SPECTRAL-TEXTURAL IMAGE CLASSIFICATION IN A SEMIARID ENVIRONMENT

Philippe Maillard,

Depto. Cartografia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 31270-901, Brazil. philippe@cart.igc.ufmg.br

KEYWORDS: Land Cover, Classification, Texture, Vision, Landsat, SPOT, High resolution.

ABSTRACT:

Image classification can benefit from incorporating texture by enabling an increased number of classes and improving thematic accuracy. Incorporating texture also involves special attention in a number of aspects that range from the texture source to the evaluation of accuracy through pre-processing, training strategy and choosing a texture extraction paradigm and a classifier. Without special care in these aspects, classification results can be very unpredictable, especially when mixing spectral and textural features in the classification. This is mainly due to the spatial dependency of texture features.

The present article aims at analyzing these aspects (six in all) through a review of the concepts involved and a demonstration with two sample image data sets in a complex semiarid environment in Brazil. The data sets were formed with texture features from a SPOT-5 panchromatic image and spectral features from LANDSAT 7 ETM+ data. Results suggest that useful texture features can be extracted from SPOT-5 panchromatic data and that a mixed classification scheme is generally better than either approaches (spectral or textural). They also suggest that a non parametric classifier (Fisher linear discriminant) performs better for sets incorporating spectral and textural features and is less affected by edges and borders.

RÉSUMÉ:

La classification d'image peut bénéficier de l'ajout de textures qui permet d'augmenter le nombre de classes et la précision thématique. L'incorporation de la texture implique une attention spéciale à un nombre d'aspects qui vont du choix de la source des textures à l'évaluation des résultats en passant par le pré-traitement, la stratégie d'entraînement ainsi que le choix d'un paradigme d'extraction de texture et d'un classificateur. Sans se soucier de ces aspects, les résultats de la classification peuvent être imprévisibles, surtout si la classification incorpore un mélange d'éléments spectraux et de texture, cette dernière étant un processus à dépendance spatiale.

Le présent article vise l'analyse de ces aspects (six en tout) à travers une revue des concepts impliqués et une demonstration à l'aide de deux échantillons d'image provenant d'un environnement semi-aride du Brésil. Les données sont formées d'éléments de texture extraits d'une image panchromatique SPOT-5 et de bandes spectrales d'une image LANDSAT 7 ETM+. Les résultats suggèrent que les textures extraites d'image SPOT-5 peuvent être utili à la classification mixte (spectrale et de texture) qui s'est avérée meilleure que les approches individuelles. Ils suggèrent également que le classificateur non paramétrique ("Fisher linear discriminant") s'avère meilleur pour la classification de données mixtes (spectrales et de texture) et est moins affecté par les arêtes et bordures.

1. INTRODUCTION

The concept of texture analysis in remote sensing is still only marginally used in the classification process and a robust method for incorporating it remains elusive (Ferro and Warner, 2002). Most papers using texture in a classification process show that its use can bring significant improvement to the results and many off-the-shelf remote sensing software offer some kind of texture extraction method. So why not use it more? One reason might well be the difficulties encountered in choosing the settings and measurement types. Another reason may lie in the fact that classification can be a very sensitive process and the incorporation of features from spatially dependant processes can bring a strong bias to the results. One of the most popular classifiers assumes variables to have a normal distribution; texture features usually do not. Even field work should take into account that texture will be used and avoid proximity to edges and linear features.

The objective of the present paper is to assess the different aspects of including texture analysis into a classification process and to demonstrate the procedure using medium-high resolution data. These aspects are represented by six sequential steps. Two small complex scenes are used to test the mixed (spectral and textural) classification process and demonstrate the process. Results obtained with two sub-images are presented and discussed in a separate section. Concluding remarks make up the last section.

2. BACKGROUND

Texture in remote sensing has generated an extensive body of literature for which a review is beyond the scope of this paper. Of the many approaches that this wealth of studies has proposed, the GLCM appears to be the most commonly used (Franklin, 2001) and has proven amongst the most powerful methods for many situations of texture classification (Clausi, 2000; Maillard, 2003).

