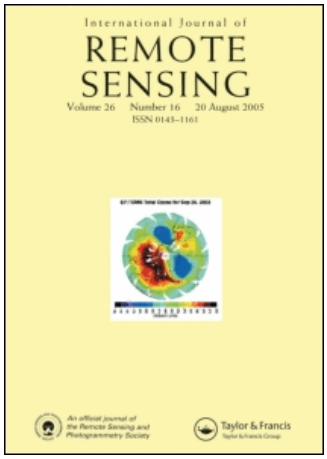

International Journal of Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713722504

The use of texture for image classification of black & white air photographs

C. M. R. Caridade [a]; A. R. S. Marçal [b]; T. Mendonça [b]
[a] Instituto Superior de Engenharia de Coimbra Rua Pedro Nunes, Portugal
[b] Faculdade de Ciências, Universidade do Porto DMA, 4169-007 Porto, Portugal

First Published on: 25 October 2007
To cite this Article: Caridade, C. M. R., Marçal, A. R. S. and Mendonça, T. (2007) 'The use of texture for image classification of black & white air photographs', International Journal of Remote Sensing, 29:2, 593 - 607
To link to this article: DOI: 10.1080/01431160701281015
URL: http://dx.doi.org/10.1080/01431160701281015

International Journal of Remote Sensing
Vol. 29, No. 2, 20 January 2008, 593–607

The use of texture for image classification of black & white air photographs

C. M. R. CARIDADE†, A. R. S. MARÇAL*‡ and T. MENDONÇA‡
†Instituto Superior de Engenharia de Coimbra Rua Pedro Nunes, Quinta da Nora 3030-199 Coimbra, Portugal
‡Faculdade de Ciências, Universidade do Porto DMA, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal

(*Received 14 August 2006; in final form 8 February 2007*)

The use of black & white (B&W) air photographs for the production of historic land cover maps can be done by image classification, using additional texture features. In this paper we evaluate the importance of a number of parameters in the image classification process based on texture, such as the window size, angle and distance used to produce the texture features, the number of features used, the image quantization level and its spatial resolution. The evaluation was performed using five photographs from the 1950s. The influence of the classification method, the number of classes searched for in the images and the post-processing tasks were also investigated. The effect of each of these parameters for the classification accuracy was evaluated by cross-validation. The selection of the best parameters was performed based on the validation results, and also on the computation load involved for each case and the end user requirements. The final classification results were good (average accuracy of 85.7%, $k=0.809$) and the method has proven to be useful for the production of historic land cover maps from B&W air photographs.

1. Introduction

The characteristics of the remote sensing datasets available for land cover mapping have improved significantly in the last few years. Also the gap between space-borne and airborne data has narrowed considerably, with a number of satellite sensors acquiring very high-resolution images and airborne sensors providing digital data, in the form of multispectral or hyperspectral images. In parallel with the ever increasing quality and diversity of new remote sensing data sources, there is a growing interest in historic data, such as military declassified satellite images and old air photographic surveys. The production of historic land cover maps is an example where black & white (B&W) photographs are a valuable, sometimes unique, source of information. The land cover maps are traditionally produced from B&W photographs by photo interpretation, which is a highly time consuming process. An alternative to the manual process of photo interpretation is to perform automatic image classification, in a similar way to what is done with multispectral images. Image classification is the process of assigning thematic labels to each image pixel.

*Corresponding author. Email: andre.marcal@fc.up.pt

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2008 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160701281015

In multispectral image classification, the spectral signature of each image pixel in the multidimensional feature space is used in the discrimination process. However, when a single panchromatic image is available (such as a greyscale image obtained by digitizing a B&W air photo) this approach is not possible. One effective form to classify greyscale images is to make use of the texture information present in the image. Texture contains significant information about the structural arrangement of objects and surfaces and their relationship with the surrounding environment. This type of approach has been used in remote sensing and many other image processing areas, such as quality inspection (Basset *et al.* 2000; Herrero *et al.* 2004), medical imaging (Gupta and Undrill 1995; Tahir *et al.* 2004) or product characterization (Carrión *et al.* 2004). In remote sensing there are various examples of the use of texture features for the classification of panchromatic satellite imagery from IKONOS (Chen and Gong 2004), Landsat ETM+ (Lu and Weng 2005) and SPOT (Sali and Wolfson 1992) or airborne data (Coburn *et al.* 2004). In all these studies, the inclusion of texture improved the final classification accuracy.

The purpose of this work is to evaluate the ability of texture based image classification methods to produce historic land cover maps from B&W air photos. The greyscale digital images used in this study were obtained by scanning a set of air photographs from the 1950s. The images are approximately 6000×6000 pixels, in eight-bit format (0–255 grey levels).

2. Texture features

The most commonly used approach for image texture analysis is based on the statistical properties of the intensity histogram. The statistical texture descriptors are calculated from the normalized Gray-Level Co-occurrence Matrix (GLCM) produced for each pixel using a neighbourhood window of $D \times D$ pixels (Haralick *et al.* 1973). The GLCM corresponds to the number of pairs of grey levels encountered in the search window, within a predefined direction and distance.

