Computer Vision and Image Understanding **84,** 62–76 (2001)
doi:10.1006/cviu.2001.0941, available online at http://www.idealibrary.com on IDEAL®

Analysis of Irregularly Shaped Texture Regions[1]

Pedro García-Sevilla

Departamento de Lenguajes y Sistemas Informáticos, Universitat Jaume I, 12071 Castellón, Spain

and

Maria Petrou

School of Electronic Engineering, Information Technology and Mathematics, University of Surrey Guildford, GU2 7XH, United Kingdom

Received September 15, 1999; accepted August 27, 2001

Four different texture classification methods (wavelet-based, co-occurrence matrices-based, 1D-histograms-based, and 1D Boolean model-based) are systematically compared and evaluated with respect to their performance in identifying textures from small and irregular samples. Two sets of 135 complex shape masks (symmetric and nonsymmetric) are created using Fourier shape descriptors, and used to clip out regions of images that must be classified. Two main series of experiments are carried out: one in which the clipped test image is from the same full image included in the database, and one where the clipped test image is from a different realization of the full image included in the database. In both occasions the best performing method was the 1D sum and difference histogram-based method.

Key Words: image retrieval by texture; small sample analysis; texture analysis; texture classification; wavelet transform; co-occurrence matrices; sum and difference histograms; Boolean model.

1. INTRODUCTION

There have been many comparative studies of texture classification methods done in the past [2, 3, 7, 9, 11]. In all these studies, the experiments carried out concerned the characterization of rectangular texture samples. Even when they segmented images composed of several irregular regions, the texture for each pixel was characterized by features derived from a square neighborhood. In other words, it was assumed that a rectangular sample of significant size of the particular texture was available. In practical applications, however,

[1] This work was partly supported by grant P1-1B2000-01 Fundació Caixa Castelló-Bancaixa.

1077-3142/01 $35.00

this is seldomly the case. For example, in a case of database search by content, one may wish to use texture as a cue for the presence of a particular object in an image. It is unlikely that the particular texture sought will be present in the form of a rectangular shape of significant size.

In this paper we assume that the textured regions of an image have been identified manually or by some generic texture boundary detection method, e.g., [8]. The identified texture regions may subsequently have to be classified by comparison with entries from a texture database.

We chose to compare four different texture classification methods in this context. The first method is that proposed by Chang and Kuo [1], and it is based on the tree-structured wavelet transform. The second method is based on the well-known co-occurrence matrices defined by Haralick *et al.* [6]. The third one was proposed by Unser in [10] and it is based on the sum and difference histograms as an alternative to the usual co-occurrence matrix. Finally, the fourth method has been proposed in [4, 5] and relies on the calculation of features from various 1D strings that can be formed from a texture sample. To investigate the ability of each method to recognize a texture from samples of arbitrarily irregular shapes and sizes, we create two sets of masks that we use to isolate samples of textures from our texture database. All methods are trained on rectangularly shaped large samples, and then are tested on their ability to recognize textures from samples created by clipping test images using the sets of previously defined masks. The experiments carried out provided a direct comparison of the different methods for texture classification as all of them were run using the same image database and under identical conditions.

2. CREATING IRREGULAR MASKS

To create systematic sets of masks of various sizes and irregularities, we used the Fourier descriptors of a shape. If we measure the radius of a shape along rays emanating from the shape's center of gravity, we have a periodic function $r(\theta)$ where θ is the orientation of each ray with respect to some reference direction. This periodic function can be expanded into a Fourier series and the coefficients of the expansion can be used as shape descriptors.

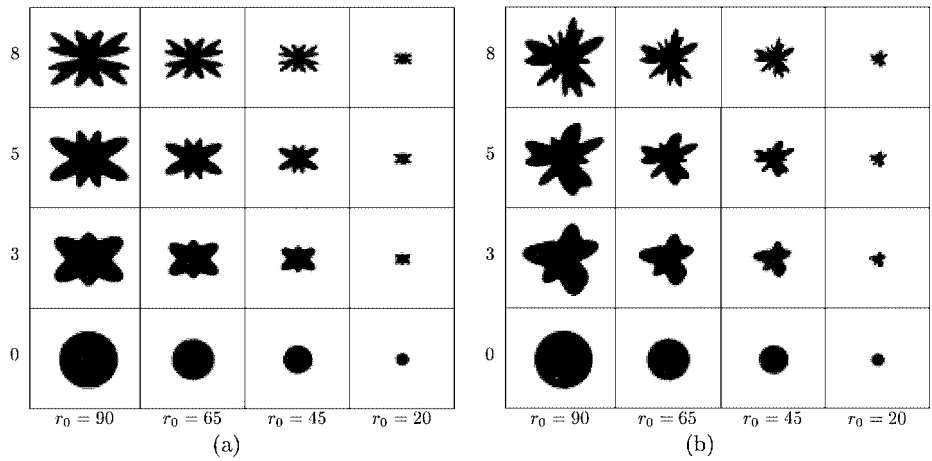

FIG. 1. Some of the shapes in the databases generated as described in Section 2. Each column corresponds to a different radius (from 90 down to 20 pixels). Each row corresponds to a number of nonnull Fourier coefficients (from 0 to 8). (a) First set: symmetric masks. (b) Second set: nonsymmetric masks.

For a first set, we decided to use masks with quadruple symmetry so that only the even coefficients of the cosine terms in the Fourier series were assumed to be nonzero. Each shape in this set, therefore, is given by

$$S(r_0, N, \theta) \equiv \sum_{n=0}^{N} r_0 f(n) \cos(2n\theta), \tag{1}$$

where r_0 is the mean radius of the shape, N is the number of nonzero terms, and $f(n)$ is a pseudo-random factor with values between 0 and 1, with $f(0) = 1$. The role of $f(n)$ is to determine the coefficients of the shape harmonics in a random way. Parameter N can be