The GLCM was first used by Julesz (1962) and proposed by Haralick et al. (1973) as an approach to extracting textural features for image classification purposes. The method is based on the assumption that grey tones are spatially dependant (conditional joint probabilities) and that their dependency can be expressed through a co-occurrence matrix. Haralick et al. have therefore proposed a series of measurements taken from such matrices that relate to various aspects of texture (i.e. homogeneity, contrast, entropy, etc.). Only one pixel pair distance is usually used to construct the co-occurrence matrix. There appears not to be any known rigorous way to determine this sampling distance (Clausi, 2000) but an analysis of the semi-variogram (Maillard, 2003) or the local variance (Cao and Lam, 1997) can help to identify the distance with greater texture contrast.

In most classification projects the distribution of spectral measurements is considered Gaussian. Texture features have distributions that can not always be considered Gaussian and so classification schemes that use mixtures of spectral and textural features can produce unpredictable results. For that reason it can be wise to use a non-parametric classifier such as the Fisher linear discriminant (Duda and Hart, 2000).

3. METHOD

The methodology proposed has been divided into six successive steps that roughly correspond to the normal steps taken when classical spectral-only classification is performed. Each one of these steps requires special attention and successful texture classification is highly dependant upon how cautiously the analyst follows these special points of care.

The image data used in this article come from an ongoing research project for mapping semiarid vegetation forms in the Peruaçu region, North of Minas Gerais - Brazil.

3.1. Step 1: determining source of texture

Unless the objects being sought are fractal in nature, texture is a highly scale-dependant process (Cao and Lam, 1997). Objects that are distinct at a resolution of one meter (e.g. houses in a residential area) might become a coarse texture at a resolution of 30 meters or even a smooth texture with a 100 meter pixel size.

Although texture extracted from classical multi-spectral data (e.g. Landsat, Aster, Spot-Mx) has shown to bring some improvement to classification results in a variety of contexts (Franklin, 2001; Ferro and Warner, 2002), it is the perspective of using or integrating high-resolution data (ten meters or less) that gives texture its full potential. It can be perceived as a way to exploit high within-class variations as a favorable asset instead of a recurring problem when using per-pixel classification algorithms on high resolution data.

The Peruaçu project was initially planned with Landsat ETM+ data which are used here as the multispectral component. Given the size of the study area and our scale requirements (1:25 000) we chose Spot-5 five meter panchromatic data to be the most suited. Other data sources qualify at least as well as possible sources of good visual texture (e.g. IRS-P6, Ikonos, Quickbird) depending on the type of texture sought (i.e. H-resolution vs L-resolution). A half image (40 km x 40 km) was acquired from the Spot Image Corporation that had less than a year in time difference with the Landsat data. The image dataset used here is made up of five texture features derived from the Spot image (contrast, angular second moment, entropy, inverse difference moment and correlation) and six spectral features from the Landsat image (bands 1 to 5 and 7).

3.2. Step 2: pre-processing channel for texture

While classical multispectral classification usually exploits differences in first-order probability densities in a multi-dimensional space, texture is usually associated with second- and higher-order probability densities. These probability densities being spatially-dependent, pre-processing usually involves some kind of filtering either in the spatial (convolution) or spectral (Fourrier) domain. The pre-processing used here consisted in transforming the image function into an independent identically distributed process, a transformation similar to eliminating low frequencies.

Since texture features are computed from a co-occurrence matrix that is in turn computed over an image ”window” of a certain size (typically equal or larger than 5×5), borders between regions of homogeneous texture will tend to appear blurred or to have a texture of their own. This will be aggravated if the two regions have significantly different expectations. This undesirable effect can be reduced by making sure the local mean is constant throughout the image. This was performed by first applying a low-pass filter (31×31 neighborhood averaging) and then subtracting the results to the original image. It is important to use a rather large kernel window in order to leave a wide range of frequencies unaltered. Figure 1 illustrates this operation and Figure 2 shows the effect of eliminating low frequencies from some computed texture features. One can clearly see that the texture features computed from the unfiltered image have a much more “busy” appearance than the filtered image on which regions are more homogeneous.

3.3. Step 3: texture processing

Regardless of the texture feature extraction method chosen (GLCM in this case) one has to decide if these features should preserve or not the anisotropy of the textures observed in the image. While anisotropic texture features have shown to yield better classification results (Franklin, 2001; Maillard, 2003) they also involve a larger data set (to account for the various directions) and the separate classification of similar textures having different orientation. An alternative approach that still integrates some anisotropy consists in first computing the mean value for all directions considered and then calculating a measure of variation (i.e. standard deviation, variance, maximum difference, etc.). In the present case, only the mean value for four orientations (0, 45, 90 and 135°) was used since most textures present are somewhat ill-defined and have little or no preferential orientation.