Figure 1 illustrates the process of building the GLCM for one image pixel, using a neighbourhood of 7×7 pixels. In this example the image has only three bits, thus the GLCM is an 8×8 matrix. For eight-bit images, the size of the GLCM produced for each pixel is 256×256. The GLCM for a 0° angle and a distance 1 can be thought of as the counting of the right (east) neighbouring pairs. For example, in the case presented in figure 1, there are nine pixels with value 0 with a right neighbour 0 (pair 0-0) and two pixels with value 1 with a right neighbour 2 (pair 1–2). Figure 1 also shows the GLCM for 0° angle with distance 2 and 45° angle with distance 1. The GLCM produced is then modified to assume symmetry, by adding the transposed version of the matrix, and normalized to assume that the matrix represents a probability function. The normalized symmetric GLCM matrices for the example described are presented in figure 1 (bottom line).

The eight texture descriptions used are presented in equations (1)–(8), where N is the number of grey levels, P is the normalized symmetric GLCM of dimension $N \times N$ and $P_{i,j}$ is the (i, j)th element of P (Haralick *et al.*, 1973).

- The most basic texture descriptions are the **mean** (MEAN) and **standard deviation** (SD) of the grey levels in the texture window used for each image

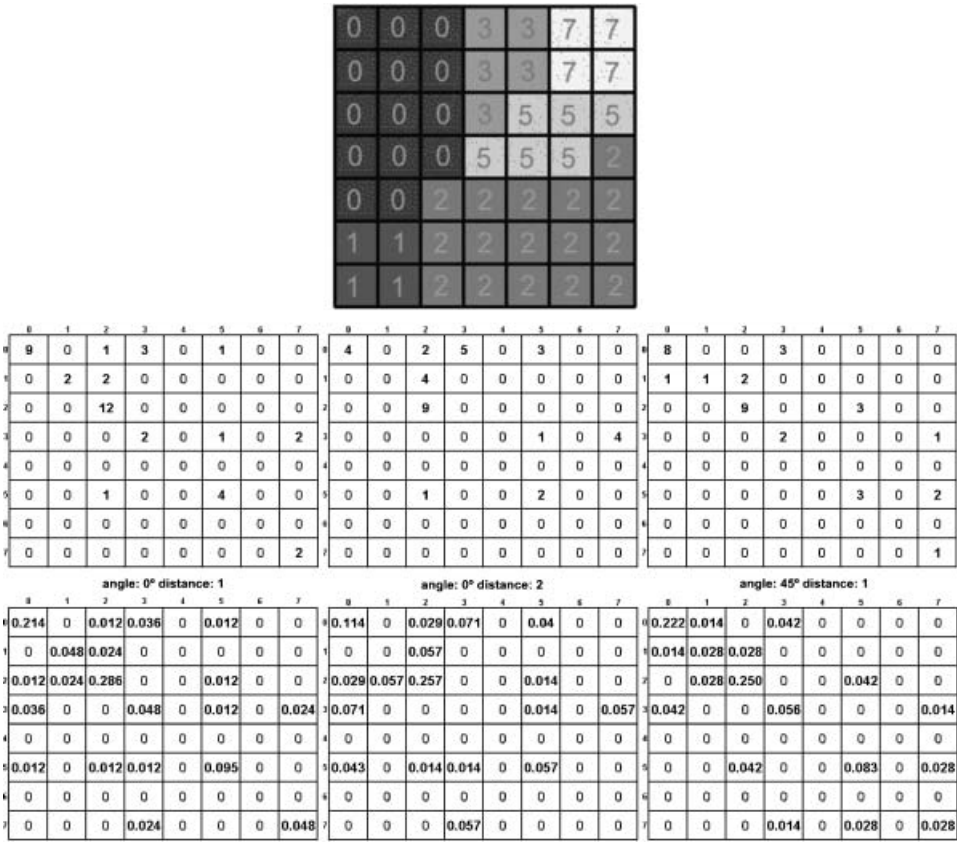

Figure 1. Example of the construction of the Gray Level Co-occurrence Matrix (GLCM) for one pixel. Original image (top); GLCM (middle); normalized symmetric GLCM (bottom).

pixel:

$$\text{MEAN} = \mu_I = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} iP_{i,j} \tag{1}$$

$$\text{SD} = \sigma_I = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j}(i-\mu_I)^2} \tag{2}$$

- The **angular second moment** (ASM) measures the local uniformity of the grey levels. In uniform images only a few transitions of grey levels exist within the texture window reaching area. That is, high values of ASM occur when the distribution of the grey level values is constant or periodic within the search window:

$$\text{ASM} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j}^2 \tag{3}$$

- The **homogeneity** (HOM) evaluates the presence of near diagonal elements in the GLCM, and results in a large value if the elements of the GLCM are

concentrated along its main diagonal:

$$\mathrm{HOM} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{4}$$

- The **contrast** (CON), **dissimilarity** (DIS) and **entropy** (ENT) measure the amount of local variation of grey levels. Small values of these variables mean that the grey levels are centred around the GLCM diagonal, otherwise there is a more even distribution of the grey levels in the GLCM:

$$\mathrm{CON} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j}(i-j)^2 \tag{5}$$

$$\mathrm{DIS} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j}|i-j| \tag{6}$$

$$\mathrm{ENT} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} -P_{i,j} \ln P_{i,j} \tag{7}$$

- The **correlation** (COR) measures the linear dependency of the grey levels. High correlation values indicate a certain local order of the grey levels:

$$\mathrm{COR} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{(i-\mu_I)(j-\mu_I)P_{i,j}}{\sigma_I^2} \tag{8}$$

Figure 2 shows an example of these eight texture features produced from one air photo. Both the original image (top left corner) and the texture images were originally in five bits (32 levels of grey), converted to an eight-bit greyscale (256 levels). Linear histogram enhancements were used for visualization purposes.

