CONSEQUENCES OF THE HUGHES PHENOMENON ON SOME CLASSIFICATION TECHNIQUES

María C. Alonso, Professor
José A. Malpica, Professor
Alex Martínez de Agirre, Postgraduate Student
School of Geodesy and Cartography.
University of Alcalá
28871 Madrid, Spain
mconcepcion.alonso@uah.es
josea.malpica@uah.es
alejandro.martineza@uah.es

ABSTRACT

In supervised classification, each pixel of the image is labeled as representing a particular ground cover or class, taking information provided by the training samples, since the parameters of the classifier are estimated from the training samples. This training can be established using maps, site visits, or aerial photography.

As the number of spectral bands or dimensions increases, the separability of classes also increases, but so does the number of statistical parameters defining the classes. Since there are only a fixed number of training samples for deriving the statistical parameters, at some point the accuracy of the estimation must begin to decrease. An optimal value of dimensions and training samples is shown to exist in any given practical circumstance and, in general, more dimensions do not necessarily lead to better results.

This paper's aim is to evaluate the behavior of two classifiers, Mahalanobis and SVM, as the number of dimensions increase. It is done using a heuristic analysis combining different numbers of training and dimensions on both synthetic and real hyperspectral data. It shows the relationship between training and dimensionality. Several tables are presented to show the superiority of SVM over Mahalanobis classifiers based on both synthetic and real hyperspectral data. These findings demonstrate that the curse of dimensionality occurs for as few as ten dimensions for Mahalanobis; however, this phenomenon has no effect on SVM.

KEYWORDS: Clustering, curse of dimensionality, dimensionality reduction, high dimensionality, hyperspectral data, supervised classification, support vector machine (SVM).

INTRODUCTION

A feature is any aspect, quality, or characteristic of an object; several features of an object make up a feature vector. The set of available feature vectors spans the feature space. 2D or 3D subsets of, or projections on, feature space can be visualized as scatter plots. In Figure 1, we can observe an example of a feature vector, its feature space, and a scatter plot in 2D.

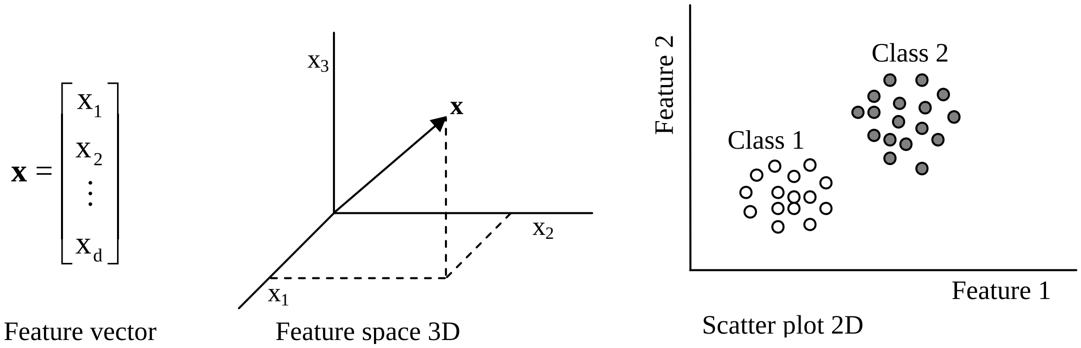

Figure 1: Relationship between feature vector and feature space.

If the objects of representation are the pixels in a image, a remote sensing image can be modeled with a Euclidean space, where the number of bands is the dimension of the space and the pixels in the image are represented as points in that space.

In supervised classification, each pixel of the image is labeled as representing a particular ground cover or class, taking information provided by the training samples. The training can be established using maps, site visits, or aerial photography. The parameters of a particular classifier algorithm are calculated from these training samples.

In recent years, remote sensing data has become increasingly larger in both number of pixels per image (high spatial resolution) and number of bands (high spectral resolution). As the number of spectral bands or dimensions increases, the separability of classes also increases, but so does the number of statistical parameters defining the classes. Since there are only a fixed number of training samples for deriving the statistical parameters, at some point the accuracy of the estimation must begin to decrease. An optimal value of dimensions and training samples is shown to exist in any given practical circumstance, and more dimensions do not necessarily lead to better results.

