is presented in Table 1. The search strategy yielded 133 relevant publications on hyperspectral image processing for anomaly detection in remote sensing applications that were published in 41 sources (Table 1). The number of citations per document, on average, was 72.65, or 8.14 citations per document per year, respectively. In the selected publications, a total of 4276 documents were cited. Most publications that fulfilled the search strategy were journal articles (118), with a smaller fraction of conference papers (15). There were only five documents created by a single author, while the average number of co-authors per document was 3.71. In the end, the collaboration index [24] was calculated as the ratio of the total number of authors of multi-authored documents and the total number of multi-authored documents, and scored the value of 2.3.

Table 1. Summary bibliometric statistics of the relevant publications on hyperspectral image processing for anomaly detection in remote sensing applications acquired by the presented search strategy (Figure 1).

Main Information	Result
Time span	2000–2020
Sources	41
Total number of documents	133
Average years from publication	7.41
Average citations per documents	72.65
Average citations per year per doc	8.14
References	4276
Document types	
Article	118
Conference paper	15
Authors and collaboration	
Authors	299
Authors of single-authored documents	5
Authors of multi-authored documents	294

Co-Authors per Documents	3.71
Collaboration Index	2.3

The core sources of relevant documents, by Bradford's law [25], were *IEEE Transactions On Geoscience And Remote Sensing* and *IEEE Geoscience And Remote Sensing Letters*, which accounted for a total of 54 publications (Table 2).

Table 2. The most relevant sources (first two zones by Bradford's law [25]), sorted in descending order by the number of documents.

Rank	Source Name	Documents	Zone[1]
1	IEEE Transactions On Geoscience And Remote Sensing	39	1
2	IEEE Geoscience And Remote Sensing Letters	15	1
3	IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing	15	2
4	Remote Sensing	9	2
5	Proceedings Of SPIE - The International Society For Optical Engineering	5	2
6	Remote Sensing Of Environment	3	2
7	Eurasip Journal On Advances In Signal Processing	2	2
8	IEEE Access	2	2

[1] By Bradford's law [25].

The top journal (Table 2) was appreciated among the first publications on this topic, and the appearance of the consequent sources followed their popularity (Figure 2a). The first document was published in the *IEEE Geoscience And Remote Sensing Letters* (rank 2) in 2005, in the *IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing* in 2012, and in the *Remote Sensing* in 2014. Researchers in this scientific field were most productive in 2018 (Figure 2b). However, the articles published in 2002 and 2005 attracted the most attention from the scientific community (Figure 2b).
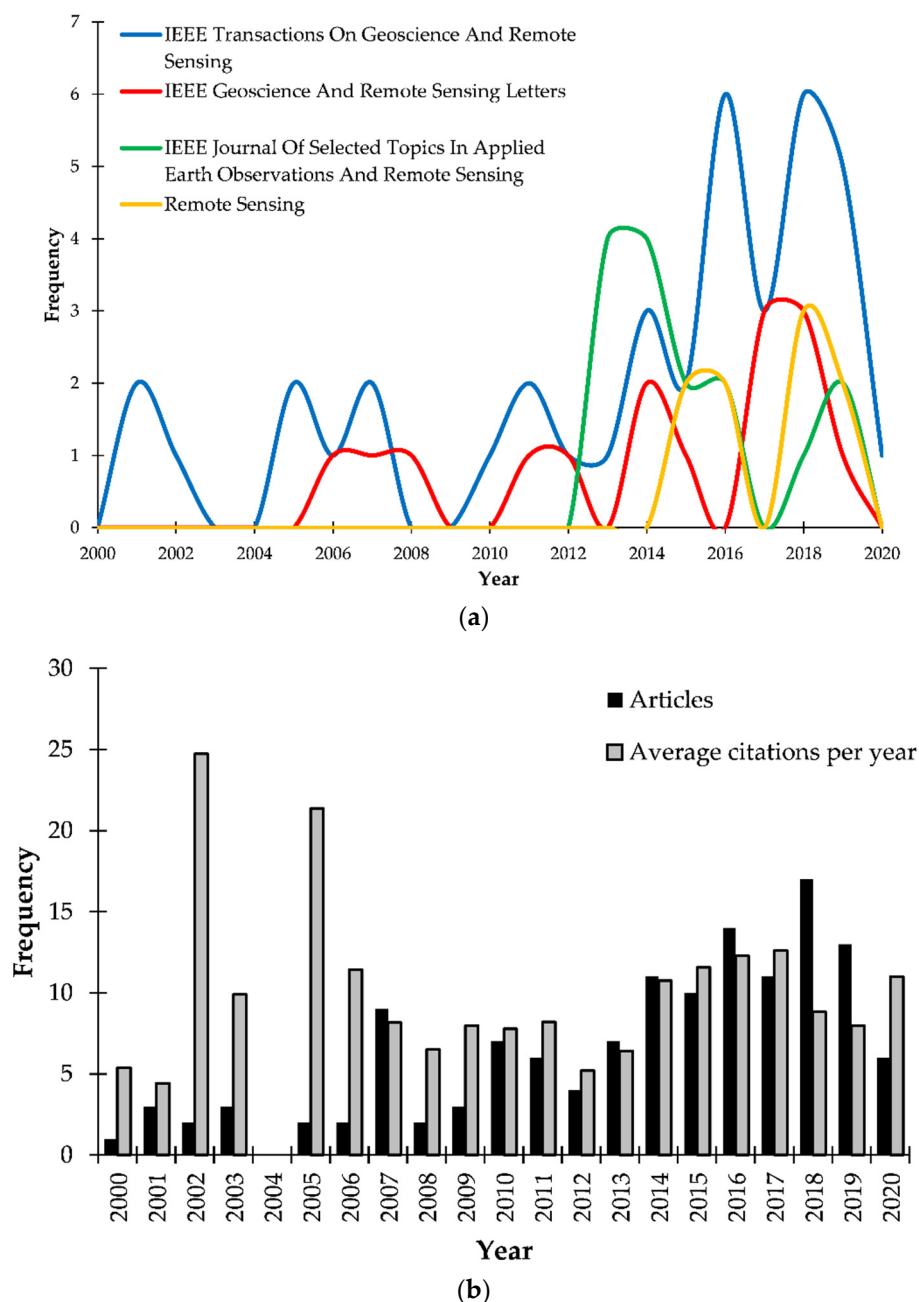
(**a**)

(**b**)

Figure 2. (**a**) Source trends for top 4 journals by number of documents (Table 2). (**b**) Trends in scientific production depicted by the number of articles and average citations per year.

2.2. Authors Analysis

The 20 most relevant authors, sorted by number of total citations and respective bibliometric statistics, are listed in Table 3. Chein-I Chang was the most cited author, with 1259 citations from 10 publications. He was followed by Du Qian with 1044 citations, who was the second most productive author with 12 publications. The most productive author, with 13 publications, was Zhang Liangpei.

Scientific production over the time span of the research (Table 1) for the top 10 authors by number of articles and total number of citations per year is presented in Figure 3. The most cited author has also been the most persistent author in the field, with his first publication originating back in 2001 and the last published in 2019. It is interesting to see that the second most cited author has also been the second most persistent, publishing

articles from 2007 to 2019. It can be noticed that some authors listed in Figure 3 are not listed in Table 3, and the reason is that sorting in Figure 3 included the normalization of citations per year. Hence, some authors had recently published quality papers with high citation rates, but the total number of citations was still not high enough to earn them a place in Table 3.

Table 3. The statistics of the 20 most relevant authors in the topic, sorted by total number of citations.

Author	SCOPUS Author ID	H-index	Total Citations	Number of Publications	First Publication (Year)
Chang CI	35253647700	10	1259	10	2001
Du Q	7202060063	11	1044	12	2007
Zhang L	8359720900	13	962	13	2011
Du B	55020400300	9	799	9	2011
Nasrabadi NM	7006312852	3	724	3	2003
Stocker AD	7006884172	2	698	2	2002
Kwon H	7401838362	3	611	3	2003
Diani M	7003735775	6	597	6	2010
Matteoli S	24076749300	6	597	6	2010
Beaven SG	57206689538	1	554	1	2002
Hoff LE	7005107977	1	554	1	2002
Schaum AP	57207501822	1	554	1	2002
Stein DWJ	7401616297	1	554	1	2002
Winter EM	7102040936	1	554	1	2002
Fowler JE	7402370679	4	513	4	2007
Chiang SS	7201472110	2	511	2	2001
Li J	24481713500	5	496	5	2014
Corsini G	7103074007	5	486	5	2010
Li W	56215159000	4	442	4	2015
Plaza A	7006613644	5	420	5	2010
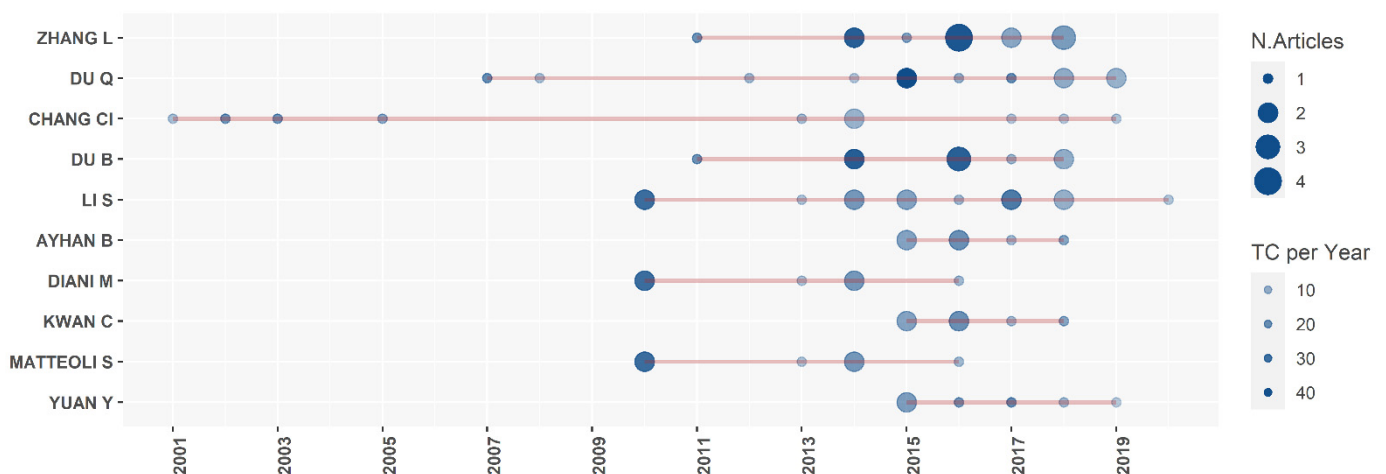
Figure 3. Scientific production of top 10 authors in the field over the time span of the research, determined and sorted by number of articles (N. articles) and total number of citations per year (TC per year). The figure was created using the bibliometrix R-package [23].

Tables 4 and 5 contain the top documents measured by global and local citations, respectively. Table 4 relates to the publications that were selected within the search strategy, sorted by global citations. Table 5 refers to documents that were highly cited locally but not enclosed by the search strategy results.

The work of Stein et al. [15] received the highest number of global citations and the second-highest number of local citations. In the analyzed document list (Figure 1), the work of Kwon and Nasrabadi [26] was appreciated the most, but it also attracted significant attention from the general scientific community, with 470 global citations. The more recent work on collaborative representation for hyperspectral anomaly detection by Li and Du [27] should be underlined, as it received great attention, both globally or locally.

Table 4. The top 10 documents with the highest number of global citations.

Document	Reference	Global Citations	Local Citations
Stein, D.W.J.; Beaven, S.G.; Hoff, L.E.; Winter, E.M.; Schaum, A.P.; Stocker, A.D. Anomaly detection from hyperspectral imagery. IEEE Signal Process Mag 2002, 19, 58-69, doi:10.1109/79.974730	[15]	554	54
Kwon, H.; Nasrabadi, N.M. Kernel RX-algorithm: A non-linear anomaly detector for hyperspectral imagery. IEEE Trans Geosci Remote Sens 2005, 43, 388-397, doi:10.1109/TGRS.2004.841487	[26]	470	56
Chang, C.I.; Chiang, S.S. Anomaly detection and classification for hyperspectral imagery. IEEE Trans Geosci Remote Sens 2002, 40, 1314-1325, doi:10.1109/TGRS.2002.800280	[28]	386	51
Ren, H.; Chang, C., I.. Automatic spectral target recognition in hyperspectral imagery. IEEE Trans. Aerosp. Electron. Syst. 2003, 39, 1232-1249, doi:10.1109/TAES.2003.1261124.	[29]	360	5
Matteoli, S.; Diani, M.; Corsini, G. A tutorial overview of anomaly detection in hyperspectral images. IEEE Aerosp Electron Syst Mag 2010, 25, 5-27, doi:10.1109/MAES.2010.5546306.	[11]	322	33
Du, Q.; Fowler, J.E. Hyperspectral Image Compression Using JPEG2000 and Principal Component Analysis. IEEE Geoscience and Remote Sensing Letters 2007, 4, 201-205, doi:10.1109/LGRS.2006.888109.	[30]	321	2
Banerjee, A.; Burlina, P.; Diehl, C. A support vector method for anomaly detection in hyperspectral imagery. IEEE Trans Geosci Remote Sens 2006, 44, 2282-2291, doi:10.1109/tgrs.2006.873019.	[31]	286	38
Li, W.; Du, Q. Collaborative representation for hyperspectral anomaly detection. IEEE Trans Geosci Remote Sens 2015, 53, 1463-1474, doi:10.1109/tgrs.2014.2343955.	[27]	251	42
Penna, B.; Tillo, T.; Magli, E.; Olmo, G. Transform Coding Techniques for Lossy Hyperspectral Data Compression. IEEE Trans Geosci Remote Sens 2007, 45, 1408-1421, doi:10.1109/TGRS.2007.894565.	[32]	241	3
Du, B.; Zhang, L. A Discriminative Metric Learning Based Anomaly Detection Method. IEEE Trans Geosci Remote Sens 2014, 52, 6844-6857, doi:10.1109/TGRS.2014.2303895.	[33]	220	27

Although the search strategy did not cover it, the paper published in 1990 by Reed and Yu [34] received by far the highest number of local citations. In that paper, the foundations of the well-known (Reed-Xiaoli, RX) hyperspectral anomaly detector were established. This algorithm is still popular; it is widely regarded as the benchmark hyperspectral anomaly detection technique and is often used in the comparison of newly developed anomaly detectors.

Manolakis and Shaw (Table 5) are among the pioneers of hyperspectral target and anomaly detection, and their work is also highly cited. The diligent work of Nasrabadi has been recognized, accounting for as many as four out of the 10 top locally cited publications in Table 5.

Table 5. The top 10 documents with the highest number of local citations that were not selected within the applied search strategy (Figure 1).

