

Supervised classification of hyperspectral images using a combination of spectral and spatial information

Dirk Borghys, Michal Shimoni, Christiaan Perneel

Royal Military Academy, Signal & Image Centre,
Renaissancelaan 30, B-1000 Brussels, Belgium

ABSTRACT

This paper describes a new method for classification of hyperspectral images for extracting cartographic objects. The developed method is intended as a tool for automatic map updating. The idea is to use an existing map of the region of interest as a learning set. The proposed method is based on logistic regression. Logistic regression (LR) is a supervised multi-variate statistical method that finds an optimal combination of the input channels for distinguishing one class from all the others. LR thus results in detection images per class. These can be combined into a classification image. The LR method that is used here is a step-wise optimisation that also performs a channel selection. The results of the LR are further improved by taking into account spatial information by means of a region growing method. The parameters of the region growing are optimised for each class of interest. For each class the optimal set of parameters is determined. The method is applied on a HyMap hyperspectral image of an area in Southern Germany and the results are compared to those of classical methods. For the comparison a ground truth image was created by combining data from a cadaster map and a digital topographic map.

Keywords: hyperspectral image classification, region growing, logistic regression, channel reduction

1. INTRODUCTION

For many remote sensing applications e.g. emergency cartography, disaster monitoring, damage assessment, etc., it is important to be able to quickly obtain an overview of the current situation. In particular the detection and classification of man-made objects is very important. This paper presents part of the work done in a project that aims to classify man-made objects using hyperspectral images and to investigate the complementarity between hyperspectral and SAR data. The paper describes a new method for classification of hyperspectral data for updating maps. The proposed method uses a statistical method - logistic regression - for channel selection and to obtain a first classification. This classification is then improved using spatial information in a region growing approach.

The method is supervised in the sense that existing digital maps are used. The maps serve as the sole basis to construct the learning and validation databases. The idea is to avoid having to undertake an extensive ground truth mission for constructing the learning set. The classification is based on a multi-variate statistical technique called logistic regression¹ (LR). LR combines the different channels in a way that optimises the distinction between one class and all the others. It thus results in a detector for each class. The detection results are then combined into a classification image. The LR uses a step-wise method that only adds a channel to the used set of channels if the improvement in detection caused by this addition is statistically significant. LR thus implicitly performs a channel selection. The results of the detection and channel selection methods are compared to those that are obtained by conventional hyperspectral methods. The developed method does not require the use of laboratory spectra or extensive ground truth. It only requires an existing map on which a limited number of points are identified for learning. In the conventional method for hyperspectral image classification, the first step is a channel reduction based on Minimum Noise Fraction² (MNF) or Principal Component Analysis (PCA). These methods use the statistical variation in the data set but they do not take into account information about the classes of interest to lead the channel selection. In order to compare the results of our channel selection with the one obtained by MNF and PCA, a matched filter³ (MF) is applied to

Email D. Borghys: Dirk.Borghys@elec.rma.ac.be

detect each of the classes of interest on the channels selected by the different methods. The results of these matched filters is also compared to the results of the LR.

In the literature on classification of hyperspectral images, often pixel-wise methods are used because one is interested in classes of very small spatial extent (typically 1 pixel or less). As we are interesting in cartographic applications of the hyperspectral images, the classes of interest are much larger and cover several pixels. Therefore it is possible to use spatial information for improving the detection results of the LR as well as the classification results. This is the purpose of the region growing. The idea is to select reliable starting points for each of the classes and to grow these seed regions by comparing the spectra inside the already selected regions to neighbouring pixels. The comparison of spectra is based on spectral distance measures. These are also used in classification tools such as the various minimum distance classifiers (MD). Commonly used MD's are based on the Euclidean or Mahalanobis distance or on the spectral angle. However, in MD the spectra of pixels to be classified are compared to the average spectrum of each class, as calculated from the complete learning set. In the approach we propose, the average spectrum is estimated locally using the pixels that were already classified with high degree of certainty. The starting pixels for the region growing are found by applying a threshold to the detection image obtained from the LR. The region growing gradually merges surrounding pixels to these starting kernels. In the paper we investigate a number of parameters for the region growing and for each class determine the best combination. Considered parameters are the threshold T1, the spectral distance measure used as a growing criterion and the normalisation method that is applied to the spectra prior to comparing them. Evaluation of results for the different combination of parameters is based on a Figure-of-Merit (FOM) for target detection. For the evaluation the complete validation set is used.

