Comparing Texture Analysis Methods through Classification

Philippe Maillard

Abstract

The development and testing of two techniques of texture analysis based on different mathematical tools—the semi-variogram and the Fourier spectra—are presented. These are also compared against a benchmark approach: the Gray-Level Co-occurrence Matrix. *The three methods and their implementation are briefly described. Three series of experiments have been prepared to test the performance of these methods in various classification contexts. These contexts are simulated by varying the number, type and visual likeness of the texture patches used in classification tests. More specifically, their ability to correctly classify, separate, and associate texture patches is assessed. Results suggest that the classification context has an important impact on performance rates of all methods. The variogram-based and the* Gray-Tone Dependency Matrix *methods were generally superior, each one in particular contexts.*

Introduction

As scientists and researchers of the remote sensing community began to use high spatial resolution data, it soon became clear that spectral-based methods of computer classification and segmentation were doomed to yield unsatisfactory results. At high resolution, conceptual objects like forests or pasture usually show significant variations in their pixel values (Strahler *et al.*, 1986). Stationary in nature, these variations can give rise to an apparently regular spatial pattern referred to as *texture* (Kittler, 1983). One of the key elements that the interpreters use to identify and analyze images is clearly the spatial arrangement of color and tone that form natural visual entities: visual texture (Haralick *et al.*, 1973; Pratt *et al.*, 1978).

Because there is no universally accepted definition of visual texture, one has to choose a definition that best reflects the objective or the results being sought. The definition adopted here was given by Pratt (1991, p. 505): *natural scenes containing semi-repetitive arrangements of pixels*. The problem of analyzing and classifying texture has generated a wealth of studies and techniques that are seldom compared in a sys tematic way. This study is an experimental analysis of the problem of classifying texture using different mathematical tools. In particular, the specific classification context is analyzed in terms of the effect of *between-class variation* and *number of classes* on classification accuracy. To achieve the latter, a special experimental framework has been prepared and experimental results are presented and discussed.

The paper is organized in six sections. A short background review of feature extraction methods for texture analysis follows the introduction. Then the three approaches are described individually and compared through sample data. The fourth section describes the experimental framework and the data used for the experiments. Next comes the results and their analysis followed by the main conclusions.

Background

Reed and du Buf (1993) claim that most development in texture has been concentrated on *feature extraction methods* (sometimes called *channel-based* methods) which seek to extract relevant textural information and map it onto a special dedicated channel called a *feature*. The authors classified the various *fea ture extraction methods* as belonging to one of three possible classes: *feature-based*, *model-based*, or *structural*. Cocquerez and Philipp (1995) have used a similar classification of image segmentation methods which they compare in varioussituations (including textured images).

In feature-based methods, characteristics of texture (such as orientation, spatial frequency, or contrast) are used to classify homogeneous regions in an image. Model-based methods rely on the hypothesis that an underlying process governs the arrange ment of pixels (such as Markov chains or Fractals) and try to extract the parameters of such processes. Structural methods assume that a texture can be expressed by the arrangement of some primitive element using a placement rule. Feature-based, model-based, and hybrid methods have overwhelmingly dominated the scene in the last 20 years or so. One of their findings was that, although so many different methods have been devel oped, no rigorous quantitative comparison of their results had ever been done, which is a major theme of the present work.

Because Bela Julesz (1965) has shown evidence that human perception of texture could be modeled using second-order statistics (although he would later change his theory for the "texton" approach; see Julesz (1981)) many researchers have explored second-order statistics as possible features for texture analysis. Among the most common second-order statistics that have been used are the co-occurrence matrix, the spatial-autocorrelation, the covariogram, and the semi-variogram.

The frequency domain approach, also referred to as the Fourier Spectra approach, has been a long time favorite for texture analysis. From the early attempts at using it as a texture analysis tool by Rosenfeld (1962) to the recent use of Gabor functions as filters in the frequency domain to create frequency- and orientation-specific texture features (e.g., Fogel and Sagi, 1989; Jain and Farrokhinia, 1991; Manjunath and Ma, 1996), the Fourier transform offers infinite possibilities not only for texture analysis but for applications requiring the analysis of spatial frequencies and their orientation.

In order to evaluate a technique, it is necessary to have some base for comparison. In this research, the comparison will take

Universidad Federal de Minas Gerais, Departamento de Carto grafiá, Av. Antônio Carlos, 6627, Belo Horizonte MG 31270-091, Brazil (philippe@cart.igc.ufmg.br).

Photogrammetric Engineering & Remote Sensing
Vol. 69, No. 4, April 2003, pp. 357–367.

0099-1112/03/6904–357$3.00/0

the form of another technique that has already been widely accepted by the scientific community and obviously performs well. This method was proposed by Haralick *et al.* (1973) who have named it *Gray-Tone Spatial-Dependence Matrices*, also known as *Gray-Level Co-occurrence Matrices* (GLCM). Not only do almost all the authors in visual texture analysis quote the GLCM, but many have already used it as a comparison technique. Among them, Davis *et al.* (1979), Conners and Harlow (1980), Pratt (1991), Bonn and Rochon (1992), Wu and Chen (1992), Reed and du Buf (1993), Dikshit (1996), Franklin *et al.* (2001) and Zhang (2001) have either used the GLCM method as a comparison or have described it in their review.

It was decided that all three methods be *rotation-invariant* so that the particular orientation of texture would not be considered even though it was observed that considering particular orientations can increase classification accuracy (Franklin and Peddle, 1989; Maillard, 2001). This was important especially for the *crops* and *waves* classes for which the factors controlling their orientation are difficult to predict.

Description of the Three Texture Feature Extraction Methods

The Variogram Approach

Many authors have already shown the potential of the variogram as a texture analysis approach (Serra, 1982; Woodcock *et al.*, 1988; Ramstein and Raffy, 1989; Miranda *et al.*, 1992; Atkinson, 1995; St-Onge and Cavayas, 1995; Lark, 1996). On the one hand, the variogram is related to other statistical approaches like the autocorrelation function and the fractal Brownian motion (Xia and Clarke, 1997). On the other hand, it is computationally simple and easy to interpret as a graph. One point in which the variogram appears more appropriate is that only weak stationarity is assumed, in other words, the expectation only has to be constant locally (Woodcock *et al.*, 1988).

It appears, however, that most techniques using the variogram do so in the geostatistical manner, i.e., a model is usually applied whose parameters are taken as a way of describing the semi-variogram curve. In Remote Sensing images some texture-based variograms might be best modeled using the *spherical* model while others are best represented with an *exponential* or even *sinusoidal* model. This poses a problem in terms of creating a systematic approach. One solution would be to use the "best" model type, selected as a texture feature. But using a nominal scale feature would cause problems further down the classification process. This would also imply that a battery of models would have to be fitted for all pixels of all texture samples, and the cost in terms of computing would be high. For these reasons and because others have already pursued that line of research, the "traditional" function representation of *sill* and *range* has not been considered here.

Another point that has received attention is the alternate use of the *mean square-root pair difference* (SRPD(h)) function proposed by Cressie and Hawkins (1980) as a semi-variance estimator which is resistant to outliers. Lark (1996) has also shown that for four different classes of texture (urban, farmland, woodland, and meadow), when tested for normality, the SRPD(h) function scored much better than the γ(h) function, a fact confirmed by an earlier study by the author (Maillard, 2001).

Considering these findings, a number of considerations were taken to implement the texture feature extraction based on the variogram:

- a rather large window had to be used in order to cover larger distance lags (up to 32 pixels),
- the texture feature set had to be rotation-invariant but had to preserve anisotropy, and
- the behavior of the SRPD graph near the origin had to receive special attention because it bears a special significance in terms of micro-texture (Serra, 1982; Jupp *et al.*, 1989; Xia and Clarke, 1997).