Document	Reference	Local Citations
Reed, I.S.; Yu, X. Adaptive Multiple-Band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution. IEEE Trans. Acoust. Speech Sign. Proces. 1990, 38, 1760-1770, doi:10.1109/29.60107.	[34]	90

Carlotto, M.J. A cluster-based approach for detecting man-made objects and changes in imagery. IEEE Trans Geosci Remote Sens 2005, 43, 374-387, doi:10.1109/TGRS.2004.841481	[35]	32
Manolakis, D.; Shaw, G. Detection algorithms for hyperspectral imaging applications. IEEE Signal Process Mag 2002, 19, 29-43, doi:10.1109/79.974724.	[9]	32
Nasrabadi, N.M. Hyperspectral target detection: An overview of current and future challenges. IEEE Signal Process Mag 2014, 31, 34-44, doi:10.1109/MSP.2013.2278992	[18]	21
Harsanyi, J.C.; Chang, C.I. Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach. IEEE Trans Geosci Remote Sens 1994, 32, 779-785, doi:10.1109/36.298007	[36]	19
Kerekes, J. Receiver operating characteristic curve confidence intervals and regions. IEEE Geoscience and Remote Sensing Letters 2008, 5, 251-255, doi:10.1109/lgrs.2008.915928.	[37]	17
Manolakis, D.; Marden, D.; Shaw, G.A. Hyperspectral Image Processing for Automatic Target Detection Applications. Lincoln laboratory journal 2003, 14, 79-116	[38]	16
Nasrabadi, N.M. Regularization for spectral matched filter and RX anomaly detector. In Proceedings of Proc SPIE Int Soc Opt Eng, 2008	[39]	16
Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Sparse Representation for Target Detection in Hyperspectral Imagery. *IEEE J. Sel. Top. Signal Process.* 2011, 5, 629-640, doi:10.1109/jstsp.2011.2113170	[40]	16
Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral Image Classification Using Dictionary-Based Sparse Representation. IEEE Trans Geosci Remote Sens 2011, 49, 3973-3985, doi:10.1109/tgrs.2011.2129595.	[41]	16

Figure 4 and Table 6 depict the temporal development of research on hyperspectral image processing for anomaly detection in remote sensing applications. Table 6 is based on the use of the author's keywords, and Figure 4, on the document title's keywords. In selecting the author's keywords presented in Table 6, trivial keywords such as hyperspectral, anomaly, detection, target, background, and similar terms were omitted. Then, the most relevant keywords were manually selected from the remaining keywords based on keyword frequency and global citation of the respective document. For the creation of Figure 4, the most frequent title keywords per year are displayed with the constraint of a maximum of three keywords per year. The respective frequency (for the specific year) is displayed on a vertical axis in a logarithmic scale, and as can be seen, no trivial keywords were omitted, nor was any manual filtering performed.

Table 6. The topic development expressed by the key author's keywords over time. The most relevant keywords were filtered and manually selected by authors.

Year	Document
2001	Competitive Region Growth, Elliptically Contoured Distributions, Evolutional Algorithm, Kurtosis, Projection Pursuit, Spherically Invariant Random Vectors
2002	Causal RXD, Correlation Matched-Filter-Based Measure, Target Discrimination Measure
2003	Clustering Algorithms, Dual Window, Eigen Separation Transform, Embedded Computing
2005	Kernels, Linear Discriminant Analysis, Orthogonal Subspace Projection/AD, RX Detector, Signal Parameter Estimation
2006	Bhattacharyya Distance, Signal Subspace Processing, Support Vector Data Description (SVDD)
2007	Detection Index, Minimum Description Length, Real-Time (R-T) Processing, Self-Organising Maps, Separability Index, Signal-Subspace Rank, Singular Value Decomposition, Wavelets
2008	Karhunen-Love-Transform, Principal Component Analysis (PCA), Kernel PCA, Signal Detection, Spectral Decorrelation
2009	GPU Processing, Generalized Least Squares, Maximum Autocorrelation Factors, Multivariate Normal Mixture Model, Principal Autocorrelation Factors
2010	Cluster-Based Approach, Feature Selection, Kernel-Based Learning, Quasi-Local Covariance Matrix, Regularization, Robust Locally Linear Embedding
2011	Embedded Systems, Gaussian Kernel, Independent Component Analysis, ROC Space, Sparse Matrix Transform, Support vector machine (SVM)
2012	Clustering, Compressed Sensing, PCA, Segmentation-Based AD, Sparse Kernel-Based Ensemble Learning, Spectral Unmixing
2013	Bayesian Learning, Dual Window-Based Eigen Separation Transform, Finite Mixture Model, Kernel density estimation (KDE), Multicore Platforms, Multiple-Window AD, Nonlinear PCA

| 2014	Dimensionality Reduction, High-Order Statistics, Local Sparsity Divergence, Low-Rank (L-R) And Sparse, Matched Filter, Robust Regression Analysis, Superpixels, Variable Bandwidth KDE, Weighted-RXD
2015	Graph Theory, High Order Statistics, L-R Approximation, Manifold Learning, R-T processing, Residual Analysis, Robust Background Estimation
2016	Cluster Kernel RX, Dual Clustering, Kernel Collaborative Representation (CR), Local Summation Strategy, Locally Linear Embedding, ROC, Robust PCA, Sparse Representation (SR), Sparsity Divergence Index, Spectral-Spatial Integration, Tensor Representation
2017	3-D ROC, Band Subset Selection, Convolutional Neural Network (NN), Differential Morphology, Edge-Preserving Filtering, Joint SR, K-SVD, Multiple Graphs, Autoencoders, Tensor Decomposition
2018	A Posteriori AD, Band Selection, Deep Learning, Feature Extraction, Inverse PCA, Iterative AD, L-R Representation, Multiple Dictionaries, R-T Applications, Sparse Coding, Structured SR
2019	Adaptive Weighting, Constrained SR, Deep Brief Network, Dictionary Learning, Fractional Fourier, Local Summation, Low Dimensional Manifold Model, Structure Tensor
2020	Density Peak Clustering, Isolation Forest, Radiative Transfer Modeling
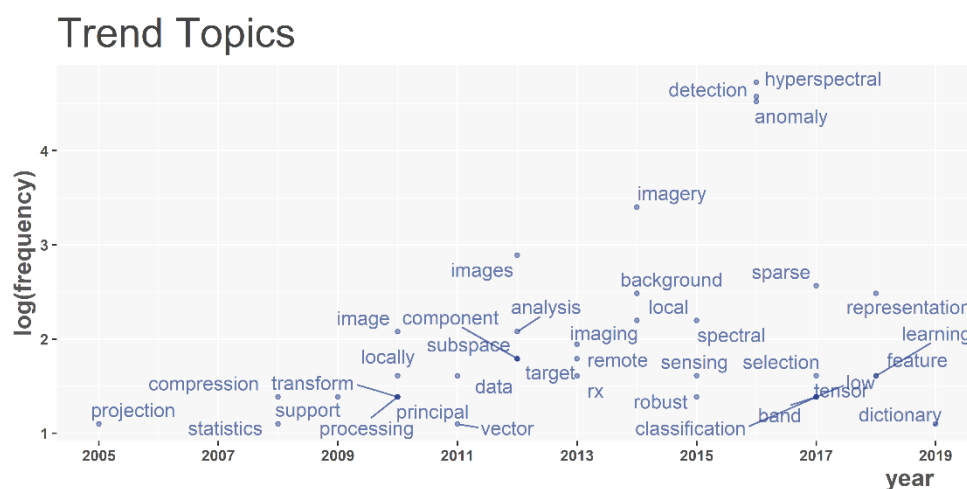
Figure 4. The development of the theme described by the title keywords. The graph represents the use of the most frequent title keywords (their frequencies are expressed on a logarithmic scale) over time. The figure was created using the bibliometrix R-package [23].

The Word cloud presented in Figure 5 enables clear insight into the trends and future scientific research in hyperspectral image processing for anomaly detection in remote sensing applications. The figure was generated using the document title keywords, similarly as in Figure 4, but here, the total frequency over the analyzed time span was considered. The total frequencies determined the respective size of words, so the most frequent keywords were written with more prominent characters.

By analyzing Figures 4 and 5 and Table 6, a recent trend in the development of representation-based techniques can be identified. Low-rank, spare, collaboration, joint, and dictionary are terms that can be related to these techniques. Hence, a special section is dedicated to these methods in Part B of this paper.

Figure 5. Word cloud generated from the most frequent keywords found in the titles of the analyzed documents. The size of the word is proportional to its frequency. The figure was created using the bibliometrix R-package [23].

3. Part B: An Overview of Hyperspectral Image Processing for Anomaly Detection in Remote Sensing Applications

This section provides a mathematical framework for anomaly detection, followed by the description of specific anomaly detection techniques that are generally divided into methods that adopt structured background models or unstructured background models.

4. Mathematical Framework for Anomaly Detection

The mathematical framework for anomaly detection arises from the signal detection theory [42], and is based on the field of statistical or binary hypothesis testing. A hyperspectral image or a hyperspectral cube can be regarded as a rank-3 tensor arranged in two spatial dimensions and one spectral dimension. If we consider a hyperspectral cube having N pixels and K spectral bands ($\underline{X} \in \mathbb{R}^{N \times K}$), then a pixel spectrum could be represented as the realization of a random vector ($\underline{X}$) that can be denoted as $\underline{x} = [x_1, x_2, \ldots, x_K]^T$. Given an observed pixel $\underline{x}$, it needs to be decided between two disjunct premises:

$$\widehat{H}(\underline{x}) = \begin{cases} H_0: \text{ target absent } (\underline{x} \text{ is a background pixel}) \\ H_1: \text{ target present } (\underline{x} \text{ is a target pixel}) \end{cases}. \tag{1}$$

In general, a statistical hypothesis test is a rule for division of input K-dimensional feature space in two segments: space (A) consistent with the null hypothesis H_0 and its complement (B) that contains values consistent with the alternative hypothesis H_1. In signal detection theory terminology, it can be stated that a decision about the proper hypothesis is determined by a test or detection statistic ($\Lambda, \Lambda(\underline{x}) = \lambda$) whose detection threshold or critical value (η) splits input feature space into segments A ($H_0: \Lambda(\underline{x}) \leq \eta$) and B ($H_1: \Lambda(\underline{x}) > \eta$). The detection statistic usually maps input n-dimensional feature space (K-dimensional in our specific case) into 1-dimensional space, thus enabling separation of spaces A and B on a line instead of some high dimensional space.

Upon hypothesis testing, two types of erroneous inferences could arise: a type I error occurs if a null hypothesis H_0 is rejected while being true, and a type II error occurs if the H_0 is accepted when the alternative hypothesis H_1 is true. The probability of a type I error is usually denoted with α, commonly called the significance level. It can be mathematically formulated as [11]:

$$\alpha = P\{\Lambda(\underline{x}) > \eta | H_0\} = \int_{\lambda > \eta} f_{\Lambda|H_0}(\lambda)\, \mathrm{d}\lambda = \int_{\underline{x}:\Lambda(\underline{x})>\eta} f_{\underline{X}|H_0}(\underline{x})\, \mathrm{d}\underline{x}. \tag{2}$$

In hyperspectral anomaly detection, a type I error occurs if an observed pixel is proclaimed as a target (anomaly) when it actually belongs to the background. Therefore, a type I error is also called a false alarm ($P_{FA} = \alpha$). The probability of a type II error is usually denoted with β, and it can be expressed as [11]:

$$\beta = P\{\Lambda(\underline{x}) \le \eta | H_1\} = \int_{\lambda \le \eta} f_{\Lambda|H_1}(\lambda)\, \mathrm{d}\lambda = \int_{\underline{x}:\Lambda(\underline{x})\le\eta} f_{\underline{X}|H_1}(\underline{x})\, \mathrm{d}\underline{x}. \tag{3}$$

A type II error happens if an observed pixel is proclaimed as a background pixel when it is a target (anomaly), thus resulting in missed detection. When β is known, the probability of anomaly detection (P_D) can be directly determined as in [11]:

$$P_D = 1 - \beta$$

$$P_D = P\{\Lambda(\underline{x}) > \eta \,| H_1\} = \int_{\lambda > \eta} f_{\Lambda|H_1}(\lambda)\, \mathrm{d}\lambda = \int_{\underline{x}:\Lambda(\underline{x})>\eta} f_{\underline{X}|H_1}(\underline{x})\, \mathrm{d}\underline{x} \tag{4}$$

The ideal detection statistic could make the probabilities of both error types arbitrarily small, but no such statistical test exists. In statistical decision theory, it is well-accepted that decisions based on the likelihood ratio test (LRT) are optimum over numerous performance criteria [10,42]. Let $f(\underline{X}|H_0)$ and $f(\underline{X}|H_1)$ be conditional probability distributions of observing $\underline{x}$ under H_0 and H_1 respectively. Then, the LRT between $(\underline{X}|H_0)$ and $f(\underline{X}|H_1)$ can be derived as [10]:

$$\Lambda(\underline{x}) = \frac{f_{\underline{X}|H_1}(\underline{x})}{f_{\underline{X}|H_0}(\underline{x})} \overset{H_1}{\underset{H_0}{\gtrless}} \eta. \tag{5}$$

The classical approach in selecting optimum detection statistics is based on the well-known Neyman-Pearson (NP) lemma [43]. The NP lemma for a given significance level α obtains the test that has the lowest possible probability of missed detection β for all parameters defined by the alternative hypothesis H_1. In other words, NP is the optimum criterion that maximizes the probability of detection for any desired false alarm rate [42]. Solution of (5), in an NP sense, is only possible if both hypothesis H_0 and H_1 are simple hypotheses. Simple hypotheses have parameters of their conditional probability functions $f(\underline{X}|H_0)$ and $f(\underline{X}|H_1)$ known, but unfortunately, these requirements are rarely fulfilled in practical applications. In those cases, sub-optimum criteria that are based on the generalized likelihood ratio test (GLRT) are implemented. In GLRTs, unknown parameters are substituted with their maximum likelihood estimates (MLE) [11,38]:

$$\Lambda_{GLRT}(\underline{x}) = \frac{f_{\underline{X}|H_1}(\underline{x};\hat{\underline{\vartheta}}_1)}{f_{\underline{X}|H_0}(\underline{x};\hat{\underline{\vartheta}}_0)} = \frac{max_{\underline{\vartheta}_1}\, f_{\underline{X}|H_1}(\underline{x};\hat{\underline{\vartheta}}_1)}{max_{\underline{\vartheta}_0}\, f_{\underline{X}|H_0}(\underline{x};\hat{\underline{\vartheta}}_0)} \overset{H_1}{\underset{H_0}{\gtrless}} \eta, \tag{6}$$

where $\hat{\underline{\vartheta}}_0$, $\hat{\underline{\vartheta}}_1$ represent the vectors of unknown parameters. MLEs of unknown parameters are usually determined from the test and/or reference samples that should be independent and identically distributed (IID).

In anomaly detection, no presumptions about the target model are made. Instead, background models are developed. In doing so, a wide variety of creative approaches were suggested, including the adoption of multivariate normal models [34], subspace projections [36], and kernel methods [26]. Although attempts have been made to provide the in-depth taxonomy of complex hyperspectral anomaly detection methods [10,11] or background modeling for detection of anomalies in hyperspectral remotely sensed imagery

[17], we suggest a simple distinction of anomaly detection methods in two basic categories: methods that implement unstructured background models and methods that adopt structured background models.

Besides background models, methods could be further specified by the implemented spatial model. The spatial model can be one of the following: global, local, or quasi-local. If a global model is used, then all or most available hyperspectral pixels are used to characterize the background. In a local model, only the spatial neighbors of PUT are used for the description of the background. Quasi-local methods present a trade-off solution to global and local models.

5. Unstructured Background Models

Fundamental anomaly detection methods implement unstructured background models. They are often specified as data-driven, statistical, or probabilistic. These models do not assume any specific structure on the hyperspectral cubes based on a-priori knowledge, but integrate additive noise in the model.