Four parameters have to be set for the GLCM method: 1) the pixel pair sampling distance, 2) the window size to analyze, 3) the number of quantization levels and 4) the number and type of measurements. The sampling distance can be set by trial and error or by computing the local variance (Cao and Lam, 1997; Maillard, 2003) or the semi-variogram (Franklin, 2001; Maillard, 2003). The semi-variograms of the two sub-images are shown in figure 3. It appears quite clear that a lag of two represents a significant break in the curves and justify our choice of a pixel pair sampling distance of two pixels. The window size is also very important and since here we are using a dual-resolution data set (Landsat with 30 m and Spot with 5m), the window size was chosen based on multiples of the worse resolution (30 m). Window sizes of 18×18 ($3 \times [30m/5m]$) and 30×30 ($5 \times [30m/5m]$) have been tested and texture measures appeared to be more stable with the latter. Texture feature extracted from larger resolution cells should use smaller windows; 5×5 or 7×7 for Landsat 30 m cell for instance.

The number of quantization levels was set to 16 as it was observed that more levels (32 or 64) tend to lower the measurements taken and reduce the between-class distance. The last consideration is the number and type of measurements taken from the co-occurrence matrices. Here, five different measurements were considered: 1) contrast (also known as inertia), 2) angular second moment, 3) entropy, 4) inverse difference moment and 5) correlation. These were selected from previous work by the author (Maillard, 2003) and by “general popularity” in the liter-

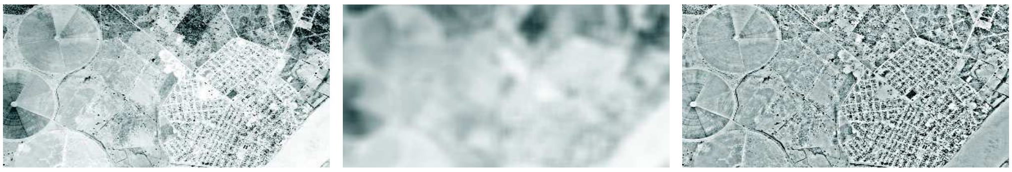

Figure 1: Process of eliminating low frequencies through low-pass filtering and image algebra. (a) original image, (b) image resulting from a low-pass convolution filter (31x31) and (c) result of subtracting the filtered image from the original one.

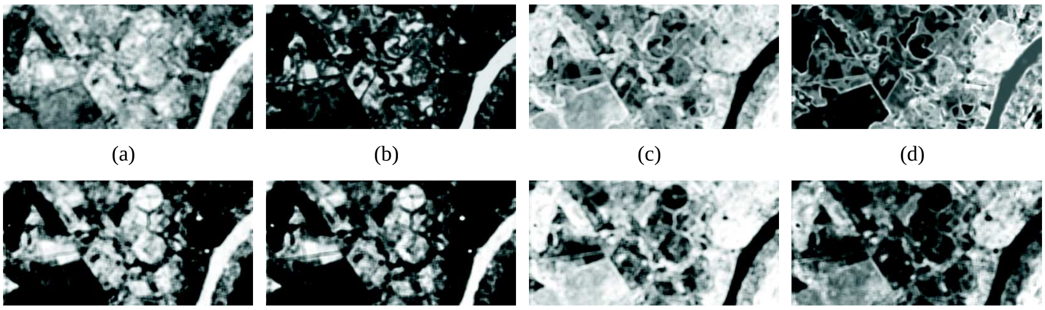

Figure 2: Illustration of the effect of eliminating low frequencies prior to extracting texture features (bottom line) compared to maintaining them (top line): a) inverse difference moment, b) angular second moment, c) entropy, d) correlation.

ature (Haralick et al., 1973; Conners and Harlow, 1980; Clausi, 2000). Some examples of the texture features derived from these measurements can be seen in Figure 2 (bottom line).

3.4. Step 4: training strategy

The training strategy have already been pinpointed as crucial elements of any supervised classification process (see Congalton (1988) for a review). An effort has been made to respect these guidelines. Other considerations have also been taken to account for the particular situation of using texture features and a dual resolution set.

The fact that texture is measured over a window of considerable size makes these measurements particularly sensitive to edges and borders so that these should be avoided as training sites. Some region's edges can be used as test sites (see step 6) to give them some representation. However, during field work, care should be taken not to choose training sites too close to object boundaries or linear elements (roads) as is often the case when access is difficult.