High dimensional data is difficult to work with for several reasons; among them, we can say that a lot of features increase the noise factor, and hence the error factor, that there are not enough observations to get good estimates, or that most data is scattered.

As an interesting note about the previous observation, consider a sphere of radius *r* inscribed inside a hypercube of dimension *d* and sides of length 2*r*. The volume of the hypercube is $(2r)^d$, where *d* is the number of dimensions. It is possible to find that the volume of the sphere is $\dfrac{(2r)^d \pi^{d/2}}{d\Gamma\left(\frac{d}{2}\right)}$. Therefore, the proportion of the volume of the square that is inside the sphere is

$$\lim_{d\to\infty} \frac{\pi^{d/2}}{d\Gamma\left(\frac{d}{2}\right)} \to 0$$ (Stibor et al, 2006).

It looks like, intuitively, that in high dimensionality the data accumulate in the corners. Bellman (1961), referring to the computational complexity of searching the neighborhood of data points in high-dimensional settings, was the first to put forward the term *curse of dimensionality* in order to describe the problem of data sparseness. Hughes (1968) conducted a statistics analysis, showing how the accuracy of a classifier depends of the number of training samples. Therefore, the curse of dimensionality is also known as the Hughes effect or the Hughes phenomenon. Many works have dealt with this dimensionality phenomenon for the last four decades; most recently, Lavergnea and Patilea (2008) proposed a general nonparametric method trying to avoid or reduce the Hughes effect; Gheyas and Smith (2010) present a hybrid algorithm since, as they say, no existing algorithm is entirely satisfactory in isolation, but that a carefully designed combination can overcome the weaknesses of each. Diani et al. (2008) present a methodology for band selection for hyperspectral sensors tailored to target detection applications, which chooses a subset of bands that maximizes an objective function suitable for target detection. Mojaradi et al. (2009) propose two methods for dimensionality reduction of hyperspectral data via spectral feature extraction, and compared them to the traditional methods for finding relevant channels in order to determine optical regions; moreover, instead of optimizing separability criteria, the overall classification accuracy of a validation dataset is used to decide which disjoint optical regions yield maximum accuracy.

This paper's aim is to evaluate the robustness of two classifiers, Mahalanobis distance and the Support Vector Machine (SVM), used with remote sensing data under a twofold condition of high dimensionality and minimal training. Regarding dimensionality, we have concentrated mostly on a dozen or fewer dimensions, since this is the amount most frequently used in practice. Hyperspectral imagery is usually reduced to a dozen dimensions with new remote-sensing sensors taking imagery with approximately half a dozen bands, in what is called multispectral imagery. A relationship is shown between the number of training samples needed and the complexity of the classifier to be used. Fukunaga (1989) discussed the effect of finite sample size parameter estimates on the evaluation of a family of classifiers, and studied the relationship between the number of features and the number of training samples and their effect on measuring separability, due to mean and covariance shifts, using high-dimensional data.

MATERIAL AND METHODS

These experiments have been run on synthetic images. Two synthetic images of 14 bands were created with 250×250 pixels; they were constructed generating Gaussian random points with different means and covariance matrices, without any correlation between the different bands.

(a) (b)

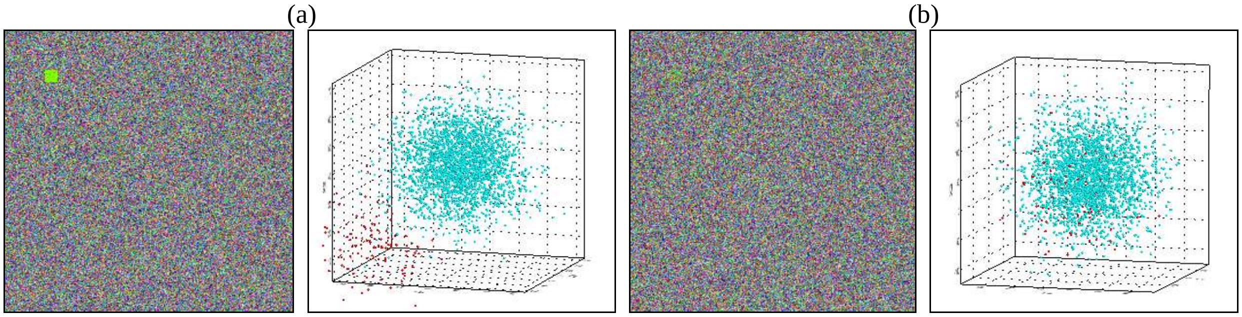

Figure 2: A synthetic image with two classes having a Bhattacharyya separability of 2 (a), and a synthetic image with two classes having a separability of 1.9248 (b).