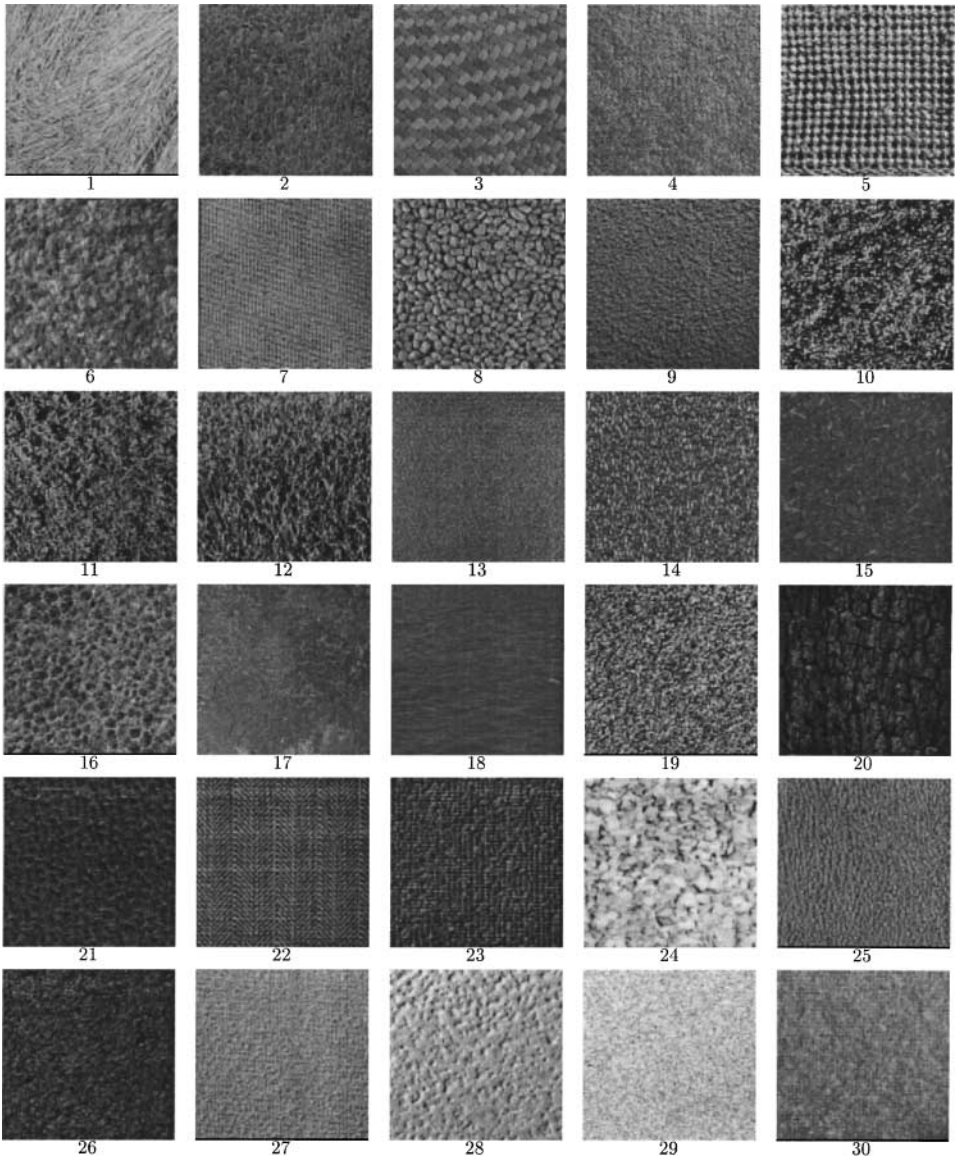

FIG. 2. Texture images used in the experiments.

seen as a measure of how irregular the shape is: the higher the number of nonzero terms (value of N), the more irregular the shape becomes.

A second set of masks was created in a similar way, but avoiding symmetry, so only the odd coefficients of the cosine terms in the Fourier series were assumed to be nonzero. Now each shape is given by

$$S(r_0, N, \theta) \equiv r_0 + \sum_{n=0}^{N} r_0 f(n) \cos((2n+1)\theta + \phi(n)), \tag{2}$$

where r_0 is the mean radius of the shape, N is the number of nonzero terms, $\phi(n)$ is a pseudo-random phase, and $f(n)$ is a pseudo-random factor with $f(0) = 0$.

Parameter N varied from 0 (perfect circle) to 8, while the radius r_0 varied from 90 down to 20 pixels, with decrements of 5 pixels. That is, a total of 135 shapes were included in each set. Figure 1 shows some of the shapes included in each set. Each cell in the tables represents an area of 256×256 pixels in size.

Once the sets of irregular shapes were created, we were able to create irregular texture areas by simply using the binary shapes as clipping masks of the texture images. The image database of gray-level textures used in [5] was considered for our experiments.[2] In total we had 30 texture classes with four nonoverlapping images per class, where each of them was 256×256 pixels in size. We will refer to each image in a class as a *quadrant*, as they were obtained by splitting a larger image into 4. Figure 2 shows some texture classes in the database.

3. METHODOLOGY

As we are interested in studying the influence of the size and irregularity of the area to be analyzed to the calculation of the texture parameters, we run a first series of experiments in which the whole square images were used as models and a series of irregular areas obtained directly from these models were used for testing. In this case, the irregularly shaped patch to be classified contains exactly the same texture as the models. Therefore, the differences in the parameters between the test shapes and the corresponding model images must be only due to the irregularity and different size of the texture area, and to nothing else.

One image from each texture class was selected, and its corresponding parameters were computed. From each image, irregularly shaped regions were clipped using the sets of masks defined. Each region was classified according to the features computed from it using the nearest-neighbor rule with the proper distance definition for the texture classification method considered.

Figure 3a shows an example of the results obtained for the texture areas derived from image number 24 in the database using the 1D Boolean model with bit slicing. The vertical axis corresponds to the number of nonnull Fourier coefficients used to create the shapes while the horizontal axis refers to their average radius. Each cell is color-coded to represent whether the shape was correctly classified (black or gray) or not (white). In order to simplify the information contained in this sort of graph, we search for a compact block of correctly classified shapes in such a way that we can draw a line that separates recognizable from unrecognizable shapes. To find this limit we move through each row of the graph from left to right (decreasing size) until we find the first error. Then the rest of the shapes on this row

[2] All texture images and masks used in this paper can be obtained from `ftp://marmota.act.uji.es/pub/research/texture/`.