3. Experiment description

There are a number of issues that influence the performance of the statistical texture classification, such as the quantization level of the digital images, the angle, distance and window size used to compute the GLCM, the feature selection criteria, the spatial resolution, the number of defined classes, the classification method and the pos-classification processing of the image. In this study we performed an evaluation of the importance of each of these issues, in the context of air photo classification for the production of land cover maps.

3.1 *Test images*

A set of scanned B&W air photos from 1958 was used, from the national park of Peneda-Gerês in northwest Portugal. The photographs were acquired by a photogrammetric camera of 230×230 mm format, at a scale of about 1 : 15000. Each eight-bit greyscale image is approximately 6000×6000 pixels, corresponding to an area of about 3×3 km on the ground. Only five images were used in this test, but

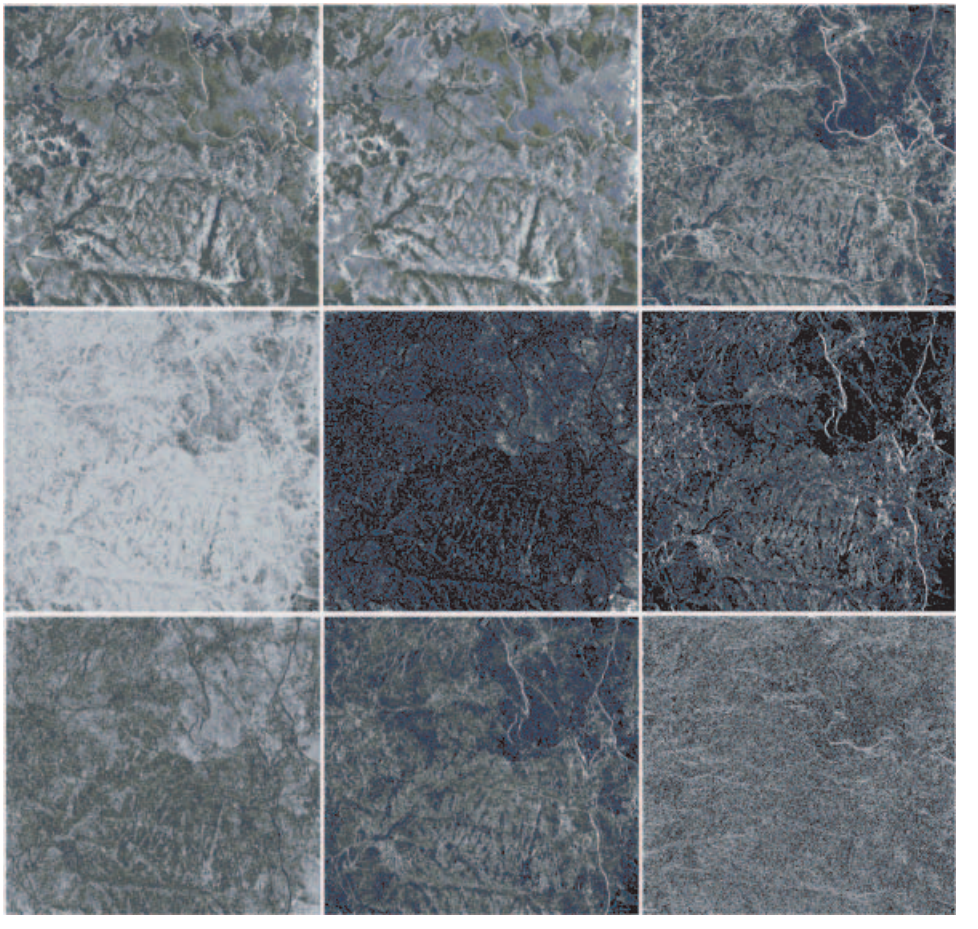

Figure 2. The eight textures of the first image from left to right. Top row: original image, MEAN, SD; middle row: ENT, ASM, CON; bottom row: HOM, DIS, COR. The images SD, ENT, DIS, HOM, COR, ASM are presented after a linear enhancement was applied.

nearly 100 are required to cover the whole park of Peneda-Gerês at this scale. The large number of images is certainly a strong motivation for having an automatic classification methodology. Five land cover classes were selected: water, bare ground, trees, hight scrub and grassland. Two simplified versions were also established, with four classes (high scrub and trees merged in a single class: scrub + trees) and with only three classes (scrub + trees and grassland also merged in a broad vegetation class). Training areas were identified by selecting 10 areas for each class in the group of images. For some classes, there is a slight overlap between some of these areas. Each area is typically around 10000 pixels in size (100×100).