After numerous tests using different ways to transform the variogram into texture features, the most promising approach was found to be the averaging of selected distance lag intervals. The SRPD texture feature extraction routine can be summarized in the following steps:

- For every pixel in the image, a neighboring window (32 by 32 pixels) is considered and four directional variograms (0°, 45°, 90°, and 135°) are computed for all possible combinations in that window.
- The maximum lag size is equal to one half the window size.
- The *mean Square-Root Pair Difference* is used as a semi-variogram estimation.
- Six values (features) of the SRPD are computed by giving more weight to the values corresponding to the smaller lags; in other words, by computing the SRPD features over regular intervals on a logarithmic scale. Figure 1a illustrates such a scale with six lag ranges while Figures 1b and 1c illustrate the semi-variance graph before and after the averaging, respectively.
- These values are computed for all four directions for a total of 24 features
- The 24 directional features are then transformed to 18 rotation-invariant features: for each lag, the *mean*, *standard deviation* and *sum of perpendicular ratios* ($\Sigma[\gamma_0/\gamma_{90} + \gamma_{90}/\gamma_0 + \gamma_{45}/\gamma_{135} + \gamma_{135}/\gamma_{45}]$ where γ is the estimate of variance) are computed where the latter two parameters are meant to preserve anisotropy in the data.

The Fourier Approach

The Fourier-based method proposed by Stromberg and Farr (1986) was surely the first really successful use of the Fourier Transform through the application of a series of ring (band-pass) filters before applying the inverse transforms from which the texture features resulted. This was consistent with the findings of Richard and Polit (1974) and Harvey and Gervais (1978) in their psychophysical experiments. But, as Caelli (1982) would later show, it lacked the orientation component. This was largely corrected by the method proposed by Wilson and Spann (1988), which divided the spectral domain into a series of band-pass/orientation-pass filters defined by a *finite prolate spherical* function. Others then suggested the use of a bank of Gaussian filters applied locally (as opposed to the whole image) in the frequency domain to create the textural features (Gorenic and Rotman, 1992). The most popular Fourier-related texture analysis method is without doubt the "wavelet" transform which uses a similar approach but replaces the Gaussian filters by the FT of a Gabor function, also applied locally (Reed and Wechsler, 1990; Jain and Farrokhinia, 1991; Dunn *et al.*, 1994; Manjunath and Ma, 1996). The method proposed here makes use of these most recent findings but also attempts to simplify significantly these methods and make them more computationally efficient.

In an effort to simplify the approach and to avoid a two-dimensional transform, the Fourier approach was implemented through appending all lines of pixels in the window in four directions (0°, 45°, 90°, and 135°). In this approach, the frequencies between zero and the number of lines appended are artifact and irrelevant because the appending of lines of image pixels is an artificial process. These artifact frequencies were eliminated. The following steps summarize the computation of the Fourier texture features:

- For every pixel in the image, a neighboring window is considered and four directional transforms (0°, 45°, 90°, 135°) are computed for all appended lines; frequencies between 0 and 32 were eliminated; a window size of 32 by 32 pixels is used for the 0° and 90° directions and of 64 by 64 pixels for the 45° and 135° directions.
- Six Gaussian filters are applied to the transform to create the following texture frequency features: *very low* (centered on 1 cycle per window of 32 pixels), *low* (2 c/w), *mid-low* (4 c/w), *mid-high* (8 c/w), *high* (14 c/w), and *very high* (25 c/w). An

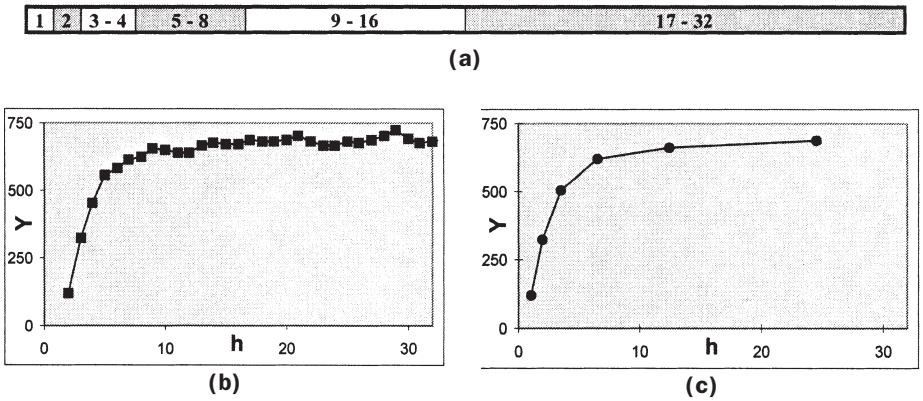


Figure 1. The construction of texture features for the Variogram approache to texture classification. (a) The logarithmic scale used to average SRPD values. (b) The original SRPD graph showing all the values. (c) The six SRPD values based on averaging the semi-variance according to a logarithmic scale.

average value is computed for each filtered result of each direction for a total of 24 directional features. Figure 2 illustrates the process of filtering; in Figure 2a the Gaussian filters are presented while Figure 2b shows the effect of applying the filters on a sample transform of a forest image.

- The 24 directional features are then transformed to 18 rotation-invariant features: the *mean*, *standard deviation* and *sum of perpendicular ratios* are computed for each frequency band.

The Gray-Level Co-occurrence Matrices

The Gray-Level Co-occurrence Matrices (GLCM) method was implemented in a manner similar to its original form (in Haralick *et al.*, 1973). Because many of the features first described by Haralick are highly correlated among themselves (Haralick *et al.*, 1973), a pre-selection was done to reduce the 14 possible measures to less than half. The selection was done by combining all the features used by many different research teams that have used the GLCM method. Table 1 gives a listing of the authors considered and the texture features they have used. Analyzing the table revealed that the most commonly used features are in decreasing order of popularity: *Angular Second Moment*, *Entropy*, *Inertia* (initially contrast), *Correlation*, and *Inverse Difference Moment*.

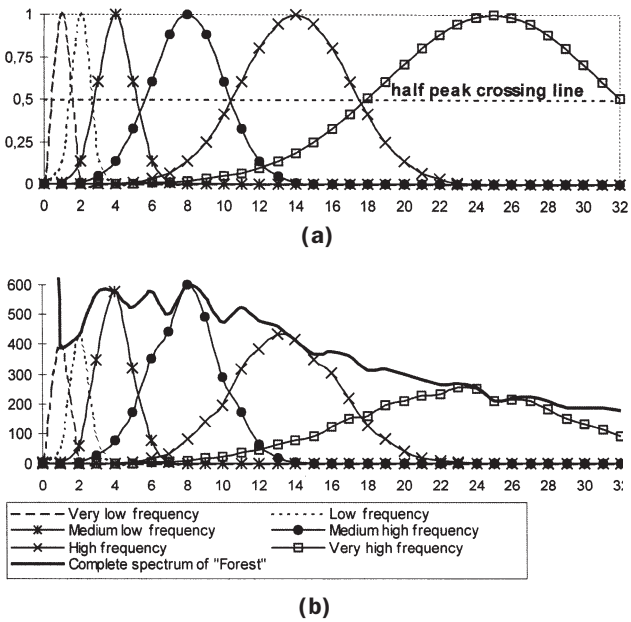


Figure 2. Filtering the Fourier transform of a single (forest) pixel. (a) The six Gaussian filters. (b) Result of applying the six filters on the Fourier transform of the neighboring window of a forest pixel (0° direction).

Apart from the texture features used, pixel pair sampling distances have to be chosen with respect to the expected spatial frequencies present in the images. The choice of sampling distance is as important as the types of measurements. In order to be as objective as possible, the sampling distances have been chosen based on the visual analysis of the semi-variograms of the sample texture patches. This analysis yielded the following distances: three, six, and twelve pixels. In their original setting, Haralick *et al.* would choose a particular sampling distance and then rotate it by steps of 45 degrees so that, for a distance of three pixels, the *x*,*y* sampling distances setting would be (3,0), (3,3), (0,3), and (−3,3) for the 0°, 45°, 90°, and 135° orientations, respectively. Then, for each sampling distance, the mean and standard deviation would be computed over the four orientations instead of using each orientation separately. Therefore, the features are not orientation-specific but still account for some effect of anisotropy. This approach is meant to obtain rotation-invariant features similar to those adopted for the variogram and Fourier methods.