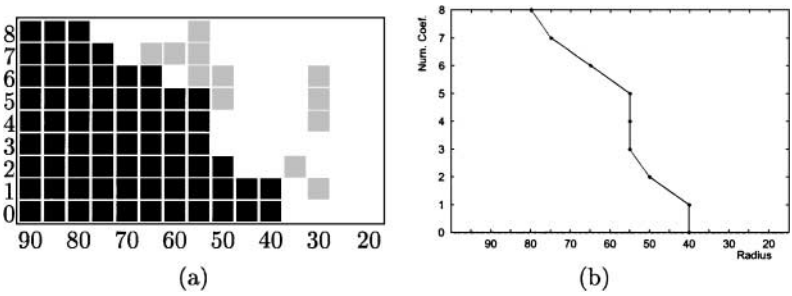

FIG. 3. Classification results for shapes derived from class number 24 (quadrant 1), using the 1D Boolean model with bit slicing. (a) Black, correctly classified shapes; gray, shapes that were correctly classified, but ignored; white, errors. (b) Chosen border between recognizable and unrecognizable shapes.

are ignored. Similarly, we move through each column of the graph (increasing number of nonnull coefficients) and once we find the first error, the rest of the shapes are ignored. In Fig. 3a the gray cells refer to shapes that, although were correctly classified, were finally ignored. Figure 3b shows the final line considered as a border between recognizable and unrecognizable shapes of the particular texture. We consider the successful recognition of shapes marked with gray in Fig. 3a as coincidental. For example, if shape with $N = 2$ and $r_0 = 55$ could not be recognized, we do not expect that an even smaller sample of the texture, indicated by $N = 2$ and $r_0 = 45$, could be reliably recognized.

As we have four different images per texture class, the previous steps can be repeated a total of four times, once per each quadrant. Thus, we will find four different curves that represent the border between recognizable and unrecognizable shapes. Figure 4a shows the four lines obtained for image number 24 using the 1D Boolean model. In order to obtain only one boundary line per image, we combined the results obtained for each quadrant into one global line that summarizes all the information for that texture class. This global border is chosen by taking the most restrictive result of the four experiments (that is, the worst one). Figure 4b shows the global border obtained for image number 24 using the 1D Boolean model.

In order to obtain a quantitative comparison between the methods tested, we may compute for each method the percentage of irregular shapes that fall in the recognizable side of the previous graph. As we used 135 irregular shapes per image and there were 30 textures in

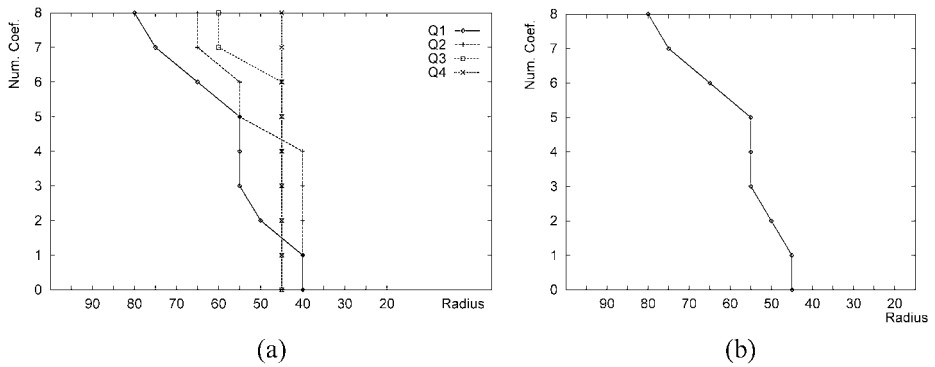

FIG. 4. Boundaries between recognizable and unrecognizable shapes derived from class number 24 using the 1D Boolean model with bit slicing. (a) Boundaries for the four quadrants in the class. (b) Global boundary defined for the class.

the database, a total of 4050 shapes applied to four different quadrants were considered in each experiment.

All texture characterization methods were run using several different parameter settings.

4. EXPERIMENTAL RESULTS

First we present the results obtained by using the masks of Fig. 1a.

When we use the tree-structured wavelet transform (TSWT), a maximum of four decompositions were used. As the TSWT must deal always with square images whose size is a power of 2, each irregularly shaped texture patch was enclosed into a minimum square of the appropriate size. The pixels that did not belong to the texture were filled with the mean gray value of the texture area. This was so that the sharp edges created at the border of the shape were as insignificant as possible. When the energy of a channel was computed, it was normalized, depending not only on the size of the channel but also on the ratio of pixels used in the irregular shape (pixels that belong to the irregular patch) with respect to the total number of pixels in the square image.

Four different filter banks were considered for the experiments, namely the Daubechies filters of size 4, 12, and 20, as well as the Haar filter bank. The number of channels considered varied from 10 to 255 (the maximum number when four decompositions were used). Figure 5a shows the percentage of correctly classified samples for each filter bank as a function of the number of channels used in the classification.

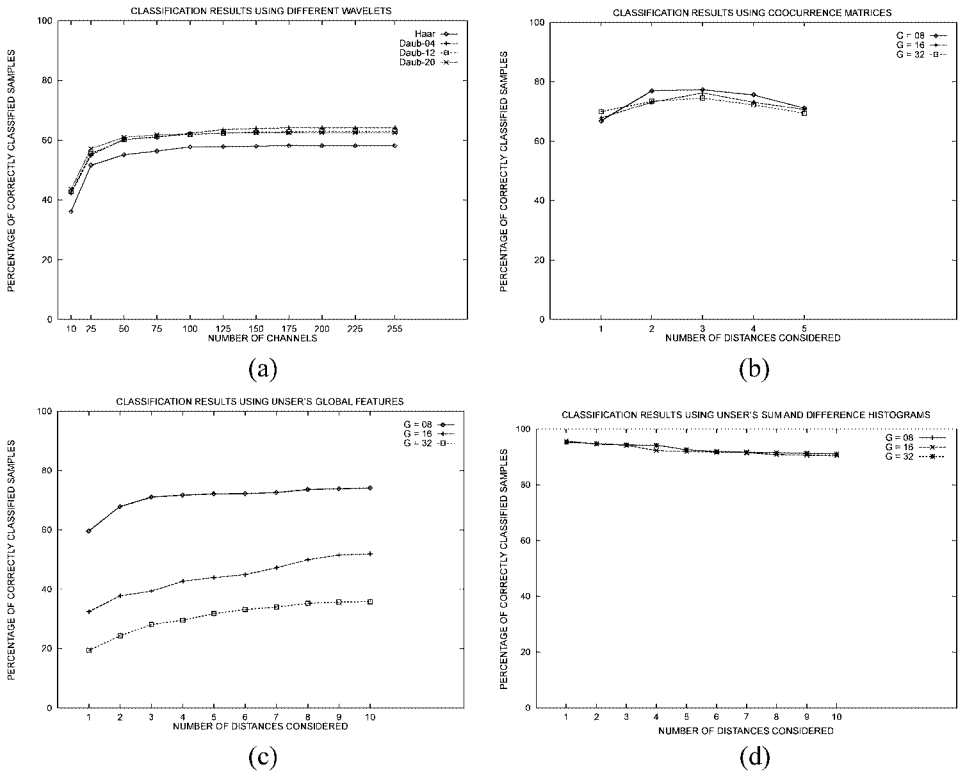

FIG. 5. Percentages of correctly classified samples when test samples were derived from model samples. (a) TSWT using four different filter banks. (b) Co-occurrence matrices. (c) Global features derived from the sum and difference histograms. (d) Sum and difference histograms.