The algorithm has been tested using hyperspectral imagery from the Airborne Hyperspectral Scanner AHS sensor acquired by the Spanish National Institute for Aerospace Technology—INTA (Instituto Nacional de Técnica Aeroespacial). The AHS is an airborne scanner, of the whiskbroom type, with 80 bands in the electromagnetic spectrum. The AHS image used here was obtained in April 2004. This dataset was taken at a height of approximately 1,300 m, with a 2,700 m cross-track and 14 -km along-track, and a resolution of approximately three- and-a-half meters. The experiment was performed on the original radiance data, and no correction was performed. Out of the 80 bands, 20 were removed because they were characterized by some noise, and from the 60 remaining bands 14 were chosen equidistantly, so the algorithms were run on these 14 bands to compare with the synthetic experiment that also ran with 14 bands. The image in Figure 3 was made using bands 6, 27, and 72 for red, green, and blue colors, respectively. The size of the image is 512×512 pixels.

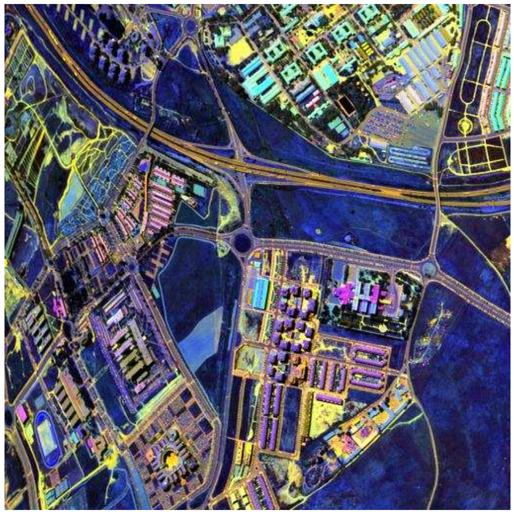

Figure 3: AHS imagery of the university campus at Alcala de Henares.

Since there is no single method of classification that clearly outperforms all the other methods in all problem situations, we have selected two classifiers: one based on the Mahalanobis distance (as a parametric classifier), and the other on SVM (as a non-parametric classifier). Parametric classifiers rely on assumptions of data distribution. The performance of a parametric classifier depends largely on how well the data match the pre-defined models, and on the accuracy of the estimation of the model parameters. They suffer from the Hughes phenomenon (the curse of dimensionality), and consequently it might be difficult to have a significant number of training pixels.

The SVM is a supervised classification method (Burges, 1998) based on structural risk minimization. The key idea of this technique is to estimate a separator boundary or surface between the spectral classes. This surface, which maximizes the margin between classes, uses limited numbers of boundary pixels (support vectors) to create the decision surface. This technique has been applied in different contexts of data classification, as well as remote sensing applications and hyperspectral images (Antony Gualtieri and Cromp, 1998).

The idea for SVM initially appeared in an article by Boser et al. (1992), where they applied it to optical character recognition problems. They demonstrated the superior generalization of SVM compared to other learning algorithms. SVM maximizes the margin between the training patterns and the decision boundary.

The SVM has been compared to other classification methods for remote sensing imagery, such as the Neural Networks, Nearest Neighbor, Maximum Likelihood, and Decision Tree classifiers, and have surpassed them all in robustness and accuracy for remote sensing data (Melgani & Bruzzone, 2004). These results lead us to apply SVM instead of others classifiers.

RESULTS AND DISCUSION

Synthetic Imagery

The two synthetic images were created with two classes with different separability, the first in Figure 2a, with two classes and Bhattacharyya separability 2, and the second in Figure 2b, with two classes and a separability of 1.9248.

In Table 1, we can see the results with the training of the Mahalanobis algorithm with 100% (121 and 1560 pixels) and with 10% (12 and 156 pixels); the evaluation was made with the total of the ROI pixels in both cases. We have calculated the mean of the accuracy using all possible combinations of the 14 bands taken by two bands, the same taken by three bands, and so on. As can be observed, the results are quite robust because of the excellent separability of the classes (2.0). In practical cases, it would not be necessary to take a large number of pixels for the training set.

