Analysis of Irregularly Shaped Texture Regions: A Comparative Study

P. García-Sevilla
Department of Computer Science
Universitat Jaume I
Castellón, Spain

M. Petrou
School of Electronic Engineering
University of Surrey
Guildford GU2 7XH, UK

Abstract

Most of texture classification techniques are evaluated using large rectangular samples of each texture. This, however, is unrealistic, especially as samples of a certain texture may be used as cues in searching image databases or identifying objects in a scene. In this paper, four different texture classification methods (wavelets, co-occurrence matrices, sum and difference histograms, and 1D Boolean models) are systematically compared and evaluated with respect to their performance in identifying textures from small and irregular samples.

1. Introduction

Most of the algorithms that have been proposed in the literature for texture classification and texture recognition have been tested assuming that a rectangular sample of significant size of the particular texture is available. In practical applications, however, 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 by some generic texture boundary detection method, e.g. [7]. The identified texture regions may subsequently have to be classified by comparison with entries from a texture database. Alternatively, one may consider medical or forensic applications where small texture samples are available, possibly manually segmented, which have to be characterised and identified.

We chose to compare four different texture classification methods in this context: the tree-structured wavelet transform [1], co-occurrence matrices [5], sum and difference histograms [9], and features derived from the discrete 1D Boolean model [3][4]. To investigate the ability of each method to recognise a texture from samples of arbitrarily irregular shapes and sizes, we create a set of masks which we use to isolate samples of textures from our texture database. All methods are trained on rectangularly shaped large samples, and then are tested for their ability to recognise textures from samples not previously seen by the system, created by clipping test images using the set 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. Although there have been many comparative studies of texture classification methods done in the past (eg. [10][2][6][8]), it is the first time a systematic study of the role played by the available texture sample size has been undertaken.

2. A set of irregular masks

To create a set 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 centre 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. For simplicity 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 non-zero. Each shape in the set, therefore, is given by:

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

where r_0 is the mean radius of the shape, N is the number of non-zero 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 seen as a measure of how irregular the shape is: the higher the number of the non-zero terms (value of N), the more irregular the shape becomes.

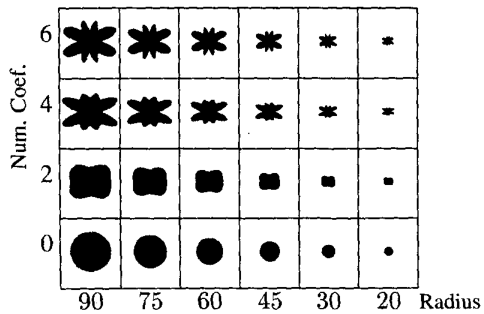

Figure 1. Some of the shapes in the database generated as described in section 2.

Shapes obtained using different values of parameters N and r_0 were included in the set. 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 the set. Fig. 1 shows some of the shapes created. Each cell in the table represents an area of 256 $\times$ 256 pixels in size.

Once the set of irregular shapes was created, we were able to create irregular texture areas by simply using the binary shapes as clipping masks of the texture images.

An image database containing 30 Brodatz-like grey level textures was used in our experiments. We had 4 non-overlapping images per class, where each of them is 256 $\times$ 256 pixels in size. We will refer to each image in a class as a *quadrant*, as they were obtained by splitting a bigger image into 4. Therefore, a total of 120 square grey level images were included in the database.

3. Methodology

As we are interested in studying the influence of the size and irregularity of the area to be analysed to the calculation of the texture parameters, we run three series of experiments in which the whole square images were used as models and a series of samples were used for testing. In the first series of experiments, the irregular samples were obtained directly from the model images. In this case, the irregularly shaped patch to be classified must contain exactly the same texture as the models, and therefore, the differences in the parameters between the test samples and the corresponding model images must be only due to the irregularity and different size of the texture area, and to nothing else. In the second series of experiments, the irregular samples are derived from images that are not used as models by the classifier, so they provide a more realistic situation, where differences between model and test parameters may be due to the shape and size of the sample as well as to the texture that appears into each image. Finally, the third series of experiments considers the traditional situation where square samples are classified against square models. This case is to be used as a benchmark for the other two cases.

For the first series of experiments, the methodology applied can be described as follows:

Thirty square grey level images were selected from the database, one from each texture class and their corresponding parameters were computed. We took each one of those 30 images independently and applied over it all the irregular masks previously defined. That is, we obtained 135 irregular areas per class. The corresponding textural parameters were computed for all these areas. Then, the parameters of each area were compared with those of the square models and for each sample we found whether it could be correctly classified or not.