The best results of all experiments using the TSWT were obtained when the Daub-04 filter bank was considered, and 175 channels were used according to their energy. A 64.12% of correctly classified samples was achieved in this case. Chang and Kuo [1] reported that only 10 channels are enough for a good characterization. They used 100 overlapping images that were 256×256 pixels in size, obtained from a single 512×512 image. Then they used the leave-one-out method to compute the classification accuracy. We carried out the same experiment and obtained the same result (almost 100% accuracy in the classification). However, in our problem there is no overlapping between test and model samples and there are also differences in size and shape, so more channels are needed in order to obtain a good characterization. The edge effects must also be taken into account, although we tried to reduce them as much as possible. Figure 6 shows the boundaries found for the 30 texture classes in this case.

Experiments using features obtained from co-occurrence matrices were run using the same parameters described in [7]; that is, four directions were considered and four textural features were computed for each matrix. These features were energy, contrast, correlation, and entropy. Each feature was normalized to a distribution with zero mean and unit standard deviation. To deal with irregular areas, only those pairs of pixels that fell inside the irregular

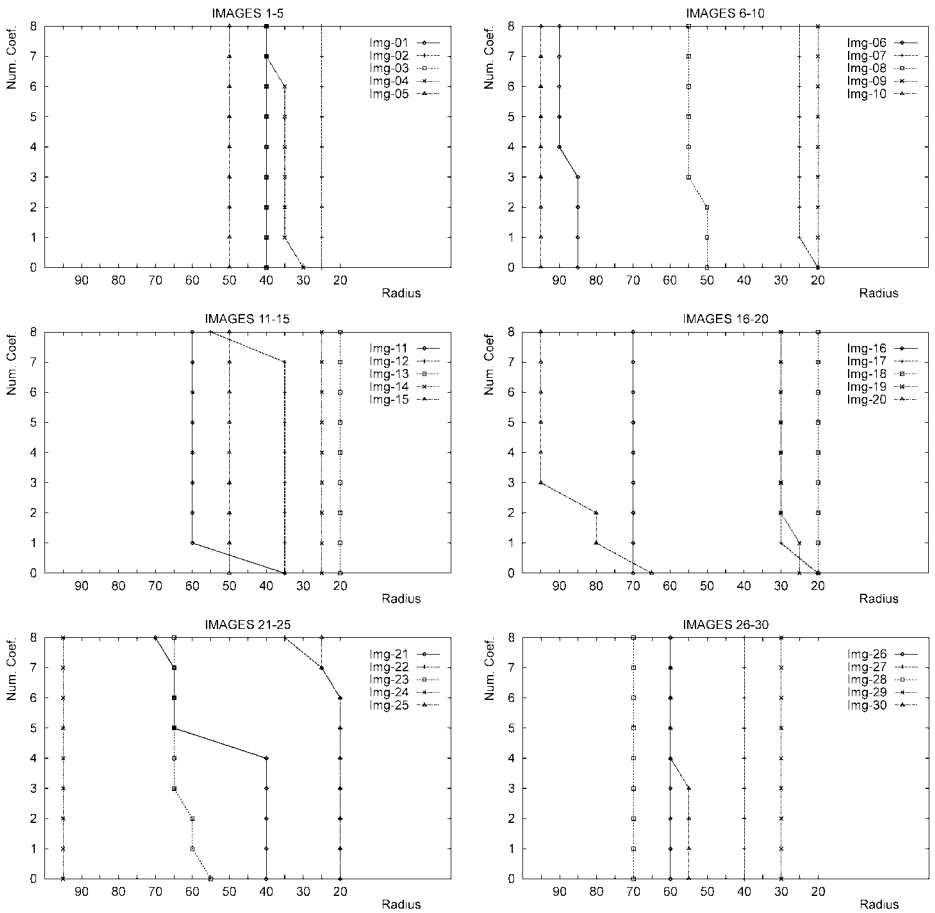

FIG. 6. Boundaries between recognized and unrecognized shapes for all texture classes using the Tree-Structured Wavelet Transform.

shape were taken into account, and this number of pairs was also used to normalize the textural features.

The distances d considered were 1, 2, 3, 4, and 5, while the number of gray levels was quantized to $G = 8$, 16, and 32. Figure 5b shows the percentage of correctly classified samples for each value of G as a function of the number of distances d considered. Note that the performance deteriorates as the number of distances considered increases. This is because for large values of d there are not enough pairs of pixels falling inside the shape to produce reliable statistics.

It can be noticed that all the parameters performed similarly. The best results of all experiments using the co-occurrence matrices were obtained when three distances were considered and the gray images were quantized to 8 ranges. An accuracy of 77.33% of correctly classified samples was achieved in this case.

Figure 7 shows the boundaries found between recognized and unrecognized shapes for the 30 texture classes when the textural features obtained from the co-occurrence matrices with $d = 3$ and $G = 8$ were used to characterize the textures. In this case, the results are better in general than those obtained when the TSWT was used.