The following steps summarize the implementation of the GLCM method:

- For every pixel in the image, a neighboring window (32 by 32 pixels) is considered and matrices are computed for three different pixel pair sampling distances (3, 6, and 12 based on analysis of the variograms) and four directions (0°, 45°, 90°, and 135°).
- Five measurements are computed for each matrix: *contrast*, *angular second moment*, *inverse difference moment*, *entropy*, and *correlation* for a total of 60 features.
- The 60 "directional" features are transformed into 30 rotation-invariant features by computing the *mean* and *standard deviation* over the four original directions.

Comparison of the Three Approaches

Displaying the texture features from the three methods as images would be of little help in understanding how these

TABLE 1. LIST OF THE GLCM FEATURES USED IN SOME TEXTURE ANALYSIS STUDIES

Author(s) and Year	Feature Used (See Legend at Foot of Table)
Haralick *et al.*, 1973–Three Experiments	(a) *ASM*, *Con*, *Cor* (b) *ASM*, *Con*, *Cor*, *Ent* (c) *ASM*, *Con*, *Cor*, *SSq*, *IDM*, *SAv*, *SV*, *SEnt*, *Ent*, *DV*, *DEnt*
Haralick, 1979	*ASM*, *Con*, *Cor*, *IDM*, *Ent*
Davis *et al.*, 1979	*ASM*, *Con*, *Cor*, *Ent*
Conners and Harlow, 1980	*ASM*, *Cor*, *IDM*, *Ent*
Rosenfeld and Kak, 1982	*ASM*, *Con*, *IDM*, *Ent*
Gonzalez and Wood, 1992	*ASM*, *Cor*, *IDM*, *Ent*, *Other*
Sali and Wolfson, 1992	*ASM*, *Con*, *IDM*, *Ent*, *DEnt*
Wu and Chen, 1992	*ASM*, *Con*, *Cor*, *IDM*, *SAv*, *SV*, *SEnt*, *Ent*, *DV*, *DEnt*, *IMC*
Anys and He, 1995	*SSq*, *ASM*, *Ent*, *Con*, *Cor*, *Others*
Dikshit, 1996	*Ent*, *Con*, *IDM*, *ASM*
Hay *et al.* (1996)	*ASM*, *Ent*, *IDM*, *SV*

Legend: *ASM*: Angular Second Moment, *Con*: Contrast (or inertia), *Cor*: Correlation, *SSq*: Sum of Squares, *IDM*: Inverse Difference Moment, *SAv*: Sum Average, *SV*: Sum Variance, *SEnt*: Sum Entropy, *Ent*: Entropy, *DV*: Difference Variance, *DEnt*: Difference Entropy, *IMC*: Information Measure of Correlation, *MCC*: Maximum Correlation Coefficient, *Other*: Other measures.

multi-dimensional spaces can help separate textures of various origin. A graph representation was preferred for its simplicity and ease of interpretation. Figure 3 illustrates the data generated by each method as a series of graphs for six texture classes: i.e., *forest*, *residential*, *desert*, *crops*, *shrub*, and *waves* (samples of these classes are presented in Figure 4).

The semi-variogram (Figure 3a) displays variance as a function of lag distance. The sill reached by each texture

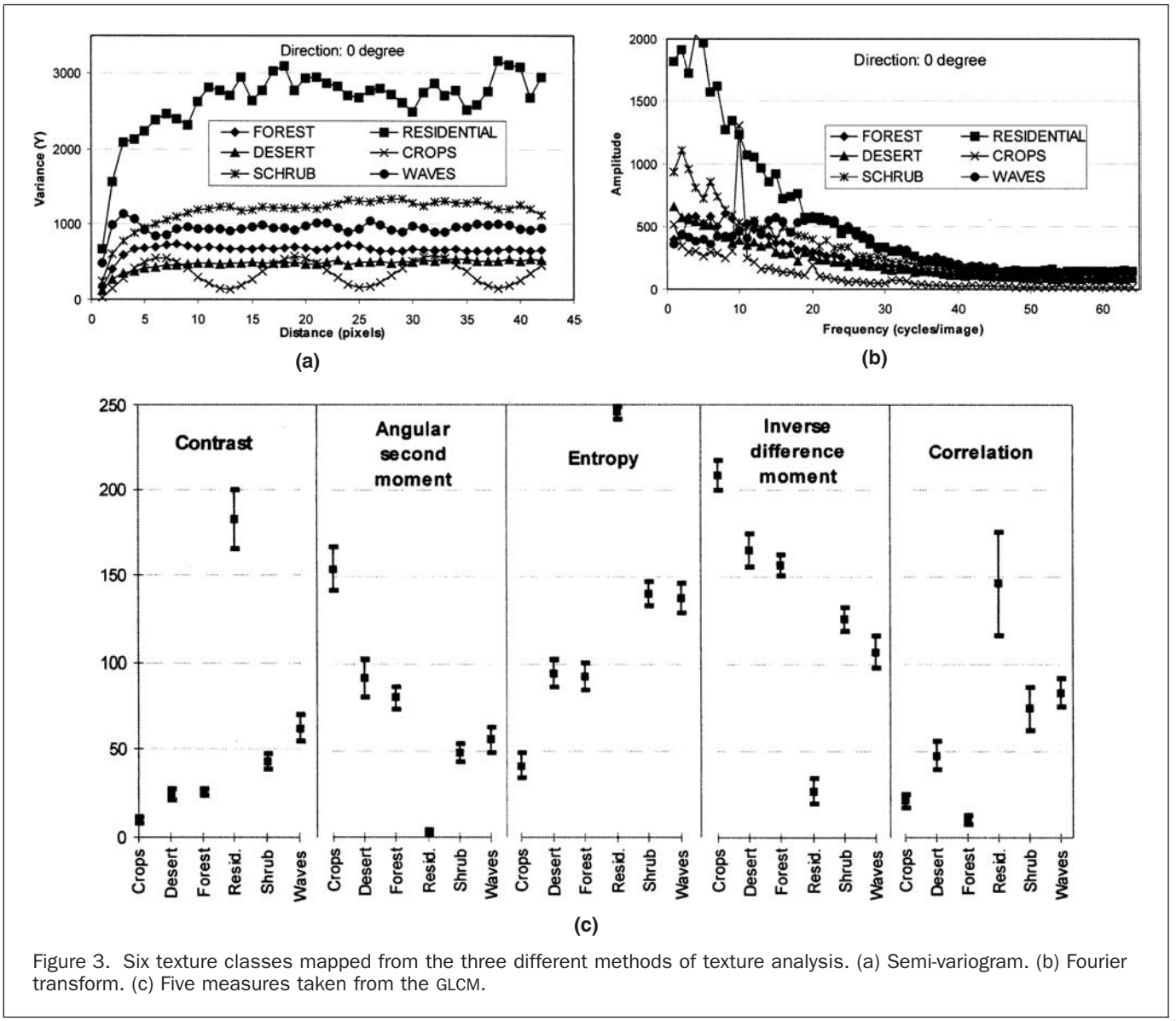


Figure 3. Six texture classes mapped from the three different methods of texture analysis. (a) Semi-variogram. (b) Fourier transform. (c) Five measures taken from the GLCM.

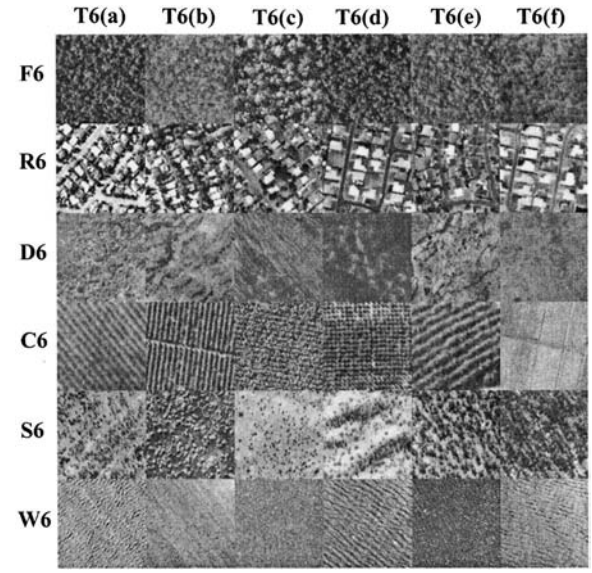


Figure 4. Data set composed of 36 square patches of texture from six generic classes: forest (F6), residential areas (R6), desert (D6), crops (C6), shrub (S6) and waves (W6). Each column (T6(a) to T6(f)) contains one sample of each generic class.

clearly separates them from each other, but it is the behavior of the curve near the origin that is most determinant as a discriminating factor. The most striking features are the height and irregularity of the *Residential* class, the cyclic shape of the *Crops* and, to a lesser extent, *Waves* classes. Interestingly, the latter reaches its peak at a distance of three pixels and then drops and stabilizes with an undulating pattern. Apart from the height of the sill and the level of regularity of the curve, it is the slope described by the first few points that better differentiate the six texture classes. The fact that these first few data points have received more attention in generating the texture features made it possible to preserve this important aspect.

