New Developments and Challenges in Remote Sensing, Z. Bochenek (ed.)
©2007 Millpress, Rotterdam, ISBN 978-90-5966-053-3

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

C.M.R. Caridade
Instituto Superior de Engenharia de Coimbra, Portugal

A.R.S. Marçal & T. Mendonça
Faculdade de Ciências, Universidade do Porto, Portugal

Keywords: image classification, textures, land cover mapping

ABSTRACT: The use of black & white 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 quantization level, the window size used to produce the texture features, the feature selection criteria and the image spatial resolution. The evaluation was performed using 4 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 importance 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. An average accuracy of 85.2% was achieved for 4 land cover classes.

1 INTRODUCTION

The Image classification is the process of assigning thematic labels to each image pixel. This is a frequently used methodology to produce land cover maps from air-photographs or satellite images. Image classification is usually performed on RGB multi-spectral or hyper-spectral images, where the spectral signature of each pixel in the multi-dimensional feature space is used in the discrimination process. Although the new satellite and airborne sensors can provide imagery with a large number of spectral bands, there are cases where only a single panchromatic image (greyscale image or black & white photograph) is available and thus this approach is not possible.

The classification of greyscale images has to be based in other characteristics instead of the color or multi-dimensional signature of each pixel. One effective form to classify greyscale images is to make use of the texture information present in the image. The texture contains significative information about the structural arrangement of objects and surfaces and their relationship with the surrounding environment. This type of approach has been used in different areas of image processing such as quality inspection (Herrero *et al.* 2004), medical imaging (Thir *et al.* 2005) and remote sensing (Durrieu *et al.* 2006; Zhang *et al*. 2003; Zhang *et al.* 2004). Another example where the use of texture for image classification can be valuable is the production of land cover maps from historic air photographic surveys.

The purpose of this work is to evaluate the ability of texture based image classification methods to produce historic land cover maps from Black & White (B&W) air-photos. The grey scale 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 8-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 Grey-Level Co-occurence Matrix (GLCM) produced for each pixel using a neighbourhood window of D×D pixels (Haralick *et al.* 1973). The GLCM corresponds to the number of pairs of grey levels encountered in the search window, within a pre-defined direction and range. The horizontal direction 0° with a range of 1 (nearest neighbour) was used in this work. The 8 texture descriptions used are presented in equations (1) to (8), where N is the number of grey levels, P is the normalized symmetric GLCM of dimension N×N and $\mathrm{P}_{\mathrm{i,\ j}}$ 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 pixel.

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

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

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 occurs when the distribution of the grey level values is constant or periodic within the search window. The **Homogeneity** (HOM) measures the sensitivity to 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.

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

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

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 centered around the GLCM diagonal, otherwise there is a more even distribution of the grey levels in the GLCM.

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

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

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

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

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

where $\mu_J = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} jP_{i,j}$ and $\sigma_J = \sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{i,j}(j-\mu_J)^2}$

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 window size used to compute the GLCM, the feature selection criteria, the spatial resolution, the number of defined classes, the classification method and 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. The direction and range used on the GLCM to produce the texture features was also evaluated. However, the results are not presented here as no significant changes were observed for both the direction and range.

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 8-bit greyscale image is approximately 6000×6000 pixels, corresponding to an area of about 3×3 km

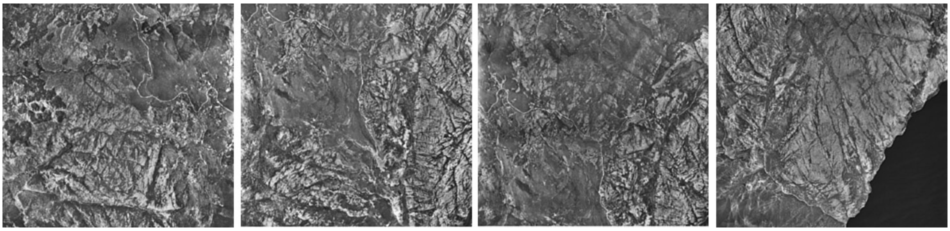

Figure 1. Test images used, from the national park of Peneda-Gerês, Northwest Portugal.

on the ground. Figure 1 shows the 4 images that were used in this test, but nearly 100 are required to cover the whole park of Peneda-Gerês. 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, Forest, High Scrubs and Grassland.

Two simplified versions were also established, with 4 classes (High Scrubs and Forest merged in a single class: Scrubs + Trees) and with only 3 classes (Scrubs + 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. Each area is typically around 10000 pixels in size (100×100), with a total between 76240 and 322538 pixels selected for training each class.

3.2 *Evaluation methodology*

The evaluation of the classification result was done by cross validation. Only 9 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 classification tests performed, leaving at each time one of the 10 different areas for validation.

