

# MAXIMUM LIKELIHOOD TEXTURE CLASSIFICATION AND BAYESIAN TEXTURE SEGMENTATION USING DISCRETE WAVELET FRAMES

S. Liapis, N. Alvertos, and G. Tziritas  
 Institute of Computer Science - FORTH, and,  
 Department of Computer Science, University of Crete  
 P.O. Box 1470, Heraklion, Greece  
 E-mails: {liapis,alvertos,tziritas}@csd.uch.gr

**Abstract:** In this work a new approach is presented for the classification and segmentation of texture images, where a different statistical methodology and criterion for texture characterization is proposed. The scheme, in both problems, uses the concept of Discrete Wavelet Frames for the appropriate frequency decompositions, as applied to 2-D signals, and a distance measure based on the evaluation of parametric scatter matrices of the texture images to be segmented or classified. Experiments yielding excellent results are presented for both algorithms.

## 1 Introduction and problem definition

In computer vision tasks, including multimedia applications (e.g., [8]), often texture information must be classified and segmented for recognition purposes. Several statistical approaches have been proposed in the past for texture analysis [3],[6],[11], which later were enhanced in terms of the information they preserve [9],[1]. However, some disadvantages such as increased computational cost and irreversibility, which are inherent with those approaches, can be eliminated using the wavelet transform [7], [10]. In this paper, the problem of texture classification and segmentation is approached with algorithms based on the concept of wavelet frames. The aim of the analysis is to determine corresponding characteristics to each texture content so that each is uniquely defined. Such a distinction takes place in the frequency domain, where the input image is decomposed to different frequency levels using the Discrete Wavelet Frames (DWF). Once these characteristics are deduced, statistical properties are applied to conclude those features necessary to describe and classify the texture content. Although the philosophy to this approach has been introduced in the past [12], our scheme differs in the statistical methodology for evaluating texture parameters and in the criterion by which a texture point is assigned to a particular subregion of the image to be segmented or classified. The presented work is organized as follows. In the second section, the underlying theory of the basic filters, the necessary decomposition by upsampling and the use of Discrete Wavelet Frames in the form of an algorithm, as applied to 2-D signals, are described. Next, in Section 3, a classification method is introduced, which is based on the maximum likelihood criterion. The segmentation procedure, where a representative vector for the different frequency layers and the distance criterion based on the texture statistics have been defined, is explained in the fourth section. The experimental results for both classification and segmentation are presented in Section 5.

## 2 Preliminary analysis

The fundamental tools used for building the processing of texture images are a group of filters and the concept of wavelet frames. A lowpass filter  $H(z)$  and its complementary highpass  $G(z)$  form the basis for generating more filters by upsampling, so that the whole range of bands is covered. For these basic types of filters the following hold true [12], respectively:

$$\left. \begin{aligned} H(z) &= \frac{z^2+4z+6+4z^{-1}+z^{-2}}{16} \\ G(z) &= zH(-z^{-1}) \end{aligned} \right\} \quad (1)$$

in the frequency domain and

$$h(n) = \mathcal{Z}^{-1}\{H(z)\}, \quad g(n) = (-1)^{1-n}h(1-n) \quad (2)$$

in the time domain. In addition, the generated filters are characterized by locality, thus, taking advantage of the periodicity of signals. Such filters can form orthogonal wavelet base functions of the form [7]:

$$\left. \begin{aligned} \phi_{i,t}(k) &= 2^{i/2}h_i(k-2^it) \\ \psi_{i,t}(k) &= 2^{i/2}g_i(k-2^it) \end{aligned} \right\} \quad (3)$$

where  $\phi, \psi$  are the wavelet base functions,  $i$  is the scale index and  $t$  is the translation index. And so the input signal can be decomposed into wavelet coefficients corresponding to different layers of frequency resolution. In order, however, to consider characteristics of texture, such as periodicity and translational invariance, the Discrete Wavelet Frames (DWF) are used to define a vector representing the filters necessary for decomposition at the different frequency levels. All of the above should be extended into 2-D so that it becomes functional for images with texture, the features of which must be extracted. This can be accomplished by forming wavelet bases which result from the cross product of separable bases in each direction, as follows:

$$\begin{aligned} \Phi(x, y) &= \phi(x)\phi(y) & \Psi_1(x, y) &= \phi(x)\psi(y) \\ \Psi_2(x, y) &= \psi(x)\phi(y) & \Psi_3(x, y) &= \psi(x)\psi(y) \end{aligned} \quad (4)$$

where  $\Phi, \Psi_1, \Psi_2, \Psi_3$  are the 2-D wavelet base functions, and  $\phi, \psi$  are as defined in (3). Thus, the analysis is computationally less complicated, since rows and columns of the image are processed separately as though they were 1-D signals. The decomposition algorithm for images (2-D) is described below:

$$\left. \begin{aligned} d_{1,i+1}(k, l) &= [h]_{2^i}(k) * [g]_{2^i}(l) * s_i(k, l) \\ d_{2,i+1}(k, l) &= [g]_{2^i}(k) * [h]_{2^i}(l) * s_i(k, l) \\ d_{3,i+1}(k, l) &= [g]_{2^i}(k) * [g]_{2^i}(l) * s_i(k, l) \\ s_{i+1}(k, l) &= [h]_{2^i}(k) * [h]_{2^i}(l) * s_i(k, l) \end{aligned} \right\} (5)$$

where  $(k, l)$  is an image point,  $[\ ]_m$  is upsampling with a factor of  $m$ ,  $d_{1,i+1}, d_{2,i+1}, d_{3,i+1}$  are the details of the  $i+1$  layer and  $s_{i+1}$  the approximation of the decomposition.

### 3 Classification

The previous analysis can be applied to input texture images to distinguish  $I$  frequency layers, yielding the following representative vector:

$$y(k, l) = \langle y_1(k, l), \dots, y_{N-1}(k, l), y_N(k, l) \rangle (6)$$

where each element of  $y(k, l)$  has been determined according to the analysis in (5) and the dimension of the vector is  $N = 3I + 1$ , composed of  $N - 1$  detail components and the approximation at level  $I$  component. It is evident that in each resolution level three new feature channels are obtained. The first analysis layer corresponds to high frequencies, while an increasing order of layers represents decreasing frequencies. Thus, depending on the value of the corresponding vector coefficient, the direction and the amount of frequency contribution is deduced at a given image point  $(k, l)$ . Different textures are distinguished based on these last two characteristics.

In this work the discrimination of different textures is only based on the  $N - 1$  high frequency components. Each texture class is then characterized by the variances of the high frequency components  $y_i(k, l)$ , say  $\sigma_i^2$  ( $i = 1, \dots, N - 1$ ). Indeed, the mean value of each high frequency component, as well as the correlation coefficients between different components could be assumed to be zero. There are two reasons justifying the above: first, because a texture is best described through the frequency channels and not through the difference of the approximation and second, because two images of the same texture content may have different variances in the approximation channel only due to differences in contrast. Also, assuming Gaussian probability density functions with the previous statistics, the maximum likelihood criterion gives the distance of a test texture  $y$  from a class  $j$ ,

$$d_j(y) = \sum_{i=1}^{N-1} \sum_{k,l} \left( \frac{y_i^2(k, l)}{\sigma_{i,j}^2} + \log \sigma_{i,j}^2 \right) (7)$$

where the first sum is taken over all image points, and  $\sigma_{i,j}^2$  is the variance of the  $i$  component of class  $j$ .

### 4 Segmentation

The same image decomposition and statistical analysis could be used for texture segmentation. Nevertheless,

for the segmentation purpose all the components of the decomposed image should be used. In this work, the variances of the components are assumed as known from a previous learning process, considered in a subsequent paragraph. The mean value of the approximation component is also assumed as known.

Taking into consideration all parameters characterizing a texture content, a given point belongs to a known texture content, if its distance from the given texture is minimal. Assuming that the probability density function of the texture images is a Gaussian one, then the distance of a point  $(k, l)$  represented by the vector  $y(k, l)$  from a texture content with variances  $\sigma_{i,j}^2$  and mean value  $\mu_j$  is determined as follows:

$$d_j(y(k, l)) = \sum_{i=1}^{N-1} \left( \frac{y_i^2(k, l)}{\sigma_{i,j}^2} + \log \sigma_{i,j}^2 \right) + \frac{(y_N(k, l) - \mu_j)^2}{\sigma_{N,j}^2} + \log \sigma_{N,j}^2 (8)$$