Table 1: Results with the training of the Mahalanobis algorithm for 10% of the ROI (168 pixels) and 100% of the ROI (1681pixels) of training for the image Figure 2a

# BANDS	10%(168p)	100%(1681p)
2	85.3044	84.8180
3	92.1506	91.9535
4	95.9177	95.8690
5	97.9377	97.9597
6	98.9976	99.0380
7	99.5313	99.5714
8	99.7909	99.8231
9	99.9138	99.9340
10	99.9696	99.9788
11	99.9917	99.9949
12	99.9987	99.9993
13	100.0000	100.0000
14	100.0000	100.0000

In the same way, for the second experiment with the synthetic image (Figure 2b), we have chosen two classes with a separability of 1.9248. In Table 2, we can see the mean of the accuracy using all possible combinations of the 14 bands for different numbers of training pixels (15% of the ROI, 14% of the ROI, 12% of the ROI, etc., for each class); for all of them, the evaluation was made with the whole ROI.

Table 2: Accuracy of Mahalanobis classifier for different training sets

# BANDS	15% (30p)	14% (28p)	12% (24p)	11% (22p)	10% (20p)	7% (14p)
2	68.9670	69.0165	69.6154	68.0879	66.0604	68.0220
3	73.8915	74.5343	74.4217	72.3997	70.1621	71.4753
4	77.6044	78.4940	77.9925	75.6119	73.3492	74.4241
5	80.3064	81.4211	80.7205	78.0262	75.8721	76.8427
6	82.2056	83.7208	82.9231	79.7829	77.5846	78.5491
7	83.5061	85.6482	84.7120	81.0256	78.4744	79.3632
8	84.4478	87.3039	86.1325	81.8205	78.5821	79.2244
9	85.1054	88.7890	87.2403	82.2268	78.0704	77.8724
10	85.5325	90.0909	88.0979	82.3092	77.1978	75.2468
11	85.8063	91.3049	88.7225	82.0632	76.0852	71.1909
12	85.9725	92.4121	89.2308	81.3571	74.9176	65.1978
13	86.1071	93.3214	89.5357	80.4643	72.9643	58.5714
14	86.5000	94.5000	90.0000	77.5000	68.5000	53.5000

It is interesting to note how the Hughes phenomenon occurs for the training sets with the smaller number of pixels. It actually starts with 11% ROI at about the tenth dimension.

Real Imagery

When the experiment was done with the real AHS image (in Figure 3), we obtained the results shown in Table 3. We have also calculated the mean of the accuracy using all possible combinations of the 14 bands taken by two bands, the same taken by three bands, and so on.

Table 3: Results of applying the Mahalanobis classifier with different training sets to the AHS image with 14 bands

# BANDS	100% (2904p)	10% (291p)	1% (29p)	0.7% (20p)	0.6% (17p)	0.5% (14p)
2	62.6298	62.8966	61.2891	61.5831	60.4369	58.6149
3	66.7087	67.0529	63.2305	65.7142	65.0264	62.9457
4	69.8526	70.1527	65.0940	68.4741	68.0431	65.3898
5	72.3186	72.4677	67.1100	70.1777	69.5132	67.3315
6	74.4434	74.3291	69.1824	71.2893	70.3436	68.7405
7	76.3276	75.9313	71.1369	72.1502	70.7119	69.6001
8	77.9865	77.3496	72.9051	72.7726	70.6747	69.9629
9	79.4194	78.6372	74.4302	73.0453	70.3655	69.9329
10	80.6596	79.8390	75.6984	72.9912	70.0820	69.5655
11	81.7205	80.9881	76.6689	72.7184	70.1711	68.5455
12	82.5871	82.0925	77.4271	72.4628	70.7062	66.3098
13	83.2743	83.1193	78.1902	72.2772	71.7926	61.7449
14	83.8843	84.3320	79.0978	72.2107	73.4504	39.2906

As can be observed in the real image, the Hughes phenomenon happen for the 0.5% ROI (14 pixels) and starts about the tenth dimension.

SVM

Here the experiment has been done in order to classify the real AHS image (Figure 3). We have calculated the mean of the accuracy using some random combinations of the 14 bands up to 100 taken by two bands, the same taken by three bands, and so on. The results with different training can be seen in Table 4.