The one-dimension Fourier transform of the same six texture classes is presented in Figure 3b. The graph was constructed as described above by appending all lines (in the horizontal direction in this case) and then eliminating the artifact frequencies. Interestingly, the average amplitude of each texture frequency curve follows the same progression order as in the sill reached in the semi-variogram graph; in increasing order: *Crops*, *Desert*, *Forest*, *Waves*, *Shrub*, and *Residential*. As before, the *Crops* class has a well-defined frequency peak but so has the *Shrub* class, a feature not easily perceived in the semivariogram. The high frequency part of the curves (lower end) is much more confused and clear patterns are hard to identify.

For the *Grey-Level Co-occurrence Matrix*, a series of *mean-standard deviation* plots were constructed for each co-occurrence based measure: *contrast*, *angular second moment*, *entropy*, *inverse difference moment*, and *correlation*. The *mean* marks the center of each plot line while the height is determined by the *standard deviation*. Generally speaking, all five measures appear to separate well the six classes analyzed except for two class pairs: *Desert—Forest* and *Shrub—Waves*. The former can only be distinguished by the *correlation* and the latter by the *contrast* and the *inverse difference moment*.

Experimental Methodology

A special experimental framework is needed in order to evaluate and compare texture analysis methods (Smith, 2001; Maillard, 2001). Such a framework has already been proposed by the *MeasTex* project through three aspects: (1) the implementation of major texture processing paradigms, (2) a library of texture images, and (3) a measure of their performance (MeasTex, 2001). In this work, some further refinements of such framework are proposed through the use of a texture base exclusively dedicated to visual textures found in high-resolution images of the Earth and a series of specific experiments, both of which are described below.

Data Set Description

The texture samples were extracted from scanned aerial photographs of the Emerald (desert samples) and Brisbane areas (both in Queensland, Australia) and correspond to six different generic classes of texture: forest, residential areas, arid environment, agriculture, shrub, and ocean waves. Figure 4 presents the texture images used while the generic classes are briefly described next.

Forest includes basically all natural tree formations, both closed and open. Reforestation is not included in this class.

Residential has been preferred to Urban because it usually covers more extensive areas. Although all residential texture patches have been selected in the Brisbane area, it could incorporate other types of residential areas like shantytowns or other suburbs.

Desert is a rather broad class that includes natural areas with scarce vegetation in which the soil, rock, and low vegetation are responsible for the texture patterns.

Agriculture is another very broad class that was intended to incorporate all structured forms of agriculture including orchards. Pasture is a somewhat marginal member of this class only when it shows some artificial pattern.

Shrub was basically meant to complement the forest class for low and open naturally vegetated areas. This class justifies itself by its extent in many countries like Australia.

Waves is a generic class of bodies of water under the influence of wind or currents. Both ocean and lakes are included, but rivers remain a marginal member when wide enough to produce a visible wave pattern.

Six samples of each generic class were selected for a total of 36 texture image samples of 128 by 128 pixels each. To assess the appropriate resolution, a *local variance* was computed for a series of resolutions in the same manner suggested by Cao and Lam (1997) and yielded a graph with a first break at about two meters and a peak at about three. Scanning the 1:25,000-scale black-and-white photographs at a resolution of 450 dpi yielded a ground resolution of about 1.4 meters and ensured that most spatial variability was preserved without generating unnecessarily large image sets. All texture samples are presented in Figure 4.

As can be observed, each column of the data set contains a different sample of each generic texture class (named T6(a) to T6(f)) while each row is composed of six different samples of a single generic class (named F6, R6, D6, C6, S6, and W6). Although the choice criteria were somewhat suggestive, care was taken to pick samples with low *within-class* variability but high *within-generic* class variability. For example, each individual *Forest* sample had to be homogeneous (texture-wise), but the generic class is represented by diversified samples of forest.

The first series of classifications will be carried out on each column of texture samples separately. This was a way of simulating six different classification contexts (as, for instance, with different photographs). The second series of classifications will be performed on each row of the set to simulate thematic conditions in which one wishes to distinguish similar yet different classes of texture. The last series of classifications will include all 36 samples.

The Comparison Tool

Classification is a statistical tool used to assign an object to one of a number of possible groups or classes with the help of a decision rule (James, 1985). The classification technique is (usually) a point-dependent algorithm, meaning that only one pixel is considered at a time. In order for it to consider region-based information such as texture, this information has to be first extracted in the neighbourhood of each pixel and coded on a different plane, which is called a *texture feature* here and is performed here by one of three *texture processing* algorithms. Classification was preferred over segmentation approaches for various reasons:

- Classification algorithms have been used in Image Processing for quite some time and are probably the most widely used pattern recognition techniques,
- The decision rules of classification algorithms are simple and easy to control whereas many region-dependant segmentation algorithms use a threshold that can be hard to set, and
- The texture feature extraction (using one of the texture analysis methods) can yield an analysis of the behavior of the various textures for each method.

In the present study, the problem is not to assign each pixel to one of a group of possible classes in the test image because the class membership is already known for all pixels in the image. An alternate formulation would be: *can a given set of variables describing the texture of a pixel makes its assignment to the "right" class possible*? Or rather, to what extent can the "right" texture class be assigned to the texture patches forming the test image given one of three methods of texture analysis? To ensure the most objective approach, a "blind" technique has been used for defining the training class samples. Training was done through a mixture of *systematic* and *cluster sampling*: sixteen equally spaced clusters of 25 pixels (5 by 5) ensured that the class was well represented (2.44 percent) and that the samples were spread throughout the image. Although other studies have shown that clusters of up to 25 pixels reduce bias due to autocorrelation (Congalton, 1988), it can still be expected within the clusters because texture processing is always based on neighboring windows of pixels.

For classification to be considered a testing tool, a number of assumptions had to be made, six of which are described here:

- Bayes' rule is optimal only if minimizing Total Errors of Classification (TEC) is considered the "best" criterion; for instance, it does not take classification error cost into account (James, 1985).
- In its normal form, the Bayes' rule assumes that all variables from the measurement vector have a normal distribution. However, as was pointed out by Scheffé (1959), because the sample size is reasonably large (400) and is the same for all textures, the assumption of normality can be relaxed (Clark and Hosking, 1986). Some consideration of normality has already been taken by preferring the Square-Root Pair Difference (SRPD) to the straight semi-variance because it has been found to have a more normal distribution and is more resistant to outliers (Lark, 1996).
- TEC and its complement, Total Success of Classification (TSC), are acceptable ways to express the error measure through a confusion matrix. TEC is obtained by subtracting the number of correctly classified objects from the total number of pixels. However, the Kappa statistic also takes the mere chance factor into account (Foody, 1992) and so it was preferred as a classification accuracy value.
- It is usually assumed that the conceptual objects being classified have low *within-class* variability but, in the case of texture, such a statement would have to be reviewed because *texture variability* is not yet a well defined concept.
- All *a priori* probabilities are equal because all texture samples are of equal size.
- Overall accuracy and the edge effect. The fact that the training areas are part of the texture sample being classified brings a bias in the classification results and so these pixels were withdrawn from the overall accuracy computation. Also, the fact that no special consideration has been taken for the borders and edges of the texture samples can be seen as another source of bias. In this case the bias is partly wanted in order to assess how each method behaves with respect to texture edges. To evaluate better how these edges alter the results, a separate Kappa statistic was computed where the edges (and their *zone of influence*) have been removed from the computation. This is consistent with the fact that many authors consider that texture analysis and texture edge recognition should be approached as different tasks. Pratt (1991) suggested that boundaries should be established prior to attempting texture measurements. Serra (1982) treats the *edge effect* from what he calls the *individual-textural* paradigm. Recent texture analysis methods that integrate both *region growing* and *edge detection* such as the one proposed by Pavlidis and Liow (1990) are the consecration of such an approach.