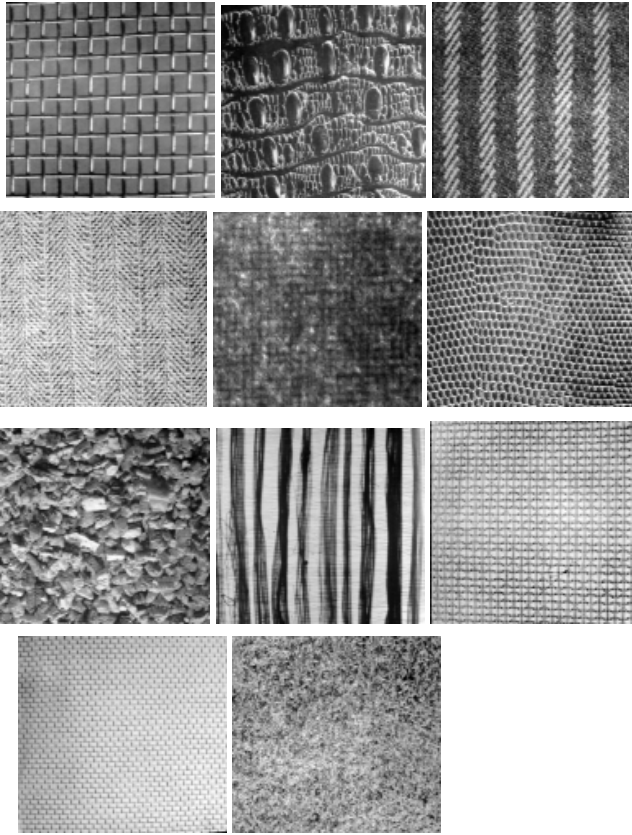
Following distance evaluation so that it is determined to which texture content each image point belongs, an algorithm for merging connected regions is used. This is necessary because of small point-classification errors of the statistical method described earlier. A bayesian approach is adopted based on a Markov random field model of the texture labels. The optimization is performed using the *Highest Confidence First (HCF)* algorithm [4].

As it was described, at the beginning of the segmentation process the the variances of the components are required. The corresponding regions from which these initial parameters are calculated can be given at the input by the user. It is possible, however, to apply hierarchical clustering algorithms [5] so that their initial parameters are estimated without the user's supervision. The only given information at the input is the number of different texture contents present in the image. To achieve this, at the initial stage the whole image is divided into non-overlapping windows of size  $32 \times 32$ . The inter distances between these windows are evaluated using the sum of squared distances for each component:

$$d_{12} = \sum_{i=1}^N (\sigma_{i,1}^2 - \sigma_{i,2}^2)^2 (9)$$

Other distance criteria can be used as well, such as a weighted sum of differences, where the size of the clusters to be compared is an important factor. In any case, those neighboring windows with the smallest distance measure are merged and the corresponding parameter is estimated. This procedure is repeated, for another such pair of windows, up to the point where there is a sufficient number of neighboring pairs of windows from almost all regions of the initial image so that each pair from a region has a common parameter. Then, those parameters, which are further from each other, are kept, based on the distance measure given above. Next, a global clustering takes place on all windows of the image using as initial parameters those resulting from the previous step. Thus, the initial parameters for the segmentation procedure are obtained. The

reason for breaking down the evaluation of these parameters into the intermediate steps described earlier is due to the appearance of large variances for neighboring windows from two different texture contents, which results into evaluation of parameters based only on neighboring windows of the same texture content.



**Figure 1. Brodatz Images: D1, D10, D11, D17, D19, D3, D5, D51, D52, D6, D9**

## 5 Experimental results

### 5.1 Classification Results

The data base used for the classification experiments consists of eleven images of different texture content from the Brodatz album [2], shown in Figure 1. The DWF algorithm is applied to analyze the images from the data base. Then, the variances  $\sigma_i^2$  are calculated in order to characterize each texture image, based on the previously described classification algorithm. In addition, so that the test data set is enlarged, the images of the data base are divided into smaller size images (blocks) of  $32 \times 32$  dimension. This results into a data set of 2580 subimages. Then, each subimage is classified as one of the database images. The statistical results of this experimental procedure are shown in Table 1. In this table the percentage of correct classifications is given, as well as the misclassifications.

In our case, the degree of difficulty of the classification task was considerably higher than other experiments, since the produced data set contained a much larger number of subimages, as each image is divided in 240 subimages. The analysis was performed at a depth value of 5,

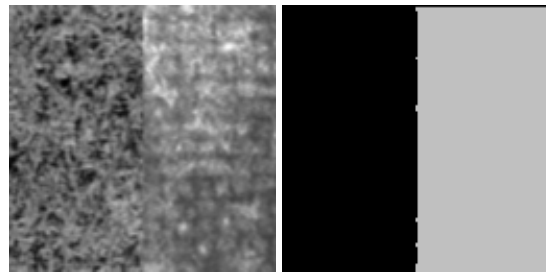
	total blocks	misclassifications	percent
D1	240	none	100.00%
D10	240	D19:2 D5:7 D9:4	94.58%
D11	225	D52:1	99.56%
D17	240	D3:1 D9:4	97.92%
D19	210	D11:1 D9:3	98.10%
D3	240	D17:1 D52:2	98.75%
D5	240	D10:4 D19:3 D9:5	95.00%
D51	240	D1:1 D10:4	97.92%
D52	225	D11:9 D3:4	94.22%
D6	240	D1:2 D17:2 D5:1	97.92%
D9	240	D10:3 D11:2 D17:6 D52:1	95.00%