2. THE DATASET

2.1. Hyperspectral data

For this project HyMap data were acquired over the villages of Oberpfaffenhofen and Neugilching in the South of Germany. The HyMap data were acquired by the German Aerospace Agency DLR and contain 126 contiguous bands ranging from the visible region to the short wave infrared (SWIR) region ($0.45-2.48\mu m$). The bandwidth of each channel is 15-18 nm. The spatial resolution is 4 m at nadir and the image covers an area of $2.6 \times 9.5 km$. Four subsequent levels of pre-processing were applied: radiometric correction, geocoding, atmospheric correction (using ATCOR4⁴) and the "Empirical Flat Field Optimized Reflectance Transformation" (EFFORT⁵). The pre-processing was done by the Flemish Institute for Technological Development (VITO). We received the data corresponding to the four different levels of pre-processing and for the moment only the last level is used. Later we intend to apply our methods to the different levels of pre-processed data in order to study the degradation of the classification results when less pre-processing steps are applied. One of the final aims of the project is to develop algorithms that require the least amount of pre-processing.

In all processing the first and last channel were discarded. The first contains too much noise while the last is saturated.

2.2. Ground truth data

For this site we also acquired cadastral data (obtained from the "Bayerischen Vermessungsverwaltung") of a part of the village of Neugilching as well as digital topographic maps of the surroundings (ATKIS data from the "Bayerisches Landesvermessungsamt"). These map data serve as a basis for building the learning and validation set. The ground truth was acquired for a $2 \times 1 km$ part of the image located in a residential area.

The topographic and cadaster maps were also used to construct the learning set for the logistic regression. For each of the six classes of interest, i.e. Roads, Buildings, Paths, Railways, Forest and Background (other vegetation), approximately 150 points were selected to constitute the learning set. In classical hyperspectral classification methods, e.g. the spectral angle mapper⁶ (SAM) or spectral unmixing methods,⁷ the classes above would be too general. A class like "Buildings" would for instance have to be sub-divided according to the color of the roofs of the buildings. This would require a ground truth mission or an extensive image interpretation by a human expert. As our intention was to use only information from existing maps, we can not create detailed classes and didn't use spectral unmixing or the SAM.

Figure 1 shows the dataset used in this paper. The left image shows part of the HyMap dataset in RGB color combination, the right image is the ground truth image used for validation. The railway on the image is in fact a tramway (an S-Bahn).

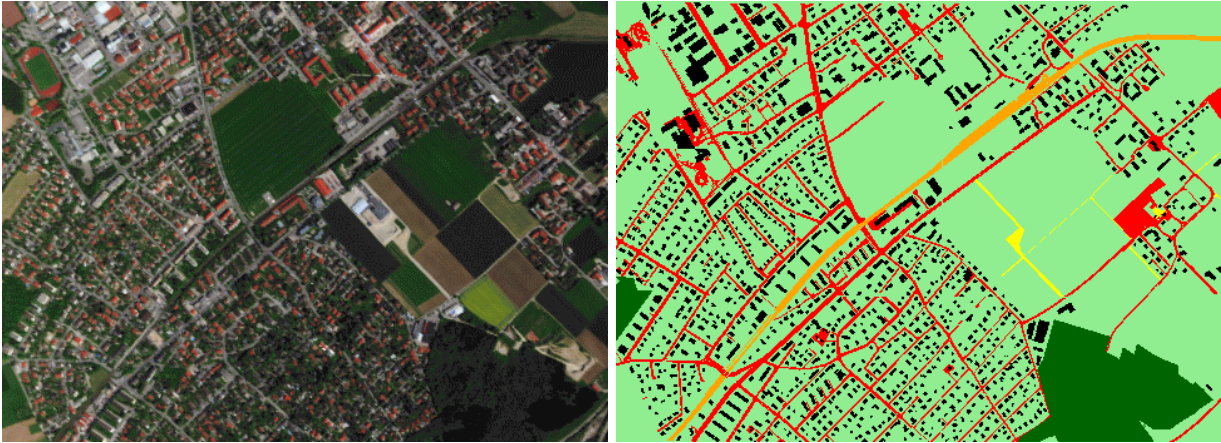


Figure 1. The dataset used in this paper. Left: HyMap image (R=635nm, G=558nm, B=497nm), Right: Validation image (Red: Roads, Black: Buildings, Yellow: Paths, Orange: Railway, Dark Green: Forest, Light Green: Background)

3. CLASSIFICATION METHOD

3.1. Overview

The first step in the classification is based on logistic regression. The logistic regression results in a detection image per class. The values in the detection image for a given class correspond to the conditional probability that the considered pixel belongs to that class. The detection images can thus be combined into a classification image by assigning to each pixel the class for which the value in the detection image is highest. However, in our approach we apply a threshold to the detection images for each class, giving the pixels that have a high probability to belong to the class. These pixels are then used as starting regions for a region growing. The region growing is based on spectral distance measures and exploits the local properties for the already selected regions.

