

STUDY OF THE INFLUENCE OF PRE-PROCESSING ON LOCAL STATISTICS-BASED ANOMALY DETECTOR RESULTS

D. Borghys

Royal Military Academy
Signal & Image Centre
Brussels, Belgium

C. Perneel

Royal Military Academy
Dept. of Appl. Mathematics
Brussels, Belgium

ABSTRACT

Anomaly detection in hyperspectral data has received much attention for various applications and is especially important for defense and security applications. Anomaly detection detects pixels in the hyperspectral data cube whose spectra differ significantly from the background spectra [1]. Most existing methods estimate the spectra of the (local or global) background and then detect anomalies as pixels with a large spectral distance w.r.t. the determined background spectra. Many types of anomaly detectors have been proposed in literature. This paper reports on a sensitivity study that tries to determine an adequate pre-processing chain for anomaly detection in hyperspectral scenes. The study is performed on a set of five hyperspectral datasets and focuses on statistics-based anomaly detectors.

Index Terms— Anomaly detection, pre-processing, data reduction, spectral normalization

1. INTRODUCTION AND METHODOLOGY

Many types of anomaly detectors (ADs) have been proposed in literature. In [2] an overview and an examination of their inter- and intra-method consistency is presented. The impact of parameter settings on the results of the detectors is also examined there. In the current paper the influence of pre-processing is investigated. For three well-established statistics-based ADs (Local RX and PCA- and KPCA AD) the influence of data reduction and spectral normalization is studied. In [3] ways to improve the estimation of the covariance matrix under the small sample size condition are examined. This is of interest for applying the Reed-Xiaoli (RX) detector locally. That paper proposes a regularization of the covariance matrix by diagonal loading. The current paper also examines two other methods: one based on singular value decomposition and one based on prior data reduction by Principal Component Analysis (PCA) with a fixed threshold on explained variance and eigenvalue. These methods are compared between them and with methods based on a prior data reduction by PCA, Maximum Noise Fraction (MNF)[4] and Independent Component Analysis (ICA)[5]. The impact

of spectral normalization is also investigated. Examined normalization methods are minimum/maximum normalization (MM) and continuum removal (CR). The comparison is performed on a set of five hyperspectral datasets. In two scenes sub-pixel anomalies were artificially inserted. The aim of this work is to determine an adequate pre-processing chain for applying these ADs.

2. DATASET

The analysis was performed on a set of 5 hypercubes of scenes with various complexity, acquired by 3 different airborne sensors. Fig. 1 shows RGB composites of each of the examined datacubes. The size of the images in the figure is not representative for their relative spatial resolution, but their aspect ratio was preserved.

Table 1 presents the main characteristics of the dataset. The first column is the name by which the scenes will be referred further in this paper. The first three datacubes were acquired by 3 different airborne sensors. The ground truth was manually indicated and consists of all vehicles (cars, trucks, airplanes). The two last datacubes (HAR and CAM) are real data from the Hymap sensor in which anomalies were inserted artificially by linearly mixing the spectra of a green paint with the original background pixel. For display purposes, in fig. 1, images with full-pixel paint spectra are shown. For the evaluation of anomaly detection results, images with a mixing ratio of 1/10 were used.

3. ANOMALY DETECTION METHODS

In this work three well-established statistics-based anomaly detectors have been used. They are briefly described below.

3.1. Local RX detector(LRX)

The RX detector [6] is a standard in anomaly detection. In RX a modified Mahalanobis distance is calculated between



Fig. 1. RGB composite of the original datacubes. From left to right: PAV, BJO, OBP, HAR, CAM

Name	Site	Sensor name	# bands	Waveband (μm)	Spat. Res. (m)	Scene Description
PAV	Pavia (It)	Rosis	102	0.430-0.834	1.3	City
BJO	Bjoerkelangen (No)	HySpex	80	0.410-0.984	0.2(0.8)	Rural village
OBP	Oberpfaffenhofen(Ge)	Hymap	126	0.44-2.45	4	Airfield with aerospace industry
HAR	Hartheim (Ge)	Hymap	126	0.44-2.45	4	Agricultural area
CAM	Camargue (Fr)	Hymap	126	0.44-2.45	4	Agricultural area