Experiments using features obtained from the sum and difference histograms were run using the four directions considered for the co-occurrence matrices, that is, 0° (horizontal),

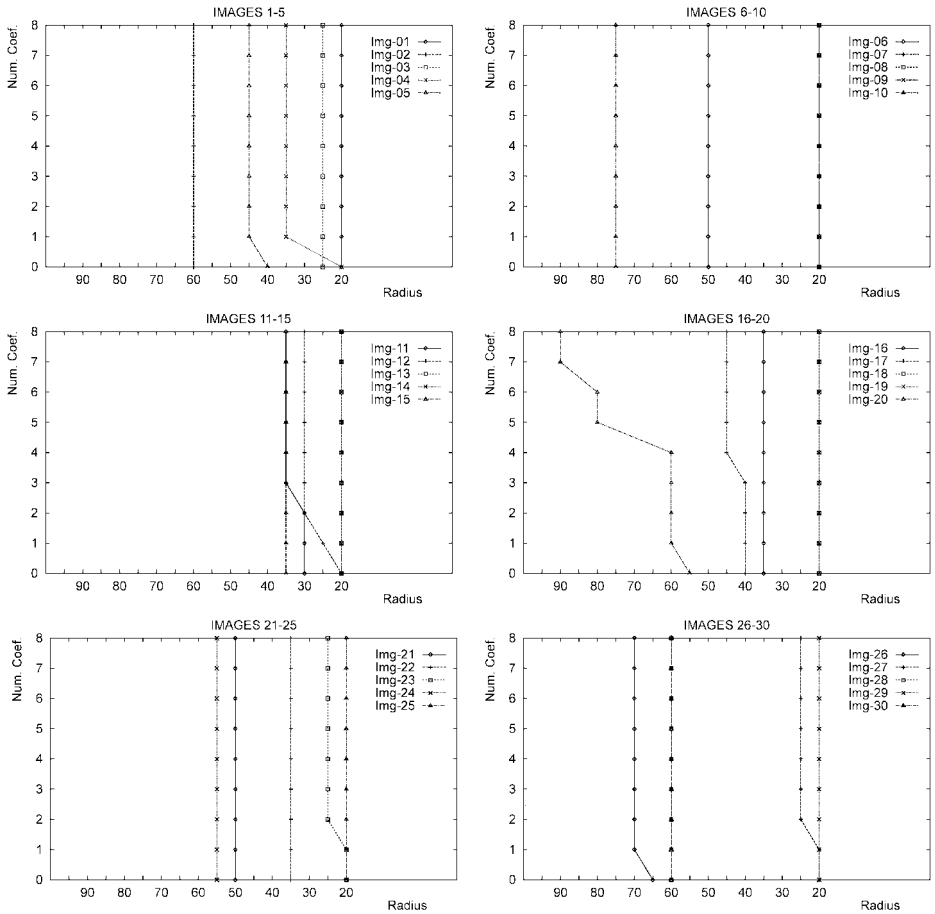

FIG. 7. Boundaries between recognized and unrecognized shapes for all texture classes using co-occurrence matrices.

45° (first diagonal), 90° (vertical), and 135° (second diagonal). To deal with irregular areas, only those pairs of pixels that fell inside the irregular shape were taken into account. The distances d considered varied from 1 to 10, while the number of gray levels in the images was quantized to $G = 8$, 16, and 32. Both classification methods described in [10] were used.

Figure 5c shows the percentage of correctly classified samples for each value of G as a function of the number of distances d considered when the global features described in [10] were used to characterize each histogram. It can be noticed that the results are highly influenced by the number of gray levels used in the quantization process. When the number of gray levels used in the quantization process was increased, the classification rate decreased significantly. However, the classification accuracy did not decrease as the number of distances considered was increased. The best results of all experiments run using global features computed from the sum and difference histograms were obtained when 10 distances were considered, using 8 ranges for the quantization process. A 74.15% of correctly classified samples was achieved in this case, although very similar results were obtained for $d = 3$ to $d = 9$.

Figure 8 shows the boundaries found between recognized and unrecognized shapes for the 30 texture classes in the best case. The results are better in general than those obtained

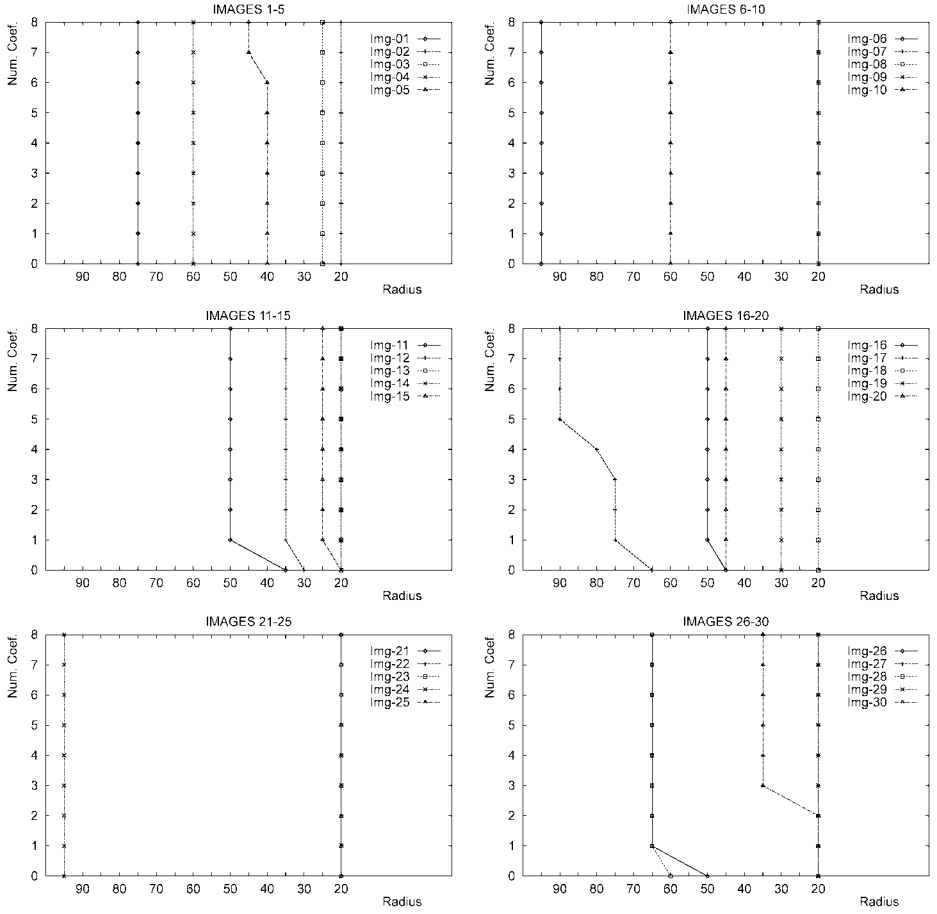

FIG. 8. Boundaries between recognized and unrecognized shapes for all texture classes using global features computed from the sum and difference histograms.

when the TSWT was used and slightly worse than those obtained using features computed from the co-occurrence matrices.