3.2. Logistic Regression

Logistic regression¹ (LR) is developed for dichotomous problems where a target class has to be distinguished from the background. The method combines the input features - the input channels in this case - into a non-linear function, the logistic function, defined as:

$$p_{x,y} \left(target \mid \vec{C} \right) = \frac{\exp [\beta_o + \sum_i C_i(x,y)\beta_i]}{1 + \exp [\beta_o + \sum_i C_i(x,y)\beta_i]} \quad (1)$$

$p_{x,y} \left(target \mid \vec{C} \right)$ is the conditional probability that a pixel (x,y) belongs to the considered class (target class) given the vector of input channels (\vec{C}) at the given pixel. $C_i(x,y)$ is the value of pixel (x,y) in the i^{th} channel of the dataset.

The LR finds a combination of the input channels that approximates optimally the $p_{x,y} \left(target \mid \vec{C} \right)$ for each class, based on learning data.

The logistic regression (i.e. the search for the β_i 's) is carried out using Wald's forward step-wise method using the commercial statistics software SPSS. In the Wald method, at each step, the most discriminant channel is added and the significance of adding it to the model is verified. This means that only the channels that

contribute significantly to the discrimination between the foreground and the background class are added to the model. The logistic regression thus gives an optimal combination of a sub-set of input parameters for separating one class from all others, implicitly performing a channel selection.

3.3. Region Growing

The approach until now is purely pixel based, i.e. only the information contained in a pixel is used to classify the pixel. In the literature on classification of hyperspectral images, often pixel-wise methods are used because one is interested in classes of very small spatial extent (typically 1 pixel or less). As we are interested in cartographic applications of the hyperspectral images, the classes of interest are much larger and cover several pixels. Therefore it is possible to use spatial information for improving the classification results. This is the purpose of the region growing (RG). The idea is to select reliable starting points for each of the classes and to grow these seed regions by comparing the spectra inside the already selected regions to neighbouring pixels. The comparison of spectra is based on spectral distance measures. These are also used in classification tools such as the minimum distance classifier (MD). However, in MD the spectra of pixels to be classified are compared to the average spectrum of each class, as calculated from the complete learning set. In the approach we propose, the average spectrum is estimated locally using the pixels that were already classified with high degree of certainty.

The starting regions are obtained by thresholding the probability images resulting from the logistic regression. In these images the different selected regions are labeled. The idea is to try and expand the labeled regions taking into account the local properties of the regions that are already classified. The principle of the region growing is illustrated in fig. 2.

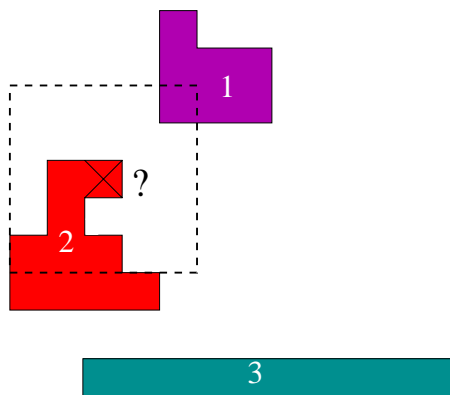


Figure 2. Principle of the region growing.

The image is scanned and as soon as a labeled pixel is encountered the region grower starts. First the average spectrum is determined in a rectangular neighbourhood around the starting pixel (the dashed rectangle around the selected pixel in region 2 in the figure). Only pixels belonging to the same region are considered in the calculation. For these pixels the maximal distance D_{max} between each of the pixel's spectra and the average spectrum is also determined. This is used as a threshold for deciding whether a neighbouring pixel should be added to the region or not. Pixels adjacent to the selected pixel that do not yet belong to the region are investigated. The spectral distance D from their spectrum t to the average spectrum r of the region is determined and if this distance is below the threshold (given D_{max} multiplied by a factor DF), the pixel is added to the region:

$$\begin{aligned}
 D(t, r) < DF * D_{max} & \text{ add the pixel} \\
 > DF * D_{max} & \text{ do not add the pixel}
 \end{aligned}
 \tag{2}$$

The average spectrum and maximal distance are constantly re-evaluated while the region expands. As a result of the region growing we save the distance measure map. In this way a classification image can be obtained by assigning to each pixel the class for which this distance is minimal. If only a binary result of the RG would be stored, conflicting classifications are more likely to occur.