Table 1. Overview of the dataset

the current pixel and the background:

$$D_{RX} = R_M^t \left(\frac{NC_B}{N+1} + \frac{R_M R_M^t}{N+1} \right)^{-1} R_M = R_M^t C_{B'}^{-1} R_M.$$

C_B and μ_B are respectively the spectral covariance matrix and the spectral mean of the background pixels, N is the number of background pixels and r and $R_M = (r - \mu_B)$ with r the spectrum of the current pixel. If N is very large, the second term can be neglected. In local RX (LRX) the background statistics are determined from the local neighborhood of the current pixel, defined by an outer window (OWS) and separated from the current pixel by a guard window (GWS).

3.2. PCA-AD and KPCA-AD

In the PCA-based detector [7] the same dual-window idea is used as with RX. The anomaly detection value is the projection separation value between OWS and the current pixel. The projection vectors are formed by the first principal component vectors calculated in the outer window (with a cut-off percentage of 99%). The kernelized version of this method was also examined [7, 8] (KPCA-AD). In the current paper a Gaussian radial basis function is used as kernel function and an adequate value for the standard deviation of the Gaussian σ_g is determined in each pixel as a function of the estimated standard deviation of the spectral values in the outer window σ_{OW} . It was empirically determined as $\sigma_g = 100 \sigma_{OW}$.

4. RESULTS AND DISCUSSION

The comparison of the different methods is based on the area under the RoC curve (AUC).

4.1. Results for LRX

The local RX requires the estimation of the covariance matrix in a small rectangle around the current pixel. For a reliable estimate of the $C_{B'}$, a number of samples of about ten times the number of bands would be desirable. However, the size of the local windows that are typically used is much too small. This leads to an ill-conditioned and unstable inverse [3]. Several ways to overcome this problem are suggested in literature. In this paper three regularization methods are examined:

- **Prior PCA (PPCA):** A PCA-based channel reduction at fixed threshold is applied prior to RX detection. In this paper the channels are cut at 99.99% of the total variance or 0.01 % of the maximal eigenvalue.
- **Diagonal Loading(DL):** a scaled identity matrix is added to $C_{B'}$ before inversion [9]. The scale factor is the median of the eigenvalues of the covariance matrix calculated on the complete image [3].
- **Singular Value Decomposition (SVD):** here the inverse is replaced by a pseudo-inverse, using SVD. In the pseudo-inverse the eigenvalues below a certain threshold are set to zero. In this paper the eigenvalues are cut at 99.9 % of the trace.

Site	Spec. Norm.	LRX results									
		Regularization type			Results after data reduction						
		DL	SVD	PPCA	PCA		MNF		ICA		
			AUC/BestAUC	%	AUC/BestAUC	%	AUC/BestAUC	%			
PAV	None	0.76	0.64	0.72	0.74(16)/-	-	0.74(14)/-	-	0.74(14)/0.74(3,10)		1
	MM	0.69	0.76	0.52	0.71(4)/0.72(3,6)	1	0.75(8)/0.75(2,7)	0	0.57(2)/-		-
	CR	0.58	0.48	0.52	0.64(2)/-	-	0.71(8)/0.71(2,5)	-	0.68(6)/-		-
BJO	None	0.59	0.76	0.59	0.55(10)/0.56(3,10)	1	0.60(10)/0.64(3,4)	6	0.56(4)/0.61(2,3)		8
	MM	0.70	0.54	0.52	0.61(10)/0.62(3,10)	2	0.72(4)/0.73(2,7)	1	0.73(6)/-		-
	CR	0.74	0.60	0.57	0.71(8)/0.72(2,11)	2	0.63(2)/0.72(3,10)	12	0.72(4)/0.73(2,7)		1
OBP	None	0.85	0.62	0.66	0.65(2)/-	-	0.77(12)/0.77(2,11)	0	0.76(4)/0.77(2,17)		1
	MM	0.84	0.76	0.69	0.63(2)/0.65(2,3)	3	0.63(2)/-	-	0.71(2)/-		-
	CR	0.62	0.54	0.60	0.60(2)/-	-	0.69(6)/-	-	0.70(2)/-		-
HAR	None	1.00	0.47	1.00	1.00(18)/-	-	0.66(16)/-	-	0.87(2)/-		-
	MM	1.00	0.49	0.56	1.00(18)/-	-	0.93(10)/-	-	0.70(10)/0.72(3,2)		3
	CR	1.00	0.49	0.61	1.00(18)/-	-	0.92(14)/-	-	0.96(10)/-		-
CAM	None	1.00	0.52	0.55	1.00(18)/-	-	0.66(16)/-	-	0.67(18)/-		-
	MM	1.00	0.55	0.72	1.00(18)/-	-	0.68(14)/-	-	0.70(12)/0.71(2,11)		2
	CR	1.00	0.52	0.57	1.00(18)/-	-	0.91(16)/-	-	0.98(14)/-		-