Classify to Associate, Generalize, Merge, and Separate

Although classification can be defined simply as a way of assigning an object to one of a number of possible groups, it is in fact a complex operation by which a geographical meaning is given to a set of spectral and spatial measurements. According to Nyerges (1991), four abstraction types have been used to support meaning in geographical information: *classification* ("*instance-of*" relationship), *association* ("*member-of*" relationship), *generalization* ("*kind-of*" relationship), and *aggregation* ("*part-of*" relationship). Often, a classification process is expected not only to classify but also to generalize and sometimes even to associate objects depending on the amount of information available about the region being classified. Classification is used as a generalization tool by expecting it to eliminate unnecessary details, sometimes through the use of filters either prior to or after the classification process. Association is sometimes expected as in the case of opposing slopes of a hill receiving different amounts of light and having different spectral signatures but that should nonetheless be associated to the same class.

On another level of reasoning, the process of classification is one that has both the functions of joining and separating objects as belonging to the same class or not. In a general *land-use/land-cover* type classification, the different types of forest might not be known or relevant, in which case the classifier is expected to be able to join all forest types into one broad class. Not having the information on all forest types present also prohibits effective sampling. In another situation, a user might want not only to differentiate the forest types but also their health status. Can a single classification algorithm be an efficient tool for both situations? Part of the answer might lie in the choice of the number and type of features (or variables) being used. Experiments have been selected to assess both types of situation as well as the generalization problem.

Description of the Experiments

It is clear, from the observations above, that no single experiment could assess all of these problems. Mike James (1985) points out that classification methods have been criticized for working well during tests but giving poor results in a "real" situation but that this situation usually results from an inadequate test procedure. To ensure adequacy of the test procedure, three different classification experiments have been prepared in order to assess all of the above aspects. These experiments are explained below.

- First experiment: evaluating the different approaches for separating and classifying sets of texture patches having very different visual characteristics and belonging to different generic classes of objects.
- Second experiment: assessing the separation capability of the methods by comparing their performance in separating sets of similar textures belonging to the same generic class and evaluating their relative performance according to the textural context (e.g., forest versus residential).

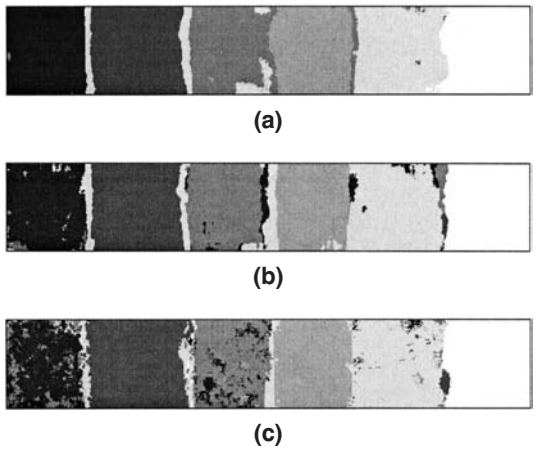


Figure 5. Classification results from all three methods using the T6a texture subset. (a) The GLCM approach. (b) The VGM approach. (c) The FFT approach.

- Third experiment: separating a mixture of both different and similar textures and testing the capacity of good association to test the performance of each method for the classification of a large number of classes (36) belonging to both very similar and very different geographic realms.

Results and Discussion

First Experiment: Separating a Set of Very Different Textures

The first experiment has been organized in the following manner: first, the gray-level dependency matrix method (GLCM) was applied, then the variogram method (VGM), and finally the Fourier-based approach (FFT). In order to give more reliability to the results obtained, the classification tests were carried out on all six similar (but different) sets of six texture patches each. Figure 5 shows the graphical results of the best results obtained with the T6a set for each method while Table 2 shows the tabular results of six sets of textures samples.

The results clearly show that all three methods of texture processing approaches have very good potential for classification purposes because, in all cases, the Kappa statistics obtained are mostly over 70 percent (one exception), and over 90 percent if edges are not computed. Table 3 presents a pairwise test of significance between the three Khat's obtained for each set. The test uses the normal deviate (Z statistics) to determine if the confusion matrices are significantly different (Cohen, 1960; Congalton, 1991; Foody, 1992). Analysis of Table 3 reveals that differences as low as 1 percent or less can still be significant given their context and the large samples used. In fact, only two pairwise comparisons did not prove significantly different at the 99 percent level of confidence (but would be at 95 percent).

Table 2. Classification Scores (Kappa Statistics) from the Classification of Six Texture Sets; White Columns Show Overall Results While Gray Columns Show the Results without Considering the Edges. Best Results Are in Bold and the Last Row Shows the Average Score

	GLCM Method		VGM Method		FFT Method	
Texture set	$\hat{K} \times 100$ Overall	No edges	$\hat{K} \times 100$ Overall	No edges	$\hat{K} \times 100$ Overall	No edges
T6a	**90.1%**	**99.8%**	86.1%	99.4%	82.2%	93.1%
T6b	79.2	98.0	**80.8**	**99.3**	73.9	95.5
T6c	76.0	99.1	**78.5**	**99.9**	76.4	97.1
T6d	**79.8**	**100.0**	75.8	95.5	67.6	95.3
T6e	**75.4**	**99.4**	74.8	98.9	74.4	92.1
T6f	**76.6**	99.3	75.2	**99.5**	70.1	93.1
Average	**79.5**	**99.3**	78.5	98.8	74.0	94.4

Table 3. Results of Pairwise Comparison of Khat Values for *Overall* Results of Table 2; S = Significant at the 99% Level of Confidence, NS = not Significant at the 99% L.O.C.

	GLCM *vs* VGM		VGM *vs* FFT		FFT *vs* GLCM	
Data set	Z Statistic	Result	Z Statistic	Result	Z Statistic	Result
T6a	−44.60	S	24.21	S	20.50	S
T6b	−24.26	S	−7.40	S	31.68	S
T6c	1.96	NS	−11.59	S	9.64	S
T6d	−53.16	S	18.31	S	34.78	S
T6e	−4.20	S	2.35	S	1.85	NS
T6f	−28.02	S	6.13	S	21.90	S

Given the slightly superior (but mostly significant) success rate of the GLCM method, the first experiment has failed to demonstrate that the alternative methods proposed can bring any improvement. These results suggest that the GLCM method, when used properly, is a very efficient approach to simple texture classification problems where the texture characteristics are visually very different.

The second best method, the VGM approach, shows slightly poorer results but the difference in Khat (4 percent at the most and about 1 percent on average) can be considered small but important, given the fact that the differences are significant and apply to four subsets out of six.

The same reasoning can be applied to the FFT results which are systematically the lowest for both kinds of result (*overall* and *no edges*). The fact that, in the frequency domain, the high frequencies (which correspond to small lags or sampling distances in the other methods) are usually regarded as noise can perhaps explain the poorer performance of the FFT method which is typically used to separate signal (larger frequencies) from noise.

In all three methods and for all six texture sets, edges account for about 20 percent of misclassification errors on average. This is a very significant number which tends to confirm the fact that texture boundaries have a major role in classification errors (Ferro and Warner, 2002) and should perhaps be extracted prior to classification as suggested by various authors (Pavlidis and Liow, 1990; Pratt, 1991; Jones, 1994). This also outlines the role of patch size, hence resolution: in a texture patch of 128 by 128, considering a texture analysis window of 32 by 32, the zone of influence of the texture edge is equal to $128^2 - [128 - 32]^2$ or 7168 pixels in this case, which is about 44 percent of the total number of pixels of the patch. This percentage drops to 24 percent for a 250 by 250 patch, 12 percent for a 500 by 500 patch, and 6 percent for a 1000 by 1000 patch.