3.4. Optimisation of the Region Growing Parameters

The result of the region growing depends on different parameters. We were interested to find the best selection of these parameters for detecting each of the classes of interest. Considered parameters are:

- T1: the threshold applied on the LR detection image to create the starting regions of the region growing
- D: the spectral distance measure that was used to compare two spectra in the region growing (sect. 3.5).
- N: the normalisation that was applied to the spectra before performing the region growing. (sect. 3.6).
- DF: the multiplication factor for the threshold on spectral distance in the region growing

A figure of merit for target detection (sect. 4) is used as optimisation parameter.

3.5. Spectral Distance and Similarity Measures

Several spectral distance or similarity measures are proposed in literature. They all measure the distance or similarity between two spectra, a reference spectrum 'r' and test spectrum 't'. The ones used in the current paper are listed below. In the expressions below N is the number of bands.

- *Euclidean Distance* The euclidean distance is defined as:

$$D_{eucl}(t, r) = \sqrt{\sum_{i=1}^N (t_i - r_i)^2} \quad (3)$$

- *Hamming Distance*

The Hamming distance D_{ham} is used in the classification method 'binary encoding'. It determines the average value of the spectra and then counts the number of channels for which both spectra are either above or below their respective averages.

- *Spectral Angle*

The spectra are represented as vectors in a N-dimensional space where N is the number of bands. The distance measure is the spectral angle⁶ between the vectors.

$$D_{sam}(t, r) = \cos^{-1} \left(\frac{\vec{t} \bullet \vec{r}}{\|\vec{t}\| \|\vec{r}\|} \right) = \cos^{-1} \left(\frac{\sum_{i=1}^N t_i r_i}{\left(\sum_{i=1}^N t_i^2 \right)^{1/2} \left(\sum_{i=1}^N r_i^2 \right)^{1/2}} \right)$$

- *Spectral correlation similarity*

This is based on the Pearson statistical correlation. It shows how two vectors are linearly correlated and should thus be independent of the strength of illumination. It is defined as:

$$\rho(t, r) = \frac{\sum_{i=1}^N t_i r_i - \sum_{i=1}^N t_i \sum_{i=1}^N r_i}{\sqrt{\left[\sum_{i=1}^N r_i^2 - \left(\sum_{i=1}^N r_i \right)^2 \right] \left[\sum_{i=1}^N t_i^2 - \left(\sum_{i=1}^N t_i \right)^2 \right]}} \quad (4)$$

For the region growing distance measure, $D_{cor}(t, r) = 1 - \rho(t, r)$ is used.

3.6. Normalisation Methods

For the parameter “spectral normalisation type” (NT) we considered the non-normalised spectra (NT=0) and two normalisation methods:

- *MinMax Normalisation (NT=1)*

This normalisation rescales the spectrum in each pixel between the minimum and maximum value. Its effect is a reduction of the influence of illumination.

- *Continuum Removal (NT=2)*

For continuum removal the convex hull of the spectrum is determined. The spectrum is rescaled between 0 and the convex hull. The effect of this normalisation is a magnification of small differences in the spectra, i.e. small absorption peaks are high-lighted.

4. EVALUATION METHOD

The evaluation of the results is based on the ground truth image shown on the right in fig. 1. The detection results were compared using ROC curves. These are plots of the probability of detection (Pd) versus the probability of false alarms (Pf).

The optimisation parameter for determining the best combination of parameters in the region growing is a figure-of-merit for target detection⁸ defined as:

$$FOM = \frac{N_{DT}}{N_{FT} + N_{TT}} \quad (5)$$

where N_{DT} is the number of correctly detected target pixels, N_{FT} the number of false alarm pixels and N_{TT} the number of true target pixels actually present in the image.

5. RESULTS AND DISCUSSION

5.1. Channel Selection

Fig. 3 shows the average spectra for the six classes as determined from the learning set. It also shows the results of the channel selection performed by the logistic regression as short lines above and below the spectra. The color of the lines corresponds to the legend. The number of channels NC selected for the different classes is given in table 1.

Class	NC	Class	NC	Class	NC
Roads	24	Buildings	19	Paths	9
Railways	16	Forests	2	Background	8

Table 1. Number of channels selected by LR for each class

Roads, buildings and railways are apparently the most difficult to detect as they require the largest number of channels. For forests only 2 channels are selected. It is the intention to find a physical explanation for the channel selection for each class. This is a topic for further work.

