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.
If the objects of representation are the pixels in an 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 $2r$. 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 $\frac{(2r)^d \pi^{d/2}}{d\Gamma\left(\frac{d}{2}\right)}$. Therefore, the proportion of the volume of the square that is inside the sphere is

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

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 on 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, Lavergne 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.

![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).](image)

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.
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 DISCUSSION

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.

Figure 3: AHS imagery of the university campus at Alcala de Henares.
Table 1: Results with the training of the Mahalanobis algorithm for 10% of the ROI (168 pixels) and 100% of the ROI (1681 pixels) of training for the image Figure 2a

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

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

<table>
<thead>
<tr>
<th># BANDS</th>
<th>15% (30p)</th>
<th>14% (28p)</th>
<th>12% (24p)</th>
<th>11% (22p)</th>
<th>10% (20p)</th>
<th>7% (14p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>68.9670</td>
<td>69.0165</td>
<td>69.6154</td>
<td>68.0879</td>
<td>66.0604</td>
<td>68.0220</td>
</tr>
<tr>
<td>3</td>
<td>73.8915</td>
<td>74.5343</td>
<td>74.4217</td>
<td>72.3997</td>
<td>70.1621</td>
<td>71.4753</td>
</tr>
<tr>
<td>4</td>
<td>77.6044</td>
<td>78.4940</td>
<td>77.9925</td>
<td>75.6119</td>
<td>73.3492</td>
<td>74.4241</td>
</tr>
<tr>
<td>5</td>
<td>80.3064</td>
<td>81.4211</td>
<td>80.7205</td>
<td>78.0262</td>
<td>75.8721</td>
<td>76.8427</td>
</tr>
<tr>
<td>6</td>
<td>82.2056</td>
<td>83.7208</td>
<td>82.9231</td>
<td>79.7829</td>
<td>77.5846</td>
<td>78.5491</td>
</tr>
<tr>
<td>7</td>
<td>83.5061</td>
<td>85.6482</td>
<td>84.7120</td>
<td>81.0256</td>
<td>78.4744</td>
<td>79.3632</td>
</tr>
<tr>
<td>8</td>
<td>84.4478</td>
<td>87.3039</td>
<td>86.1325</td>
<td>81.8205</td>
<td>78.5821</td>
<td>79.2244</td>
</tr>
<tr>
<td>9</td>
<td>85.1054</td>
<td>88.7890</td>
<td>87.2403</td>
<td>82.2268</td>
<td>78.0704</td>
<td>77.8724</td>
</tr>
<tr>
<td>10</td>
<td>85.325</td>
<td>90.0909</td>
<td>88.0979</td>
<td>82.3092</td>
<td>77.1978</td>
<td>75.2468</td>
</tr>
<tr>
<td>11</td>
<td>85.8063</td>
<td>91.3049</td>
<td>88.7225</td>
<td>82.0632</td>
<td>76.0852</td>
<td>71.1909</td>
</tr>
<tr>
<td>12</td>
<td>85.9725</td>
<td>92.4121</td>
<td>89.2308</td>
<td>81.3571</td>
<td>74.9176</td>
<td>65.1978</td>
</tr>
<tr>
<td>13</td>
<td>86.1071</td>
<td>93.3214</td>
<td>89.5357</td>
<td>80.4643</td>
<td>72.9643</td>
<td>58.5714</td>
</tr>
<tr>
<td>14</td>
<td><strong>86.5000</strong></td>
<td><strong>94.5000</strong></td>
<td><strong>90.0000</strong></td>
<td><strong>77.5000</strong></td>
<td><strong>68.5000</strong></td>
<td><strong>53.5000</strong></td>
</tr>
</tbody>
</table>