For each class a total of nine bundles of 324 pixels (18×18) that correspond to an area of $8100m^2$ which in turn correspond of a bundle of nine Landsat ETM pixels (3×3) were sampled from the scene. To avoid strong autocorrelation, only one in every six pixels and lines was actually sampled in order to match the Landsat resolution. This selection process yielded nine sample pixels for every bundle. These sites were selected based on a series of fieldwork episodes conducted between October 2001 and August 2003. Three of these bundles were used as training while the remaining six were reserved for testing purposes.

Classes can be quite different when texture is considered and special attention should be given to classes that can span over more than one texture. This may be the case for crops and their arrangement of furrows or even orchards that may have a different tree arrangement. It was necessary for instance to create a separate crop class to account for circular furrows (center-pivot). The same goes for pasture fields that have been separated according to the presence of trees and the presence of bare soil patches. The legend of Figure 4 lists all the classes considered for the two images.

3.5. Step 5: classification

One of the most common classifiers in remote sensing applications are the Mahalanobis distance classifier (MD) and the maximum likelihood (ML) parametric classifier based on Bayes' rule. Although these classifiers have already given good results with texture features some of their assumptions usually have to be re-

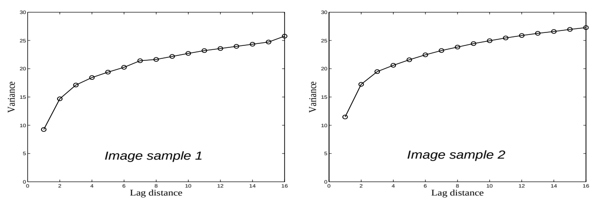

Figure 3: Isotropic semi-variograms of the two Spot-5 sub-images in Figure 4.

laxed, particularly the assumption of normality. The MD and ML classifiers can be sensitive to large feature space such as when spectral and textural features are used. Typically, the absolute probabilities generated with the MD classifier in the present case were below $1e-10$ (usually about 75% of the image). The Fisher linear discriminant (FLD) classifier being non-parametric is comparatively less sensitive to large feature spaces (Clausi, 2000). The FLD classifier uses the training data to construct a linear function that combines all the features (a one-dimension feature space) to maximize the variance between classes and minimize the variance within class (Duda and Hart, 2000).

3.6. Step 6: evaluating classification accuracy

Any map produced from indirect means such as remote sensing should be tested for accuracy. This is normally performed using ground truth samples independent from those used as training. Acceptable sampling schemes can be simple random, stratified random, systematic or stratified systematic unaligned sampling; they can use single or small clusters of pixels (Congalton, 1988). Clusters are used mainly in order to limit the costs of acquiring ground truth data or to account for positioning errors. Ideally these samples should be checked on the field but, for practical reasons, many studies choose to rely on indirect methods such as maps or photointerpretation. In the present case, field checking test samples was not possible for a number of reasons (e.g. distance, inaccessible terrain) so that our first approach was to split our initial field samples in two sets and to reserve one set for testing purposes. This however would produce an optimistically biased accuracy estimate since only relatively large homogeneous and unambiguous areas were chosen (Steele et al., 2003). Still these samples can be used in a comparison approach between classifiers. In order to give a more reliable estimate of accuracy, a simple random sampling scheme was used to select a total 100 clusters. These clusters were interpreted and the results of the interpretation was used to evaluate map accuracy. The McNemar test was used for comparing the results since the same samples were used for all classification tests and were not therefore independent as would require a Kappa difference test (Foody, 2004). The McNemar test computes a z statistic from a two by two matrix based on correctly and incorrectly classified pixels in both classifications as follows:

$$Z = \frac{f12 - f21}{\sqrt{f12 - f21}} \quad (1)$$

where $f12$ represent the pixels that are correctly classified in the first classification and incorrectly classified in the second classification and $f21$ represents the opposite situation. Z values of 1.96 and 2.58 were considered for the 95 and 99% levels of confidence respectively.

4. RESULTS OBTAINED WITH TWO SAMPLE IMAGES

4.1. Overall results

Results obtained with the two image samples shown in Figure 4 (only the Spot panchromatic band is shown) are presented in both tabular (Table 1 through 3) and graphical form (Figure 4). For all two image samples, both the MD and FLD classifiers have been tested and are compared in the discussion. Since prior probabilities were unknown, the ML classifier has been omitted from the present work. Finally the effect of pre-processing the panchromatic channel prior to texture feature extraction is shown at the end of this section (Table 3).