The same experiments were run using now the second classification method described in [10]; that is, the whole histograms were used to describe a texture class and a maximum likelihood decision rule was used to classify samples. Figure 5d shows the percentage of correctly classified samples obtained for each value of G as a function of the number of distances d considered.

In this case, we note that the results are much better than those obtained using all previous methods. Also, it is clear that there is almost no difference in the classification rate for different values of G (the number of gray levels used in the quantization process). However, the larger the number of gray levels considered, the larger the histograms computed, and therefore, a larger number of features are considered to characterize a texture. Nevertheless, the percentage of correctly classified samples when only 8 gray levels were used in the quantization process is more then 90%, while the previous methods never provided more than 77% of correctly classified samples.

In this case, the best results were obtained when only one distance was considered, using 16 ranges for the quantization process, which provided 95.68% of correctly classified

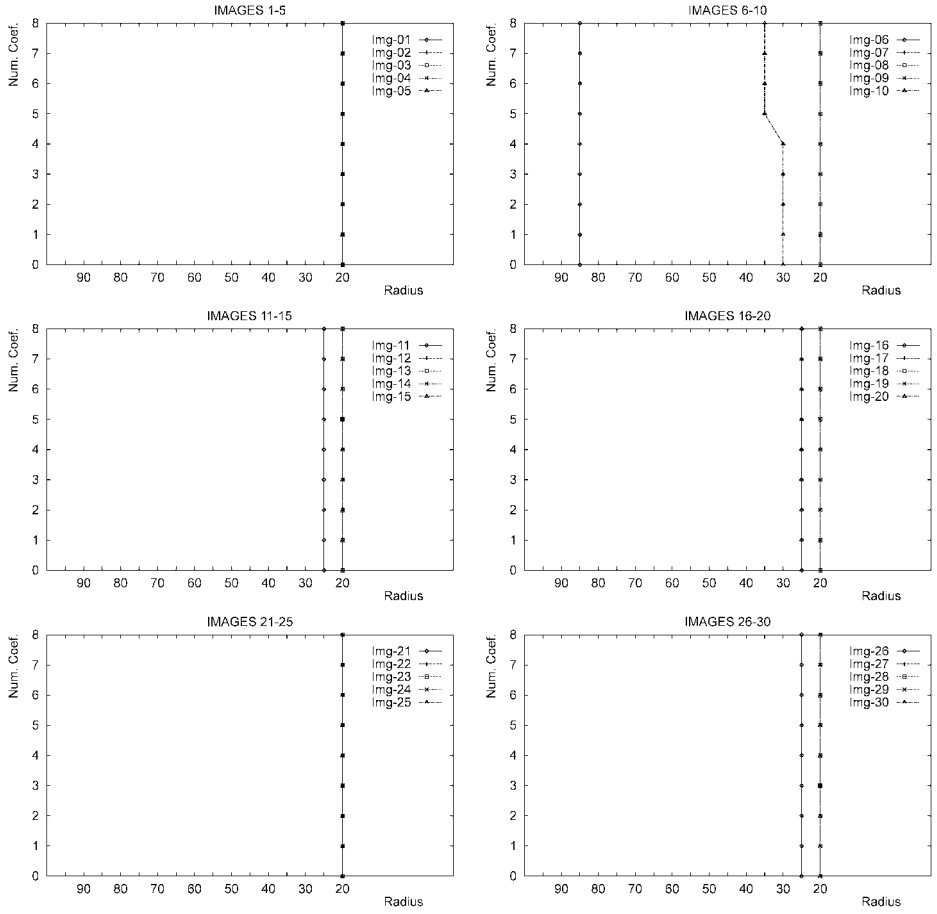

FIG. 9. Boundaries between recognized and unrecognized shapes for all texture classes using all values from the sum and difference histograms.

TABLE 1
Percentage of Correctly Classified Shapes for Each Experiment Run Using the One-Dimensional Boolean Model

Method	Percentage (%)
Bit slicing (8 planes)	77.23
Bit slicing (3 planes)	72.81
Quantization (8 planes)	72.40
Quantization (3 planes)	62.18

Note. Test samples were derived from model samples.

samples. Figure 9 shows the boundaries found between recognized and unrecognized shapes for the 30 texture classes.

Experiments were run using the Boolean model and different methods for splitting a gray image into a set of binary planes, namely bit slicing and gray-level quantization [5]. Two