Figure 3 shows the five images used and the training areas identified. A statistical analysis of the training data is presented in table 1. The average, minimum and maximum values of each feature are presented for each class, based on the training data. An inspection of table 1 reveals how distinctive classes are and how relevant are the features. For example, the signature of water is clearly different from the signatures of the other classes (land) in seven out of eight features (the exception is COR). Features ASM, HOM and DIS are redundant, as the signatures of the four

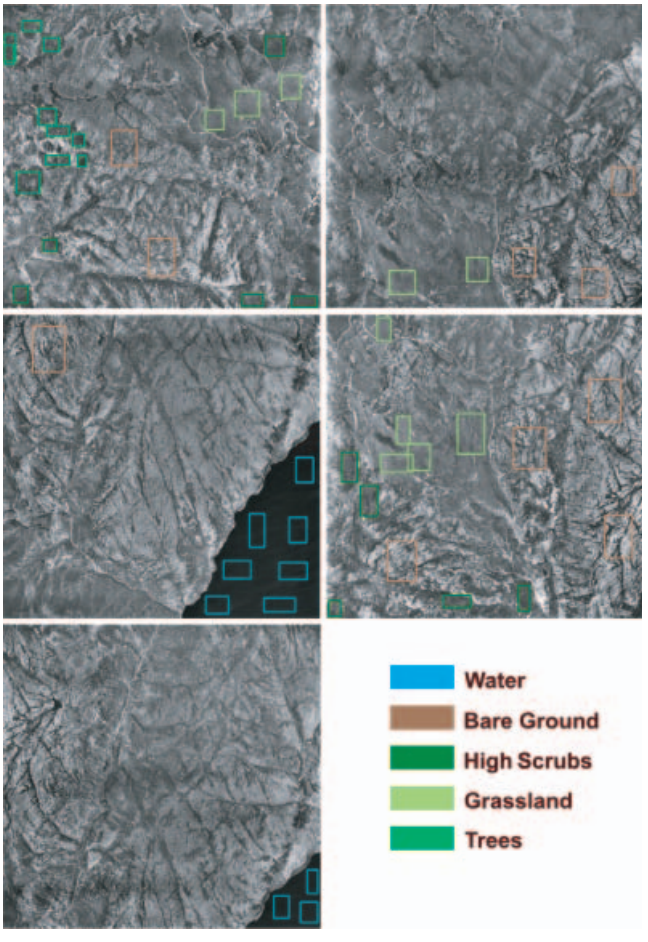

Figure 3. Training areas for the five land cover classes.

classes are nearly the same. The remaining features are slightly better to discriminate the four land classes, but there is still considerable overlap between the signatures of these classes in any single feature.

3.2 *Evaluation methodology*

The evaluation of the classification results was done by cross-validation. Only nine of the 10 areas available for each class are used for training; the remaining area is used to validate the classification result. The process is repeated 10 times with each of the 10 areas left out of the training stage. The results presented are a weighted average of the 10 classifications tests performed, leaving at each time one of the 10 different areas for validation.

A mode filter was used after the classification process in order to smooth the results (Gonzalez *et al.* 2004). Various window dimensions were used for this filter with different classification accuracies, as will be shown in §4. For each classification test, the average and overall accuracies are computed as well as the k coefficient (Richards and Jia 2006), using equation (9). The k coefficient ranges from -1 to 1 with a value close to 0 when the results are nearly random. High classification

Table 1. Statistical characterization of the eight textures for the five land cover classes.

Features		Water	Bare ground	High scrub	Grassland	Trees
MEAN						
	average	24.56	114.37	102.43	103.91	95.10
	min	19.60	90.35	78.48	87.39	73.12
	max	28.15	132.12	122.06	118.44	117.49
SD						
	average	3.46	13.49	12.76	9.78	10.79
	min	3.20	11.32	7.62	6.17	8.58
	max	3.63	18.38	15.80	15.86	13.37
ASM						
	average	4.83	5.20	5.21	5.17	5.19
	min	4.76	5.17	5.16	5.09	5.14
	max	4.87	5.25	5.25	5.25	5.23
HOM						
	average	0.040	0.028	0.028	0.029	0.029
	min	0.039	0.027	0.027	0.027	0.027
	max	0.043	0.029	0.029	0.031	0.030
CON						
	average	13.80	148.35	190.64	114.22	112.50
	min	10.36	78.67	77.89	39.98	72.61
	max	15.47	349.54	266.97	272.88	163.36
DIS						
	average	0.31	0.12	0.11	0.14	0.13
	min	0.29	0.08	0.08	0.08	0.10
	max	0.34	0.15	0.15	0.19	0.16
ENT						
	average	2.78	8.96	10.02	7.68	7.98
	min	2.41	6.76	6.72	4.81	6.38
	max	2.97	14.04	12.46	12.55	9.86
COR						
	average	0.48	0.59	0.44	0.46	0.50
	min	0.45	0.42	0.37	0.36	0.41
	max	0.53	0.67	0.48	0.51	0.56

accuracies will correspond to k values close to 1:

$$k = \frac{P\sum_k x_{kk} - \sum_k x_{k+}x_{+k}}{P^2 - \sum_k x_{k+}x_{+k}} \tag{9}$$

where $x_{i+} = \sum_j x_{ij}$, $x_{+j} = \sum_i x_{ij}$ and P is the number of observations.