Table 4: SVM applied to AHS with different training sets

# BANDS	100% (2904p)	10% (291p)	1% (29p)	0.7% (20p)	0.5% (14p)
3	78.5145	73.6357	63.1777	63.8502	64.3264
4	81.4697	76.9170	63.5751	64.1105	64.6019
5	83.6388	80.2218	63.5888	64.4797	64.1570
6	85.4236	81.9890	63.7166	64.9132	64.9298
7	86.4056	82.9277	64.7693	67.4366	65.6780
8	88.0207	84.3244	65.8970	68.7307	65.8543
9	89.0551	85.5275	66.8271	70.9632	66.1546
10	89.6481	85.8957	66.9508	73.3877	67.0351
11	90.3309	86.6236	67.7056	75.9136	68.0795
12	90.8243	87.2238	68.6825	77.5478	69.7897
13	91.3002	87.7755	69.2591	79.0388	71.7016
14	91.8388	88.2920	69.2149	80.0620	74.0014

The important thing here is that the Hughes phenomenon was not observed with the SVM algorithm. This experiment has been also performed with alternative ROIs (though not presented here), and the results were very similar.

In order to know if the Hughes phenomenon has any effect over the SVM algorithm, we extended the space to 30 dimensions, (i.e. considering 30 bands instead of 14 in the AHS image); to do this we chose 30 bands without noise from the available 80 AHS bands, which were separated more or less equally in frequency. The results can be observed in Table 5.

Table 5: SVM with 30 bands

# BANDS	0.5% (14p)	0.25% (8p)
3	63.2428	50.6505
4	63.2765	54.3189
5	63.8626	58.2269
6	65.5771	58.4298
7	66.3109	60.8681
8	68.3994	61.2868
9	68.0758	63.7252
10	69.3182	65.4043
11	70.2989	64.4022
12	70.6643	65.3268
13	71.9229	66.5193
14	71.4628	66.5909
15	72.0957	68.2738
16	72.6942	68.6546

# BANDS	0.5% (14p)	0.25% (8p)
17	72.2187	68.5523
18	73.5844	68.2634
19	75.0944	70.3151
20	75.4576	71.3729
21	75.9280	72.1932
22	76.9105	73.3388
23	76.8516	74.5424
24	76.8202	76.4284
25	77.9101	76.8027
26	78.0372	76.7317
27	78.4032	77.7903
28	79.9084	78.3747
29	79.2676	78.3822
30	79.3733	78.4435

Still, no Hughes effect could be detected even with as small a number of pixels as 14 or 8. This experiment cannot be performed with the Mahalanobis classifier, since it needs the covariance matrix to be estimated, and this matrix will have a dimension of 30×30, which is impossible to be estimated with only 8 or 14 pixels.

Comparison of Mahalanobis and SVM

In this section we are going to compare the two classification methods: Mahalanobis and SVM, whose results with different training methods can be seen in Table 6. We have also drawn the maximum and minimum accuracy for each combination band. In Table 6 we can see the results, using the whole ROI for training and the same ROI for evaluation.

Table 6: Comparison of the Mahalanobis and SVM classifiers

	MAHALANOBIS DISTANCE			SUPPORT VECTOR MACHINE		
	100% ROI (2904 pixels)			100% ROI (2904 pixels)		
# BANDS	**MAX**	**MEAN**	**MIN**	**MAX**	**MEAN**	**MIN**
3	83.4022	66.7087	55.3375	85.8127	78.5145	68.7328
4	82.7824	69.8526	57.2658	87.5689	81.4697	73.8292
5	83.4366	72.3186	58.5744	88.2231	83.6388	78.1336
6	84.2975	74.4434	58.8499	89.1873	85.4236	77.9959
7	84.8485	76.3276	59.7796	90.5303	86.4056	80.0964
8	85.0895	77.9865	62.9821	91.2879	88.0207	84.5730
9	85.1584	79.4194	66.9077	91.9766	89.0551	85.1928
10	85.0551	80.6596	71.8320	92.2865	89.6481	86.6391
11	84.8140	81.7205	75.8953	92.2521	90.3309	87.0868
12	84.7107	82.5871	78.9256	92.2521	90.8243	88.7052
13	84.5041	83.2743	79.9587	91.8044	91.3002	89.6006
14	83.8843	83.8843	83.8843	91.8388	91.8388	91.8388

As can be observed, SVM is in all cases superior to Mahalanobis, and is more robust, since the range of the variability between maximum and minimum is narrower in SVM than in Mahalanobis.