Table 2. Overview of LRX results for the different scenes and pre-processing methods.

Columns 3 to 5 of table 2 show the results of these three pre-processing method for LRX, for the 5 scenes and the three spectral normalization methods: None, Min-Max (MM) and Continuum Removal (CR). Presented results show a large difference between the LRX results for the three different pre-processing methods. DL gives the best results in most cases. In only one scene SVD gives better results (BJO). For verifying whether prior data reduction improves results, three different data reduction methods were applied. The number of bands, kept after data reduction, was varied between 2 and 20. The first number in columns 6,8 and 10 presents the highest obtained AUC and the number between brackets is the corresponding number of selected bands. We also examined whether it is better to discard the first or second band after data reduction. If this gives better results, the result is given after the slash and between bracket is the starting band and the number of bands. Columns 7,9 and 11 give the percentage of gain in AUC when the starting band is changed from 1 to 2 or 3. For the LRX results after data reduction, in most cases the best results are obtained starting from the first band and the percentage gained when starting from band 2 or 3 is mostly quite small. More important is the fact that none of the data reduction based methods for LRX give better results than the DL. The only exception is BJO where the best results are obtained after the SVM-based regularization. In all cases the best results were obtained without applying any spectral normalization. No consistent link could be found between the best number of bands and the explained variance. This may be intrinsic to the problem of anomaly detection because the properties of anomalies are not related to the scene's global statistics.

4.2. Results for PCA-AD and KPCA-AD

Table 3 shows the results obtained for the PCA-AD and KPCA-AD. Column three is the result obtained without any data reduction, the other columns are analogous to the last columns in table 2 and present the results after data reduction. Results for HAR and CAM are not shown due to paper length limitations, but the conclusions are similar as for the three other scenes. Contrary to the LRX case, for PCA-AD and KPCA-AD the best results are obtained after MinMax normalization except for the PCA-AD in the BJO scene. For the results after PCA and MNF data reduction, it is preferable for both detectors to discard the first one or two bands. For ICA this is less important. The data reduction method that gives the best results depends on the scene.

5. CONCLUSIONS

The influence of data reduction and spectral normalization on the results of three local statistics-based anomaly detectors was examined.

For local RX the method various covariance matrix regularization methods were also examined. The best results for LRX are obtained using regularization by diagonal loading and without any data reduction or spectral normalization. Only in one dataset the SVD-based pseudo-inverse gives better results; also without prior data reduction or normalization.

For the PCA- and KPCA-based detector the best results are mainly obtained after MinMax normalization. For PCA and MNF data reduction the first couple of bands should be discarded when applying these two detectors. The data re-