A set of reference values was established for the various parameters, which were maintained constant while varying each of the others at a time. The reference values were: GLCM computed on a 5×5 window with 0° angle and distance 1, Bayes classifier, five-bit radiometric image resolution, use of only four texture features (MEAN, SD, ASM, CON), 25% spatial resolution, four land cover classes and 9×9 dimension mode filter.

4. Results

The first parameter tested was the classifier. Three distance classifier methods were tested: Euclidean, Mahalanobis and Bayes (Gonzalez *et al.* 2004). The classification accuracy (% of pixels classified in the correct class), obtained from cross-validation, is presented in table 2 for each classifier and class. As previously stated, all the

Table 2. Classification accuracies for the three classifiers tested (in %). The reference value is underlined and the best choice is in bold.

Land cover class	Euclidean classifier	Mahalanobis classifier	Bayes classifier
Water (%)	100.0	86.8	100.0
Bare ground (%)	73.1	74.5	76.0
Scrub + trees (%)	64.5	76.4	48.6
Grassland (%)	89.9	66.9	90.9
Average accuracy	**81.9**	76.1	78.9
Overall accuracy	**80.8**	75.1	80.2
k coefficient	**0.758**	0.682	0.718

remaining parameters were maintained fixed at their reference values. The average accuracy, overall accuracy and k coefficient are also presented in table 2. The Euclidean classifier performed slightly better overall than the Bayes method (k=0.758, k=0.718, respectively) and is the classifier requiring less computational effort. It is worth noting that the Mahalanobis distance classifier performance was surprisingly poor, particularly for the class water, which should be easily distinguishable.

The second parameter tested was the radiometric resolution. The original images in eight-bit were degraded to 7, 6, 5, 4 and 3-bit and the classifiers applied to all of these. The results are presented in table 3. The average accuracy and the k coefficient vary considerably with the radiometric resolution. The overall accuracy is not so significant, as the number of training pixels for water is the highest and the classification accuracy for this class is 100% for all cases. The same happens for most parameters tested here, thus the overall accuracy is not as meaningful as the average accuracy and the k coefficient. The best results are obtained for four-bit images, which is particularly convenient as the computation burden is greatly dependent on this factor. The computational time required to produce the GLCM matrix and the texture features is proportional to 2^n, where n is the number of bits. So, using the four-bit version of the image, instead of the original eight-bit format, besides increasing the classification accuracy it also reduces the running time by a factor of 16.

The third parameter tested was the window dimension of the mode filter applied after the classification process. The results are presented in table 4. Again, the classification accuracy is 100% for water on all cases. Although the average accuracy

Table 3. Classification accuracies for the various radiometric resolutions tested (in %). The reference value is underlined and the best choice is in bold.

Land cover class	Three-bit	Four-bit	Five-bit	Six-bit	Seven-bit	Eight-bit
Water (%)	100.0	100.0	100.0	100.0	100.0	100.0
Bare ground (%)	67.6	70.5	76.0	70.5	66.9	66.6
Scrub + trees (%)	80.6	81.3	48.6	21.8	20.0	20.1
Grassland (%)	56.3	66.4	90.9	89.3	48.6	42.3
Average accuracy	76.1	**79.6**	78.9	70.4	58.9	57.2
Overall accuracy	72.3	**76.1**	80.2	73.8	62.1	60.5
k coefficient	0.682	**0.728**	0.718	0.605	0.452	0.430

Table 4. Classification accuracies for the window dimension of mode filter (in %). The reference value is underlined and the best choice is in bold.

Land cover class	3 × 3	5 × 5	7 × 7	9 × 9	11 × 11	13 × 13	15 × 15	17 × 17	19 × 19	21 × 21
Water (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Bare ground (%)	69.6	71.8	74.0	76.0	77.9	79.9	81.7	83.2	84.6	86.0
Scrub + trees (%)	45.4	46.4	47.4	48.6	49.7	50.6	51.3	51.9	52.5	53.4
Grassland (%)	86.7	88.0	89.4	90.9	92.3	93.4	94.3	94.9	95.5	96.0
Average acc.	75.4	76.5	77.7	**78.9**	80.0	81.0	81.8	82.5	83.2	83.9
Overall acc.	75.9	77.3	78.8	**80.2**	81.6	82.9	83.9	84.9	85.7	86.6
k coefficient	0.672	0.687	0.703	**0.718**	0.733	0.747	0.758	0.767	0.775	0.785

and the *k* coefficient both increase with the size of the window, some caution should be taken in the selection of the optimal size for the mode filter. A too large window will lead to an over-smooth result, which might not be satisfactory from the end user's perspective. An inspection of the graphs of the first and second derivate of the function accuracy versus window size was carried out, as well as visual analysis of the classified images. The window size 9 × 9 was thought to be the most appropriate for the mode filter. However, the decision of the final selection is a compromise between the classification accuracy and the end user requirements for spatial detail.