A mode filter was used after the classification stage in order to smooth the results (Gonzalez *et al.* 2004). Various window dimensions were used for this filter with different accuracies, as it will be shown in Section 4. However, the decision of the final selection is a compromise between the classification accuracy and the end user requirements for spatial detail. For each classification test, the average and overall accuracies are computed as well as the k coefficient (Richards & Jia 2006). The k coefficient ranges from -1 to 1 with a value close to 0 when the results are nearly random. High classification accuracies will correspond to k values close to 1. 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: Bayes classifier, 5-bit radiometric image resolution, 5×5 window dimension for the computation of the texture features, use of only 4 texture features (MEAN, SD, ASM, CON), 25 % spatial resolution, 4 classes defined, 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

Table 1. Classification accuracies for the 3 classifiers tested. Reference value underlined; best choice in bold.

Land cover class	Euclidean Classifier	Mahalanobis Classifier	Bayes Classifier
Water	100.0%	82.4%	100.0%
Bare Ground	73.2%	74.5%	76.0%
Scrubs + Trees	64.3%	76.4%	48.6%
Grassland	89.9%	66.9%	90.9%
Average accuracy	**81.8%**	75.1%	<u>78.9%</u>
Overall accuracy	**86.1%**	76.5%	<u>85.7%</u>
k coefficient	**0.758**	0.667	<u>0.718</u>

presented in Table 1 for each classifier and class. As previously stated, all the remaining parameters were left fixed at their reference values. The average accuracy, overall accuracy and *k* coefficient are also presented in Table 1. The Euclidean classifier performed slightly better overall than the Bayes method ($\mathrm{k} = 0.758, \mathrm{k} = 0.718$ respectively) and is the classifier requiring less computational time. It is worth nothing that the Mahalanobis distance classifier performance was surprisingly poor, particularly for the class water, which should be easily distinguished.

The second parameter tested was the radiometric resolution. The original images in 8-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 2. 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 highest and the classification accuracy for this class is 100% for all cases. As the same happens for most parameters tested here, the overall accuracy is not as meaningful as the average accuracy and the *k* coefficient. The best results are obtained for 4-bit images, which is particularly convenient as the computation burden is greatly dependent on this factor. The computation time is proportional to 2^{n}, where n is the number of bits. So using the 4-bit version of the image reduces the running time by a factor of 16 and increases the classification accuracy.

The third parameter tested was the window dimension of the mode filter applied after the classification process. The results are presented in Table 3. Again, the classification accuracy is 100% for water on all cases. Although the average accuracy and the *k* coefficient both increase with the size of the window, some caution should be taken in

Table 2. Classification accuracies for the various radiometric resolutions tested. The reference value is underlined; the best choice is in bold.

Land cover class	3-bit	4-bit	5-bit	6-bit	7-bit	8-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%
Scrubs + Trees	80.7%	81.1%	48.6%	21.8%	20.0%	20.1%
Grassland	56.3%	66.4%	90.9%	89.0%	48.4%	42.2%
Average accuracy	76.2%	**79.5%**	<u>78.9%</u>	70.3%	58.8%	57.2%
Overall accuracy	80.0%	**82.7%**	<u>85.7%</u>	81.0%	72.5%	71.4%
k coefficient	0.682	**0.727**	<u>0.718</u>	0.605	0.451	0.430

Table 3. Classification accuracies for the window dimension of mode filter. The reference value is underlined; the best choice is in bold.

class	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%
Bare Ground	71.9%	74.0%	76.0%	77.9%	80.0%	81.7%	83.2%	84.7%	86.1%
Scrubs + Trees	46.3%	47.4%	48.6%	49.7%	50.6%	51.3%	51.9%	52.6%	53.3%
Grassland	88.0%	89.4%	90.9%	92.3%	93.4%	94.3%	94.9%	95.5%	96.0%
Average acc.	76.5%	77.7%	**78.9%**	80.0%	81.0%	81.8%	82.5%	83.2%	83.9%
Overall acc.	83.6%	84.6%	**85.7%**	86.7%	87.6%	88.4%	89.0%	89.7%	90.3%
k coefficient	0.687	0.702	**0.718**	0.733	0.747	0.758	0.767	0.776	0.785

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 1st and 2nd derivate graphs was done, as well as visual analysis of the classified images, and the size of 9×9 was thought to be the most appropriate for the mode filter.

The dimension of the window used to compute the texture features was also evaluated. The reference value for this parameter is 5×5, and a range of values between 3×3 and 13×13 was tested. The results, presented in Table 4, show that the average accuracy and *k* coefficient are highest for a window of 7×7, sharply decreasing for sizes of 9×9 or higher. The computation burder increases with the increasing window, so these are good reasons to keep it reasonably small at 7×7.