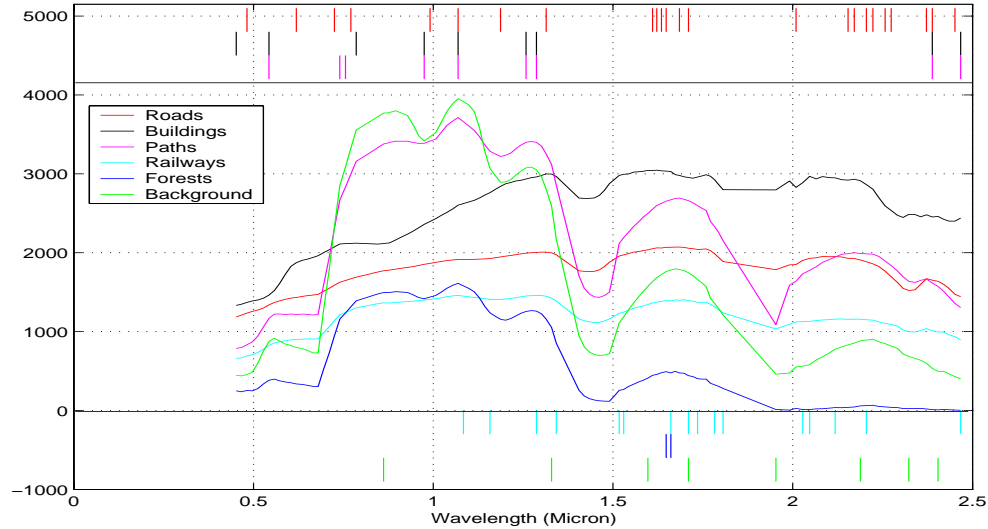


Figure 3. Spectra derived from the learning set with the channels selected by LR for each of the classes.

5.2. Detection Results

In order to evaluate the detection results of the logistic regression based approach, we compare it with the results of matched filtering. In fig. 4, for each class of interests, several ROC curves are shown:

1. The result after logistic regression
2. The result after MF using the channels selected by LR
3. The result of MF after PCA, using the channels corresponding to $> 99.99\%$ of the information (21 channels)
4. The results of MF after inverse PCA with the 22 channels of the PCA and using the channels given by the LR. Note that the IPCA reconstructs a hyperspectral image with 124 channels, but reduces the noise in the different channels.
5. The results of MF after MNF, using the channels corresponding to eigenvalues larger than 1 (30 channels)

For the calculation of MNF and PCA and for applying the MF, the commercial software ENVI⁵ was used. From fig. 4 it appears that for roads, buildings, railways and background, the method based on logistic regression gives the best results. For paths the results of matched filtering after PCA and MNF are better than all those where the channel selection provided by LR is used. For paths, the channel selection provided by LR is therefore inappropriate. For forests, MF after MNF gives the best results but the LR results and the two other methods where the LR selected channels are used, are much better than the results of MF after PCA.

5.3. Results of the Region Growing

The results of the region growing using different distance measures and different normalisation methods are compared in an objective manner, using the FOM, calculated from the validation image. Fig. 5 shows the results of the optimisation. Each plot represents the results for one class for a fixed value of $T1 = 0.95$ except for the class forests where $T1 = 0.55$ in order to have a reasonable detection level. In the plot the x-axis represents the 4 spectral distances and the y-axis is the FOM. The three normalisation methods are represented by different symbols (\circ : NT=0, $+$: NT=1, \square : NT=2). The parameter DF was varied from 1 to 4. This is represented by different colors as well as a small horizontal shift, i.e. DF=1 is the most left for each spectral distance, while DF=4 is the most to the right. It was verified that the choice of T1 doesn't change the appearance