Table 1A shows the tabular results based on the arbitrarily selected *in situ* samples with the two image samples for both the MD and FLD classifiers. Table 1B shows the same classification results but based this time on a simple random sampling scheme in which most sites were interpreted and not directly visited. The two tables are divided into three categories according to the type of features used (spectral, textural or mixed). The first difference that appears is the relatively large difference between both sets of results. While the results of the first table are probably optimistically biased from choosing homogeneous and unambiguous sites, the second is pessimistically biased by adding interpretation errors to the classification errors. More realistic values are bound to stand somewhere in between but could not be calculated at this point. Apart from the performance level being much lower for the random samples, the two sets of results are consistent with one another and tend to reveal the same trends as to the most effective classifier and classification scheme (spectral, textural or mixed).

In nine out of twelve classifications (using both classifiers) the mixed classifications (spectral and textural) represent an improvement to both spectral only and textural only classifications. In all significantly different results (ten in all) the FLD classifier proved superior. Results from the MD classifier were slightly better in four cases but the difference could not be considered significant (H_0 rejected at the 95% confidence level). The graphical results for the mixed classification (Figure 4) show that the FLD classifier is not nearly as much affected by edges and borders as the MD classifier. This can partly be explained by the fact that the MD classifier makes assumptions about the normality of the distribution of the data and therefore can sometime expand a class probability far beyond what can be observed in the data. Edges and borders will typically have marginal values that will be "absorbed" by classes having high covariance values. The FLD classifier does not make assumption about the normality of the data features and is therefore less sensitive to these situations.

4.2. Individual per image results

Only the results of the mixed classification using the FLD classifier are presented in this section since this combination has proven to yield better results in the present research.

The first image comes from a complex blend of urban, rural and natural environments having very different textures. For most classes, user's and producer's accuracy scores (Table 2, top part) are fairly high for both sampling schemes except for the three classes of pasture which were eventually merged to increase the overall success rate. The classified image (top of Figure 4) produced with the FLD classifier is much more consistent and transition between classes is "cleaner" than with the MD classifier which tends to "insert" a thin band between objects. This observations tend to make the FLD classifier generally superior for dealing with textural and mixed classifications.

The second image sample is located in a complex karstic landscape where the rugged relief is responsible for much of the apparent textures. Riparian forest is constrained to two narrow canyons dominated by shadow. Spectrally, the differences between vegetation types are subtle. The classification results for this image (Table 2, bottom part) show that the best results were obtained with the FLD classifier and the mixed classification scheme. In this case however, the spectral classification scored much better than the textural one, perhaps because of the greater similarity between textures of natural vegetation. As in the first image sample, the graphical results of the mixed classification tend to prove the FLD superior for dealing with edges.

	A. Arbitrarily selected in situ samples								B. Random samples (interpreted)							
	Image 1				Image 2				Image 1				Image 2			
Feature set	$\hat{\kappa_1}$ MD	$\hat{\kappa_2}$ FLD	$\|Z\|$	sig.	$\hat{\kappa_1}$ MD	$\hat{\kappa_2}$ FLD	$\|Z\|$	sig.	$\hat{\kappa_1}$ MD	$\hat{\kappa_2}$ FLD	z	sig.	$\hat{\kappa_1}$ MD	$\hat{\kappa_2}$ FLD	z	sig.
Spectral	72.42	76.75	2.68	yes	75.45	75.49	0.0	no	46.44	52.18	6.19	Yes	57.02	56.20	1.20	no
Textural	74.43	83.23	5.03	yes	60.27	64.34	1.87	no	43.48	53.45	8.59	Yes	43.27	46.08	1.86	no
Mixed	88.04	87.21	0.65	no	79.07	84.95	3.99	yes	54.52	68.07	10.80	yes	56.91	65.44	7.72	yes

Table 1: Comparison of overall results obtained for the MD and FLD classifiers with spectral, textural and mixed classifications for the two images using two different sampling schemes. Z statistics are based on the McNemar test for related samples.