Site	Spec Norm.	Data Reduction Method							
		None	PCA		MNF		ICA		
PCA-AD results									
			AUC/BestAUC	%	AUC/BestAUC	%	AUC/BestAUC	%	
PAV	None	0.76	0.76(18)/0.78(3,2)	2	0.75(12)/0.78(3,6)	3	0.79(12)/0.79(3,10)	1	
	MM	0.86	0.86(18)/0.87(2,3)	1	0.81(8)/0.83(3,6)	2	0.68(2)/-	-	
	CR	0.65	0.65(2)/-	-	0.64(10)/0.69(3,4)	7	0.69(6)/0.70(3,4)	1	
BJO	None	0.80	0.80(16)/-	-	0.74(10)/0.76(3,2)	23	0.66(2)/-	-	
	MM	0.70	0.70(14)/0.79(2,13)	11	0.78(18)/ 0.80(2,15)	1	0.71(18)/-	-	
	CR	0.78	0.76(16)/ 0.81(2,17)	6	0.79(10)/ 0.80(2,9)	1	0.77(2)/-	-	
OBP	None	0.87	0.87(18)/0.88(3,16)	1	0.88(18)/0.88(2,15)	0	0.90(16)/0.91(3,14)	1	
	MM	0.89	0.89(16)/ 0.93(2,17)	4	0.87(16)/0.88(3,14)	1	0.90(6)/0.91(2,5)	1	
	CR	0.52	0.51(4)/0.55(2,3)	7	0.80(18)/0.82(3,16)	2	0.85(16)/0.85(3,14)	1	
KPCA-AD results									
PAV	None	0.83	0.80(2)/-	-	0.81(8)/-	-	0.74(12)/-	-	
	MM	0.86	0.85(2)/-	-	0.83(8)/-	-	0.53(2)/-	-	
	CR	0.82	0.69(2)/-	-	0.72(10)/-	-	0.70(6)/-	-	
BJO	None	0.75	0.76(2)/-	-	0.76(12)/0.77(3,8)	1.8	0.63(10)/-	-	
	MM	0.65	0.66(8)/0.76(2,5)	13	0.77(6)/ 0.82(3,4)	7	0.70(16)/-	-	
	CR	0.74	0.70(16)/0.76(2,3)	8	0.74(10)/0.78(3,8)	5	0.76(4)/0.80(3,2)	4	
OBP	None	0.88	0.69(18)/0.77(3,4)	10	0.73(18)/0.85(3,2)	14	0.81(10)/0.82(2,9)	2	
	MM	0.91	0.74(6)/0.82(2,17)	10	0.78(10)/0.81(2,9)	4	0.95(6)/0.96(3,4)	1	
	CR	0.78	0.46(2)/0.68(3,2)	32	0.74(12)/0.80(2,11)	8	0.85(16)/0.85(3,14)	0	

Table 3. Overview of the results of the PCA- and KPCA-based ADs for the different scenes and pre-processing methods.

duction method that gives the best results for PCA-AD and KPCA-AD is scene dependent.

6. ACKNOWLEDGMENTS

The PAV scene was provided by Prof Gamba of Pavia Univ. and acquired in the HySens campaign by DLR. Dr Skauli of the Norwegian FFI provided the BJO data. HAR and CAM were provided to us by Dr Achard of the French ONERA.

7. REFERENCES

- [1] D.W.J. Stein, S.G. Beaven, L.E. Hoff, E.M. Winter, A.P. Schaum, and A.D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Proc. Mag.*, vol. 38, pp. 58–69, Jan 2002.
- [2] D. Borghys, E. Truyen, M. Shimoni, and C. Perneel, "Anomaly detection in complex environments: Evaluation of the inter- and intra-method consistency," in *Proc. 1st IEEE WHISPERS*, Grenoble, Aug 2009.
- [3] S. Matteoli, M. Diani, and G. Corsini, "Different approaches for improved covariance matrix estimation in hyperspectral anomaly detection," in *Proc. Annual Meeting. GTTI*, 2009.
- [4] A.A. Green, M. Berman, P. Switzer, and M.D. Graig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE-TGRS*, vol. 26, no. 1, pp. 65–74, 1988.
- [5] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," Tech. Rep., Lab. of Computer and Information Science, Helsinki Univ. of Technology, Helsinki, Finland, 2000.
- [6] I.S. Reed and X. Yu, "Adaptive multiband cfar detection of an optical pattern with unknown spectral distribution," *IEEE ASSP*, vol. 38, no. 10, pp. 1760–1770, Oct 1990.
- [7] H. Goldberg and N. Nasrabadi, "A comparative study of linear and nonlinear anomaly detectors for hyperspectral imagery," in *Algorithms and Techn. for Multi-, Hyper-, and Ultraspectral Imagery XIII*. 2007, vol. 6565, SPIE.
- [8] B. Scholkopf, A.J. Smola, and K.R. Muller, "Kernel principal component analysis," *Neural Computation*, vol. 10, pp. 1299–1319, 1999.
- [9] B.D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. El.Syst.*, vol. 24, no. 4, pp. 397–401, Jul 1988.