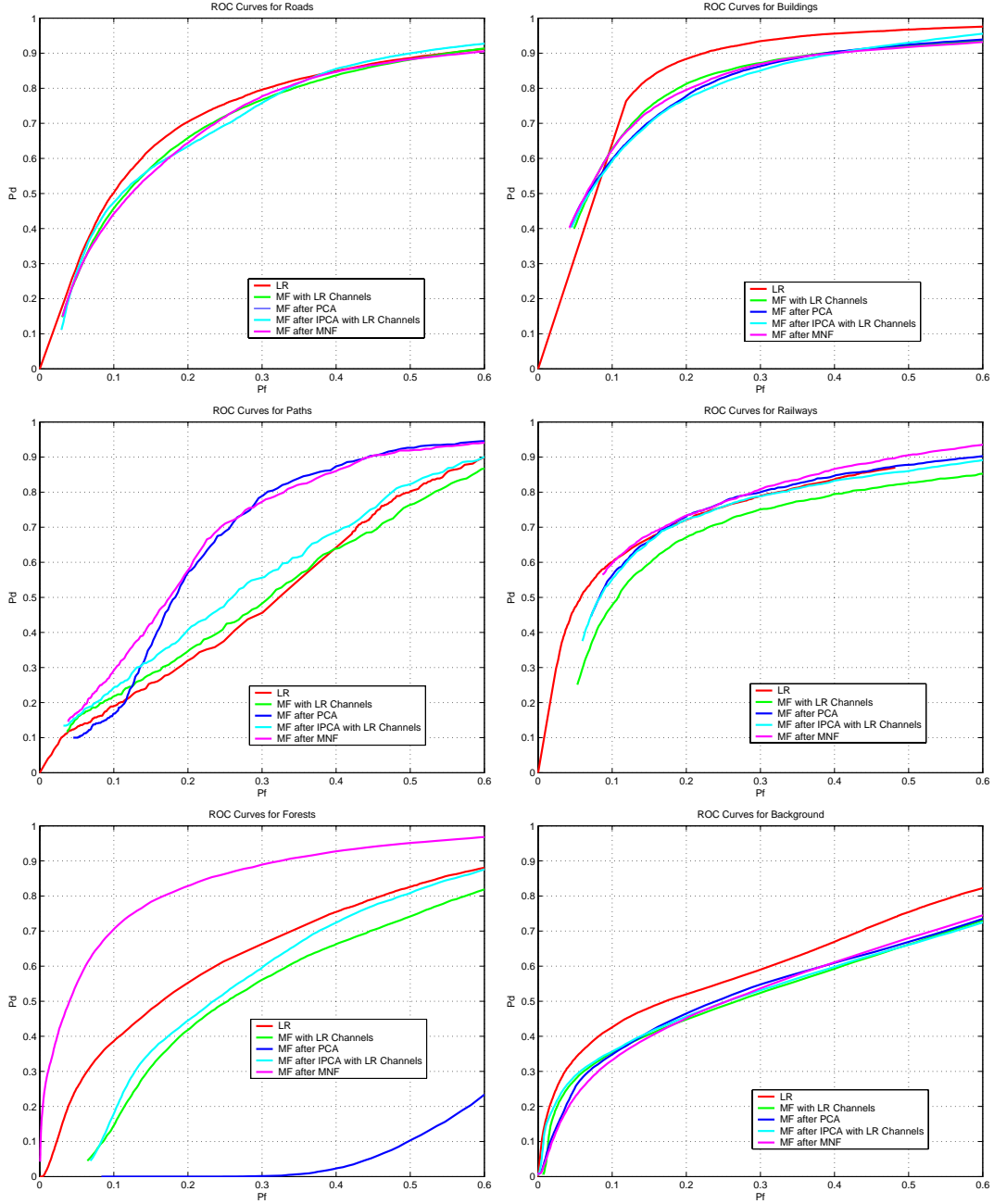


Figure 4. ROC curves for the different classes of interest for the different detection methods

of the plots; only the scale of the FOM changes when T_1 is changed. The best choice for T_1 depends on the class.

It is very hard to draw any general conclusions from the figures. The obtained FOM for the class “paths” is very low. This probably means that the method based on logistic regression is not appropriate for this class. We therefore discard the class in the discussion on the parameters of the region growing.

Concerning the choice of the spectral distance measure, the Hamming distance generally leads to the worst results. The distance measures based on spectral angle (D_{sam}) and correlation (D_{cor}) are equivalent for all