The parameters used to compute the GLCM were also evaluated. The reference value for the window dimension is 5 × 5, and a range of values between 3 × 3 and 15 × 15 was tested. The results, presented in table 5, show that the average accuracy and *k* coefficient are highest for a window of 7 × 7, sharply decreasing for sizes of 11 × 11 or higher. Moreover, the computation burden increases with the increasing window size, so these are good reasons to keep it reasonably small at 7 × 7. The reference values for the angle and distance used for the pairings for the GLCM were 0° and 1. The distance was varied between 1 and 4 and the angles tested were 0°, 45°, 90°, 135° and 180°. The results of these tests are presented in tables 6 and 7. The best choice for distance was the reference value, but for the angle the value of 90° was the best performer, favouring the pairing direction north.

Since some of the texture features are highly correlated, usually not all of them are needed to perform the classification. Haralick *et al.* (1973) showed that some features, such as ASM, CON, COR and ENT, are more important than others. Zhang *et al.* (2003) demonstrated the advantage of using a combination of only four or five texture features instead of all eight. A total of 13 feature combinations were

Table 5. Classification accuracies for the various window dimensions used to produce the texture features (in %). The reference value is underlined and the best choice is in bold.

Land cover class	3 × 3	5 × 5	7 × 7	9 × 9	11 × 11	13 × 13	15 × 15
Water (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Bare ground (%)	69.7	76.0	77.4	76.1	70.5	75.6	76.7
Scrub + trees (%)	32.3	48.6	53.2	50.9	13.8	22.8	22.7
Grassland (%)	95.9	90.9	89.1	89.5	73.7	70.8	71.0
Average accuracy	74.5	78.9	**79.9**	79.1	64.5	67.3	67.6
Overall accuracy	76.4	80.2	**81.0**	80.2	69.0	71.8	72.3
k coefficient	0.660	0.718	**0.732**	0.722	0.527	0.564	0.568

Table 6. Classification accuracies for GLCM distance (in %). The reference value is underlined and the best choice is in bold.

Land cover class	Distance 1	Distance 2	Distance 3	Distance 4
Water (%)	100.0	100.0	100.0	100.0
Bare ground (%)	76.0	67.1	63.7	64.9
Scrub + trees (%)	48.6	21.4	19.3	14.8
Grassland (%)	90.9	81.6	45.7	30.4
Average accuracy	**78.9**	67.5	57.2	52.5
Overall accuracy	**80.2**	70.4	59.9	56.1
k coefficient	**0.718**	0.567	0.429	0.367

tested, six with four features (#1 to #6), six with five features (#7 to #12) and one with all eight features (#13). The feature combination #1 (MEAN, SD, ASM, CON) was used as the reference for the test of the other parameters. The results are presented in table 8, and confirm the views of Haralick *et al.* (1973) and Zhang *et al.* (2003), as combinations #1 and #2 with four features and combinations #7 and #8 with five features have about the same accuracy, both just slightly higher than the combination #13 using all eight features. In this case the variability of the classification average accuracy and the *k* coefficient were small, for different feature combinations. However, as the computational time increases linearly with the number of features, the best selection is the combination of four features #2 (MEAN, SD, ENT, CON).

Another parameter evaluated was the image spatial resolution. The original image size was reduced to 50%, 25% and 10% (both in lines and columns). The classification was performed on the reduced version of the images, with the results presented in table 9. The accuracies obtained using the images reduced to 50% and 25% are nearly the same, but the performance with the images reduced to 10% is considerably worst. The image size greatly affects the computation running time. A reduction by a factor f in the image size will reduce the computational time by f^2. For this particular case the reduction of the spatial resolution up to 10% of the original format is not considered to be a major disadvantage as the spatial resolution of the images (approx. 0.5 m) is too high for the end user requirements (approx. 5 m). The images reduced to 50% and 25% are both good choices, but the images at 25% size were considered to be the most suitable ones in this case, as the accuracy is only marginally inferior to the 50% version and the computational effort is much smaller.

Table 7. Classification accuracies for GLCM angle (in %). The reference value is underlined and the best choice is in bold.

Land cover class	Angle 0	Angle 45	Angle 90	Angle 135	Angle 180
Water (%)	100.0	100.0	100.0	100.0	100.0
Bare ground (%)	76.0	69.5	74.5	71.2	75.4
Scrub + trees (%)	48.6	23.1	53.0	27.9	42.6
Grassland (%)	90.9	93.4	93.3	95.9	91.6
Average accuracy	78.9	71.5	**80.2**	73.8	77.4
Overall accuracy	80.2	74.5	**80.7**	76.5	79.3
k coefficient	0.718	0.620	**0.736**	0.650	0.699

Table 8. Classification accuracies for the texture features combination (in %). Combinations #1 to #6 use four features, #7 to #12, five features, and #13 all eight features. The reference value is underlined and the best choice is in bold.

Land cover class	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13
Water (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Bare ground (%)	76.0	76.3	64.7	65.9	63.9	65.0	76.0	76.3	64.7	65.9	63.9	65.0	76.5
Scrub + trees (%)	48.6	48.3	82.0	82.3	82.1	82.3	48.6	48.3	82.0	82.3	82.1	82.3	47.3
Grassland (%)	90.9	91.1	53.4	54.3	52.4	53.3	90.9	91.1	53.4	54.3	52.4	53.3	91.3
Average acc.	78.9	**78.9**	75.0	75.6	74.6	75.2	78.9	78.9	75.0	75.6	74.6	75.2	78.8
Overall acc.	80.2	**80.4**	70.4	71.2	69.8	70.6	80.2	80.4	70.4	71.2	69.8	70.6	80.4
k coefficient	0.718	**0.719**	0.667	0.675	0.661	0.669	0.718	0.719	0.667	0.675	0.661	0.669	0.717

Table 9. Classification accuracies for the spatial resolution (in %). The reference value is underlined and the best choice is in bold.