**Table 1. Results of classification for  $32 \times 32$  blocks.**

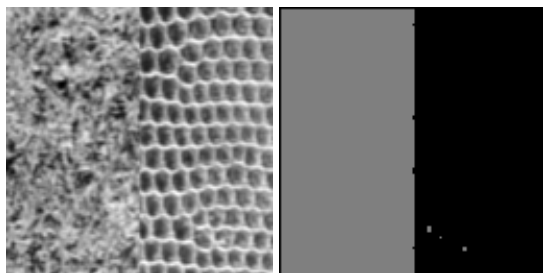
which corresponds to fifteen element feature vectors since the approximation value was omitted. The total number of misclassified texture subimages was 73 out of 2580, corresponding to 97.17% correct classification. The same results are given (Table 2) for  $64 \times 64$  blocks, where, as expected, the percentage of correct classifications is higher, and for our method near to 100% (99.66%).

	total blocks	misclassifications	percent
D1	56	none	100.00%
D10	56	none	100.00%
D11	49	none	100.00%
D17	56	D9:1	98.21%
D19	49	none	100.00%
D3	56	none	100.00%
D5	56	none	100.00%
D51	56	D10:1	98.21%
D52	49	none	100.00%
D6	56	none	100.00%
D9	56	none	100.00%

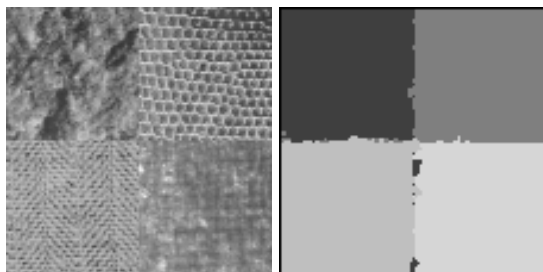
**Table 2. Results of classification for  $64 \times 64$  blocks.**



**Figure 2. Segmentation for the synthetic image which contains two textures D9, D19**



**Figure 3. Segmentation for the synthetic image which contains two textures D9 D3**



**Figure 4. Segmentation for the synthetic image which contains four textures D2 D3 D17 D19**

## 5.2 Segmentation Results

The segmentation algorithm described in the second section of this paper was applied on several images containing two or four different textures. One initial image to be segmented (composed of textures D9 and D19) and the result of the segmentation are shown in Figure 2. Another example is illustrated in 3 where the image consists of textures D9 and D2. Finally, an additional synthetic image which, in this case, is composed of four different textures (D2, D3, D17, D19) and the resulting segmented output image are depicted in Figure 4.

## 6 Conclusion

The problem of texture classification and segmentation is addressed, where the concept of Discrete Wavelet Frames is used for decomposing the image into different frequency levels. Both procedures use the same statistical methodology for evaluating texture parameters and the same form of criterion by which a texture point is assigned to a particular subregion of the image to be segmented or classified. Both algorithms (segmentation and classification) were demonstrated using images of different texture content, where the results were very satisfactory (e.g., 97.17% correct classification) and more encouraging compared to other works. An improvement of the classification approach could be achieved if a normalization in terms of the image variance takes place for all images contained in the data base and in the test set at the beginning of the process. In that case, the resulting effect would be a set of images with similar contrast.

## References

- [1] A.C. Bovic, M. Clark, and W.S. Geisler. Multi-channel texture analysis using localized spatial filters. *IEEE Transactions on PAMI*, 12:55–73, January 1990.
- [2] P. Brodatz. *A Photographic Album for Artists and Designers I*. Dover, New York, 1966.
- [3] P.C. Chen and T. Pavlidis. Segmentation by texture using correlation. *IEEE Transactions on PAMI*, 5:64–69, January 1983.
- [4] P. Chou and C. Brown. The theory and practice of bayesian image labeling. *International Journal of Computer Vision*, 4:185–210, 1990.
- [5] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. New York: J. Wiley & Sons, 1973.
- [6] R.L. Kashyap, R. Chellappa, and A. Khotanzad. Texture classification using features derived from random field models. *Pattern Recognition Letters*, 1:43–50, 1982.
- [7] S. G. Malat. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Patt. Anal. Machine Intell.*, 11:674–693, January 1989.
- [8] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *M.I.T. Media Laboratory Perceptual Computing Technical Report No. 255*, November 1993.
- [9] M. Porat and Y. Y. Zeevi. Localized texture processing in vision: Analysis and synthesis in gaborian space. *IEEE Trans. Biomed. Eng.*, 36:115–129, 1989.
- [10] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Signal Processing Mag.*, 8:11–38, October 1991.
- [11] M. Unser. Local linear transforms for texture measurements. *Signal Processing*, 11:61–79, July 1986.
- [12] M. Unser. Texture classification and segmentation using wavelet frames. *IEEE Trans. on Image Processing*, 4:1549–1560, November 1995.