Second Experiment: Separating Sets of Similar Textures

In the second series of experiments, six texture sets composed of similar texture samples were created (Figure 4) in order to test the capacity of the three methods to separate textures belonging to the same generic classes. Additionally, this series of experiments helps characterize the particular aspects of texture that are better taken into account by the three different methods.

Table 4 shows the compiled results for all six texture sets for both *overall* and *no edges* results. In four out of six texture sets, the VGM method performed better by yielding superior TSC scores of ≈1% (F6), ≈4% (D6), ≈2.5% (C6), and ≈2% (W6), all

TABLE 4. CLASSIFICATION SCORES (KAPPA STATISTICS) FROM THE CLASSIFICATION OF THE SIX TEXTURE SETS EACH COMPOSED OF TEXTURE PATCHES BELONGING TO THE SAME GENERIC CLASS. BEST RESULTS FOR EACH ARE IN BOLD

	GLCM Method		VGM Method		FFT Method	
Texture Set	$\hat{K} \times 100$ Overall	No edges	$\hat{K} \times 100$ Overall	No edges	$\hat{K} \times 100$ Overall	No edges
F6 (forest)	77.7%	93.4%	**78.6%**	93.3%	72.3%	87.0%
R6 (residential)	**76.7**	**93.2**	60.1	81.9	46.7	64.3
D6 (desert)	74.1	91.7	**77.9**	**95.4**	75.3	87.5
C6 (crops)	92.6	100.0	**95.0**	100.0	85.4	98.4
S6 (shrub)	**83.1**	**97.3**	73.7	92.4	69.4	87.4
W6 (wave)	86.9	98.2	**88.8**	**99.6**	84.4	95.3

of which were significant at the 99 percent confidence level. As for the other two cases (R6 and S6), the GLCM method was superior by about 16.6 percent and about 9.4 percent, respectively. In all cases, the VGM and the GLCM methods were significantly better than the FFT approach by a margin of 2.5 percent to 30 percent in Khat scores. The results for *no edge* scores are as high as 100 percent in some cases. The difference between *overall* and *no edge* results are about 14 percent on average for the three methods which can be considered high given the sample size.

The first texture set for which the GLCM proved superior, R6, is also characterized by relatively low Khat for all three methods. One conclusion that this brings is that the R6 texture set is a poor candidate with ill-defined textures. Another observation is that the fact that the GLCM method includes a broader variety of measurements is possibly the reason why it scores better whereas the other methods are more "specialized." In the second texture set for which the GLCM shows superior results, S6, the situation is different but still keeps similar elements. Although the Khat scores are higher (roughly between 65 percent and 80 percent), a visual inspection of the individual texture patches (see Figure 4) reveals that some of them are not very homogeneous, having sometimes a dual textural characteristic (patches #3, #4, and #6 in particular), which tends to give more weight to cues other than the simple spacing of apparent objects on the background scene (which is the basis for the variogram approach). In this regard, the GLCM method has a definite superiority over the other two. This suggests that measurement type diversity can prove an important asset for textures that are not necessarily blessed with a homogeneous visual aspect.

As for the FFT method, its generally poorer performance can be attributed to two different facts. The first one is inherent to the approach (or its implementation) that was chosen, involving the appending of consecutive lines in a semi two-dimensional approach instead of the full two-dimensional Fourier transform. This approach might have created undesired artifacts (for instance, the phase of the frequencies is not respected in this approach). The second one is that, unlike the real time series for which the Fourier transform was developed, spatial frequencies in these texture sets are ill-defined and often require a complex set of sine-like waves to describe square-like shapes (as in the case of residential areas).

Third Experiment: Separating a Mixture of Both Different and Similar Textures

In the first part of this experiment, the whole texture set of Figure 4 has been classified to assess the capacity for each method to deal with a complex situation where both different and similar texture samples are mixed. In the second part, a reclassification has been performed to assess the *good association* capability by observing the nature of the errors of the first phase: i.e.,whether the wrongly classified pixels were at least within the good generic class or not.

The classification of the set of 36 textures yielded results that are much poorer than those which had been achieved in the other experiments (Table 5). This was predictable to a certain extent because, by increasing greatly the number of classes, the chance of misclassification was also increased. Both the VGM and GLCM methods produced Khat results of about 65 percent to 67 percent. The FFT came in last with a score approximately 8 to 10 percent lower. If these results are overall quite similar, the detailed analysis of their graphical counterpart shows that the behavior of each method can be different. Figure 6 shows the difference image between the classification results of the three methods and what would be the ideal classification. The most striking difference lies in the size and frequency of wrongly classified pixels and in the patches they form. In the GLCM results, these patches are relatively large, infrequent, and more concentrated around the edges of the texture patches; hence, the Khat of 93.8 percent for the *no edge* result. In the VGM classified image, these patches appear smaller on average but more frequent and sometimes give a speckled impression. However, the edges and borders account for a significant part of errors (a Khat difference of 26.2 percent). The classification result generated with the FFT feature set appears to suffer even more from a *salt and pepper* look: the patches are mostly small but very frequent. While about 23 percent of misclassified pixels are attributable to edges and borders, another 20 percent are found within the central parts of the texture patches. In all cases the differences between the three methods are quite significant, as can be seen in the pairwise test of significance of Table 6.

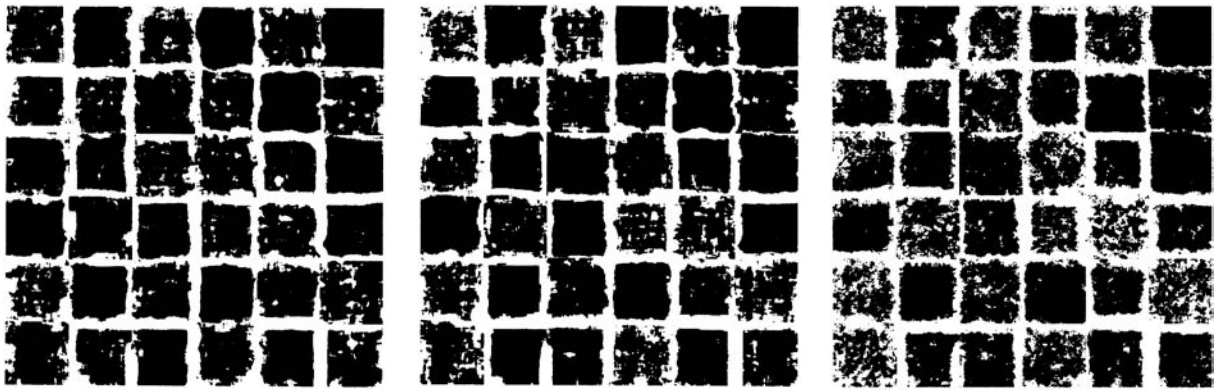
The observations above suggest the following facts:

- The GLCM method scores higher than the other two methods for complex classification situations,
- The VGM method yields comparable scores but the difference from the GLCM is significant,
- The fact that patches of misclassified pixels are generally larger but less frequent for the GLCM method suggests that the method is not easily affected by small differences and is spatially more consistent, and
- The FFT feature sets are more likely to be affected by small variations in textures than the GLCM approach.

Reclassification of the set of 36 textures. Table 7 presents the Khat results obtained after reclassification of the classified results of Table 5 (third experiment), and Figure 7 shows the difference image of the generic classes reclassification. Although, in the overall classification of the set of 36 textures

TABLE 5. KHAT SCORES FROM THE CLASSIFICATION OF THE SET OF 36 TEXTURES. GRAY COLUMNS SHOW *NO EDGE* RESULTS. BEST RESULTS ARE IN BOLD

GLCM Method		VGM Method		FFT Method	
$\hat{K} \times 100$ Overall	No edges	$\hat{K} \times 100$ Overall	No edges	$\hat{K} \times 100$ Overall	No edges
67.2%	**93.8%**	65.4%	91.6%	57.1%	80.3%



(a) (b) (c)

Figure 6. Difference image for classification results of the 36 textures set (misclassified pixels are shown in white, correctly classified one in gray). (a) The GLCM approach. (b) The VGM approach. (c) The FFT approach.