Let $\underline{\mu} = [\mu_1, \mu_2, \cdots, \mu_K]$ be the mean vector of a hyperspectral cube with K bands, where μ_1 represents the average of all pixels in the first spectral band and μ_K the average of the last spectral band. The mean vector $\underline{\mu}$ is often called the background centroid or background prototype. In this respect, a squared Euclidean distance between PUT ($\underline{x}$) and background centroid can be used as a simple detection statistic:

$$d_E^2(\underline{x}, \underline{\mu}) = \left(\underline{x} - \underline{\mu}\right)^T \cdot \left(\underline{x} - \underline{\mu}\right). \tag{7}$$

In hyperspectral images, due to their fine spectral resolution, spectral adjacent bands are usually highly correlated. Therefore in (7), a weighting matrix in the form of the inverse covariance matrix ($\underline{\underline{\Gamma}}^{-1}$) could be added:

$$d_M^2(\underline{x}, \underline{\mu}) = \left(\underline{x} - \underline{\mu}\right)^T \cdot \underline{\underline{\Gamma}}^{-1} \cdot \left(\underline{x} - \underline{\mu}\right). \tag{8}$$

Expression (8) represents the widely known squared Mahalanobis distance [44] between PUT and background centroid. It should be noticed that distance in (8) is proportional with negative log-likelihood of Gaussian or normal distribution:

$$d_M\left(\underline{x}, \underline{\mu}\right) \propto -\log p_{\mathcal{N}}\left(\underline{x}|\underline{\mu}, \underline{\underline{\Gamma}}\right), \tag{9}$$

where

$$p_{\mathcal{N}}(\underline{x}|\underline{\mu}, \underline{\underline{\Gamma}}) = \frac{1}{(2\pi)^{\frac{K}{2}} \cdot |\underline{\underline{\Gamma}}|^{\frac{1}{2}}} exp\left(-\frac{1}{2}\left(\underline{x} - \underline{\mu}\right)^T \cdot \underline{\underline{\Gamma}}^{-1} \cdot \left(\underline{x} - \underline{\mu}\right)\right). \tag{10}$$

Therefore, if a background can be adequately characterized with the multivariate normal distribution, then the squared Mahalanobis distance could be an appropriate detection statistic for anomaly detection. The multivariate normal model for anomaly detection can be written as:

$$\begin{aligned} \underline{X}|H_0 &= \underline{B} & \in \mathcal{N}\left(\underline{\mu}, \underline{\underline{\Gamma}}\right) \\ \underline{X}|H_1 &= \underline{s} + \underline{B} & \in \mathcal{N}\left(\underline{s}, \underline{\underline{\Gamma}}\right)' \end{aligned} \tag{11}$$

where $\underline{B}$ denotes the spectral vector belonging to the background (which incorporates additive noise), $\underline{s}$ is the spectral vector of the target, and $\underline{\underline{\Gamma}}$ is the background covariance matrix that is assumed to be the same for the background and target class. The most popular anomaly detector for hyperspectral remote sensed images is based on the multivariate normal model, and it is named the Reed-Xiaoli (RX) algorithm [34].

5.1. Reed-Xiaoli (RX) Algorithm

The RX algorithm [34] can be considered as the reference in hyperspectral image processing for anomaly detection in remote sensing applications, and it has become the standard with which new anomaly detectors are compared [45–49]. The algorithm idea stemmed from the test developed for the detection of radar targets [50], as well as most of the detection theory in remote sensing that arose from the processing of radar data. RX is an adaptive constant false alarm rate (CFAR) hyperspectral anomaly detection algorithm developed from the GLRT [34]. CFAR algorithms exhibit the property that the probability of false alarm does not depend on any unknown parameter, and it is a highly desirable property in the design of anomaly detectors. It should be noticed that in (11) the covariance matrices for background and target class are assumed to be the same. This condition needs to be fulfilled in order to make the RX detector optimum in a Neyman-Pearson sense (5), i.e., to hold the CFAR property [38].

To date, many implementations of the RX algorithm have been proposed: global, local, and quasi-local. Although RX in its original form is a local algorithm, we will describe its global variant first as it is the simplest one. The local and quasi-local RX models will follow afterward.

Hyperspectral pixels are considered as random spectral vectors that are IID. These pixels are then used for estimation of unknown parameters of the multivariate normal distribution: $\underline{\mu}$ and $\underline{\underline{\Gamma}}$. Consider a hyperspectral cube with a total of N pixels, namely N random spectral vectors with K elements (K being the number of spectral bands). The global RX detector can be defined as [34]:

$$D_{RX}(\underline{x}|\mathcal{B}) = D_{RX}(\underline{x}|\underline{B}) = \left(\underline{x} - \underline{\hat{\mu}}\right)^T \cdot \underline{\hat{\Gamma}}^{-1} \cdot \left(\underline{x} - \underline{\hat{\mu}}\right) \underset{H_0}{\overset{H_1}{\underset{<}{\gtrless}}} \lambda, \qquad (12)$$

where $\mathcal{B}$ represents the theoretical background model, while MLEs $\underline{\hat{\mu}}$ and $\underline{\underline{\hat{\Gamma}}}$ are determined as:

$$\underline{\hat{\mu}} = \frac{1}{N} \sum_{i=1}^{N} \underline{x}_i$$

$$\underline{\underline{\hat{\Gamma}}} = \frac{1}{N-1} \sum_{i=1}^{N} \left(\underline{x}_i - \underline{\hat{\mu}}\right) \cdot \left(\underline{x}_i - \underline{\hat{\mu}}\right)^T \qquad (13)$$

Theoretical foundations for the local (and native) RX detector were set by Hunt and Cannon [51]. They suggested that an optical digital image can be modeled as a nonstationary multivariate Gaussian random process with a highly space-varying spatially nonstationary mean vector and spatially stationary or much less space-varying covariance matrix. The local RX detector uses the same mathematical methodology as the global RX, though it calculates the MLE of unknown parameters (13) using the selected pixels found in the PUT's spatial neighborhood (Figure 6). The local neighborhood is usually defined using one or more sliding windows (annulus). There are various strategies in selecting the number of windows and their sizes. If only one window is used, it presents an outer window used to estimate all unknown parameters. The most accepted strategy includes the use of three distinct windows: guard window and windows for estimation of $\underline{\hat{\mu}}$ and $\underline{\underline{\hat{\Gamma}}}$, with their respective sizes (expressed in widths) w_g, $w_{\underline{\hat{\mu}}}$ and $w_{\underline{\underline{\hat{\Gamma}}}}$ (Figure 6). The guard window is the smallest window, and it defines the area from which samples are not taken, as there is a possibility that it contains pixels that could be anomalies or pixels that are spectral mixtures with anomalies. Inclusion of these pixels in the calculation of $\underline{\hat{\mu}}$ and $\underline{\underline{\hat{\Gamma}}}$ would lead to lower anomaly detection statistic scores and finally decrease the probability of detection. Therefore, its size should be set to accommodate for expected target sizes. Windows for $\underline{\hat{\mu}}$ and $\underline{\underline{\hat{\Gamma}}}$, defined as hollow boxes bounded with guard window from the

inside are used for estimation of $\hat{\underline{\mu}}$ and $\hat{\underline{\Gamma}}$. It was shown in [52] and [53] that local Gaussianity on the image can be forced by subtracting the nonstationary mean. In [34], it is suggested that residual background is a zero-mean Gaussian and independent in spatial domain:

$$
\begin{aligned}
\underline{X}|H_0 &= \underline{B} && \in \mathcal{N}\left(\underline{0}, \hat{\underline{\Gamma}}\right) \\
\underline{X}|H_1 &= \underline{s} + \underline{B} && \in \mathcal{N}\left(\underline{s}, \hat{\underline{\Gamma}}\right).
\end{aligned}
\tag{14}
$$

The difference in window sizes for $\hat{\underline{\mu}}$ and $\hat{\underline{\Gamma}}$ arises from the stationarity of these unknown parameters [51]; the mean should vary more frequently than the covariance matrix.
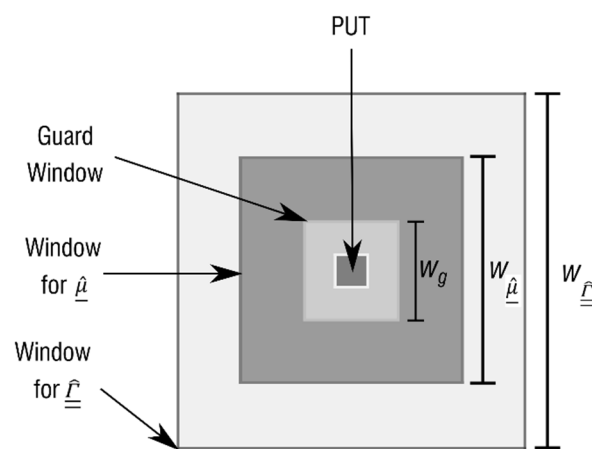
Figure 6. Geometrical principle of local RX detector with three specific sliding windows: guard window and windows for calculation of $\hat{\underline{\mu}}$ and $\hat{\underline{\Gamma}}$, with their respective widths w_g, $w_{\hat{\underline{\mu}}}$ and $w_{\hat{\underline{\Gamma}}}$. Detection statistic is calculated in a convolutional manner using the squared Mahalanobis distance.

In selecting the largest window size, a balance between two opposing requirements should be found. The window for $\hat{\underline{\Gamma}}$ should include enough samples for stable determination of the inverse of the covariance matrix. Swain and Davis [54] argue that a reliable estimate of covariance matrix requires at least 10 times and preferably 100 times more samples than the number of spectral bands in a hyperspectral image. The opposing requirement states that the window size should be as small as possible to capture a homogenous background and reduce the computational load needed for inverse calculation of the covariance matrix.

Quasi-local implementation of the RX detector includes the use of one global covariance matrix but performs local demeaning. This approach requires the calculation of only one covariance matrix using all available samples, which leads to a more stable inverse and lower computational complexity than the local implementation of the RX detector.

Advantages of the global RX detector are simple implementation and fast computation, as there is only one covariance matrix to be inverted. If the assumptions about the local multivariate normal model (14) and $\underline{\Gamma}_B = \underline{\Gamma}_s = \underline{\Gamma}$ are fulfilled, then the squared Mahalanobis distance follows a non-central chi-squared distribution $\chi_K^2(\beta)$ with K degrees of freedom and non-centrality parameter β [38]:

$$D_{RX}(\underline{x}|\mathcal{B}) \sim \begin{cases} \chi_K^2(0), \text{under } H_0 \\ \chi_K^2(d_M^2), \text{under } H_1 \end{cases}, \tag{15}$$

thereby enabling the CFAR property.

The limitation of the local RX detector is that the covariance matrix is determined from a small number of high-dimensional and highly correlated samples. That can result in rank deficiency of covariance matrix or often leads to instability of its inverse. Matrix inversion needs to be calculated for every pixel, which makes the local RX detector computationally intensive. The local RX detector suffers from an increased number of false alarms because some pixels can be anomalies in the local background but not on the entire image (e.g., isolated tree). The main limitation of RX detectors is that in hyperspectral imaging in remote sensing applications, the assumption of Gaussianity is often violated due to the presence of multiple materials in the scene or in the local background, which adversely impacts detection performance.

Improved Variants of the RX Detector

Many authors have tried to tackle the outlined shortcomings of the RX detector and improve its detection performance. First of all, it should be remarked that there is a confounding aspect to the naming of the new anomaly detectors. In our opinion, the RX detector should only refer to the local algorithm defined and explained above, with the exceptions of the global and quasi-local spatial implementations. The anomaly detection methods that use the squared Mahalanobis distance as detection statistics are often named RX variants [15,18,26].

To mitigate the computational cost of the (local) RX detector, a variety of parallel implementations for multicore platforms (CPU and GPU) have been developed [55–58]. Molero et al. [57] proposed optimized versions of the RX detectors based on the Cholesky decomposition of the correlation matrices with parallel implementations on high-performance computing platforms: multicore and GPU.

Manolakis [10] stated that the normal distribution model accurately fits the body of the data but not its tails. That is particularly important in anomaly detection, as the distribution tails have the most influence on the false alarms. Therefore, the family of multivariate elliptically contoured *t* distribution, which can model heavier distribution tails, is suggested in lieu of the multivariate normal distribution [59–61].

The quasi-local (QL) RX detector is a compromise between the local and global RX detector and should not be confused with the quasi-local implementation of the RX detector. The original QL idea was applied to achieve the QL covariance matrix [62–64], but in later publications, this approach is referred to as the QL RX detector [65,66]. The basic idea of the QL RX detector is to decompose the global covariance matrix using the eigenvalue decomposition:

$$\underline{\underline{\Gamma}} = \underline{\underline{E}} \cdot \underline{\underline{\Lambda}} \cdot \underline{\underline{E}}^T. \tag{16}$$

The eigenvectors in the $\underline{\underline{E}}$ are forwarded to the detector, but the eigenvalues in the $\underline{\underline{\Lambda}}$ are replaced by the larger of the local and the global variance [62]:

$$\lambda_{QL}^i = max(\lambda_{LOCAL}^i, \lambda_{GLOBAL}^i). \tag{17}$$

That enables lower detection scores in the image areas with higher variance, and it thus may lead to a lower probability of false alarms. The main benefits of the QL approach are a more stable estimation of the covariance matrix and a much less expensive way of computing its inverse.

Due to the high correlation of the spectral bands, the high dimensionality of the hyperspectral data, and a limited number of samples for estimation of the local covariance matrices, they often suffer from ill-conditioning. A widely accepted method for improving the estimation of the inverse covariance matrix issue is shrinkage [67–70]. The most

straightforward method for the shrinkage of the covariance is the addition of a scaled identity matrix, which is often applied in the ridge regression technique [71]. The ridge regularized (RR) squared Mahalanobis distance (SMD) detector is then formulated as [39]:

$$D_{RR\,SMD}\left(\underline{x}|\mathcal{B}\right) = d^2_{M\,RR}(\underline{x}, \hat{\underline{\mu}}) = \left(\underline{x} - \hat{\underline{\mu}}\right)^T \cdot \left(\hat{\underline{\varGamma}} + \delta\underline{I}\right)^{-1} \cdot \left(\underline{x} - \hat{\underline{\mu}}\right) \overset{H_1}{\underset{H_0}{\gtrless}} \lambda, \qquad (18)$$

where δ denotes the regularization parameter and $\underline{I}$ is the identity matrix. The same detector (18) is referred to as ridge-regularized RX in [18]. Shrinkage performs small perturbations of the estimated covariance matrix, but that can accomplish significant effects on the invertibility if $\hat{\underline{\varGamma}}$ is near-singular. Many other methods for shrinkage can be found in the literature [39,72,73].

Chang and Chiang [28] presented the causal RX detector, which incorporates a sample correlation matrix instead of a covariance matrix and enables real-time processing of the RX detector. In the context of [28], real-time processing refers to processing the pixel as it is received, i.e., in an online data processing approach. Davidson and Ben-David [74] argued for the use of covariance or the correlation matrix and determined that the use of the correlation matrix could offer an improvement over the covariance matrix, but only in a relatively small part of the parameter space. They concluded that the potential performance gain could be modest, yet the potential performance loss could be devastating.

Real-time processing with the RX algorithm has preoccupied numerous scientists [28,75–77]. To the best of our knowledge, the first operational implementation of real-time hyperspectral detection was executed in Dark HORSE 1 (Hyperspectral Overhead Reconnaissance and Surveillance Experiment 1) [77]. In that research, it was shown that it is possible to autonomously detect military ground targets using visible hyperspectral images in real-time. In [75], the Woodbury matrix identity is introduced, which could be used to update the previously computed inverse matrix when new data needs to be considered.