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

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

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

<table>
<thead>
<tr>
<th># BANDS</th>
<th>100% (2904p)</th>
<th>10% (291p)</th>
<th>1% (29p)</th>
<th>0.7% (20p)</th>
<th>0.6% (17p)</th>
<th>0.5% (14p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>78.5145</td>
<td>73.6357</td>
<td>63.1777</td>
<td>63.8502</td>
<td>64.3264</td>
<td><strong>64.3264</strong></td>
</tr>
<tr>
<td>4</td>
<td>81.4697</td>
<td>76.9170</td>
<td>63.5751</td>
<td>64.1105</td>
<td>64.6019</td>
<td><strong>64.6019</strong></td>
</tr>
<tr>
<td>5</td>
<td>83.6388</td>
<td>80.2218</td>
<td>63.5888</td>
<td>64.4797</td>
<td>64.1570</td>
<td><strong>64.1570</strong></td>
</tr>
<tr>
<td>6</td>
<td>85.4236</td>
<td>81.9890</td>
<td>63.7166</td>
<td>64.9132</td>
<td>64.9298</td>
<td><strong>64.9298</strong></td>
</tr>
<tr>
<td>7</td>
<td>86.4056</td>
<td>82.9277</td>
<td>64.7693</td>
<td>67.4366</td>
<td>65.6780</td>
<td><strong>65.6780</strong></td>
</tr>
<tr>
<td>8</td>
<td>88.0207</td>
<td>84.3244</td>
<td>65.8970</td>
<td>68.7307</td>
<td>65.8543</td>
<td><strong>65.8543</strong></td>
</tr>
<tr>
<td>9</td>
<td>89.0551</td>
<td>85.5275</td>
<td>66.8271</td>
<td>70.9632</td>
<td>66.1546</td>
<td><strong>66.1546</strong></td>
</tr>
<tr>
<td>10</td>
<td>89.6481</td>
<td>85.8957</td>
<td>66.9508</td>
<td>73.3877</td>
<td>67.0351</td>
<td><strong>67.0351</strong></td>
</tr>
<tr>
<td>11</td>
<td>90.3309</td>
<td>86.6236</td>
<td>67.7056</td>
<td>75.9136</td>
<td>68.0795</td>
<td><strong>68.0795</strong></td>
</tr>
<tr>
<td>12</td>
<td>90.8243</td>
<td>87.2238</td>
<td>68.6825</td>
<td>77.5478</td>
<td>69.7897</td>
<td><strong>69.7897</strong></td>
</tr>
<tr>
<td>13</td>
<td>91.3002</td>
<td>87.7755</td>
<td><strong>69.2591</strong></td>
<td>79.0388</td>
<td>71.7016</td>
<td><strong>71.7016</strong></td>
</tr>
<tr>
<td>14</td>
<td>91.8388</td>
<td>88.2920</td>
<td>69.2149</td>
<td>80.0620</td>
<td><strong>74.0014</strong></td>
<td><strong>74.0014</strong></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th># BANDS</th>
<th>0.5% (14p)</th>
<th>0.25% (8p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>63.2428</td>
<td>50.6505</td>
</tr>
<tr>
<td>4</td>
<td>63.2765</td>
<td>54.3189</td>
</tr>
<tr>
<td>5</td>
<td>63.8626</td>
<td>58.2269</td>
</tr>
<tr>
<td>6</td>
<td>65.5771</td>
<td>58.4298</td>
</tr>
<tr>
<td>7</td>
<td>66.3109</td>
<td>60.8681</td>
</tr>
<tr>
<td>8</td>
<td>68.3994</td>
<td>61.2868</td>
</tr>
<tr>
<td>9</td>
<td>68.0758</td>
<td>63.7252</td>
</tr>
<tr>
<td>10</td>
<td>69.3182</td>
<td>65.4043</td>
</tr>
<tr>
<td>11</td>
<td>70.2989</td>
<td>64.4022</td>
</tr>
<tr>
<td>12</td>
<td>70.6463</td>
<td>65.3268</td>
</tr>
<tr>
<td>13</td>
<td>71.9229</td>
<td>66.5193</td>
</tr>
<tr>
<td>14</td>
<td>71.4628</td>
<td>66.5909</td>
</tr>
<tr>
<td>15</td>
<td>72.0957</td>
<td>68.2738</td>
</tr>
<tr>
<td>16</td>
<td>72.6942</td>
<td>68.6546</td>
</tr>
</tbody>
</table>

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 $30 \times 30$, 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

<table>
<thead>
<tr>
<th># BANDS</th>
<th>MAHALANOBIS DISTANCE</th>
<th>SUPPORT VECTOR MACHINE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAX</td>
<td>MEAN</td>
</tr>
<tr>
<td>3</td>
<td>83.4022</td>
<td>66.7087</td>
</tr>
<tr>
<td>4</td>
<td>82.7824</td>
<td>69.8526</td>
</tr>
<tr>
<td>5</td>
<td>83.4366</td>
<td>72.3186</td>
</tr>
<tr>
<td>6</td>
<td>84.2975</td>
<td>74.4434</td>
</tr>
<tr>
<td>7</td>
<td>84.8485</td>
<td>76.3276</td>
</tr>
<tr>
<td>8</td>
<td>85.0895</td>
<td>77.9865</td>
</tr>
<tr>
<td>9</td>
<td>85.1584</td>
<td>79.4194</td>
</tr>
<tr>
<td>10</td>
<td>85.0551</td>
<td>80.6596</td>
</tr>
<tr>
<td>11</td>
<td>84.8140</td>
<td>81.7205</td>
</tr>
<tr>
<td>12</td>
<td>84.7107</td>
<td>82.5871</td>
</tr>
<tr>
<td>13</td>
<td>84.5041</td>
<td>83.2743</td>
</tr>
<tr>
<td>14</td>
<td>83.8843</td>
<td>83.8843</td>
</tr>
</tbody>
</table>
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