Land cover class	10% of full resolution	25% of full resolution	50% of full resolution	100%
Water (%)	100.0	100.0	100.0	100
Bare ground (%)	80.0	76.0	70.5	59.3
Scrub + trees (%)	23.1	48.6	80.3	62.3
Grassland (%)	98.1	90.9	65.4	15.7
Average accuracy	75.3	**78.9**	79.1	59.3
Overall accuracy	89.1	**80.2**	75.7	56.3
k coefficient	0.670	**0.718**	0.721	0.458

The final parameter tested was the number of land cover classes looked for in the images. Table 10 shows the accuracies obtained for the classification into three, four and five classes. The accuracy results are obviously better for a reduced number of classes, but the information provided is less valuable. In this case, four was considered to be the most reasonable value, considering the accuracy of the classification and the end user requirements.

The tests carried out indicated that the best parameters to classify the park Peneda-Gerês images are: Euclidean classifier (table 2), four-bit radiometric image resolution (table 3), 9×9 dimension of the mode filter (table 4), GLCM computation using 7×7 windows size (table 5), with dimension 1 (table 6) at 90° direction (table 7), four texture features (MEAN, SD, ENT, CON) (table 8), image size reduced to 25% (table 9) using four different classes (water, bare ground, scrub + trees and grassland). This selection of parameters was used to perform a final classification of the five test images. The results from cross-validation with the best choice of parameters were: average accuracy 83.9%, overall accuracy 82.0% and the *k* coefficient 0.786. A final classification process was carried out, using all training data available and the best selection for all the parameters, presented in figure 4. The confusion matrix for this final classification task is presented in table 11. The average accuracy of the final classification was 85.7%, the overall accuracy 83.4% and the *k* coefficient 0.809.

Table 10. Classification accuracies for the number of land cover classes (in %). The reference value is underlined and the best choice is in bold.

Land cover class	Three classes	Four classes	Five classes
Water (%)	100.0	100.0	100.0
Bare ground (%)	75.9	76.0	77.2
Vegetation (%)	81.3	…	…
Grassland (%)	…	48.6	67.9
Scrub + trees (%)	…	90.9	…
Trees (%)	…	…	22.8
High scrub (%)	…	…	77.2
Average accuracy	85.7	**78.9**	69.0
Overall accuracy	82.3	**80.2**	74.8
k coefficient	0.786	**0.718**	0.613

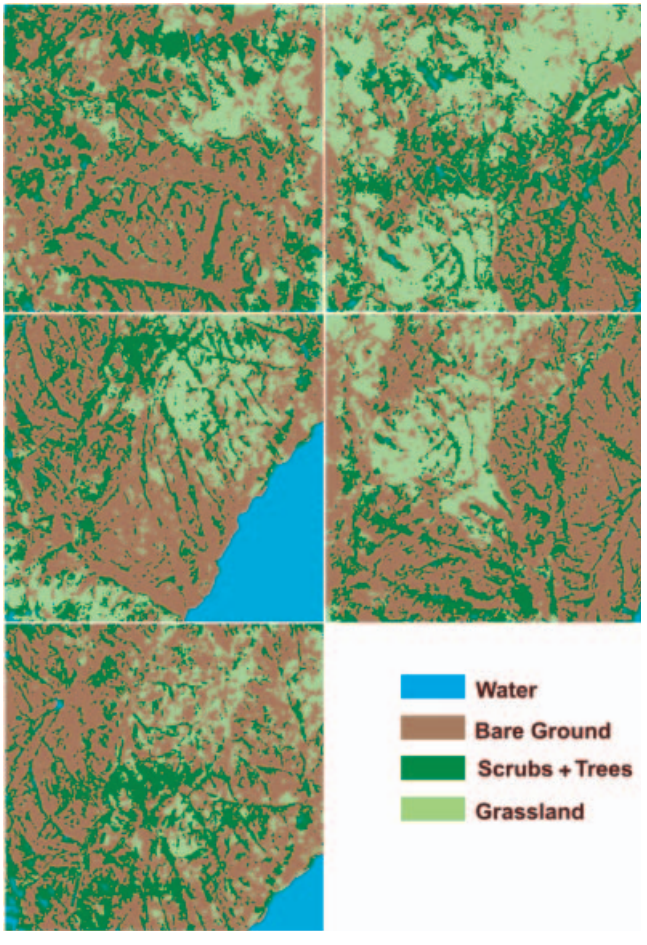

Figure 4. Final classified images (four classes).

5. Conclusions

The aim of this work was to evaluate the ability of texture based image classification methods to produce historic land cover maps from black & white (B&W) air photos of the national park of Peneda-Gerês.