To deal with different anomaly sizes, Liu et Chang [78] proposed a multiple-window approach. If the anomalies come in various sizes, as can happen in real applications, the detection performance of the local RX detector is limited. Although the idea is presented in general form, they proposed three specific multiple-window anomaly detectors (MW AD), of which one is the MW variant of RX detector MW-RXD. MW-RXD basically corresponds to the result of the local RX detector that is executed with *K* different window sizes. Finally, an overall MW-RXD detection map is obtained by a fusion of *K* detection maps, i.e., a summation of the *K* local RX detector results [72]:

$$D_{MW-RXD}\left(\underline{x}|\mathcal{B}\right) = \max_{1 \le i \le K} D_{RX}(\underline{x}|\mathcal{B})_i. \qquad (19)$$

In [79], a superpixel-based dual window RX (SPDW RX) detector is presented to address the same issue. The SPDW RX uses superpixel segmentation to adaptively determine the dual windows, instead of the two fixed-size windows of the local RX detector. Firstly, the hyperspectral image is divided into superpixels. Then for every superpixel, a minimum bounding rectangle is defined. The background statistics are then determined based on these minimum bounding rectangles, which are further used to calculate the detection statistic. The authors showed that SPDW RX could provide a small increase in detection performance but a large increase in processing speed compared to the local RX detector.

5.2. Nearest Neighbor Detectors

Besides the family of the RX detectors, the detectors based on the principle of the spectrally nearest neighbors (NN) [80,81] can be categorized as the unstructured background anomaly detectors. For the background characterization, a spectral distance of

PUT to the *C*-nearest neighbor can be used as a detection statistic, an average distance of the PUT to the *C*-nearest neighbors, or a distance of the PUT to the average of the *C*-nearest neighbors [82]. A choice between various distance metrics could be made: Euclidean, Mahalanobis, spectral angle [83], Manhattan [84], or Chebyshev [85].

The basic NN anomaly detector can be mathematically formulated as follows. Let w_i be the weight, which depends on the distance from the PUT $\underline{x}$ and background spectral vectors $\{\underline{x}_i\}$. Consider that background vectors are sorted by some distance metric to $\underline{x}$ (such that $\underline{x}_1$ is the closest). Now, a simple weight vector $\underline{w}$ for one PUT can be determined as [82]:

$$w_i = \begin{cases} 1, & \text{if } i \leq C \\ 0, & \text{if } i > C \end{cases}, \tag{20}$$

where *C* denotes the chosen number of the nearest neighbors. The detection statistic for the *C*-NN detector that uses the distance from the PUT ($\underline{x}$) to the average of *C*-NN can now be formulated as [82]:

$$D_{NN}(\underline{x}|\mathcal{B}) = \left\| \underline{x} - \frac{1}{N}\sum_{i=1}^{N} w_i \cdot \underline{x}_i \right\| \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \lambda \tag{21}$$

where $\mathcal{B} = \{w_i, \underline{x}_i\}_{i=1}^{N}$ presents the background model determined by a weight vector $\underline{w}$ and spectral background vectors $\underline{x}_i$, and $\| \quad \|$ denotes the distance operator.

The *C*-NN detector enables simple implementation, but its complexity depends on the selection of a distance metric. Additionally, for every PUT, the spectral distance to every background pixel needs to be calculated and sorted, which is computationally expensive. Finally, the criterion for the selection of optimal number *C* is not intuitively determined. It should be greater than the overall number of expected anomaly pixels, which could be difficult to foresee in real applications.

A Euclidean distance transformation for anomaly detection in spectral imagery has been proposed by Schlamm and Messinger [86]. They introduced the nearest neighbor transformation (NNT), in which the spectral *k*-nearest neighbors for every pixel in the HS image are determined using the ATRIA [87]. ATRIA is the algorithm, based on the Delaunay triangulation, that offers an efficient determination of a selection of the nearest neighbors. The NNT creates a new *k*-dimensional image where every *i-th* band contains the Euclidean distance of every pixel to its *i-th* spectrally nearest neighbor. A similar approach can also be found in the work of Zhao and Saligrama [88]. The standard RX detector or some other detection statistic can then be applied to the data transformed by the NNT. The NNT can also be regarded as the preprocessing step for the anomaly detectors that use the subspace models, which are explained later in the paper.

5.3. Kernel-Based Models

If the background and the anomalies can not be adequately separated in the data space, it may be useful to seek the simple decision boundary in the higher dimensional feature space. That is the basic idea of the kernel-based anomaly detection methods that rely on the so-called "kernel trick" [89]. We want to replace complex anomaly detection models in the data space with much simpler models in the higher dimensional feature space, which is generated using the non-linear mapping function $\underline{\Phi}(\cdot)$. The mapping can be done to *M*-dimensional feature space $\mathcal{F}$, where the dimensionality of the $\mathcal{F}$ can be indefinite (but usually $M \gg K$, where *K* denotes the number of spectral bands of the input hyperspectral image). The goal is to find the appropriate feature space $\mathcal{F}$ where the background and the anomaly class can be easily and more accurately separated. Simple decision boundaries in the higher dimensional space project to more complex boundaries in the lower dimensional space. That is the main benefit of the kernel-based methods, as they are able to reduce a non-linear algorithm in the data space to a linear one in the higher

dimensional $\mathcal{F}$. However, it is not computationally feasible to directly implement any algorithm in the $\mathcal{F}$, due to its high dimensionality. Luckily, there is a way to implicitly compute dot products in the feature space $\mathcal{F}$ without actually performing the non-linear mapping $\underline{\Phi}(\cdot)$ of the input spectral vectors $\{\underline{x}_i\}$. It is called the kernel trick, and it is an effective method to implement the dot product in the feature space using the kernel functions. The dot products in $\mathcal{F}$ can be kernelized as [26]:

$$k(\underline{x}_i, \underline{x}_j) = \underline{\Phi}(\underline{x}_i)^T \cdot \underline{\Phi}(\underline{x}_j). \tag{22}$$

From (22), it can be seen that the dot product in $\mathcal{F}$ can be replaced by a non-linear kernel function k, which can be computed without explicitly defining the mapping function $\underline{\Phi}(\cdot)$. One of the most commonly used kernel functions is the Gaussian radial basis function (RBF) kernel [18]:

$$k(\underline{x}_i, \underline{x}_j) = \exp\left(\frac{-\left\|\underline{x}_i - \underline{x}_j\right\|^2}{\sigma^2}\right) \tag{23}$$

where σ denotes the kernel bandwidth parameter.

5.3.1. Kernel RX detector

In [26], a non-linear anomaly detector that adopts a normal model in a higher dimensional feature space is presented. The RX algorithm in the feature space can be represented as [26]:

$$D_{FS-RX}[\underline{\Phi}(\underline{x})|\mathcal{B}] = \left(\underline{\Phi}(\underline{x}) - \hat{\underline{\mu}}_\Phi\right)^T \cdot \hat{\underline{\underline{\Gamma}}}_\Phi^{-1} \cdot \left(\underline{\Phi}(\underline{x}) - \hat{\underline{\mu}}_\Phi\right) \underset{H_0}{\overset{H_1}{\underset{<}{\gtrless}}} \lambda \tag{24}$$

where $\hat{\underline{\underline{\Gamma}}}_\Phi$ and $\hat{\underline{\mu}}_\Phi$ are the estimated covariance matrix and mean vector of the background in the feature space that can be estimated with the same spatial models as the RX detector. The equation (24) can not be explicitly implemented in the feature space due to the non-linear mapping function $\underline{\Phi}(\cdot)$, which projects data in a higher-dimensional space. To avoid doing so, equation (24) can be kernelized using the aforementioned kernel trick [18,26]:

$$D_{K-RX}(\underline{x}|\mathcal{B}) = \left(\underline{k}_{\underline{x}} - \underline{k}_{\hat{\underline{\mu}}}\right)^T \cdot \hat{\underline{\underline{K}}}^{-1} \cdot \left(\underline{k}_{\underline{x}} - \underline{k}_{\hat{\underline{\mu}}}\right) \underset{H_0}{\overset{H_1}{\underset{<}{\gtrless}}} \lambda \tag{25}$$

where $\underline{k}_{\underline{x}} = \underline{\Phi}\left(\underline{X}\right)^T \cdot \underline{\Phi}(\underline{x})$ represents the empirical kernel map of the test pixel $\underline{\Phi}(\underline{x})$, $\underline{k}_{\hat{\underline{\mu}}} = \underline{\Phi}\left(\underline{X}\right)^T \cdot \underline{\Phi}\left(\hat{\underline{\mu}}\right)$ denotes the corresponding empirical kernel map of the background mean $\underline{\Phi}\left(\hat{\underline{\mu}}\right)$, and $\hat{\underline{\underline{K}}} = \underline{\Phi}\left(\underline{X}\right)^T \underline{\Phi}\left(\underline{X}\right)$ is the centered kernel Gram matrix of the mean-removed background pixels $\underline{\Phi}\left(\underline{X}\right)$ in the feature space. That enables implementation of (25) without knowledge of the mapping function $\underline{\Phi}(\cdot)$; the only requirement that remains is selecting the kernel function k, which produces a positive definite Gram matrix. However, this can not be easily foreseen in real applications. Performance of the kernel RX detector is generally limited by the following: 1) it is sensitive to background contamination with anomalous pixels and noise, and 2) the inverse of $\hat{\underline{\underline{K}}}$ is usually rank-deficient [90]. A Gaussian background purification approach adapted to background data samples probability distribution and an inverse-of-matrix-free method based on kernel principal component analysis (PCA) [91] were proposed in [90] to address the aforementioned kernel RX limitations. To improve the memory and time efficiency of the kernel RX detector, two families of techniques for approximation of the kernel function with either the data-

independent random Fourier features or the data-dependent basis with the Nyström approach were proposed in [92].

5.3.2. Kernel Density Estimate of the Background Distribution Models

Kernel density estimation (KDE) is the well-known technique for nonparametric estimation of an unknown probability density function (PDF) based solely on the given dataset [93,94]. For estimation of the background PDF of hyperspectral images, a multivariate KDE can be used [45,95–99]. The multivariate KDE can be represented as [94]:

$$p_b(\underline{x}|\mathcal{B}) = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\left\|\underline{\underline{H}}(\underline{x},\underline{x}_n)\right\|} \kappa \left[\underline{\underline{H}}^{-1}(\underline{x},\underline{x}_n) \cdot (\underline{x}-\underline{x}_n) \right] \tag{26}$$

where $\underline{\underline{H}}(\cdot)$ is the bandwidth matrix (contains the kernel function widths) and $\kappa(\cdot)$ is the kernel function centered at each of the sample data $\{\underline{x}_n\}_{n=1}^{N}$. A simple strategy in determining the bandwidth matrix could be to set the same bandwidth to all spectral bands. That means that $\underline{\underline{H}}$ is equal to the scaled identity matrix $\underline{\underline{H}} = h \cdot \underline{\underline{I}}$, which makes the contours of the kernel function spherically symmetric [17]. With this simplification, (26) becomes [17]:

$$p_b(\underline{x}|\mathcal{B}) = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{h^d(\underline{x},\underline{x}_n)} \kappa \left[\frac{\underline{x}-\underline{x}_n}{h(\underline{x},\underline{x}_n)} \right]. \tag{27}$$

The respective detector for (27) is given in [11,17] as the background PDF log-likelihood function:

$$D_{FKDE}(\underline{x}|\mathcal{B}) = -\log \left[\frac{1}{N} \sum_{n=1}^{N} \frac{1}{h^d(\underline{x},\underline{x}_n)} \kappa \left[\frac{\underline{x}-\underline{x}_n}{h(\underline{x},\underline{x}_n)} \right] \right] \overset{H_1}{\underset{H_0}{\gtrless}} \lambda. \tag{28}$$

Equation (28) represents a fixed form of KDE (FKDE) [94], often called Parzen windowing [11]. It was shown in [100,101] that FKDE could be seen as a Euclidean distance detector applied in a higher dimensional kernel-induced feature space. The major impact on the performance of the FKDE detector has the selection of the kernel bandwidth h value. Numerous techniques have been proposed for its selection [94,96,98], but a unique h value that escapes over-smoothing the PDF body and simultaneously under-smoothing PDF tails may not exist. This problem motivated the development of the variable-bandwidth KDE (VKDE) [94,102]. In [95,97,99], it has been shown that VKDE achieves better background estimation in comparison with the FKDE. According to [17], there are two distinct types of variable bandwidth selection techniques: the balloon estimator (BE) and the sample point estimator (SPE). The BE varies the bandwidth for every test pixel $h(\underline{x},\underline{x}_n) = h(\underline{x})$ and the SPE varies the bandwidth at each sample data $h(\underline{x},\underline{x}_n) = h(\underline{x}_n)$.

5.3.3. Support Vector Data Description (SVDD)

Most anomaly detectors on hyperspectral images try to model or estimate the PDF of the background. However, one could instead try to directly estimate the size and shape of the background support region for a given dataset. That is the basic point of the SVDD based anomaly detector [31,103], which is basically a single-class support vector machine (SVM) classifier [104–106]. It avoids a-priori assumptions about the underlying background PDF and directly estimates the region of support for the background. SVMs are large-margin techniques that achieve good generalization of high-dimensional non-Gaussian data by directly estimating a maximum separability decision boundary [31]. SVMs showed great potential in classifying hyperspectral images [107,108], and in [31], are extended for the anomaly detection problem. The SVDD benefits for anomaly detection are listed in [31]: 1) it is nonparametric (data-driven), 2) requires a few training sam-

ples for background characterization, 3) avoids overfitting and provides good generalization, and 4) can model nontrivial multimodal distributions by applying the kernel trick. In a geometrical sense, the SVDD finds a minimum enclosing hypersphere that includes the background data in either original data space or in the high-dimensional feature space. In the former case, the linear SVDD is applied, while in the latter case, the non-linear kernel-based SVDD is used.

As the derivation of both SVDD algorithms is essentially the same, and the only change is in the non-linear mapping $\underline{\Phi}(\cdot)$, only the algorithm for the non-linear SVDD is presented. The linear SVDD can be derived by omitting the non-linear mapping function. The need for mapping in the higher dimensional space arose as the hypersphere in the original data space does not provide a tight representation of the complex distributions found in the background. A minimum enclosing hypersphere in higher dimensional feature space corresponds to a much more complex boundary in the input data space. The smallest enclosing hypersphere in the feature space $S = \left\{ \underline{\Phi}(\underline{x}) : \left\| \underline{\Phi}(\underline{x}) - \underline{c} \right\|^2 < R^2 \right\}$ that contains the set of mapped training data $T = \left\{ \underline{\Phi}(\underline{x}_i), \ i = 1, \dots, M \right\}$, where $\underline{c} = \sum_i \alpha_i \cdot \underline{\Phi}(\underline{x}_i)$ represents the center of the hypersphere that corresponds to the center of gravity of the support vectors given the optimal weights α_i. The optimal weights are scalars (Lagrange multipliers) that need to satisfy sum-to-one and non-negativity constraints. The center $\underline{c}$ and the radius R of the minimum enclosing hypersphere are determined by optimizing the Lagrangian, the optimal solution of which must satisfy Karush-Kuhn-Tucker (KKT) conditions [109]. The decision statistic for the non-linear SVDD of the test pixel $\underline{x}$ can be now formulated as [31]:

$$
\begin{aligned}
SVDD_\Phi(\underline{x}|\mathcal{B}) &= \left\| \underline{\Phi}(\underline{x}) - \underline{c} \right\|^2 \\
&= \left\| \underline{\Phi}(\underline{x}) - \sum_i^N \alpha_i \cdot \underline{\Phi}(\underline{y}_i) \right\|^2 \\
&= \underline{\Phi}(\underline{x})^T \cdot \underline{\Phi}(\underline{x}) - 2 \cdot \sum_{i=1}^N \alpha_i \cdot \underline{\Phi}(\underline{x})^T \cdot \underline{\Phi}(\underline{y}_i) + \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \cdot \underline{\Phi}(\underline{y}_i)^T \cdot \underline{\Phi}(\underline{y}_j) \underset{H_0}{\overset{H_1}{\underset{<}{>}}} R^2
\end{aligned}
\tag{29}
$$

where $\underline{y}$ denote the training data, N is the number of examples in each training set, R is the hypersphere radius, and α_i are the Lagrange multipliers. It should be noted that the optimization of Lagrangian function L with respect to α will typically end with a large fraction of α_i to become zero. The training examples (background pixels) with non-zero α_i are called support objects or support vectors. Expression (29) can be kernelized as [31]:

$$
SVDD_\Phi(\underline{x}|\mathcal{B}) = K(\underline{x}, \underline{x}) - 2 \cdot \sum_{i=1}^N \alpha_i \cdot K(\underline{x}, \underline{y}_i) + \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \cdot K(\underline{y}_i, \underline{y}_j) \underset{H_0}{\overset{H_1}{\underset{<}{>}}} R^2 \tag{30}
$$

In [31], Gaussian RBF (23) with free parameter σ was applied as the kernel function. For its estimation, the authors proposed a minimax approach that minimizes an approximate upper bound on the average false alarm rate (FAR) over the entire image [31]:

$$
\begin{aligned}
\hat{\sigma} &= \min_\sigma \frac{1}{M} \sum_{i=1}^M P_{FA_i} \\
&\approx \min_\sigma \left\{ \frac{1}{M} \sum_{i=1}^M \frac{\#SV_i}{N} \right\}
\end{aligned}
\tag{31}
$$

where P_{FA} denotes the probability of false alarm (FAR), M is the number of training sets, N is the number of examples in each training set, and $\#SV_i$ is the number of support vectors in the i-th training set.