Fig. 2(Left) shows an example of the results obtained for the irregular masks derived from a class. The axes are the same as those of Fig. 1, i.e. along the vertical axis we have a measure of the irregularity of the shape, while along the horizontal axis we have a measure of the size of the sample. Each cell is colour coded to represent whether the shape was correctly classified (black or grey) 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 recognisable from unrecognisable shapes. Grey cells refer to shapes that, although were correctly classified, were finally ignored. We consider the successful recognition of shapes marked with grey in Fig. 2(Left) as coincidental. For example, if shape with $N = 2$ and $r_0 = 55$ could not be recognised, we do not expect that an even smaller sample of the texture, indicated by $N = 2$ and $r_0 = 45$ could be reliably recognised. Fig. 2(Right) shows the final line considered as a border between recognisable and unrecognisable shapes of the particular texture.

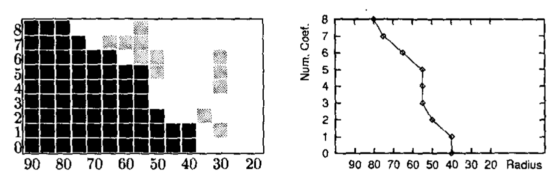

Figure 2. Classification results for shapes derived from one class.

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 recognisable and unrecognisable shapes. In order to obtain only one boundary line per image, we combined the results obtained for each quadrant into one global line that summarises 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). In order to obtain a quantitative comparison between the texture characterisation methods tested, we may compute for each method the percentage of irregular shapes that fall in the recognisable side of the previous graph. As we used 135 irregular shapes per image and there were 30 textures in the database, a total of 4050 shapes applied to four different quadrants were considered in each experiment.

4. Experiments

4.1. Tree-Structured Wavelet Transform

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 grey 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 normalised 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). Fig. 3(a) shows the percentage of correctly classified samples for each filter bank as a function of the number of channels used in the classification for the first experiment. It can be noticed that all the Daubechies coefficients performed similarly, while the Haar filter bank performed slightly worse.

4.2. Co-occurrence Matrices

Experiments using features obtained from co-occurrence matrices were run using the same parameters described in [6], that is, four directions were considered and four textural features were computed for each matrix. Each feature was normalised 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 shape were taken into account, and this number of pairs was also used to normalise the textural features.

The distances d considered varied from 1 to 5, while the number of grey levels was quantised to $G = 8$, 16, and 32. Fig. 3(b) shows the percentage of correctly classified samples for each value of G as a function of the number of distances d considered for the first experiment. 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. Furthermore, the curse of dimensionality must be taken into account because the more distances considered, the more features are used to describe a texture.

4.3. Unser's sum and difference histograms

Experiments using features obtained from the sum and difference histograms were run using the same four directions as for the co-occurrence matrices. 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 grey levels in the images was quantised to $G = 8$, 16, and 32. Both classification methods described in [9] were used.

Fig. 3(c) shows the percentage of correctly classified samples obtained in the first experiment for each value of G as a function of the number of distances d considered. In this case, the global features described in [9] were used to characterise each histogram. It can be noticed that the results are highly influenced by the number of grey levels used in the quantisation process. However, the classification accuracy did not decrease as the number of distances considered was increased.

The same experiments were run using now the whole histograms to describe a texture class and a maximum likelihood decision rule was used to classify samples [9]. Fig. 3(d) shows the percentage of correctly classified samples obtained in the first experiment for each value of G as a function of the number of distances d considered.

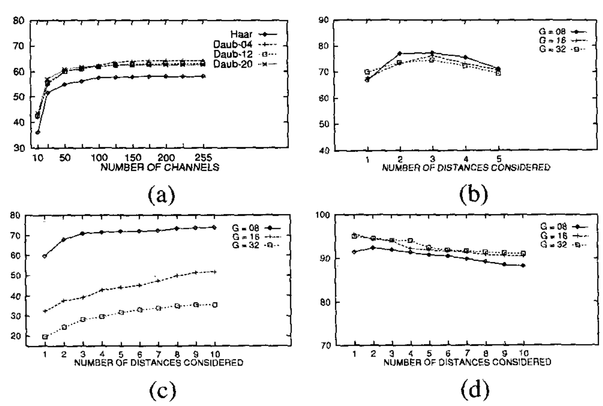

Figure 3. Classification results (a) TSWT (b) Co-occurrence matrices (c) Unser's global features (d) Unser's histograms.