IMAGE SAMPLE 1

Class	Accuracy based on arbitrary samples (field data)			Accuracy based on random samples (interpreted data)		
	Kappa statistics			Kappa statistics		
	All classes:		87.21%	All classes:		68.07%
	Merging pasture types:		92.15%	Merging pasture types:		84.13%
	Sample size	Producer's accuracy	User's accuracy	Sample size	Producer's accuracy	User's accuracy
Urban/commercial	54	98.15	82.81	90	93.33	60.87
Urban/residential	54	83.33	88.24	88	75.00	71.74
Dry sclerophy. forest (deg.)	54	100.00	90.00	280	82.50	87.83
Dry sclerophy. forest (pre.)	54	90.74	98.00	79	93.67	74.00
Semi-deciduous forest	54	96.30	92.86	104	84.62	87.13
Pasture (w. trees)	54	70.37	69.09	237	26.58	41.45
Pasture (no trees)	54	100.00	80.60	233	66.52	69.20
Pasture (bare soil)	54	68.52	71.15	327	66.67	57.37
Crops (w. furrows)	54	100.00	100.00	81	100.00	87.10
Central-pivot	54	68.52	100.00	70	74.29	89.66
Sand bars/bare soil	54	81.48	91.67	95	76.84	85.88
Water	54	100.00	100.00	84	97.62	100.00

IMAGE SAMPLE 2

	All classes:		84.95%	All classes:		65.44%
Dry sclerophylous forest	108	100.00	90.76	586	90.44	72.11
Dense savanna	45	88.89	85.11	147	48.30	86.59
Scarce savanna	54	100.00	78.26	106	79.25	60.43
Regenerating savanna	54	55.56	96.77	63	58.73	51.39
Transition forest	63	65.08	68.33	135	40.74	50.00
Riparian forest	34	94.12	96.97	86	59.30	96.23
Rock outcrops (limestone)	54	100.00	100.00	99	98.99	100.00
Pasture (with trees)	54	90.74	87.50	105	71.43	60.00
Pasture (no trees)	54	68.52	100.00	126	49.21	100.00
Bare soil	31	93.55	52.73	52	88.46	45.54
Eroded/degraded land	54	83.33	93.75	144	38.19	77.46
Populated area	54	92.59	100.00	78	93.59	92.41

Table 2: Overall and class specific estimates of accuracy for Image 1 (top) and 2 (bottom).

4.3. The effect of eliminating low spatial frequencies

In an effort to measure the effect of pre-processing the images prior to extracting texture, classifications have been performed on data with and without low frequencies using both classifiers (MD and FLD). Table 6 shows the mixed classification results obtained with the raw data compared to pre-processing the panchromatic channel by eliminating low frequencies prior to texture feature extraction. The McNemar comparison test was used to compare the results that suggest that for both the MD and FLD classifiers, improvement can be achieved by eliminating low frequencies but that in half of the cases, this improvement is not significant. One can note however that in the case of the first image sample, the increase in precision is fairly important and that this image has the most predominant texture component. Further testing will be needed to quantify more thoroughly the effect of pre-processing images prior to extracting texture.

5. SUMMARY AND CONCLUSIONS

Six levels of consideration were presented in order to successfully incorporate texture in digital image classification. They are: 1) source of texture, 2) pre-processing channel for texture, 3) texture processing, 4) training strategy, 5) classification and 6) evaluating accuracy. Although the Grey-Level Co-occurrence Matrix approach to texture analysis was used in this paper, these considerations should be mostly independent of the texture approach used. The proposed methodology can be used as an effective way to merge higher resolution panchromatic data (Spot-5, five meter data here) with medium scale multispectral data (Landsat-7 TM here) to increase classification accuracy and/or the number of classes considered. The results are presented for spectral-only, textural-only and mixed classification schemes; the latter being superior in most cases depending on the classifier used and the relative importance of texture in the classes considered. The following conclusions can be outlined:

- In the context of merged Landsat ETM and Spot-5 data, mixed classifications generally brought improvement to either spectral or textural classification alone.
- The relative importance of each aspect (spectral and textural) varies according to the context of the scene.
- Eliminating low frequencies makes the texture features more stable and generally improves classification results.
- Training should avoid proximity to object boundaries especially if differences in texture can be observed.
- Because the Fisher Linear Discriminant classifier is nonparametric, it appears to respond much better to texture features and texture boundaries than the Mahalanobis Distance classifier.

	Kappa statistics ($\hat{\kappa}$)			
	Image sample 1		Image sample 2	
Classifier	MD	FLD	MD	FLD
Raw texture channel	44.91	60.02	56.33	63.88
Pre-processed texture channel	54.52	68.07	56.91	65.44
$\|Z\|$ (McNemar)	8.81	6.92	0.18	1.56
Significant ? (99%)	yes	yes	no	no

Table 3: Compared results between mixed classifications from pre-processed and raw texture channel using the random sampling scheme.