Expressions (29) and (30) will lead to the computation of the optimal hypersphere radius for every PUT. That may prevent the distances between PUT and the hypersphere centroid from being used to compare multiple pixels to their local backgrounds mutually. Therefore, a normalized detection statistic is also proposed in [31]:

$$SVDD_{\Phi\_normalized}(\underline{x}|\mathcal{B}) = \frac{SVDD_{\Phi}(\underline{x})}{R^2} \overset{H_1}{\underset{H_0}{\gtrless}} \lambda. \tag{32}$$

In the SVDD algorithm, unknown parameters can be estimated globally or locally using the double-sliding window. The global implementation offers better computational efficiency, but the local implementation could lead to better detection performance. The kernel RX algorithm and the SVDD algorithm are related techniques but with two key differences. The kernel RX is a generative model that assumes a Gaussian distribution in the feature space, while SVDD is based on a discriminative model that avoids making such assumptions. The SVDD does not require inversion of the large covariance matrices, which is characteristic of every detector using the Mahalanobis distance (such as the kernel RX). Nevertheless, it should not be overseen that in the optimization step, SVDD requires the inversion of Gramm matrices of the training data, the size of which depends on the number of support vectors.

6. Structured Background Models

Structured background models, on the basis of a-priori knowledge, assume a specific structure of the background. In the case of the spectral data, this assumption generally arises from the physical principles of the observed data. We can presume that a pixel spectrum is a mixture of the pure spectral signatures (endmembers) of the objects or materials found on the Earth's surface. If we assume the linearity in the spectra mixing, we could also claim that a pixel spectrum actually lies in the subspace spanned by the vectors (spectra) of the unique materials found in the hyperspectral scene. Subspace models, cluster or mixture-based models, and representation-based models are the most prominent techniques for anomaly detection in hyperspectral images for remote sensing applications that utilize structured background models. Nevertheless, all listed techniques generally employ a linear mixture model (LMM) that incorporates additive noise [10]:

$$\underline{X} = \sum_{i=1}^{L} \delta_i \cdot \underline{u}_i + \underline{W}$$
$$\delta_i \geq 0, \ \sum_{i=1}^{L} \delta_i = 1 \tag{33}$$

where $\underline{W}$ denotes the model fit error (the spectrum fraction that is not modeled as a mixture or noise), $\underline{u}_i$ represents the *i*-th endmember, background eigen or basis vector, cluster or segment centroid, L is their total number, and $\delta_i, i = 1, \ldots, L$ are their respective abundances. The linear subspace model can be derived from LMM if the abundance constraints in (33) are relaxed [42].

6.1. Subspace Models

Subspace projection models determine those vectors that span the subspace of the background without explicitly defining their physical meaning. In doing so, it is possible to seek an orthogonal subspace or a signal subspace of the background.

6.1.1. Orthogonal Subspace Models

An orthogonal subspace of the background can be characterized directly or by the singular vectors of the input hyperspectral cube $\underline{X}$ [110], or indirectly by the eigenvectors of the correlation matrix $\hat{\underline{\underline{R}}} = \frac{1}{N-1} \cdot \underline{\underline{X}}^T \cdot \underline{\underline{X}}$ [91]. Linear methods such as singular value decomposition (SVD) [110] can be used to determine the singular vectors. The SVD method

performs factorization of the input hyperspectral image on two unitary matrices and a diagonal matrix of singular values. The columns of the unitary matrices are formed by the left and respectively right singular vectors of the input hyperspectral image. The singular values of the input hyperspectral image, which are found in the diagonal matrix, are sorted in descending order. It can be interpreted that the first singular values and singular vectors ($\underline{b}_i$) comprise the highest amount of "information" included in the hyperspectral image. Therefore, if the frequency of anomalies is very low in relation to the background, then they should be characterized by farther components of the SVD. Then the background can be described by the subspace spanned by the first singular vectors [11]:

$$\text{SVD} \rightarrow \underline{\underline{\Psi}} = \left[\underline{b}_1, \dots, \underline{b}_M\right]$$
$$\underline{\underline{P}} = \underline{\underline{\Psi}} \cdot \underline{\underline{\Psi}}^T \qquad . \tag{34}$$

where $\underline{\underline{\Psi}}$ refers to the matrix with first M singular vectors determined by the SVD, and $\underline{\underline{P}}$ is the projection matrix on the background subspace. The subspace projection vector $\underline{\hat{x}}$ is often called reconstruction or approximation of the $\underline{x}$ by the $\underline{\underline{\Psi}}$, and it is determined as:

$$\underline{\hat{x}} = \underline{\underline{P}} \cdot \underline{x}. \tag{35}$$

The residual of the $\underline{x}$ reconstructed with the $\underline{\hat{x}}$ is defined as:

$$\underline{r} = \underline{x} - \underline{\hat{x}} = \underline{\underline{P}}^{\perp} \cdot \underline{x}$$
$$\underline{\underline{P}}^{\perp} \cdot \underline{x} = \underline{\underline{I}} - \underline{\underline{P}} \cdot \underline{x} \tag{36}$$

For the indirect approach of the background characterization, principal component analysis [91] is commonly used. Both SVD and PCA do not necessarily result in singular vectors or eigenvectors that correspond to spectra of physical material.

In [10], the independent component analysis (ICA) [111] method is suggested to be applied instead of the PCA as it may provide spectra that are closer to physically observed ones.

Orthogonal subspace projection (OSP) inhibits the influence of the dominant background structures on the pixel spectra, which should then lead to improved detection performance of the anomaly detection techniques [36,112]. After the background impact has been suppressed by the OSP, the decision hypotheses can be set [11]:

$$\underline{\underline{P}}^{\perp} \cdot \underline{\underline{X}} | H_0 \approx \underline{\underline{P}}^{\perp} \cdot \underline{\underline{W}}$$
$$\underline{\underline{P}}^{\perp} \cdot \underline{\underline{X}} | H_1 \approx \underline{\underline{P}}^{\perp} \cdot \delta_s \cdot \underline{s} + \underline{\underline{P}}^{\perp} \cdot \underline{\underline{W}} \tag{37}$$

Then, as the detection statistic, various distance measures can be used, e.g., Mahalanobis, Euclidean, or other, combined with different spatial implementations (global, local, or quasi-local). For example, the local RX algorithm could be used on the background suppressed hyperspectral cube. Furthermore, one could use the square of the residual reconstruction vector $\underline{r}$ as the detection statistic, which is sometimes called distance from the feature space [82]:

$$D_{DFFS}(\underline{x}|\mathcal{B}) = \left\|\underline{r}\right\|^2 = \underline{x}^T \cdot \underline{\underline{Q}} \cdot \underline{x} = \underline{x}^T \cdot (\underline{\underline{P}}^{\perp})^T \cdot (\underline{\underline{P}}^{\perp}) \cdot \underline{x} \underset{H_0}{\overset{H_1}{\underset{<}{\gtrless}}} \lambda \tag{38}$$

The performance of the OSP techniques primarily depends on the quality of the background reconstruction. There should not be leakage from the target space to the background subspace [10], as it will surely impair the detection performance. Both PCA and SVD techniques are susceptible to that issue, so removing target-like pixels from the input

data is beneficial before applying the PCA or SVD. That problem motivated the development of techniques that are more robust and less sensitive to an outlier presence; these are presented in the section on representation-based models.

6.1.2. Signal Subspace Models

In [113], the anomaly detection method that works within a subspace of the original signal space is presented. The method is called the signal subspace processing anomaly detector (SSPAD), and it operates within a subspace spanned by the spectral vectors in the immediate vicinity of the PUT. SSPAD is able to detect local anomalies without requiring covariance estimation or calculation of its inverse. SSPAD determines the finite impulse response (FIR) filter coefficients for the neighboring spectral vectors that minimize the mean square error of the PUT reconstruction. For selecting the signal subspace, in [113], a guard window is implemented, and then four boxes centered at horizontal and vertical axes originating from the PUT are proposed. The boxes are placed just outside of the guard window to prevent the projection of the target spectra to the signal subspace. The pixels contained in a specific box are the inputs for the FIR filter. The SSPAD detection statistic in the matrix form is derived in [11]:

$$
\begin{aligned}
SSPAD(\underline{x}|\mathcal{B}) &= \min_{k=1,2,3,4}\{\|\underline{e}_k\|\} \\
&= \min_{k=1,2,3,4}\{\|\underline{\underline{P}}_k^\perp \cdot \underline{x}\|\} \\
&= \min_{k=1,2,3,4}\left\{\left\|\left(\underline{\underline{I}} - \underline{\underline{V}}_k \cdot \underline{\underline{V}}_k^\#\right) \cdot \underline{x}\right\|\right\} \underset{H_0}{\overset{H_1}{\underset{<}{\gtrless}}} \lambda
\end{aligned}
\tag{39}
$$

where $\underline{\underline{P}}_k^\perp$ is the projection matrix onto the subspace orthogonal to the signal subspace spanned by the spectral vectors of the k-th box, $\underline{\underline{V}}_k$ denotes the matrix containing those vectors, and $\underline{\underline{V}}_k^\#$ is the pseudoinverse of the $\underline{\underline{V}}_k$.

The anomalies in the SVDD are the pixels that significantly deviate from the local background determined by the four FIR filters. In the algorithm implementation, care should be exercised regarding the appropriate size of the guard window and the signal subspace boxes. In order to improve computational efficiency, the box size should be small, and also, it is important not to employ too many constraints on the projection [11]. If the anomalies are spatially grouped and co-aligned in the vertical or horizontal direction, this could lead to bad SSPAD detection performance. In that case, a different strategy for a spatial sampling of the signal subspace samples should be applied.

6.2. Cluster or Mixture-based Models

Cluster or mixture-based models aim to directly solve the expression (33). They seek the background endmembers and their abundances in the pixel spectrum. As they allow a pixel to be a mixture of the spectra, they are adapted to detect unresolved or sub-pixel targets (anomalies). A plethora of automatic endmember extraction and respective abundance estimation techniques have been developed [114], such as N-FINDR [115] or OASIS [116]. Enforcing the constraints in the LMM (33) leads to the convex hull model (CHM) [117] that provides physically related endmembers [10]. Cluster or mixture models conform to the convex hull model. The detection statistic can be derived as the distance of PUT spectral vector $\underline{x}$ from a convex hull of the endmembers [82]:

$$
D_{CHM}(\underline{x}|\mathcal{B}) = \left\|\underline{x} - \underline{\underline{U}} \cdot \underline{\delta}\right\|^2 \underset{H_0}{\overset{H_1}{\underset{<}{\gtrless}}} \lambda
\tag{40}
$$

where vector $\underline{\delta}$ is the least-squares solution of the equation $\underline{\underline{U}} \cdot \underline{\delta} = \underline{x}$ with non-negativity and sum-to-one constraints:

$$0 \leq \delta_i \leq 1$$
$$\sum \delta_i = 1 \tag{41}$$

and matrix $\underline{\underline{U}}$ is composed of the background endmembers u_i: $\underline{\underline{U}} = [\underline{u}_1 \cdots \underline{u}_L]$.

Besides the presented approach, anomaly detection in the LMM sense can be carried out if the background convex hull (determined by the endmembers) is first subtracted from the input hyperspectral image, and some of the RX variants are then run on the residual image.

Considering that the global or local Gaussian models (RX detectors) have shown inadequate modeling nonhomogeneous backgrounds [11], image clustering or the use of the more complex models is suggested [82]. Therefore, models based on Gaussian-mixture and cluster or segmentation anomaly detection have been developed.

6.2.1. Gaussian-Mixture Model

A more complex Gaussian-mixture model (GMM) can more closely describe the non-homogeneous backgrounds (i.e., model the presence of different materials in the scene) [15,35]. The basic idea of the GMM is as follows: the scene is divided into a set of mixtures, each of which follows a Gaussian distribution. Suppose that there are *L* Gaussian distributions on the scene (*L* mixtures), then the probability of the occurrence of each pixel can be shown as the weighted sum of the probabilities of those distributions:

$$p_{GMM}(\underline{x}|\mathcal{B}) = \sum_{i=1}^{L} w_i p_{\mathcal{N}}(\underline{x}|\underline{\mu}_i, \underline{\underline{\Gamma}}_i) \tag{42}$$

where $\mathcal{B} = \left\{ w_i, \underline{\mu}_i, \underline{\underline{\Gamma}}_i \right\}_{i=1}^{L}$.

In the case of a multivariate Gaussian distribution, the unknown parameters $\underline{\mu}_i$ and $\underline{\underline{\Gamma}}_i$ are determined directly. In the GMM, the determination of w_i, $\underline{\mu}_i$, and $\underline{\underline{\Gamma}}_i$ is done using algorithms such as expectation-maximization (EM) or some variants thereof, such as stochastic expectation-maximization (SEM) [118,119] or classification expectation-maximization [119]. We consider anomalies to be those hyperspectral pixels that have a low probability of occurrence. Therefore, the detector statistic can be formulated as [82]:

$$D_{GMM}(\underline{x}|\mathcal{B}) = \sum_{i=1}^{L} w_i p_{\mathcal{N}}(\underline{x}|\underline{\mu}_i, \underline{\underline{\Gamma}}_i) \underset{H_0}{\overset{H_1}{\underset{>}{\lessgtr}}} \lambda \tag{43}$$

It should be noticed that the inequality signs in the expression (43) are oppositely oriented concerning other detectors. The weights w_i are usually determined as the a-priori probabilities of the Gaussian distributions found in the background and can be determined, for example, by using SEM. Due to the introduction of the weighting coefficient w_i The GMM is less sensitive to overestimating the number of distributions found in the background than cluster or segmentation-based methods. That is because the mixtures with a small number of elements will also have a low weight value w_i. GGMs are global detectors, and as they do not use a sliding window, they can detect targets of any size or shape. However, for the GMM to achieve successful results, the frequency of occurrence of targets needs to be low relative to the background (so that the targets do not create a separate class), and the anomaly spectra need to be significantly different from the background.