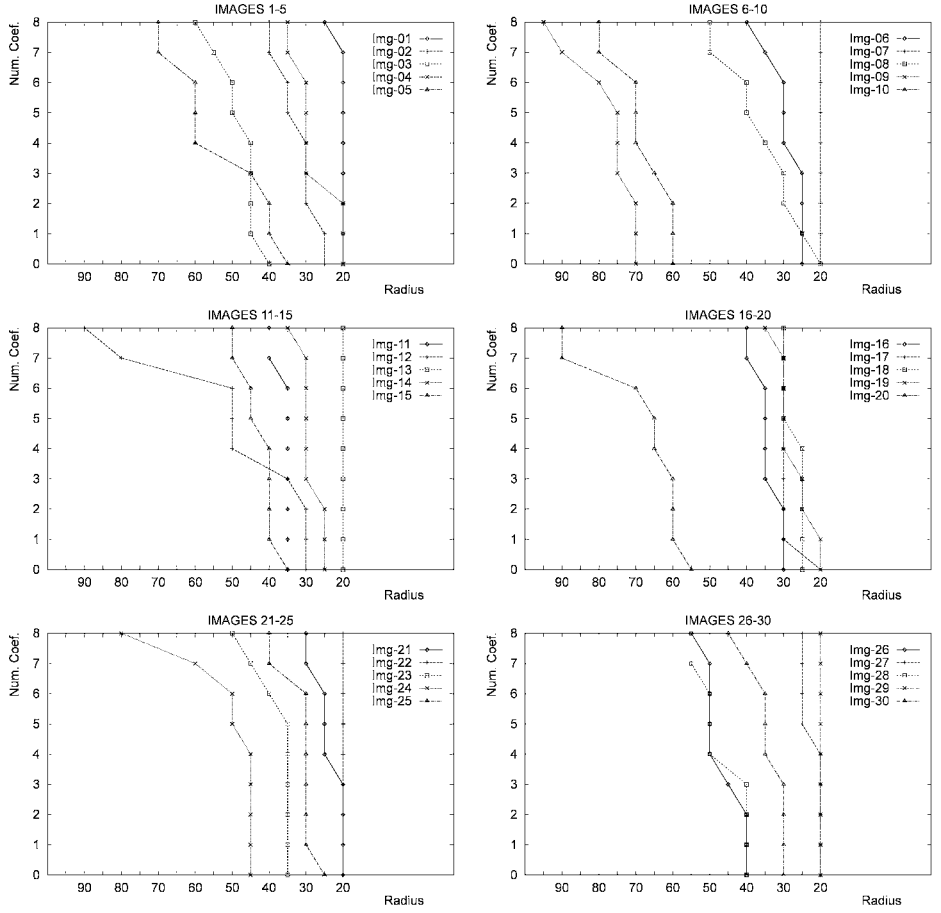

FIG. 10. Boundaries between recognized and unrecognized shapes for all texture classes. Splitting method: Bit slicing. Number of binary planes: 8.

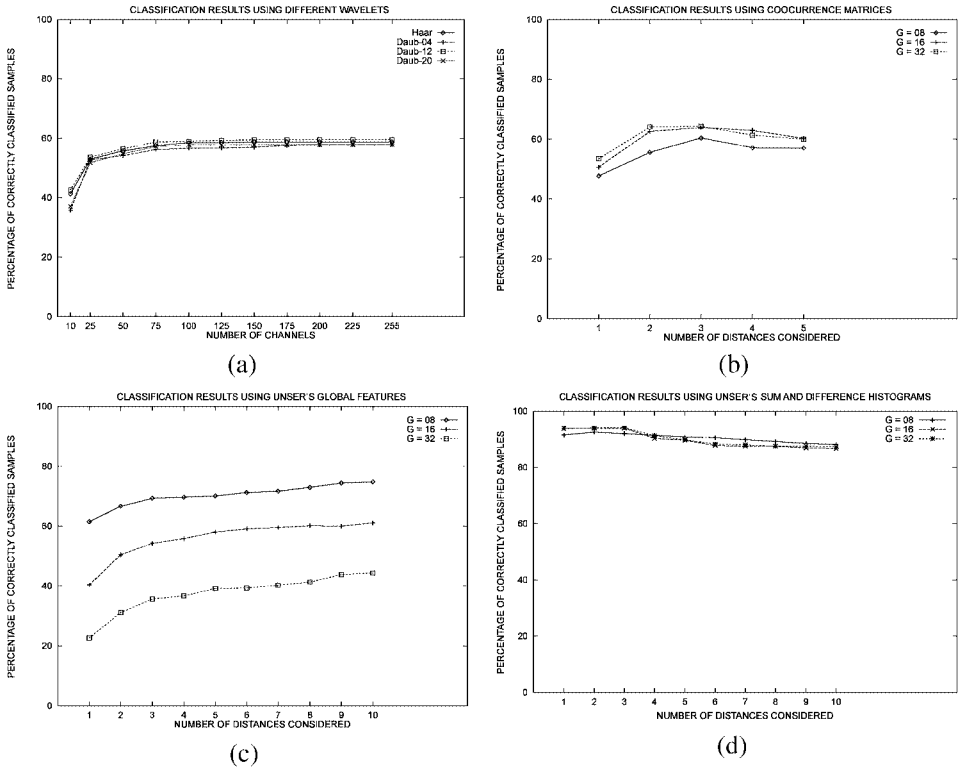

FIG. 11. Percentages of correctly classified samples when the test samples are different from the model samples. (a) TSWT using four different filter banks. (b) Co-occurrence matrices. (c) Global features derived from the sum and difference histograms. (d) Sum and difference histograms.

different numbers of binary planes were considered for each method: 3 and 8. When bit slicing was used with only 3 binary planes, the three most significant bits were used to create the binary planes. Each irregular shape was raster scanned and only complete runlengths were used to calculate the Boolean model statistics (i.e., runlengths that were interrupted by the region boundary were ignored).

Table 1 shows the percentage of correctly classified samples for the experiments run using the one-dimensional Boolean model. Figure 10 shows the recognition boundaries found for the 30 texture classes when the bit slicing method was used for the binarization of the gray images, and all binary planes were used in the classification process.

It can be noted that in all previous methods the boundaries between recognized and unrecognized samples were mostly vertical lines. This means that the texture characterization is much more influenced by the sample size than by the sample shape. However, the texture boundaries here are not vertical lines. This is because in the Boolean model only the whole runlengths are considered in the statistics. Therefore, the more irregular the sample is, the fewer complete runlengths are available and the less reliable the calculation of the statistics is. Thus, we conclude that the texture characterization using the Boolean model is not only influenced by the sample size but also by its shape.

In addition, we run all experiments again for the more realistic case when the test sample is obtained from an instantiation of the texture different from that used for training. Now the errors are expected to be larger as they are not only due to the shape of the tested region,

TABLE 2

Percentage of Correctly Classified Shapes for Each Experiment Run Using the One-Dimensional Boolean Model

Method	Percentage (%)
Bit slicing (8 planes)	76.47
Bit slicing (3 planes)	69.33
Quantization (8 planes)	72.96
Quantization (3 planes)	66.96

Note. Test samples are different from model samples.