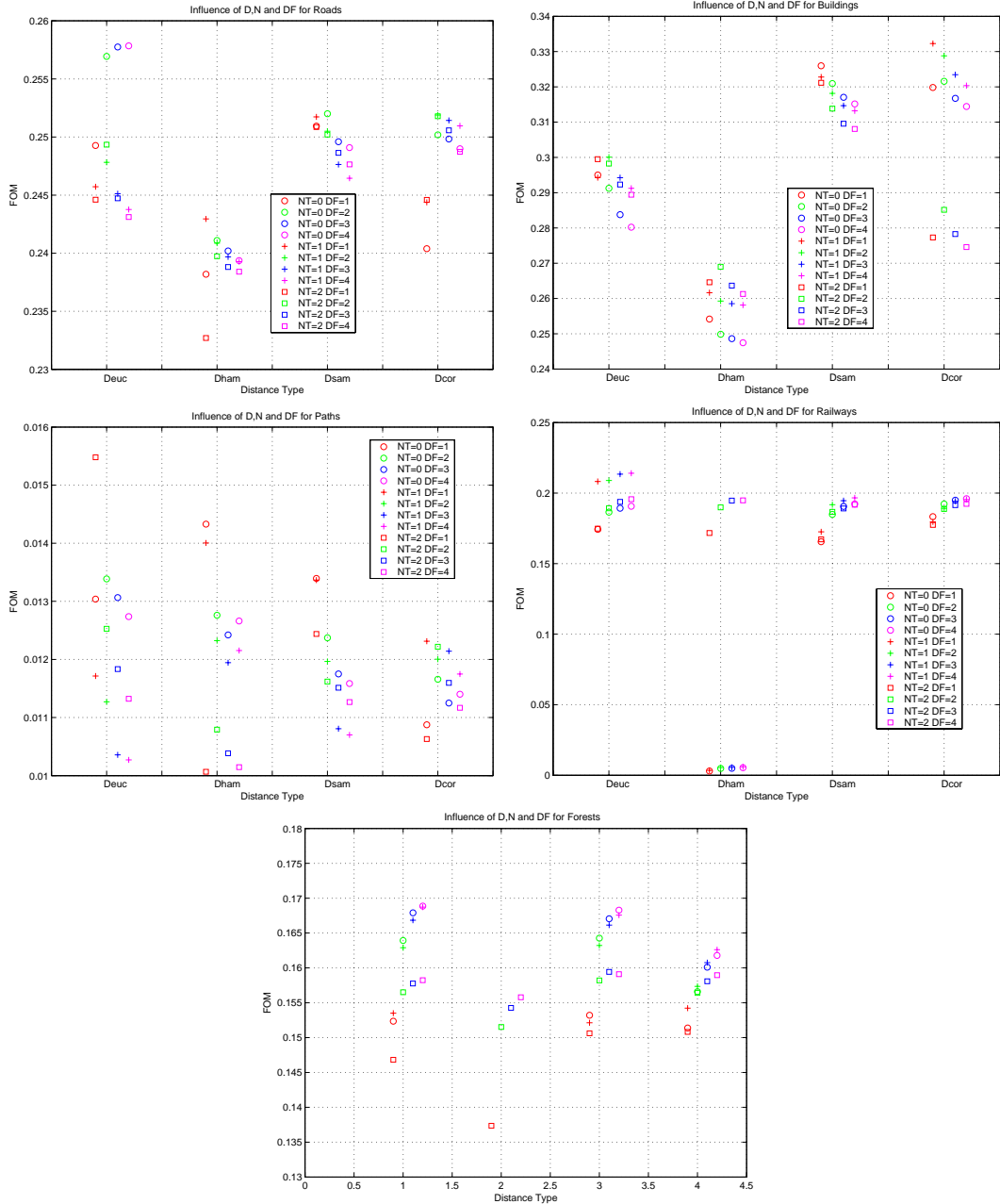


Figure 5. Results of the optimisation of the region growing for the different classes

classes. The Euclidean distance (D_{euc}) is more sensitive to the choice of the other parameters than the D_{sam} and D_{cor} . The best normalisation method depends on the spectral distance that is used as well as on the considered class. In many cases however, the continuum removal leads to lesser results. The results without normalisation and with the MinMax normalisation, are equivalent. The influence of DF depends mainly on the class: for buildings the best results are obtained for a low DF while for railways and forests the best results are obtained for a high DF. This means that the initial points selected by the logistic regression are representative for the spectra of the buildings, but they are too restricted for forests and railways. For roads the best results are obtained for DF=2. These results will need to be analysed further.

Class	T1	Class	T1	Class	T1
Roads	0.80	Buildings	0.95	Paths	0.50
Background	0.50	Railways	0.80	Forests	0.55

Table 2. Values of T1 used in the region growing for the different classes

5.4. Classification Results

For combining the results of the region growing into a classification image based on the images of spectral distance, the same set of parameters has to be applied for each class. Because no single set of optimal parameters can be identified for all classes, we ranked the best 10 combinations of parameters for each class and selected the combination that appeared the most frequently. The resulting combination was (D_{sam} , NT=0, DF=2). This was applied for constructing the classification image after region growing. For T1 the best value was determined for each class. Table 2 lists the values for T1 used for the different classes.

The result was compared to that of the logistic regression as well as a classification using the “rule-images” of the matched filter after MNF and PCA. Figure 6 shows the results of the classification for the different methods. It appears from the figure that the proposed methods (LR and LR+RG) are both better than the matched filter after MNF and after PCA. The PCA has many false alarms of the class railways, while MNF also has many false alarms for paths. Results of LR and RG appear similar. In the RG results the grey color is “unclassified”. Note that the LR presents many false alarms due to the classes forests and paths between the houses in the village. As it is likely to find trees and paths in gardens, some of these might not be false alarms.

In order to obtain a quantitative evaluation, in table 3 the User’s (UA) and Producer’s accuracy (PA) for the different classes and methods are presented. PA is the number of pixels correctly classified as a given class to the total number that actually belongs to that class. PA is thus related to the probability of detection. UA is the ratio of the number of pixels correctly classified as a given class to the total number of pixels classified as that class. UA is related to the complement of the probability of false alarms. The table confirms that the proposed methods are indeed better than the methods based on the matched filter. The region growing slightly improves UA while it decreases PA for most classes. Contrary to what one could expect, the region growing thus decreases the number of false alarms at the expense of the probability of detection. Note that after the region growing most false alarms of paths and forests in the built-up area have become “unclassified”.