6.2.2. Cluster or Segmentation Based Models

Cluster [35] or segmentation [66,120] based anomaly detection methods generally perform unsupervised clustering (classification) or automatic segmentation of the hyperspectral image and then analyze the spectral distances from PUT to cluster or segment centroids. In the unsupervised clustering step, hard clustering techniques like K-means [121] or soft clustering like Fuzzy C-means are used [122–124]. Hard clustering techniques assign an integer value that corresponds to the membership of a pixel to a certain cluster, while soft clustering methods assign a value m_{kl} for the membership of the observed pixel to each of the clusters $m_{kl} \in [0,1]$. Of soft clustering methods, fuzzy-based techniques [125,126] may show potential for anomaly detection applications. Once the cluster statistics have been determined (usually cluster mean vector and cluster covariance matrix), a distance can be computed, which can be, for example, the Euclidean or Mahalanobis distance between the PUT and each cluster centroid. As the detection statistic, the minimum distance (distance to the spectrally closest cluster centroid) is used. If the squared Mahalanobis distance is chosen, it will lead to the so-called class-conditional GLRT [11]:

$$D_{CBAD}\left(\underline{x}|\mathcal{B}\right) = -\log\left\{p_{\mathcal{N}}(\underline{x}|\hat{\underline{\mu}}_i,\hat{\underline{\underline{\Gamma}}}_i)\right\} \underset{H_0}{\overset{H_1}{\underset{<}{>}}} \lambda$$

$$\sim \left(\underline{x} - \hat{\underline{\mu}}_i\right)^T \cdot \hat{\underline{\underline{\Gamma}}}_i^{-1} \cdot \left(\underline{x} - \hat{\underline{\mu}}_i\right) \underset{H_0}{\overset{H_1}{\underset{<}{>}}} \lambda \tag{43}$$

where *i* denotes the index of the cluster to which the PUT $\underline{x}$ has been assigned (the spectrally nearest cluster).

The performance of these detectors strictly depends on the selection of the appropriate number of background clusters [11]. If this number is underestimated, then pixels that naturally belong to a larger number of clusters will be "squeezed" into one cluster. That will result in a higher variance of the resulting clusters, which will adversely affect the probability of anomaly detection. If the number of background clusters is overestimated, there is a possibility that anomalies will create a separate class. In that case, it may not even be possible to detect anomalies. In order to determine the optimal number of clusters, Akaike or Bayes information criteria were proposed [127]. The application of artificial neural networks such as self-organizing fields (self-organizing maps) [128] has been investigated in hyperspectral anomaly detection [129,130], and is less sensitive to the selection of the appropriate number of clusters.

6.3. Representation-based Models

The rising paradigm of compressed sensing [131–133] propelled the popularity of representation-based anomaly and target detection [27,40,134–139]. Consider having a set of $n \ll N$ labeled samples (pixels) $\underline{D} = \{\underline{x}_i\}_{i=1}^{n} \in \mathbb{R}^{K \times n}$ where *K* refers to number of spectral bands. In the following expressions, input HS image ($\underline{X}$) is regarded as a $K \times N$ matrix ($\underline{X} \in \mathbb{R}^{K \times N}$) which is a transpose of the definition set in section 4 ($\underline{X} \in \mathbb{R}^{N \times K}$). A dataset with all available samples $\underline{D}$ is usually called a dictionary, which is constructed of atoms (labeled pixels). The labeled datasets are not available in anomaly detection problems, but they can be reconstructed from the data [140–142].

The idea of decomposing the hyperspectral image into a low-rank background matrix and a sparse anomaly matrix was implemented in the robust principal component analysis (RPCA) [138,143,144]. This model did not account for the presence of noise in the input dataset, which adversely impacted detection performance. An improvement over RPCA was achieved in [145,146], where an additional noise factor was implemented in the low-rank and sparse matrix decomposition (LRaSMD) algorithm. The low-rank representation (LRR) [138,147,148] allows the background reconstruction using the multiple

subspaces (unlike RPCA [149]) and therefore needs a dictionary to separate the anomalies from the background. LRR enables global background characterization, as it finds the lowest rank representation of all the hyperspectral pixels simultaneously. The adequacy of the LRR model for hyperspectral data modeling is nicely outlined in [142]: background pixels can be adequately represented as linear combinations of endmembers (described in a subspace), and anomalies are spatially sparse [146]. The LLR model for anomaly detection can be formulated as [142]:

$$\min_{\underline{S},\underline{E}} \left\| \underline{S} \right\|_* + \lambda \left\| \underline{E} \right\|_{2,1} \text{ such that } \underline{X} = \underline{D} \cdot \underline{S} + \underline{E}, \tag{44}$$

where the HS image $\underline{X}$ is decomposed into a background $\underline{D} \cdot \underline{S}$ and an anomaly component $\underline{E} = [\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_N]$. $\underline{D}$ represents the dictionary, and $\underline{S}$ contains the representation coefficients. $\| \quad \|_*$ is a nuclear norm, which is a good substitute for the rank function used in the original LRR model [138] because of the convex optimization problem it causes. A tradeoff parameter $\lambda > 0$ is used to balance the effects of the background and anomaly part. $\| \quad \|_{2,1}$ is the $\ell_{2,1}$ norm defined as the sum of ℓ_2 norm of the columns of the matrix:

$$\left\| \underline{E} \right\|_{2,1} = \sum_{i=1}^{N} \sqrt{\underline{e}_i^T \underline{e}_i} \tag{45}$$

where $\underline{e}_i$ refers to a column of the matrix $\underline{E}$. The role of the $\ell_{2,1}$ norm is to encourage the columns of $\underline{E}$ to be zero, indicating that anomalies are column-wise sparse or "sample-specific". Niu and Wang [150] show the hyperspectral AD based on the LRR and learned dictionary: LLRaLD AD. They propose using the basic detectors (such as global RX) on the sparse matrix for the detection statistic. An approach that implements the spatial similarity between pixels in local regions is displayed in the work of Tan et al. [147]. They suggest incorporating spatial constraints in the detection model to improve the detection performance of the LRR model. Blind source component separation by unmixing was presented by Wang et al. [151] to identify anomalous components.

The sparse representation model assumes that a hyperspectral signal (pixel) can be adequately represented by a sparse linear combination of dictionary atoms, i.e., a pixel can be reconstructed by only a few atoms [135]. The PUT is sparsely represented using the ℓ_0 or ℓ_1 norm regularization. The main goal of the sparse representation is to determine the sparse (weight) vector $\underline{\alpha}_{SR}$ such that $\left\| \underline{x} - \underline{D} \cdot \underline{\alpha}_{SR} \right\|_2$ is minimized while $\left\| \underline{\alpha}_{SR} \right\|_l \leq L_0$ [134]. Namely, vector $\underline{\alpha}_{SR}$ can be determined by solving the optimization problem:

$$\underline{\alpha}_{SR} = \arg\min_{\underline{\alpha}_{SR}} \left\| \underline{x} - \underline{D} \cdot \underline{\alpha}_{SR} \right\|_2 \text{ such that } \left\| \underline{\alpha}_{SR} \right\|_l \leq L_0, l \in \{0,1\} \tag{46}$$

where L_0 is the constant (regularization) parameter that balances the $\underline{x}$ reconstruction error and the sparsity of $\underline{\alpha}_{SR}$, and l represents the choice of ℓ_0 or ℓ_1 norm. The listed optimization problem (46) is NP-hard if the ℓ_0-norm is applied; it can be solved by greedy-pursuit based algorithms such as the orthogonal matching pursuit [152] or a subspace pursuit [153]. If ℓ_1-norm is used, the $\underline{\alpha}_{SR}$ can be determined by convex relaxation algorithms such as in [154–156], or by pursuit methods such as basis pursuit [157] or basis pursuit denoising [158]. Li et al. [134] proposed the AD based on the joint sparse representation (JSR) framework. It uses a dual sliding window approach to estimate an active dictionary, and the detection statistic is determined by the length of the matched projection on the orthogonal complementary background subspace (estimated by the JSR). An adaptive weighted sparse representation and background estimation-based AD is presented in [145]. It uses the endmember extraction method to characterize their spectra and respective abundances. The sparse representation was adaptively weighted on both global and local domains, and the detection statistic was determined as the residual of PUT reconstructed by the background dictionary. Discriminative feature learning with

multiple-dictionary was introduced in the work of [159]. The detection statistic is designed with a global multiple-view AD strategy that incorporates multiple use of global RX detectors that are mutually fused to produce the final result. An innovative sparsity score estimation was presented in [160] that implements atom usage probability, which helps in improving the discriminative power of the background dictionary. Xu et al. [138] proposed the combined use of LRR and sparse representation based on the separation of the background and the anomalies in the observed data. LRR is used to model the background, and the sparsity-inducing regularization term is introduced to the representation coefficients, which enables the description of global and local structures of the hyperspectral dataset. The final detection statistic is determined by the response of the residual matrix.

Collaborative representation techniques take the opposite approach of sparse representation methods. "Collaborative representation means that all atoms 'collaborate' in the representation of a single pixel, and each atom has an equal chance to participate in the representation." [135]. The goal is to find the weight vector $\underline{\alpha}_{CR}$ such that $\left\| \underline{x} - \underline{\underline{D}} \cdot \underline{\alpha}_{CR} \right\|_2^2$ is minimized under the constraint that $\left\| \underline{\alpha}_{CR} \right\|_2^2$ is minimized, too. This can be formalized as:

$$\hat{\underline{\alpha}}_{CR} = \arg\min_{\underline{\alpha}_{CR}} \left\| \underline{x} - \underline{\underline{D}} \cdot \underline{\alpha}_{CR} \right\|_2^2 + \lambda \cdot \left\| \underline{\alpha}_{CR} \right\|_2^2, \tag{47}$$

where λ is the regularization parameter that controls the penalty of the weight vector $\underline{\alpha}_{CR}$ ℓ_2-norm. Additionally, some authors [78,161] suggested using a distance-weighted Tikhonov regularization besides the parameter λ. The general detection statistic in collaborative representation ADs is the reconstruction error of the PUT.

Equation (47) can also be expressed as [27]:

$$\hat{\underline{\alpha}}_{CR} = \arg\min_{\underline{\alpha}_{CR}} \left[\underline{\alpha}_{CR}^T \cdot \left(\underline{\underline{D}}^T \cdot \underline{\underline{D}} + \lambda \cdot \underline{\underline{I}} \right) \cdot \underline{\alpha}_{CR} - 2 \cdot \underline{\alpha}_{CR}^T \cdot \underline{\underline{D}}^T \cdot \underline{x} \right]. \tag{48}$$

Taking the derivative of (48) with respect to $\underline{\alpha}_{CR}$ and setting it to zero returns [27]:

$$\hat{\underline{\alpha}}_{CR} = \left(\underline{\underline{D}}^T \cdot \underline{\underline{D}} + \lambda \cdot \underline{\underline{I}} \right)^{-1} \cdot \underline{\underline{D}}^T \cdot \underline{x}. \tag{49}$$

By comparing (49) with the sparse representation model, it can be seen that collaboration-representation has a much lower computational cost as it offers the solution in a closed form. If the dictionary $\underline{\underline{D}}$ is determined locally, then expressions (46–49) should be replaced by the adaptive form $\underline{\underline{D}_j}$. Li et al. [27] proposed the algorithm based on the concept that the background can be adequately represented by its spatial neighbors, but not by anomalies. They implemented a collaborative representation model, but for a detection statistic used a projection to a higher dimensional space and the use of the kernel trick. Low rank and collaborative representation AD (LRCRD) was presented in [162]. It divides the image into two parts: the background represented with the respective dictionary whose coefficients are constrained by low-rank and ℓ_2-norm minimization, and the sparse anomaly part defined as the residual matrix constrained by $\ell_{2,1}$-norm minimization.

7. Conclusions

In this review paper, the development of hyperspectral image processing for anomaly detection in remote sensing applications was presented. In the first part of the paper, scientific research trends were presented through a bibliometric analysis. The most relevant journals, authors, and their contributions were identified, and the expansion of the field was analyzed by title and author keywords. Although the documents used in this research were mostly published in the last 20 years, as the hyperspectral imaging technology is quite recent, the oldest references date back to the 1930s.

This analysis provided the foundations for the second part of the paper, in which the overview of the mathematical framework for anomaly detection on hyperspectral images was presented. Developed anomaly detection techniques were generally classified as the methods that presume an unstructured background model or, conversely, a structured background model. The latter assumes a specific background structure, while the former method does not state any a-priori assumptions about it. A plethora of innovative concepts and ideas were applied to the anomaly detection problem on hyperspectral images: generative approaches, nonlinear mapping and kernel trick, orthogonal and signal subspace projections, and representation approaches such as sparse, collaborative, or joint ones.

No doubt, every one of these approaches has positive and negative sides that could make one detector excel in some specific scenarios, but unfortunately, might not maintain the same detection performance if the application circumstances change. There is debate as to whether the best hyperspectral anomaly detector in remote sensing applications exists, but no such statement can be made. Regardless of the underlying concept, the main problem in anomaly detection performance assessment arises from the fact that the detectors are judged on the basis of a small number of experimental scenarios. The scarcity of reference hyperspectral datasets for detection performance evaluation is still an issue. A rich collection of hyperspectral images of various natural scenes would contribute more statistical significance to the comparative results of anomaly detector performance.