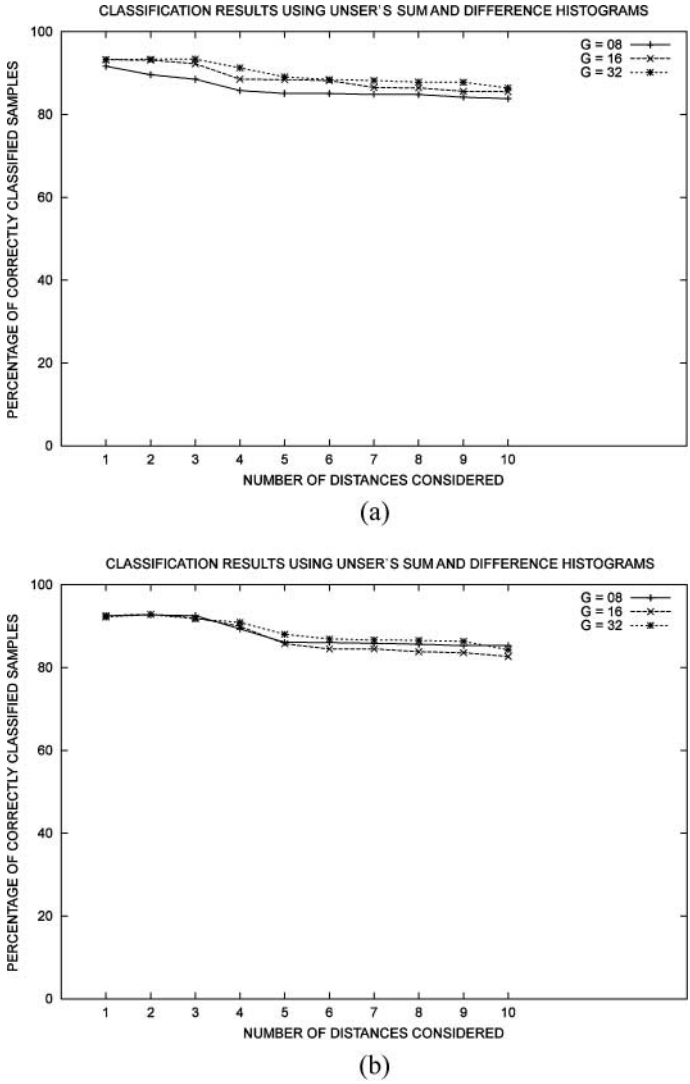

FIG. 12. Percentages of correctly classified samples using the whole sum and difference histograms when the test samples are different from model samples and obtained using the masks shown in Fig. 1b. (a) Masks are centered on the images. (b) Masks are randomly displaced.

TABLE 3
Best Results Obtained for the Different Texture Characterization Methods Compared

	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Average
TSWT	64.12	59.58	61.04	59.95	91.67	67.27
Co-occurrence matr.	77.33	64.30	62.52	63.16	98.33	73.13
Sum and dif. hist. (Global features)	74.15	74.77	75.38	70.25	100.00	78.91
Boolean model	77.23	76.47	78.47	79.14	100.00	82.26
Sum and dif. hist. (Whole histograms)	95.68	94.02	93.36	92.86	100.00	95.18

Note. Each column shows the results of a different set of experiments. The models were always square images. The test samples were different for each experiment. Exp. 1: Symmetric masks were applied over the models. Exp. 2: Symmetric masks were applied over images different from the models. Exp. 3: Same as Exp. 2 but using nonsymmetric masks. Exp. 4: Same as Exp. 3 but masks were not centered on the square image. Exp. 5: Whole square images were used as test samples.

but also due to the different version of the same texture. Indeed, in almost all cases the results were worse than before but not significantly. Figure 11 and Table 2 show the results of these experiments for all four methods. This shows that by far the most important factor in determining the performance of a texture classifier is the shape and size of the available sample rather than the different texture image, as long as, of course, that image has been obtained under the same imaging condition as that used for training. By far the most robust method was proven to be the sum and difference histograms when the whole histograms were used to characterize a texture. For this method no deterioration in performance was observed.

All the above experiments were repeated with the set of clip masks shown in Fig. 1b. The masks were applied both centered on the images and displaced at random positions (using the same displacement for all masks). The results did not show any significant variation and so we do not present them in detail. We only present in Fig. 12 the results for the sum and difference histograms method, which appears to be the best according to this study.

Finally, Table 3 summarises the best results obtained for each method in all experiments. For reference, we also included the percentage of correct classification obtained in experiments where the whole rectangular test samples were classified against the rectangular models.

5. CONCLUSIONS

A systematic way for comparing texture features was presented, and four different methods of texture analysis were studied in the context of characterizing irregularly shaped texture regions. A protocol of studying the effect of shape and size of a sample and also of characterizing the performance of each algorithm was established. In total 615 classification experiments using 30 different textures from the Brodatz-like collection were performed.

Significant differences between the methods compared were observed. The sum and difference histograms, when the whole histogram was used in the classification process, performed much better than all the other methods studied (with more than 90% accuracy in

all experiments). The method based on the Boolean model was the second best. The features computed from the co-occurrence matrices performed similarly to the Boolean model when model and test sample were identical textures, but significantly worse when model and test sample were from different realizations of the same texture. The tree-structured wavelet transform was the method that provided the worse results in all experiments, even when square models were considered. Finally, it seems that the performance of all methods studied is much more influenced by the sample size than by the sample shape, except for the Boolean model, which is influenced by the sample size and shape.

REFERENCES

1. T. Chang and C. C. J. Kuo, Texture analysis and classification with tree-structured wavelet transform, *IEEE Trans. Image Process.* **2**, 1993, 429–441.
2. R. Conners and C. Harlow, A theoretical comparison of texture algorithms, *IEEE Trans. Pattern Anal. Mach. Intell.* **3**, 1980, 25–39.
3. J. M. H. du Buf, M. Kardan, and M. Spann, Texture feature performance for image segmentation, *Pattern Recognit.* **23**, 1990, 291–309.
4. P. García-Sevilla and M. Petrou, Classification of binary textures using the 1-D Boolean model, *IEEE Trans. Image Process.* **8**, 1999, 1457–1462.
5. P. García-Sevilla, M. Petrou, and S. Kamata, The use of Boolean model for texture analysis of grey images, *Comput. Vision Image Understand.* **74**, 1999, 227–235.
6. R. M. Haralick, K. Shanmugam, and I. Dinstein, Texture features for image classification, *IEEE Trans. Systems Man Cybernet.* **6**, 1973, 609–622.
7. P. P. Ohanian and R. C. Dubes, Performance evaluation for four classes of textural features, *Pattern Recognit.* **25**, 1992, 819–833.
8. P. L. Palmer and M. Petrou, Locating boundaries of textured regions, *IEEE Trans. Geosci. Remote Sensing* **35**, 1997, 1367–1371.
9. T. R. Randen and J. H. Husoy, Filtering for texture classification: a comparative study, *IEEE Trans. Pattern Anal. Mach. Intell.* **21**, 1999, 291–310.
10. M. Unser, Sum and difference histograms for texture classification, *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 1986, 118–125.
11. J. Weszka, C. Dyer, and A. Rosenfeld, A comparative study of texture measures for terrain classification. *IEEE Trans. Systems Man Cybernet.* **5**, 1976, 269–285.