Since some of the texture features are highly correlated, usually not all of them are needed to perform the classification. Haralick showed that some features, such as ASM, CON, COR and ENT, are more important than others (Haralick *et al.* 1973). Zhang demonstrated the advantage of using a combination of only 4 or 5 texture features instead of all 8 (Zhang *et al*. 2003). A total of 13 feature combinations were tested, 6 with 4 features, 6 with 5 features and 1 with all 8 features. Only the best 8 combination results are presented in Table 5. The feature combination #1 (MEAN, SD, ASM, CON) was used as the reference for the test of the other parameters. The results presented in Table 5 confirm the views of (Haralick *et al.* 1973) and (Zhang *et al.* 2003), as combinations #1 and #2 with 4 features and combination #7 and #8 with 5 features have about the same accuracy, both just slightly higher than the combination of all 8 features.

Table 4. Classification accuracies for the various window dimensions used to produce the texture features. The reference value is underlined; the best choice is in bold.

Land cover class	3×3	5×5	7×7	9×9	11×11	13×13
Water	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Bare Ground	69.8%	76.0%	76.7%	76.1%	74.5%	75.6%
Scrubs + Trees	32.3%	48.6%	51.1%	58.3%	23.3%	22.8%
Grassland	95.9%	90.9%	89.7%	89.5%	70.5%	70.8%
Average accuracy	74.5%	78.9%	**79.4%**	79.1%	67.1%	67.3%
Overall accuracy	83.0%	85.7%	**85.9%**	85.7%	79.2%	79.6%
k coefficient	0.660	0.718	**0.725**	0.722	0.561	0.564

Table 5. Classification accuracies for features combinations. Reference value underlined; best choice in bold.

Land cover class	4 features #1	#2	#3	#4	5 features #7	#8	#10	8 feat. #13
Water	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%	76.0%	76.3%	65.9%	76.5%
Scrubs + Trees	48.6%	48.3%	81.9%	82.3%	48.6%	48.3%	82.3%	47.3%
Grassland	90.9%	91.1%	53.4%	54.3%	90.9%	91.1%	54.3%	91.3%
Average acc.	78.9%	**78.9%**	75.0%	75.6%	78.9%	78.9%	75.6%	78.8%
Overall acc.	85.7%	**85.8%**	78.6%	79.2%	85.7%	85.8%	79.2%	85.8%
k coefficient	0.718	**0.719**	0.667	0.675	0.718	0.719	0.675	0.717

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 4 features #2 (MEAN, SD, ENT, CON).

Another parameter evaluated was the image dimension. The original images were 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 6. The accuracy 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 dimension greatly affects the computation running time. A reduction by a factor f in the image size will reduce the computational time by f^2. There are also memory issues which need to be taken into account when using the full resolution images. In fact, it was not possible to perform the classification on the full resolution images (100%) in the MATLAB implementation developed (MathWorks 2002). However, this is not considered to be a major disadvantage as the spatial resolution of the images (aprox. 0.5 *m*) is too high for the end user requirements (aprox. 5 *m*). The images reduced to 25% were therefore considered to be the most suitable ones.

The final parameter tested was the number of land cover classes looked for in the images. Table 7 shows the accuracies obtained for the classification into 3, 4 and 5 classes. The results are obviously better for a reduced number of classes, but in this case

Table 6. Classification accuracies for the spatial resolution. Reference value underlined; best choice in bold.

Land cover class	10% of full resolution	25% of full resolution	50% of full resolution
Water	100.0%	100.0%	99.9%
Bare Ground	80.0%	76.0%	70.5%
Scrubs + Trees	23.0%	48.6%	80.2%
Grassland	98.1%	90.9%	65.4%
Average accuracy	75.3%	**78.9%**	79.0%
Overall accuracy	85.8%	**85.7%**	82.4%
k coefficient	0.670	**0.718**	0.720

Table 7. Classification accuracies for the class selection. Reference value underlined; best choice in bold.

Land cover class	3 classes	4 classes	5 classes
Water	100.0%	100.0%	100.0%
Bare Ground	75.9%	76.0%	77.2%
Vegetation / Forest	81.0%	-	22.8%
Scrubs + Trees / High Scrubs	-	90.0%	77.2%
Grassland	-	48.6%	67.6%
Average accuracy	85.7%	**78.9%**	69.0%
Overall accuracy	87.8%	**85.7%**	81.2%
k coefficient	0.785	**0.718**	0.612

4 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 1), 4-bit radiometric image resolution (Table 2), 9×9 dimension of the mode filter (Table 3), 7×7 windows size for the GLCM computation (Table 4), 4 texture features (MEAN, SD, ENT, CON) (Table 5), image dimension reduced to 25% (Table 6) using 4 different classes (Water, Bare Ground, Scrubs + Trees and Grassland). This selection of parameters was used to performed a final classification of the 4 test images. This resulted in the classified images presented in Figure 2. The confusion matrix of the final classification is presented in Table 8. The average accuracy was 85.2%, the overall accuracy 87.5% and the k coefficient 0.803. This final classification was performed using all the training areas. The results from cross validation were in this case 83.4% (average accuracy), 86.7% (overall accuracy) and 0.778 (k coefficient).