The specific problems in evaluating detector performance come from the sole nature of the hyperspectral images of natural scenes: they reside in highly dimensional spectral space that exhibits high spatial non-stationarity. Due to its high dimensionality, hyperspectral data processing delivers a heavy computational burden. Hence, many authors have tackled the problem of reducing or optimizing the computational complexity and real-time processing of hyperspectral images for detection purposes. Real-time processing may refer to processing the data in an online fashion or the ability to deliver detection results in real-time. These issues are still open and have attracted the scientific community's attention, as shown in recent publications focused on representation-based techniques and the implementation of neural network based approaches [163,164]. Future research in the field may continue to develop these approaches and techniques, as they offer a balance between detection performance and computational complexity. Nevertheless, hyperspectral image processing for anomaly detection in remote sensing applications is still an exceptionally worthy field of research, as the hyperspectral data carry an abundance of valuable information that may be useful in a wide variety of applications.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Adao, T.; Hruska, J.; Padua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry. *Remote Sens.* **2017**, *9*, 1110, doi:10.3390/rs9111110.
2. Gowen, A.A.; O'Donnell, C.P.; Cullen, P.J.; Downey, G.; Frias, J.M. Hyperspectral imaging—An emerging process analytical tool for food quality and safety control. *Trends Food Sci. Technol.* **2007**, *18*, 590–598, doi:10.1016/j.tifs.2007.06.001.
3. Bajić, M. Modeling and Simulation of Very High Spatial Resolution UXOs and Landmines in a Hyperspectral Scene for UAV Survey. *Remote Sens.* **2021**, *13*, 837.
4. Krtalić, A.; Bajić, M. Development of the TIRAMISU Advanced Intelligence Decision Support System. *Eur. J. Remote Sens.* **2019**, *52*, 40–55, doi:10.1080/22797254.2018.1550351.
5. Lu, G.L.; Fei, B.W. Medical hyperspectral imaging: A review. *J. Biomed. Opt.* **2014**, *19*, doi:10.1117/1.Jbo.19.1.010901.
6. Eismann, M.T.; Stocker, A.D.; Nasrabadi, N.M. Automated Hyperspectral Cueing for Civilian Search and Rescue. *Proc. IEEE* **2009**, *97*, 1031–1055, doi:10.1109/jproc.2009.2013561.
7. Krtalić, A.; Bajić, M.; Ivelja, T.; Racetin, I. The AIDSS Module for Data Acquisition in Crisis Situations and Environmental Protection. *Sensors* **2020**, *20*, 1267.
8. Govender, M.; Chetty, K.; Bulcock, H. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water SA* **2007**, *33*, 145–151.
9. Manolakis, D.; Shaw, G. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Process. Mag.* **2002**, *19*, 29–43, doi:10.1109/79.974724.
10. Manolakis, D. Taxonomy of detection algorithms for hyperspectral imaging applications. *Opt. Eng.* **2005**, *44*, 1–11, doi:10.1117/1.1930927.
11. Matteoli, S.; Diani, M.; Corsini, G. A tutorial overview of anomaly detection in hyperspectral images. *IEEE Aerosp. Electron. Syst. Mag.* **2010**, *25*, 5–27, doi:10.1109/MAES.2010.5546306.
12. Elachi, C.; Van Zyl, J.J. *Introduction to the Physics and Techniques of Remote Sensing*; John Wiley & Sons: Hoboken, NJ, USA, 2006; Volume 28.
13. Schowengerdt, R.A. *Remote Sensing: Models and Methods for Image Processing*; Academic Press: San Diego, CA, USA, 2006.
14. Alpaydin, E. *Introduction to Machine Learning*; The MIT Press: Cambridge, MA, USA, 2010.
15. Stein, D.W.J.; Beaven, S.G.; Hoff, L.E.; Winter, E.M.; Schaum, A.P.; Stocker, A.D. Anomaly detection from hyperspectral imagery. *IEEE Signal Process. Mag.* **2002**, *19*, 58–69, doi:10.1109/79.974730.
16. Huck, A.; Guillaume, M. A CFAR algorithm for anomaly detection and discrimination in hyperspectral images. In Proceedings of the 2008 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 1868–1871.
17. Matteoli, S.; Diani, M.; Theiler, J. An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2317–2336, doi:10.1109/jstars.2014.2315772.
18. Nasrabadi, N.M. Hyperspectral target detection : An overview of current and future challenges. *IEEE Signal Process. Mag.* **2014**, *31*, 34–44, doi:10.1109/MSP.2013.2278992.
19. Falagas, M.E.; Pitsouni, E.I.; Malietzis, G.A.; Pappas, G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *FASEB J.* **2008**, *22*, 338–342, doi:10.1096/fj.07-9492LSF.
20. Mongeon, P.; Paul-Hus, A. The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics* **2016**, *106*, 213–228, doi:10.1007/s11192-015-1765-5.
21. Elsevier. Scopus Content Coverage Guide. Available online: https://www.elsevier.com/__data/assets/pdf_file/0007/69451/Scopus_ContentCoverage_Guide_WEB.pdf (accessed on 24 April 2021).
22. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Altman, D.; Antes, G.; Atkins, D.; Barbour, V.; Barrowman, N.; Berlin, J.A.; et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Ann. Intern. Med.* **2009**, *151*, 264–269, doi:10.7326/0003-4819-151-4-200908180-00135.
23. Aria, M.; Cuccurullo, C. bibliometrix: An R-tool for comprehensive science mapping analysis. *J. Informetr.* **2017**, *11*, 959–975, doi:10.1016/j.joi.2017.08.007.
24. Elango, B.; Rajendran, P. Authorship trends and collaboration pattern in the marine sciences literature: A scientometric study. *Int. J. Inf. Dissem. Technol.* **2012**, *2*, 166–169.
25. Bradford, S.C. Sources of information on specific subjects. *Engineering* **1934**, *137*, 85–86.
26. Kwon, H.; Nasrabadi, N.M. Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 388–397, doi:10.1109/TGRS.2004.841487.
27. Li, W.; Du, Q. Collaborative representation for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1463–1474, doi:10.1109/tgrs.2014.2343955.
28. Chang, C.I.; Chiang, S.S. Anomaly detection and classification for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 1314–1325, doi:10.1109/TGRS.2002.800280.
29. Ren, H.; Chang, C.I. Automatic spectral target recognition in hyperspectral imagery. *IEEE Trans. Aerosp. Electron. Syst.* **2003**, *39*, 1232–1249, doi:10.1109/taes.2003.1261124.
30. Du, Q.; Fowler, J.E. Hyperspectral image compression using JPEG2000 and principal component analysis. *IEEE Geosci. Remote Sens. Lett.* **2007**, *4*, 201–205, doi:10.1109/lgrs.2006.888109.

31. Banerjee, A.; Burlina, P.; Diehl, C. A support vector method for anomaly detection in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2282–2291, doi:10.1109/TGRS.2006.873019.

32. Penna, B.; Tillo, T.; Magli, E.; Olmo, G. Transform coding techniques for lossy hyperspectral data compression. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1408–1421, doi:10.1109/tgrs.2007.894565.

33. Du, B.; Zhang, L. A discriminative metric learning based anomaly detection method. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6844–6857, doi:10.1109/tgrs.2014.2303895.

34. Reed, I.S.; Yu, X. Adaptive Multiple-Band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution. *IEEE Trans. Acoust. Speech Sign. Proces.* **1990**, *38*, 1760–1770, doi:10.1109/29.60107.

35. Carlotto, M.J. A cluster-based approach for detecting man-made objects and changes in imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 374–387, doi:10.1109/TGRS.2004.841481.

36. Harsanyi, J.C.; Chang, C.I. Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*, 779–785, doi:10.1109/36.298007.

37. Kerekes, J. Receiver operating characteristic curve confidence intervals and regions. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 251–255, doi:10.1109/lgrs.2008.915928.

38. Manolakis, D.; Marden, D.; Shaw, G.A. Hyperspectral Image Processing for Automatic Target Detection Applications. *Linc. Lab. J.* **2003**, *14*, 79–116.

39. Nasrabadi, N.M. Regularization for spectral matched filter and RX anomaly detector. *SPIE Int. Soc. Opt. Eng.* **2008**, doi:10.1117/12.773444.

40. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Sparse representation for target detection in hyperspectral imagery. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 629–640, doi:10.1109/jstsp.2011.2113170.

41. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3973–3985, doi:10.1109/TGRS.2011.2129595.

42. Kay, S.M. *Fundamentals of Statistical Signal Processing: Detection Theory*; Prentice Hall: Hoboken, NJ, USA, 1998; p. 998.

43. Neyman, J.; Pearson, E.S.; Pearson, K. IX. On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Character* **1933**, *231*, 289–337, doi:10.1098/rsta.1933.0009.

44. Mahalanobis, P.C. *On the Generalised Distance in Statistics*; National Institute of Sciences: Calcutta, India, 1936; pp. 49–55.

45. Veracini, T.; Matteoli, S.; Diani, M.; Corsini, G. An anomaly detection architecture based on a data-adaptive density estimation. In Proceedings of the 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lisbon, Portugal, 6–9 June 2011, doi:10.1109/whispers.2011.6080919.

46. Ma, N.; Peng, Y.; Wang, S.; Leong, P.H.W. An unsupervised deep hyperspectral anomaly detector. *Sensors* **2018**, *18*, 693, doi:10.3390/s18030693.

47. Su, H.; Wu, Z.; Du, Q.; Du, P. Hyperspectral Anomaly Detection Using Collaborative Representation with Outlier Removal. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 5029–5038, doi:10.1109/jstars.2018.2880749.

48. Li, F.; Zhang, X.; Zhang, L.; Jiang, D.; Zhang, Y. Exploiting Structured Sparsity for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4050–4064, doi:10.1109/tgrs.2018.2821168.

49. Taghipour, A.; Ghassemian, H. Hyperspectral anomaly detection using spectral–spatial features based on the human visual system. *Int. J. Remote Sens.* **2019**, *40*, 8683–8704, doi:10.1080/01431161.2019.1620374.

50. Kelly, E.J. An Adaptive Detection Algorithm. *IEEE Trans. Aerosp. Electron. Syst.* **1986**, *AES-22*, 115–127, doi:10.1109/taes.1986.310745.

51. Hunt, B.R.; Cannon, T.M. Nonstationary assumptions for gaussian models of images. *IEEE Trans. Syst. Man. Cybern.* **1976**, *SCM-6*, 876–882.

52. Margalit, A.; Reed, I.S.; Gagliardi, R.M. Adaptive Optical Target Detection Using Correlated Images. *IEEE Trans. Aerosp. Electron. Syst.* **1985**, *AES-21*, 394–405, doi:10.1109/taes.1985.310570.

53. Chen, J.Y.; Reed, I.S. A Detection Algorithm for Optical Targets in Clutter. *IEEE Trans. Aerosp. Electron. Syst.* **1987**, *AES-23*, 46–59, doi:10.1109/taes.1987.313335.

54. Swain, P.H.; Davis, S.M. Remote sensing: The quantitative approach. *IEEE Trans. Pattern Anal. Mach. Intell.* **1981**, 713-714.

55. Molero, J.M.; Garzon, E.M.; Garcia, I.; Plaza, A. Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 801–814, doi:10.1109/JSTARS.2013.2238609.

56. Molero, J.M.; Garzón, E.M.; García, I.; Plaza, A. Anomaly detection based on a parallel kernel RX algorithm for multicore platforms. *J. Appl. Remote Sens.* **2012**, *6*, doi:10.1117/1.JRS.6.061503.

57. Molero, J.M.; Garzon, E.M.; Garcia, I.; Quintana-Orti, E.S.; Plaza, A. Efficient implementation of hyperspectral anomaly detection techniques on GPUs and multicore processors. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2256–2266, doi:10.1109/jstars.2014.2328614.

58. Molero, J.M.; Paz, A.; Garzón, E.M.; Martínez, J.A.; Plaza, A.; García, I. Fast anomaly detection in hyperspectral images with RX method on heterogeneous clusters. *J. Supercomput.* **2011**, *58*, 411–419, doi:10.1007/s11227-011-0598-0.

59. Manolakis, D.; Marden, D. Non Gaussian models for hyperspectral algorithm design and assessment. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Toronto, ON, Canada, 24–28 June 2002; pp. 1664–1666.

60. Marden, D.B.; Manolakis, D. Modeling Hyperspectral Imaging Data. In Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX, Orlando, FL, USA, 23 September 2003; pp. 253–262.

61. Niu, S.; Ingle, V.K.; Manolakis, D.; Cooley, T. On the modeling of hyperspectral imaging data with elliptically contoured distributions. In Proceedings of the 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Reykjavik, Iceland, 14–16 June 2010; pp. 1–4.

62. Caefer, C.E.; Silverman, J.; Orthal, O.; Antonelli, D.; Sharoni, Y.; Rotman, S.R. Improved covariance matrices for point target detection in hyperspectral data. *Opt. Eng.* **2008**, *47*, doi:10.1117/1.2965814.

63. Matteoli, S.; Diani, M.; Corsini, G. Improved covariance matrix estimation: Interpretation and experimental analysis of different approaches for anomaly detection applications. In Proceedings of the Image and Signal Processing for Remote Sensing XV, Berlin, Germany, 28 September 2009, doi:10.1117/12.830445.

64. Gorelik, N.; Blumberg, D.; Rotman, S.R.; Borghys, D. Nonsingular approximations for a singular covariance matrix. In Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, Baltimore, MD, USA, 24 May 2012, doi:10.1117/12.915310.

65. Huber-Lerner, M.; Hadar, O.; Rotman, S.R.; Huber-Shalem, R. Compression of hyperspectral images containing a subpixel target. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2246–2255, doi:10.1109/jstars.2014.2320754.

66. Borghys, D.; Kasen, I.; Achard, V.; Perneel, C. Comparative evaluation of hyperspectral anomaly detectors in different types of background. In Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, Baltimore, MD, USA, 24 May 2012, doi:10.1117/12.920387.

67. Friedman, J.H. Regularized discriminant analysis. *J. Am. Stat. Assoc.* **1989**, *84*, 165–175, doi:10.1080/01621459.1989.10478752.

68. Hoffbeck, J.P.; Landgrebe, D.A. Covariance matrix estimation and classification with limited training data. *IEEE Trans. Pattern Anal. Mach. Intell.* **1996**, *18*, 763–767, doi:10.1109/34.506799.

69. Kuo, B.C.; Landgrebe, D.A. A covariance estimator for small sample size classification problems and its application to feature extraction. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 814–819, doi:10.1109/TGRS.2002.1006358.

70. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2001.

71. Manolakis, D.; Marden, D.; Kerekes, J.; Shaw, G. On the statistics of hyperspectral imaging data. In Proceedings of the Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII, Orlando, FL, USA, 20 August 2001; pp. 308–316, doi:10.1117/12.437021.

72. Hansen, P.C. *Rank-Deficient and Discrete Ill-Posed Problems*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1999.

73. Theiler, J. The incredible shrinking covariance estimator. In Proceedings of the Automatic Target Recognition XXII, Baltimore, MD, USA, 2 May 2012, doi:10.1117/12.918718.

74. Davidson, C.E.; Ben-David, A. On the use of covariance and correlation matrices in hyperspectral detection. In Proceedings of the 2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 11–13 October 2011, doi:10.1109/aipr.2011.6176374.

75. Rossi, A.; Acito, N.; Diani, M.; Corsini, G. RX architectures for real-time anomaly detection in hyperspectral images. *J. Real-Time Image Process.* **2014**, *9*, 503–517, doi:10.1007/s11554-012-0292-3.

76. Zhao, C.; Wang, Y.; Qi, B.; Wang, J. Global and local real-time anomaly detectors for hyperspectral remote sensing imagery. *Remote Sens.* **2015**, *7*, 3966–3985, doi:10.3390/rs70403966.

77. Stellman, C.M.; Hazel, G.G.; Bucholtz, F.; Michalowicz, J.V.; Stocker, A.; Schaaf, W. Real-time hyperspectral detection and cuing. *Opt. Eng.* **2000**, *39*, 1928–1935, doi:10.1117/1.602577.

78. Liu, W.M.; Chang, C.I. Multiple-window anomaly detection for hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 644–658, doi:10.1109/JSTARS.2013.2239959.

79. Ren, L.; Zhao, L.; Wang, Y. A Superpixel-Based Dual Window RX for Hyperspectral Anomaly Detection. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1233–1237, doi:10.1109/lgrs.2019.2942949.

80. Hu, X.; Hu, S.; Zhang, X.; Zhang, H.; Luo, L. Anomaly Detection Based on Local Nearest Neighbor Distance Descriptor in Crowded Scenes. *Sci. World J.* **2014**, *2014*, 632575, doi:10.1155/2014/632575.

81. Ming, Z.; Jingchao, C.; Yang, L. A Review of Anomaly Detection Techniques Based on Nearest Neighbor. In Proceedings of the 2018 International Conference on Computer Modeling, Simulation and Algorithm (CMSA 2018), Beijing, China, 22–23 April 2018; pp. 290–292, doi:10.2991/cmsa-18.2018.65.

82. Ahlberg, J.; Renhorn, I. *Multi- and Hyperspectral Target and Anomaly Detection*; Swedish Defence Research Agency, Division of Sensor Technology: Linköping, Sweden, 2004.

83. Kruse, F.A.; Lefkoff, A.B.; Boardman, J.W.; Heidebrecht, K.B.; Shapiro, A.T.; Barloon, P.J.; Goetz, A.F.H. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sens. Environ.* **1993**, *44*, 145–163, doi:10.1016/0034-4257(93)90013-N.

84. Krause, E.F. *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*; Dover: New York, NY, USA, 1986.

85. Cantrell, C.D. *Modern Mathematical Methods for Physicists and Engineers*; Cambridge University Press: Cambridge, UK, 2000.

86. Schlamm, A.; Messinger, D. A euclidean distance transformation for improved anomaly detection in spectral imagery. In Proceedings of the 2010 Western New York Image Processing Workshop, Rochester, NY, USA, 5 November 2010; pp. 26–29.