In this case, we note that the results are much better than the ones 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 grey levels used in the quantisation process). However, the larger the number of grey levels considered, the larger the histograms computed and therefore, a larger number of features are considered to characterise a texture. Note again the deterioration of performance as the number of distances considered increases.

4.4. Features derived from the Boolean model

Experiments were run using different methods for splitting a grey image into a set of binary planes, namely bit slicing and grey level quantisation. Two 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 is raster scanned and only complete runlengths are used to calculate the Boolean model statistics (i.e. runlengths that are interrupted by the region boundary are ignored) [3][4].

Table 1 shows the percentage of correctly classified samples obtained for the first experiment using the one-dimensional Boolean model.

Method	3 Planes	8 Planes
Bit Slicing	72.81%	77.23%
Quantisation	62.18%	72.40%

Table 1. Results using the 1D Boolean model.

It can be pointed out that in all experiments the boundaries between recognised and unrecognised samples in Fig. 2(Right) were almost vertical lines. This means that the texture characterisation is much more influenced by the sample size than by the sample shape.

In the second series of experiments, we considered test irregular shapes extracted by clipping images other than the ones used as models. Therefore, differences between model and test parameters may be due to the shape and size of the sample as well as to the texture that appears in each image. Although the final graphs are not shown here, the results obtained for all methods in this case were quite similar to the results obtained in the first series of experiments. In the third experiment, rectangular samples were classified against rectangular models. Table 2 summarises the best results obtained for each experiment and each texture characterisation method.

5. Conclusions

Several methods were used to characterise irregular texture areas. As the characterisation of the texture is done statistically, the area we want to analyse must contain several texture elements in order to provide valid statistics. There were significant differences between the methods compared. The sum and difference histograms when the whole histogram is used in the classification process performed much better than all the other methods studied (with more than 95% accuracy in the classification). The method based in the Boolean model was the second best. The features computed from the co-occurrence matrices performed slightly worse than the Boolean model. Finally, the TSWT was the method that provided the worse results in all experiments. This is perhaps because this method requires rectangularly shaped samples in order to characterise a texture, while all other methods can easily cope with the irregular shapes.

	Exp. 1	Exp. 2	Exp. 3
TSWT	64.12	59.58	91.67
Co-occurrence Matr.	77.33	64.30	98.33
Unser's global features	74.15	74.77	100.00
Boolean model	77.23	76.47	100.00
Unser's histograms	95.68	95.21	100.00

Table 2. Best results in each experiment.

References

[1] T. Chang and C. C. J. Kuo. Texture analysis and classification with tree-structured wavelet transform. *IEEE Trans. Image Processing*, 2(4):429–441, 1993.

[2] R. Conners and C. Harlow. A theoretical comparison of texture algorithms. *IEEE Trans. Pattern Analysis Mach. Intell.*, 3:25–39, 1980.

[3] P. García and M. Petrou. Classification of binary textures using the one-dimensional Boolean model. *IEEE Trans. Image Processing*, 8(10):1457–1462, 1999.

[4] P. García, M. Petrou, and S. Kamata. The use of Boolean model for texture analysis of grey images. *Computer Vision and Image Understanding*, 74(3):227–235, 1999.

[5] R. M. Haralick, K. Shanmugam, and I. Dinstein. Texture features for image classification. *IEEE Trans. Sys., Man, Cybern.*, 6(3):609–622, 1973.

[6] P. P. Ohanian and R. C. Dubes. Performance evaluation for four classes of textural features. *Pattern Recognition*, 25(8):819–833, 1992.

[7] P. L. Palmer and M. Petrou. Locating boundaries of textured regions. *IEEE Trans. on Geoscience and Remote Sensing*, 35(5):1367–1371, 1997.

[8] T. R. Randen and J. H. Husoy. Filtering for texture classification: a comparative study. *IEEE Trans. Pattern Analysis Mach. Intell.*, 21(4):291–310, 1999.

[9] M. Unser. Sum and difference histograms for texture classification. *IEEE Trans. Pattern Analysis Mach. Intell.*, 8(1):118–125, 1986.

[10] J. Weszka, C. Dyer, and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Trans. Syst. Man Cybern.*, 5:269–285, 1976.