The texture based classification process involves a number of parameters, including some related to the computation of the GLCM (window size, angle and distance for pairing pixels) and to the image characteristics (spatial and radiometric resolution). Other parameters were also evaluated, such as the type of classifier, the number of classes and the features used. A set of reference values was used for the

Table 11. Confusion matrix for the final classification (in %).

Land cover class	Water	Bare ground	Scrub + trees	Grassland
Water (%)	100.0	0.0	0.0	0.0
Bare ground (%)	0.2	76.5	21.8	1.5
Scrub + trees (%)	0.3	14.2	80.0	5.5
Grassland (%)	0.0	8.4	5.2	86.4

parameters. Each one of them was varied, within a suitable range of values, and its effect in the classification accuracy was evaluated by cross-validation. The final classification result was obtained with a selection of the best values for the parameters evaluated. The final selection of the combined parameters is obviously dependent of the initial reference values chosen, and is only valid for this particular dataset. However, most parameters will not significantly be affected by a different initial configuration, as long as the final selection is not very different from the reference scenario. For this particular case, the parameters that were modified from the initial reference values were: classifier method, image resolution and GLCM windows dimension and direction. One important aspect of this experiment is the fact that the training/validation was performed on the combined images, which is a crucial aspect for practical implementation, which usually require a large number of images (the national Park of Peneda-Gerês for example is covered by more than 100 images at this scale). This also prompts the issue of computational efficiency, which is strongly dependent on some of the parameters, such as image spatial and radiometric resolution, and the window size used for GLCM computation. Although eight features were produced from the original greyscale images, about the same classification accuracy was obtained with a selection of only four of these features (MEAN, SD, ENT, CON). This fact confirms the results obtained by Gonzalez *et al.* (2004) and Haralick *et al.* (1973) and proves to be more convenient from a computational efficiency point of view. It is also worth noting that the best results were obtained after reducing both the spatial and the radiometric resolution of the original images.

The final classification results (average accuracy 85.7%, overall accuracy 83.4% and $k=0.809$) are reasonably good, although only four main land cover classes were distinguished. Moreover, it is probably not realistic to expect a much better discrimination ability from the type of B&W air photograph tested, using an automatic classification process. The method can nevertheless be useful for the production of historic land cover maps from B&W air photos.

Acknowledgments

This work was done with the support of 'Centro de Investigação em Ciências Geo-Espaciais, Faculdade de Ciências da Universidade do Porto', financed by 'Fundação de Ciência e Tecnologia' POCTI/FEDER. The authors would like to thank Filipe Maia for the identification of training areas.

References

Basset, O., Buquet, B., Abouelkaram, S., Delachartre, P. and Culioli, J., 2000, Application on texture image analysis for the classification of bovine meat. *Food Chemistry*, **69**, pp. 437–445.

Carrión, P., Cernadas, E., Gálvez, J.F., Damián, M. and Sá-Otero, P., 2004, Classification of honeybee pollen using a multiscale texture filtering scheme. *Machine Vision and Applications*, **15**, pp. 186–193.

Chen, Q. and Gong, N.P., 2004, Automatic variogram parameter extraction for textural classification of the panchromatic IKONOS imagery. *IEEE Transations on Geoscience and Remote Sensing*, **42**, pp. 1106–1115.

Coburn, C., Roberts, A. and B ach, K., 2004, Spectral and spatial artifacts from the use of desktop scanners for remote sensing. *International Journal of Remote Sensing*, **22**, pp. 3863–3870.

Gonzalez, R.C., Woods, R.E. and Eddins, S.L., 2004, *Digital Image Processing using MATLAB* (Englewood Cliffs: Prentice Hall).

GUPTA, R. and UNDRILL, P.E., 1995, The use of texture analysis to delineate suspicious masses in mammography. *Physics in Medicine and Biology*, **40**, pp. 835–855.

HARALICK, R.M., SHANMUGAT, M.K. and DINSTEIN, I., 1973, Textural features for image classification. *IEEE Transaction on Systems, Man, and Cybernetics*, **3**(6), pp. 610–621.

HERRERO, J.M., ARMÁN, M.F. and CASTRO, J.L., 2004, Grading textured surfaces with automated soft clustering in a supervised SOM. *Lecture Notes in Computer Science*, **3212**, pp. 323–330 (New York: Springer).

LU, D.S. and WENG, Q.H., 2005, Urban classification using full spectral information of Landsat ETM+ imagery in Marion County, Indiana. *Photogrammetric Engineering & Remote Sensing*, **71**, pp. 1275–1284.

RICHARDS, J.A. and JIA, X., 2006, *Remote Sensing Digital Image Analysis*, 4th edn (New York: Springer).

SALI, E. and WOLFSON, H., 1992, Texture classification in aerial photographs and satellite data. *International Journal of Remote Sensing*, **13**, pp. 3395–3408.

TAHIR, M.A., BOURIDANE, A. and KURUGOLLU, F., 2004, An FPGA based coprocessor for GLCM and Haralick texture features and their application in prostate cancer classification. *Analog Integrated Circuits and Signal Processing*, **43**, pp. 205–215.

ZHANG, Q., WANG, J., GONG, P. and SHI, P., 2003, Study of urban spatial patterns from SPOT panchromatic imagery using textural analysis. *International Journal of Remote Sensing*, **24**, pp. 4137–4160.