CONCLUSION

In general, relative performance of classifiers is always influenced by the properties of application data at hand. Therefore, it is crucial to have simple, practical criteria that guarantee potential advantages of using parametric or non-parametric methods (Mahalanobis or SVM), for a given dataset. The empirical results obtained in this work, using real satellite and synthetic datasets, illustrate the usefulness of both approaches.

Since most of the budget on a remote sensing project goes to the need for terrain visits, it is important to reduce these visits as much as possible; therefore, it would be convenient to reduce the training to a few pixels. From our work we have deduced that when using the Mahalanobis classifier, using few pixels for training, it is necessary to reduce the number of dimensions in order to reduce the effect of the Hughes phenomenon. Therefore, it will be necessary to look for a good method for future selection and extraction in the case of using Mahalanobis. However, the SVM classifier is not affected by the dimensionality or Hughes phenomenon, and the robustness of SVM is superior to Mahalanobis. Furthermore, SVM is a non-parametric classifier and consequently does not demand any assumption about the statistical distribution of the pixels.

The effect of different proportions of training data was evaluated in both methods for synthetic and real data. In general, SVM is superior to Mahalanobis in accuracy and robustness. The evaluation justifies the robustness of SVM classifiers and shows that the Hughes phenomenon is not a critical issue for this technique, i.e. SVM classifies satisfactorily with very few pixels, less than the dimension of the space, and its results are better as more dimensions are used.

ACKNOWLEDGMENTS

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

REFERENCES

Anthony Gualtieri, J., and J., R. F. Cromp, 1998. Support Vector Machines for Hyperspectral Remote Sensing Classification, *27th AIPR Workshop, Advances in Computer Assisted Recognition* Oct. 1416, 1998, Washington, D.C. *Proceedings of the SPIE*, Vo1.3584.

Bellman, R. 1961. *Adaptive Control Processes: A Guided Tour*. Princeton University Press.

Boser, B.E., I. Guyon, and V.N. Vapnik, 1992. A Training Algorithm for Optimal Margin Classifiers. In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* Vol. 5 pp. 144-152, Pittsburgh. ACM Press, San Mateo, CA.

Burges, C. J. A. 1998. *Tutorial on Support Vector Machines for Pattern Recognition, in Data mining and knowledge discovery*, U. Fayyad, Ed. Kluwer Academic, pp. 1-43.

Fukunaga, K., and R. R. Hayes, 1989. Effects of Sample Size in Classifier Design. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11*(8): 873-885.

Gheyas, I.A. and L.S. Smith, 2010. Feature subset selection in large dimensionality domains. *Pattern recognition* 43:5-13

Hughes, G. F. 1968. On the Mean Accuracy of Statistical Pattern Recognizers. *IEEE Transactions on Information Theory*, IT-14:55-63.

Lavergnea, P. and V. Patilea, 2008. Breaking the curse of dimensionality in nonparametric testing. *Journal of Econometrics* 143 (1):103–122.

Melgani, F., and L. Bruzzone, 2004. Classification of Hyperspectral Remote Sensing Images With Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8), 1778 – 1790.

Mojaradi, B., H. Abrishami-Moghaddam, M. J. Valadan Zoej, R. P. W. Duin, 2009. Dimensionality Reduction of

Hyperspectral Data via Spectral Feature Extraction, *IEEE Transactions on Geoscience and Remote Sensing*, 47 (7): 2091-2105.

Stibor, T., J. Timmis, and C. Eckert, 2006. On the Use of Hyperspheres in Artificial Immune Systems as Antibody Recognition Regions. Lecture Notes in Computer Science 4163: 215-228.

Diani, M., N. Acito, M.Greco and G. Corsini, 2008. A New Band Selection Strategy for Target Detection in Hyperspectral Images. In I. Lovrek, R.J. Howlett, and L.C. Jain (Eds.): KES 2008, Part III, LNAI 5179, pp. 424 – 431. Springer-Verlag Berlin Heidelberg.