Class	Method							
	LR		Region Growing		MF after PCA		MF after MNF	
	PA	UA	PA	UA	PA	UA	PA	UA
Roads	0.5464	0.3729	0.4823	0.3851	0.3558	0.2611	0.5506	0.3182
Buildings	0.8486	0.3530	0.8128	0.3369	0.3624	0.4505	0.3305	0.4034
Paths	0.3438	0.0087	0.1828	0.0090	0.1587	0.0083	0.3175	0.0102
Railways	0.5368	0.1345	0.5863	0.2317	0.7790	0.0443	0.7484	0.0510
Forests	0.6307	0.2313	0.3138	0.3808	0.6368	0.0142	0.1226	0.0038
Background	0.3417	0.9591	0.2898	0.9645	0.2738	0.9674	0.2838	0.9593

Table 3. Evaluation of the classification results: UA and PA for the different methods for each class

5.5. Further Work

Several topics for further work are foreseen:

- Establish a physical explanation for the channel selection that is found by the logistic regression for each of the classes.
- Investigate other spectral distance measures for the region growing, e.g. the Mahalanobis and Battacharya distance or the Spectral Information Divergence.⁹



Figure 6. Classification results: upper-left: results of logistic regression upper-right: region growing results, lower-left: MF after PCA, lower-right: MF after MNF

- Investigate the influence of the pre-processing level on the results of the logistic regression
- Examine the results of the logistic regression when the parameters are learned in one region and the functions are applied to a nearby region.

6. CONCLUSIONS

A new method for the classification of hyperspectral images is presented. The aim of the method is to extract cartographic objects from the images for map updating purposes. The classification is based on logistic regression and uses existing digital maps for learning. The logistic regression gives a detection image per class. These can then be combined into a classification image. The article also investigates the improvement of the detection results by region growing based on various spectral distances. The influence of the different parameters of the region growing on the detection results were investigated. However it was not possible to identify a single best combination of parameters. It was shown that the optimal set of parameters for the region growing depends on the considered class as well as on the type of spectral distance that is used. For most classes the spectral distance based on the spectral angle gave the best results. The classification by logistic regression was compared to the results of rule-image classification of the matched filter results after MNF and PCA. The LR results were significantly better.

ACKNOWLEDGMENTS

The research presented in this paper is the result of a collaboration between three projects: HYSAR: “Man-Made object classification using fused polarimetric SAR and hyperspectral imagery data”, funded by the Belgian Government, Belgian Federal Science Policy Office, in the frame of the STEREO Program (project nr. SR/00/044),

a project funded by the EC in the frame of the PASR Program: “Advanced Space Technology to Support Security Operations (ASTRO+)” and a study funded by the Belgian MoD on the semi-automatic use of hyperspectral and SAR images for general security problems (project nr. F05/02).

REFERENCES

1. D. Hosmer and S. Lemeshow, *Applied Logistic Regression*, John Wiley and Sons, 1989.
2. A. Green, M. Berman, P. Switzer, and M. Graig, “A transformation for ordering multispectral data in terms of image quality with implications for noise removal,” *IEEE-GRS* **26**(1), pp. 65–74, 1988.
3. D. Manolakis and G. Shaw, “Detection algorithms for hyperspectral imaging applications,” *IEEE SP Mag.* **19**, pp. 29–43, Jan 2002.
4. R. Richter, “Atcor-4 user guide, v4.0,” DLR, Jan 2005.
5. *ENVI User’s Manual (v4.0)*, Research Systems Inc., Boulder, 1993.
6. F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, “The spectral image processing system (sips) - interactive visualization and analysis of imaging spectrometer data,” *Remote Sensing of Environment, Special issue on AVIRIS* **44**, pp. 145 – 163, May-June 1993.
7. N. Keshava and J. Mustard, “Spectral unmixing,” *IEEE SP Mag.* **19**, pp. 44–57, Jan 2002.
8. D. Borghys, L. Sévigny, T. Sams, R. Gabler, B. Hoeltzener, P. Schwering, J. Haddon, and J. Knecht, *Multi-Sensor Image Fusion for the Detection of Targets in the Battlefield of the Future (NATO Confidential), Final report of a Cooperative Project in Image Processing conducted by RTB(SET-TG01)*., NATO, Brussels, 1999.
9. C. Chang, “An information theoretic-based measure for spectral similarity and discriminability,” in *Proc. IEEE IGARSS*, (Hamburg), June 1999.