TABLE 6. RESULTS OF PAIRWISE COMPARISON OF KHAT VALUES FOR *OVERALL* RESULTS OF TABLE 5; S = SIGNIFICANT, NS = NOT SIGNIFICANT AT THE 99% LEVEL OF CONFIDENCE

GLCM *vs* VGM		VGM *vs* FFT		FFT *vs* GLCM	
Z Statistic	Result	Z Statistic	Result	Z Statistic	Result
19.6	S	87.5	S	−107.2	S

the GLCM method scored better, it was this approach that benefited the least from the reclassification into generic classes moving from an *all-classes* Khat statistic of 67.2 percent to 72.2 percent (a difference of 5 percent) compared with the FFT method that increased from 57.1 percent to 66.1 percent (a difference of 9 percent) or even the VGM whose TSC score increased from 65.1 percent to 71.4 percent (a difference of 6 percent). Still, the results tend to show that all three methods cannot reliably be expected to perform good association and that not having proper training areas for all classes can be costly in terms of classification errors. One could conclude that these methods are generally better at *separating* than *associating*.

It is interesting to look at which generic classes have gained the most out of reclassification because it gives an insight on the factors that might affect the texture classification accuracy. The *Residential* class gains from 31 percent to 50 percent of accuracy when accepting misclassified pixels that fall into another *Residential* texture class as correctly classified. This is understandable because this generic class stands out from the rest by its contrast and square-like objects. The *Shrub* class also gains from 18 percent to 28 percent accuracy when not considering exact class membership but, in these cases, it appears to be due to high *within-class* variability marked by variations of the *between-trees* distances. As for the *Desert* generic class, a combination of high *within-class* variability and low contrast with relation to the other texture patches might have combined to increase accuracy from 14 percent to 19 percent.

Conclusions

Three methods of texture classification have been tested in this paper, two of which have received a novel implementation: the semi-variogram and the Fourier spectra. Both have been implemented to be computationally efficient and to relate in some way to psychophysical evidence about human vision (Maillard, 2001). All three methods have proved to be powerful tools for texture classification, but the *gray-level co-occurrence matrix* has shown superior results for dealing with simple situations where the textures are visually easily separable. The *semi-variogram* was, however, slightly superior for distinguishing very similar texture patches, but more extensive testing is needed to confirm this. In complex situations (a large number of classes), the VGM and GLCM have performed almost equally but with generally poorer results (but better than the FFT). Much of this poorer performance can be attributed to borders and edges, which tends to show the importance of using a resolution finer than the "optimal" resolution as given, for instance, by a measurement like the local variance. The *good association* test proved the GLCM method slightly superior but showed that none of the methods tested can be expected to perform good association for classes not accounted for in the train-

TABLE 7. KAPPA STATISTICS OF THE *GOOD ASSOCIATION ANALYSIS* THROUGH THE RECLASSIFICATION OF THE 36 TEXTURES SET. BEST RESULTS ARE IN BOLD

	GLCM Method		VGM Method		FFT Method	
Generic Class	$\hat{K} \times 100$ correct class	correct generic class	$\hat{K} \times 100$ correct class	correct generic class	$\hat{K} \times 100$ correct class	correct generic class
Forest	**55.0%**	**66.0%**	48.9%	59.3%	43.2%	58.9%
Residential	**65.6**	**96.9**	61.3	93.1	37.4	87.5
Desert	51.3	**70.5**	**52.1**	66.5	40.8	54.4
Crops	**72.1**	**75.4**	71.7	75.0	63.3	72.8
Shrub	**63.9**	81.4	54.9	**82.5**	49.2	76.1
Waves	62.8	67.8	**68.2**	**77.0**	64.1	76.6
All classes	**67.2**	**72.2**	65.4	71.4	57.1	66.1

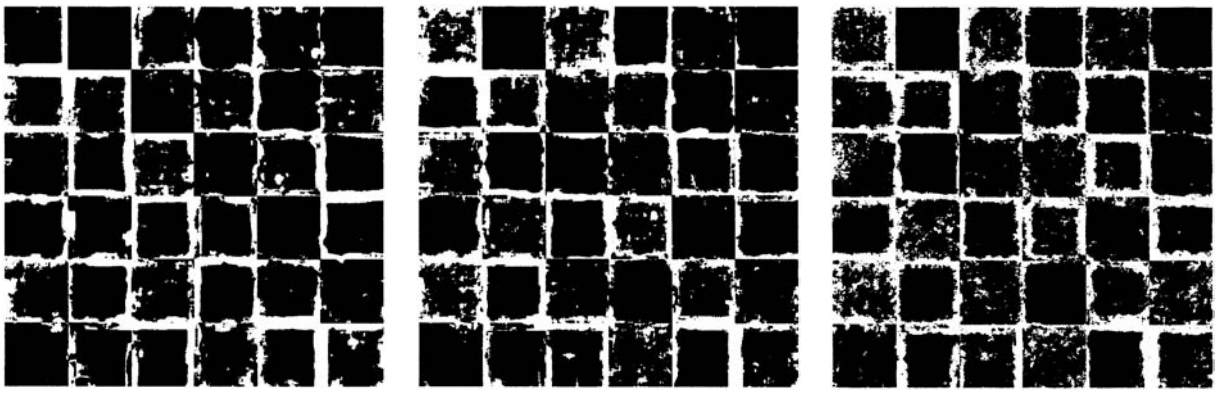


Figure 7. Difference image for reclassification of the 36 textures set according to generic classes (misclassified pixels are shown in white, correctly classified one in gray). (a) The GLCM approach. (b) The VGM approach. (c) The FFT approach.

ing phase: the methods are better at *separating* than *associating*.

The project can be said to have reached three goals: (1) new implementations of mathematical tools for texture analysis were realized with success, (2) insights on the behavior of these tools for texture analysis has provided some understanding of the characteristics of texture that are difficult to classify, and (3) a special use of classification using Bayes' theorem has proven an effective tool for testing and comparing the performance of texture analysis methods.

The proposed experimental design for testing and comparing the results proved most valuable because it clearly showed the importance of context in at least two aspects: (1) number of classes and generic groups and (2) spatial properties such as sample size and surroundings. Further studies will aim at testing these methods with other classifiers or segmentation schemes and in more realistic texture interpretation situations.

Acknowledgments

This work is part of a doctoral thesis concluded in 2001 at the University of Queensland and was made possible through the financial support of the Brazilian Federal Government (CAPES) and the Universidade Federal de Minas Gerais (UFMG).

References

Anys, H., and D.-C. He, 1995. Approche multipolarisation et texturale pour la reconnaissance des cultures à l'aide de données radar aéroporté, *Canadian Journal of Remote Sensing*, 21(2):138–157.

Atkinson, P.M., 1995. Regularizing variograms of airborne MSS imagery, *Canadian Journal of Remote Sensing*, 21(3):225–233.

Bonn, F., and G. Rochon, 1992. *Précis de Télédétection: Volume 1, Principes et Méthodes*, Presses de l'Université du Québec, Québec, Canada, 485 p.

Caelli, T., 1982. On discriminating visual textures and images, *Perception & Psychophysics*, 31(2):149–159.

Cao, C., and N. Lam, 1997. Understanding the scale and resolution effects in remote sensing and GIS, *Scale in Remote Sensing and GIS* (D.A. Quattrochi and M.F. Goodchild, editors), CRC Press, Lewis Publishers, Boca Raton, Florida, pp. 57–72.

Clark, W.A.V., and P.L. Hosking, 1986. *Statistical Methods for Geographers*, John Wiley and Sons, New York, N.Y., 518 p.

Cocquerez, J.P., and S. Philipp (editors), 1995. *Analyse d'Images: filtrage et Segmentation*, Masson, Paris, France, 457 p.

Cohen, J., 1960. A coefficient of agreement for nominal scales, *Educational and Psychological Measurements*, 20(1):37–40.

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 & Remote Sensing*, 54(5):593–600.