87. Merkwirth, C.; Parlitz, U.; Lauterborn, W. Fast nearest-neighbor searching for nonlinear signal processing. *Phys. Rev. E* **2000**, *62*, 2089–2097, doi:10.1103/PhysRevE.62.2089.

88. Zhao, M.; Saligrama, V. Anomaly detection with score functions based on nearest neighbor graphs. In Proceedings of the Advances in Neural Information Processing Systems 22, Vancouver, British Columbia, Canada, 7–10 December 2009; pp. 2250–2258.

89. Schölkopf, B.; Smola, A.J. Learning with kernels. *Learn. Kernels* **2002**.

90. Zhao, C.; Yao, X.; Yan, Y. Modified Kernel RX Algorithm Based on Background Purification and Inverse-of-Matrix-Free Calculation. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 544–548, doi:10.1109/lgrs.2017.2656251.

91. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417–441, doi:10.1037/h0071325.

92. Hidalgo, J.A.P.; Pérez-Suay, A.; Nar, F.; Camps-Valls, G. Efficient Nonlinear RX Anomaly Detectors. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 231–235, doi:10.1109/lgrs.2020.2970582.

93. Scott, D. *Multivariate Density Estimation*; Wiley: New York, NY, USA, 1992.

94. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; CRC Press: London, UK, 1986.

95. Matteoli, S.; Veracini, T.; Diani, M.; Corsini, G. Background density nonparametric estimation with data-adaptive bandwidths for the detection of anomalies in multi-hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 163–167, doi:10.1109/lgrs.2013.2250907.

96. Matteoli, S.; Veracini, T.; Diani, M.; Corsini, G. Models and methods for automated background density estimation in hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2837–2852, doi:10.1109/tgrs.2012.2214392.

97. Matteoli, S.; Veracini, T.; Diani, M.; Corsini, G. A Locally Adaptive Background Density Estimator: An evolution for rx-based anomaly detectors. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 323–327, doi:10.1109/lgrs.2013.2257670.

98. Veracini, T.; Matteoli, S.; Diani, M.; Corsini, G. Nonparametric framework for detecting spectral anomalies in hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 666–670, doi:10.1109/lgrs.2010.2099103.

99. Veracini, T.; Matteoli, S.; Diani, M.; Corsini, G.; De Ceglie, S.U. A non-parametric approach to anomaly detection in hyperspectral images. In Proceedings of SPIE 7830, Image and Signal Processing for Remote Sensing XVI, Toulouse, France, 22 October 2010; doi:10.1117/12.865073

100. Ruiz, A.; López-de-Teruel, P.E. Nonlinear kernel-based statistical pattern analysis. *IEEE Trans. Neural Netw.* **2001**, *12*, 16–32, doi:10.1109/72.896793.

101. Cremers, D.; Kohlberger, T.; Schnörr, C. Shape statistics in kernel space for variational image segmentation. *Pattern Recognit.* **2003**, *36*, 1929–1943, doi:10.1016/s0031-3203(03)00056-6.

102. Terrell, G.R.; Scott, D.W. Variable kernel density estimation. *Ann. Stat.* **1992**, *20*, 1236–1265.

103. Banerjee, A.; Burlina, P.; Meth, R. Fast hyperspectral anomaly detection via SVDD. In Proceedings of the International Conference on Image Processing, San Antonio, TX, USA, 16 September–19 October 2007; pp. IV101–IV104, doi:10.1109/ICIP.2007.4379964.

104. Schölkopf, B.; Platt, J.; Shawe-Taylor, J.; Smola, A.J.; Williamson, R.C. Estimating the support of a high-dimensional distribution. *Tech. Rep.* **1999**.

105. Tax, D.; Duin, R. Data domain description by support vectors. *ESANN* **1999**, *99*, 251–256.

106. Tax, D.M.J.; Ypma, A.; Duin, R.P.W. Support vector data description applied to machine vibration analysis. *Proc. 5th Annu. Conf. Adv. Sch. Comput. Imaging* **1999**, 398-405.

107. Gualtieri, J.A.; Chettri, S.R.; Cromp, R.F.; Johnson, L.F. Support vector machine classifiers as applied to AVIRIS data. *Summ. Eighth Jpl Airbrone Earth Sci. Workshop* **1999**, 217–227, doi:10.1.1.30.2656.

108. Gualtieri, J.A.; Cromp, R.F. Support vector machines for hyperspectral remote sensing classification. *Proc. Spie Int. Soc. Opt. Eng.* **1999**, *3584*, 221–232.

109. Boltyanski, V.; Martini, H.; Soltan, V. *Geometric Methods and Optimization Problems*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 4.

110. Golub, G.H.; Reinsch, C. Singular value decomposition and least squares solutions. In *Linear Algebra*; Springer: Berlin/Heidelberg, Germany, 1971; pp. 134–151.

111. Comon, P. Independent component analysis, a new concept? *Signal Process.* **1994**, *36*, 287–314.

112. Chang, C.I. Orthogonal Subspace Projection (OSP) revisited: A comprehensive study and analysis. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 502–518, doi:10.1109/tgrs.2004.839543.

113. Ranney, K.I.; Soumekh, M. Hyperspectral anomaly detection within the signal subspace. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 312–316, doi:10.1109/lgrs.2006.870833.

114. Winter, E.M.; Winter, M.E. Autonomous hyperspectral end-member determination methods. In Proceedings of the Sensors, Systems, and Next-Generation Satellites III, Florence, Italy, 28 December 1999; pp. 150–158.

115. Winter, M.E. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In Proceedings of the SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, Denver, CO, USA, 27 October 1999; pp. 266–275.

116. Bowles, J.; Gillis, D.; Palmadesso, P. New improvements in the ORASIS algorithm. In Proceedings of the 2000 IEEE Aerospace Conference, Big Sky, MT, USA, 25–25 March 2000; pp. 293–298.

117. Boardman, J.W. Automating spectral unmixing of AVIRIS data using convex geometry concepts. In Proceedings of the Summaries 4th Annu. JPL Airborne Geosci. Workshop, Pasadena, California, USA, 25 October 1993; pp. 11–14.

118. Belouchrani, A.; Cardoso, J.-F. Maximum likelihood source separation by the expectation-maximization technique: Deterministic and stochastic implementation. In Proceedings of the International Symposium on Nonlinear Theory and Applications NOLTA'95, Las Vegas, NV, USA, 10–14 December 1995; pp. 49–53.

119. Celeux, G.; Govaert, G. A classification EM algorithm for clustering and two stochastic versions. *Comput. Stat. Data Anal.* **1992**, *14*, 315–332, doi:10.1016/0167-9473(92)90042-E.

120. Borghys, D.; Kåsen, I.; Achard, V.; Perneel, C. Hyperspectral anomaly detection: Comparative evaluation in scenes with diverse complexity. *J. Electr. Comput. Eng.* **2012**, doi:10.1155/2012/162106.

121. Lloyd, S.P. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137, doi:10.1109/TIT.1982.1056489.

122. Dunn, J.C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J. Cybern.* **1973**, *3*, 32–57, doi:10.1080/01969727308546046.

123. Windham, M.P. Cluster Validity for the Fuzzy c-Means Clustering Algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.* **1982**, *PAMI-4*, 357-363, doi:10.1109/TPAMI.1982.4767266.

124. Alruwaili, M.; Siddiqi, M.H.; Javed, M.A. A robust clustering algorithm using spatial fuzzy C-means for brain MR images. *Egypt. Inform. J.* **2020**, *21*, 51–66, doi:10.1016/j.eij.2019.10.005.

125. Togacar, M.; Ergen, B.; Comert, Z. COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. *Comput. Biol. Med.* **2020**, *121*, doi:10.1016/j.compbiomed.2020.103805.

126. Versaci, M.; Morabito, F.C. Image Edge Detection: A New Approach Based on Fuzzy Entropy and Fuzzy Divergence. *Int. J. Fuzzy Syst.* **2021**, doi:10.1007/s40815-020-01030-5.

127. Stoica, P.; Selén, Y. A review of information criterion rules. *IEEE Signal Process. Mag.* **2004**, *21*, 36–47, doi:10.1109/MSP.2004.1311138.

128. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69, doi:10.1007/BF00337288.

129. Duran, O.; Petrou, M. A time-efficient method for anomaly detection in hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3894–3904, doi:10.1109/tgrs.2007.909205.

130. Penn, B.S. Using self-organizing maps for anomaly detection in hyperspectral imagery. In Proceedings of the IEEE Aerospace Conference Proceedings, Big Sky, Montana, USA, 9–16 March 2002; pp. 1531–1535.

131. Baraniuk, R.G. Compressive sensing. *IEEE Signal Process. Mag.* **2007**, *24*, 118–120+124, doi:10.1109/msp.2007.4286571.

132. Candes, E.J.; Wakin, M.B. An introduction to compressive sampling: A sensing/sampling paradigm that goes against the common knowledge in data acquisition. *IEEE Signal Process. Mag.* **2008**, *25*, 21–30, doi:10.1109/msp.2007.914731.

133. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306, doi:10.1109/tit.2006.871582.

134. Li, J.; Zhang, H.; Zhang, L.; Ma, L. Hyperspectral anomaly detection by the use of background joint sparse representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2523–2533, doi:10.1109/jstars.2015.2437073.

135. Li, W.; Du, Q. A survey on representation-based classification and detection in hyperspectral remote sensing imagery. *Pattern Recognit. Lett.* **2016**, *83*, 115–123, doi:10.1016/j.patrec.2015.09.010.

136. Li, W.; Du, Q.; Zhang, B. Combined sparse and collaborative representation for hyperspectral target detection. *Pattern Recognit.* **2015**, *48*, 3904–3916, doi:10.1016/j.patcog.2015.05.024.

137. Ling, Q.; Guo, Y.; Lin, Z.; An, W. A Constrained Sparse Representation Model for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2358–2371, doi:10.1109/TGRS.2018.2872900.

138. Xu, Y.; Wu, Z.; Li, J.; Plaza, A.; Wei, Z. Anomaly detection in hyperspectral images based on low-rank and sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1990–2000, doi:10.1109/tgrs.2015.2493201.

139. Zhang, Y.; Du, B.; Zhang, L. A sparse representation-based binary hypothesis model for target detection in hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1346–1354, doi:10.1109/tgrs.2014.2337883.

140. Li, S.; Yin, H.; Fang, L. Remote sensing image fusion via sparse representations over learned dictionaries. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4779–4789, doi:10.1109/tgrs.2012.2230332.

141. Sun, X.; Nasrabadi, N.M.; Tran, T.D. Task-driven dictionary learning for hyperspectral image classification with structured sparsity constraints. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4457–4471, doi:10.1109/tgrs.2015.2399978.

142. Yang, Y.; Zhang, J.; Song, S.; Liu, D. Hyperspectral Anomaly Detection via Dictionary Construction-Based Low-Rank Representation and Adaptive Weighting. *Remote Sens.* **2019**, *11*, 192, doi:10.3390/rs11020192.

143. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM* **2011**, *58*, Article 11, doi:10.1145/1970392.1970395.

144. Xu, Y.; Wu, Z.; Chanussot, J.; Wei, Z. Joint Reconstruction and Anomaly Detection from Compressive Hyperspectral Images Using Mahalanobis Distance-Regularized Tensor RPCA. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2919–2930, doi:10.1109/tgrs.2017.2786718.

145. Zhu, L.; Wen, G. Hyperspectral anomaly detection via background estimation and adaptive weighted sparse representation. *Remote Sens.* **2018**, *10*, 272, doi:10.3390/rs10020272.

146. Zhang, Y.; Du, B.; Zhang, L.; Wang, S. A low-rank and sparse matrix decomposition-based mahalanobis distance method for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1376–1389, doi:10.1109/tgrs.2015.2479299.

147. Tan, K.; Hou, Z.; Ma, D.; Chen, Y.; Du, Q. Anomaly detection in hyperspectral imagery based on low-rank representation incorporating a spatial constraint. *Remote Sens.* **2019**, *11*, 1578, doi:10.3390/rs11131578.

148. Zhang, X.; Ma, X.; Huyan, N.; Gu, J.; Tang, X.; Jiao, L. Spectral-Difference Low-Rank Representation Learning for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, doi:10.1109/TGRS.2020.3046727.

149. Liu, G.; Lin, Z.; Yu, Y. Robust subspace segmentation by low-rank representation. In Proceedings of the ICML 2010—27th International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 663–670.

150. Niu, Y.; Wang, B. Hyperspectral anomaly detection based on low-rank representation and learned dictionary. *Remote Sens.* **2016**, *8*, 289, doi:10.3390/rs8040289.

151. Wang, W.; Li, S.; Qi, H.; Ayhan, B.; Kwan, C.; Vance, S. Identify anomaly component by sparsity and low rank. In Proceedings of the Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing, Tokyo, Japan, 2–5 June 2015, doi10.1109/WHISPERS.2015.8075391:.

152. Tropp, J.A.; Gilbert, A.C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **2007**, *53*, 4655–4666, doi:10.1109/TIT.2007.909108.

153. Dai, W.; Milenkovic, O. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inf. Theory* **2009**, *55*, 2230–2249, doi:10.1109/TIT.2009.2016006.

154. Donoho, D.L.; Tsaig, Y. Fast Solution of $l1$ -Norm Minimization Problems When the Solution May Be Sparse. *IEEE Trans. Inf. Theory* **2008**, *54*, 4789–4812, doi:10.1109/tit.2008.929958.

155. Kim, S.J.; Koh, K.; Lustig, M.; Boyd, S.; Gorinevsky, D. An interior-point method for large-scale $\ell1$-regularized least squares. *IEEE J. Sel. Top. Signal Process.* **2007**, *1*, 606–617, doi:10.1109/jstsp.2007.910971.

156. Tibshirani, R. Regression shrinkage and selection via the lasso: A retrospective. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2011**, *73*, 273–282, doi:10.1111/j.1467-9868.2011.00771.x.

157. Chen, S.S.; Donoho, D.L.; Saunders, M.A. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **1998**, *20*, 33–61, doi:10.1137/S1064827596304010.

158. Gill, P.R.; Wang, A.; Molnar, A. The in-crowd algorithm for fast basis pursuit denoising. *IEEE Trans. Signal Process.* **2011**, *59*, 4595–4605, doi:10.1109/TSP.2011.2161292.

159. Ma, D.; Yuan, Y.; Wang, Q. Hyperspectral anomaly detection via discriminative feature learning with multiple-dictionary sparse representation. *Remote Sens.* **2018**, *10*, 745, doi:10.3390/rs10050745.

160. Zhao, R.; Du, B.; Zhang, L. Hyperspectral Anomaly Detection via a Sparsity Score Estimation Framework. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3208–3222, doi:10.1109/tgrs.2017.2664658.

161. Li, W.; Tramel, E.W.; Prasad, S.; Fowler, J.E. Nearest regularized subspace for hyperspectral classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 477–489, doi:10.1109/TGRS.2013.2241773.

162. Wu, Z.; Su, H.; Du, Q. Low-Rank and Collaborative Representation for Hyperspectral Anomaly Detection. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, 28 July–2 August 2019; pp. 1394–1397.

163. Jiang, K.; Xie, W.; Li, Y.; Lei, J.; He, G.; Du, Q. Semisupervised spectral learning with generative adversarial network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 5224–5236, doi:10.1109/TGRS.2020.2975295.

164. Wang, S.; Wang, X.; Zhang, L.; Zhong, Y. Auto-AD: Autonomous Hyperspectral Anomaly Detection Network Based on Fully Convolutional Autoencoder. *IEEE Trans. Geosci. Remote Sens.* **2021**, doi:10.1109/TGRS.2021.3057721.