5 CONCLUSIONS

The aim of this work was to classify a group of B&W air-photos of the national park of Peneda-Gerês to obtain historic land cover maps of the area, using the texture information associated with each class. The process of classification involves a number of parameters such as the quantization level of the digital images, the window size used to compute the GLCM, the feature selection criteria, the spatial resolution, the number of defined classes,

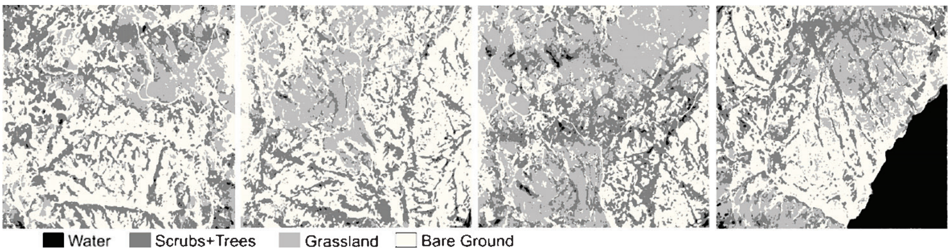

Figure 2. Final classified images from the national park of Peneda-Gerês, Northwest Portugal.

Table 8. Confusion matrix for the final classification.

Land cover class	Water	Bare Ground	Scrubs + Trees	Grassland
Water	100.0%	0.0%	0.0%	0.0%
Bare Ground	0.3%	75.2%	22.6%	1.9%
Scrubs + Trees	0.5%	13.4%	79.6%	6.5%
Grassland	0.1%	8.7%	5.0%	86.2%

the classification method and pos-classification processing of the image. The effect of each of these parameters in the classification process was evaluated by cross validation. A set of reference parameters was used, and each one of them was varied within a suitable range of values. These tests were used to find the best values for each parameter, for this particular dataset. A more careful evaluation of the existing classification methods was not performed, as this was not the main focus of this work.

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. 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. The parameters that were modified from the initial reference values were: classifier method, image resolution and GLCM windows dimension. One important aspect of this experiment is the fact that the training/validation was performed on the combined images, which is a crucial aspect as the whole national Park is covered by more than 100 images. This also prompts the issue of computational efficiency, which is strongly dependent on some of the parameters, such as image resolution, GLCM windows dimension and the combination of 4 or 5 features. Although 8 features were produced from the original greyscale images, about the same classification accuracy was obtained with a selection of only 4 of these features (MEAN, SD, ENT, CON). This confirms the results obtained by (Gonzalez *et al.* 2004) and (Haralick *et al.* 1973) and is also convenient from a computational efficiency point of view. The method proposed for the selection of parameters for the texture based classification proved to be satisfactory. The average classification accuracy achieved with the final set of parameters (85.2%) can also be considered reasonably good.

ACKNOWLEDGEMENTS

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’’ through POCTI/FEDER.

REFERENCES

Durrieu, M., Ruiz, L.A., Balaguer, A. 2006. Analysis of Geostatistical parameters for texture classification of satellite images, Proceedings of the 25th EARSeL Symposium, Global Developments In Environmental Earth Observation From Space, André Marçal (editor), Millpress, 11–18.

Gonzalez, R.C., Woods, R.E., Eddins, S.L. 2004. Digital Image Processing using MATLAB, Prentice Hall.

Haralick, R.M., Shanmuga, M.K., Dinstein, I. 1973. Textural feature for image classification, IEE Transaction on System, Man, and Cybernetics 610–621.

Herrero, J.M., Armán, M.F., Castro, J.L. 2004. Grading Textured surfaces with Automated soft Clustering in a Supervised SOM, Lecture Notes in Computer Science 3211, 323–330.

Richards, J.A., Jia, X. 2006. Remote Sensing Digital Image Analysis, Springer (4th edition)

Thir, M.A., Bouridane, A., Kurugollu, F. 2005. An FPGA Based Copressor for GLCM and Haralick Texture Features and their Application in Prostate Cancer Classification. Analog Integrate Circuits and Signal Processing 40, 205–215.

MathWorks 2002. Using Matlab, Version 6.5. The MathWorks, Inc. Natick. MA.

Zhang, Q., Wang, J., Gong, P. Shi, P. 2003. Study of urban spatial patterns from SPOT panchromatic imagery using textural analysis, Internation Journal of Remote Sensing 24, 4137–4160.

Zhang, X., Shan, T., Jiao, L. 2004. SAR Image Classification Based on Immune Clonal Feature Selection, Lecture Notes in Computer Science 3211, 504–511.