———, 1991. A review of assessing the accuracy of classification of remotely sensed data, *Remote Sensing of Environment*, 37:35–46.

Conners, R.W., and C.A. Harlow, 1980. A theoretical comparison of texture algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(3):204–222.

Cressie, N., and D.M. Hawkins, 1980. Robust estimation of the variogram, *Journal of the International Association for Mathematical Geology*, 12:115–125.

Davis, L.S., S.A. Johns, and J.K. Aggarwal, 1979. Texture analysis using generalized co-occurrence matrices, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(3):251–259.

Dikshit, O., 1996. Textural classification for ecological research using ATM images, *International Journal of Remote Sensing*, 17(5): 887–915.

Dunn, D., W.E. Higgins, and J. Wakely, 1994. Texture segmentation using 2-D Gabor elementary functions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):130–149.

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

Fogel, I., and D. Sagi, 1989. Gabor filters as texture discriminator, *Biological Cybernetics*, 61:103–113.

Foody, G.M., 1992. On the compensation for chance agreement in image classification accuracy assessment, *Photogrammetric Engineering & Remote Sensing*, 58(10):1459–1460.

Franklin, S.E., and D.R. Peddle, 1989. Spectral texture for improved class discrimination in complex terrain, *International Journal of Remote Sensing*, 10(8):1437–1443.

Franklin, S.E., A.J. Maudie, and M.B. Lavigne, 2001. Using co-occurrence texture to increase forest structure and species composition classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 67(7):849–855.

Gonzalez, R.C., and R.C. Woods, 1992. *Digital Image Processing*, Addison-Wesley Publishing Company, Reading, Massachusetts, 716 p.

Goresnic, C., and R.S. Rotman, 1992. Texture classification using the cortex transform, *CVGIP: Graphical Models and Image Processing*, 54(4):329–339.

Haralick, R.M., 1979, Statistical and structural approaches to texture, *Proceeding of the IEEE Transactions Systems, Man and Cybernetics*, 67:786–804.

Haralick, R.M., K. Shanmugan, and I. Dinstein, 1973. Texture feature for image classification, *IEEE Transactions Systems, Man and Cybernetics*, SMC-3:610–621.

Harvey, L.O., Jr., and M.J. Gervais, 1978. Visual texture perception and fourier analysis, *Perception & Psychophysics*, 24(6):534–542.

Hay, G.J., K.O. Niemann, and G.F. McLean, 1996. An object-specific Image-texture analysis of h-resolution forest imagery, *Remote Sensing of Environment*, 55:108–122.

Jain, A.K., and F. Farrokhnia, 1991. Unsupervised texture segmentation using Gabor filters, *Pattern Recognition*, 24(12):1167–1186.

James, M., 1985. *Classification Algorithms*, John Wiley and Sons, London, United Kingdom, 211 p.

Jones, G., 1994. Image segmentation using texture boundary detection, *Pattern Recognition Letters*, 15:533–541.

Julesz, B., 1965. Texture and visual perception, *Scientific American*, 212:38–48.

———, 1981. Textons, the elements of texture perception, and their interaction, *Nature*, 290:91–97.

Jupp, D.L.B., A.H. Strahler, and C.E. Woodcock, 1989. Autocorrelation and regularization in digital images, II: Simple image models, *IEEE Transactions on Geoscience and Remote Sensing*, 27:247–258.

Kittler, J., 1983. Image processing for remote sensing, *Philanthropical Transactions of the Royal Society of London*, A309:323–335.

Lark, R.M., 1996. Geostatistical description of texture on an aerial photograph for discriminating classes of land cover, *International Journal of Remote Sensing*, 17(11):2115–2133.

Maillard, P., 2001. *Texture in High Resolution Digital Images of the Earth*, Ph.D. Thesis, The University of Queensland, Brisbane, Queensland, Australia, 348 p.

Manjunath, B.S., and W.Y. Ma, 1996. Texture features for browsing and retrieval of image data, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842.

MeasTex, 2001. *Image Texture Database and Test Suite*, URL: http://www.cssip.uq.edu.au/staff/meastex/meastex.html (last accessed 22 April 2002).

Miranda, F.P., J.A. MacDonald, and J.R. Carr, 1992. Application of the semi-variogram textural classifier (STC) for vegetation discrimination using SIR-B data of Borneo, *International Journal of Remote Sensing*, 13(12):2349–2354.

Nyerges, T.L., 1991. Representing geographical meaning, *Map Generalization* (B. Buttenfield and R. McMaster, editors) Longman Scientific and Technical, Essex, England, pp. 59–85.

Pavlidis, T., and Y.-T. Liow, 1990. Integrating region growing and edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):225–233.

Pratt, W.K., 1991. *Digital Image Processing*, John Wiley & Sons, New York, N.Y., 698 p.

Pratt, W.K., O.D. Faugeras, and A. Gagalowicz, 1978. Visual discrimination of stochastic texture fields, *IEEE Transactions on Systems, Man and Cybernetics*, SMC-8(11):796–804.

Ramstein, G., and M. Raffy, 1989. Analysis of the structure of radiometric remotly-sensed images, *International Journal of Remote Sensing*, 10(6):1049–1073.

Reed, T.R., and du H. Buf, 1993. A review of recent texture segmentation and feature extraction techniques, *CVGIP: Image Understanding*, 57(3):359–372.

Reed, T.R., and H. Wechsler, 1990. Segmentation of textured images and Gestalt organization using spatial/spatial-frequency representations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):1–12.

Richard, W., and A. Polit, 1974. Texture matching, *Kybernetik*, 16: 155–162.

Rosenfeld, A., 1962. Automatic recognition of basic terrain types from aerial photographs, *Photogrammetric Engineering*, 28(1): 115–132.

Rosenfeld, A., and A.C. Kak, 1982. *Digital Picture Processing, Volume 2*, Academic Press, New York, N.Y., 351 p.

Sali, E., and H. Wolfson, 1992. Texture classification in aerial photographs and satellite data, *International Journal of Remote Sensing*, 13(18):3395–3408.

Scheffé, S., 1959. *The Analysis of Variance*, Wiley, New York, N.Y., 477 p.

Serra, J.P., 1982. *Image Analysis and Mathematical Morphology*, Academic Press, London, New York, N.Y., 610 p.

Smith, G., 2001. *MeasTex: Overview of Framework. Image Texture Database and Test Suite*, URL: http://www.cssip.uq.edu.au/staff/meastex/meastex.html, July 2001 (last accessed 22 April 2002).

St-Onge, B.A., and F. Cavayas, 1995. Estimating forest stand structure from high resolution imagery using the directional variogram, *International Journal of Remote Sensing*, 16(11):1999–2021.

Strahler, A.H., C.E. Woodcock, and J.A. Smith, 1986. On the nature of models in remote sensing, *Remote Sensing of Environment*, 20:121–139.

Stromberg, W.D., and T.G. Farr, 1986. A Fourier-based feature extraction procedure, *IEEE Transactions on Geoscience and Remote Sensing*, GE-24(5):722–731.

Wilson, R., and M. Spann, 1988. *Image Segmentation and Uncertainty*, Research Studies Press, John Wiley & Sons, Letchworth, Hertfordshire, England, 180 p.

Woodcock, C.E., A.H. Strahler, and D.L.B. Jupp, 1988. The use of variogram in remote sensing and simulated image, II: Real digital images. *Remote Sensing of Environment*, 25:349–379.

Wu, C.-M., and Y.-C. Chen, 1992. Statistical feature matrix for texture analysis, *CVGIP: Graphical Models and Image Processing*, 54(5): 407–419.

Xia, Z.G., and K.C. Clarke, 1997. Approaches to scaling of geo-spatial data, *Scale in Remote Sensing and GIS* (D.A. Quattrochi and M.F. Goodchild, editors), CRC Press, Lewis Publischers, Boca Raton, Florida, pp. 309–360.

Zhang, Y., 2001. Texture-integrated classification of urban treed areas in high-resolution color-infrared imagery, *Photogrammetric Engineering & Remote Sensing*, 67(12):1359–1365.

(Received 31 December 2001; revised and accepted 09 July 2002)