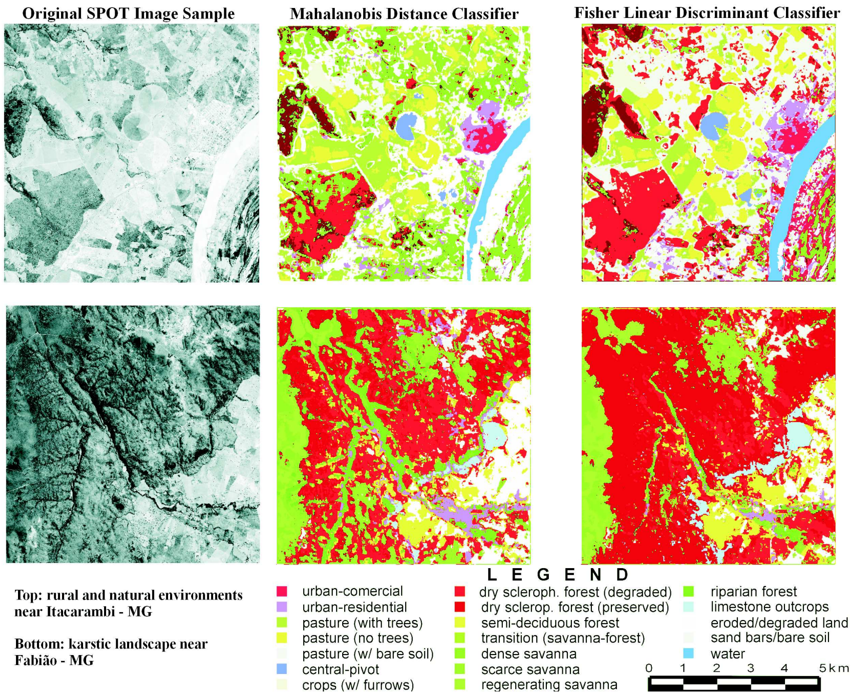

Figure 4: Classification results for the two image samples and the two classifiers tested (MD on the left and FLD on the right). Top line: image sample 1; middle line: image sample 2; bottom line: image sample 3.

The present article is intended as an effective way to incorporate texture in classification schemes for mapping purposes. It shows how to incorporate higher resolution panchromatic data with lower resolution multispectral data in a way that differs from normal fusion: through texture analysis.

References

Cao, C. and N. Lam (1997). *Scale in Remote Sensing and GIS*, Chapter Understanding the scale and resolution effects in remote sensing and GIS, pp. 57–72. Boca Raton, FLA: Lewis Publischers.

Clausi, D. A. (2000). Comparison and fusion of co-occurrence, Gabor and MRF texture features for classification of SAR sea-ice imagery. *Atmosphere-Ocean 39*(3), 183–194.

Congalton, R. G. (1988). A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated form remotely sensed data. *Photogrammetric Engineering and Remote Sensing 54*(5), 593–600.

Conners, R. W. and C. A. Harlow (1980). A theoretical comparison of texture algorithms. *IEEE Trans. Pattern Anal. Machine Intell. 2*(5), 204–221.

Duda, R. O. and P. E. Hart (2000). *Pattern Classification and Scene Analysis*. A Wiley-Interscience Publication, USA.

Ferro, C. J. S. and T. Warner (2002). Scale and texture in digital image classification. *Photogrammetric Engineering and Remote Sensing 68*(1), 51–63.

Foody, G. M. (2004). Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing 70*(5), 627–633.

Franklin, S. E. (2001). *Remote Sensing for Sustainable Forest Management*. Boca Raton, FLA: Lewis Publishers.

Haralick, R. M., K. Shanmugam, and I. Dinstein (1973). Textural features for image classification. *IEEE Trans. Sys. Man Cybern 3*, 610–621.

Julesz, B. (1962). Visual pattern discrimination. *IRE Transactionson Information Theory 8*(2), 84–92.

Maillard, P. (2003). Comparing texture analysis methods through classification. *Photogrammetric Engineering and Remote Sensing 69*(4), 357–367.

Steele, B., D. Patterson, and R. Redmond (2003). Toward estimation of map accuracy without a probability test sample. *Environmental and Ecological Statistics 10*, 333–356.