Color and Texture Image Retrieval Using Chromaticity Histograms and Wavelet Frames

Spyros Liapis and Georgios Tziritas

***Abstract*—In this paper, we explore image retrieval mechanisms based on a combination of texture and color features. Texture features are extracted using Discrete Wavelet Frames (DWF) analysis, an over-complete decomposition in scale and orientation. Two-dimensional (2-D) or one-dimensional (1-D) histograms of the *CIE Lab* chromaticity coordinates are used as color features. The 1-D histograms of the a, b coordinates were modeled according to the generalized Gaussian distribution. The similarity measure defined on the feature distribution is based on the Bhattacharya distance. Retrieval benchmarking is performed over the Brodatz album and on images from natural scenes, obtained from the VisTex database of MIT Media Laboratory and from the Corel Photo Gallery. As a performance indicator *recall* (relative number of correct images retrieved) is measured on both texture and color separately and in combination. Experiments show this approach to be as effective as other methods while computationally more tractable.**

***Index Terms*—Bhattacharya distance, color image similarity, discrete wavelet frames, texture retrieval.**

I. Introduction

THE CURRENT explosion in the size of image archives necessitates effective ways of managing (describing, indexing, and retrieving) visual information by its content [5], since a textual description of the image content may be subjective and inadequate for automatic retrieval. In order to describe image content, low level arithmetic features must be extracted that are quantitatively comparable. The MPEG-7 working groups are seeking to define and standardize image content description for automatic indexing [21], [26].

Numerous features, such as shape, color, texture, and motion, have been proposed and used to quantitatively describe visual information [5]. Many image retrieval systems have indeed been developed using all or some of these features, including QBIC [4], Photobook [20], Chabot [19], and Virage [6].

In this work, texture and color features are explored as a means of image content description. Image classification is performed using these features globally, i.e., describing the texture and color content for the image as a whole. It would also be possible to extract the same features for previously segmented objects.

Most approaches to classifying textured images proposed during the past decades fall into three broad categories: statistical, model-based, and signal processing techniques [24], [25]. Statistical approaches, such as [7], were primarily investigated in the 1980s. Model-based methods include, among others, Markov random field [9], simultaneous autoregressive [18], and Wold models [14]. Our approach could be characterized as a signal processing method, and in particular it is based on a joint spatial/spatial-frequency representation of the texture patterns. As with most signal processing methods, the image is submitted to a filter bank, following which local energy measurements are made. In this setting the Gabor filter analysis may be applied [1], [8], [22]. Having similar properties, but permitting a simpler implementation, wavelet transform representations could also be used. Pioneering work on texture classification using the wavelet packet representation is documented in [3], [12]. Also related to this approach is the quadrature mirror filter bank (QMF) decomposition presented in [10].

Manjunath and Ma [17] studied certain aspects of image processing important for extracting texture features relevant to browsing and retrieval. They propose the use of Gabor wavelet features together with an adaptive filter selection algorithm. Their dissimilarity measure was based on the Mahalanobis distance. They provided comparison over the Brodatz album [2] with the performance of three other multiscale analysis methods: the conventional pyramid-structured wavelet transform, the tree-structured wavelet transform, and the multiresolution simultaneous autoregressive model. The Gabor wavelet analysis is seen to outperform the other approaches with respect to the average recognition rate. The pyramid-structured wavelet transform was more efficient in terms of computational complexity. The MPEG-7 *homogeneous texture descriptor* is also based on a Gabor filter bank, where the frequency space is partitioned into 30 channels with equal division in the angular direction and octave division in the radial direction (five octaves) [16].

Liu and Picard [14] proposed the use of a Wold decomposition for extracting perceptual features, described as periodicity, directionality, and randomness. They compare the performance of the Wold model with that of four other methods: the multiresolution simultaneous autoregressive model, the principal component analysis, the tree-structured wavelet transform, and the Tamura features. The experiments were again performed over the Brodatz database, and the Wold model was shown to be superior, while the Tamura features and wavelet transform yield relatively poor results.

In this paper, the problems of texture classification, browsing, and retrieval are approached with algorithms based on the concept of wavelet frames. The analysis aims at determining characteristics corresponding to each texture type, so that each texture pattern is uniquely defined. Such a distinction takes place in

Manuscript received June 14, 2002; revised January 9, 2003. The associate editor coordinating the review of this paper and approving it for publicaiton was Prof. Wayne Wolf.

The authors are with the Department of Computer Science, University of Crete, GR-714 09 Heraklion, Crete, Greece (e-mail: liapis@csd.uoc.gr; tziritas@csd.uoc.gr).

Digital Object Identifier 10.1109/TMM.2004.834858

the frequency domain, where the input image is equivalently decomposed into different scale levels using the Discrete Wavelet Frames (DWF). Once these characteristics are deduced, statistical methods are applied to extract the features necessary for the description and classification of the texture pattern. Although the philosophy of this approach has been introduced in the past [27], in our work a new statistical measure is proposed for evaluating texture pattern similarities. The main advantages of the Wavelet Frames representation are that it focuses on scale and orientation texture features, it decomposes the image into orthogonal components, and it is translation-invariant.

For color features the *CIE Lab* color system was chosen, because of its perceptual uniformity. Only the chromaticity coordinates (a, b) are used to describe color. In general, color content is best described by the chromaticity distribution which is given by one-dimensional (1-D) or two-dimensional (2-D) histograms. The computational complexity is reduced if Gaussian or Laplacian models are assumed for these distributions.

We would like to point out here that in MPEG-7 color histogram description is defined by uniformly quantizing into 256 bins in the HSV (Hue, Saturation and Value) space [16]. We have opted for the *CIE Lab* color space because the quantization assumes Euclidean distances, which better correspond to color dissimilarity in such a perceptually uniform space. Our work is limited to the color histogram description, while in MPEG-7, color layout was also used for capturing the spatial distribution of color in an image [16].

In order to compare texture and color features, a common distance measure is used. This measure is chosen to be the Bhattacharya distance for its good classification properties and because it allows the combination of different features in a simple way. The performance of the features is checked according to a retrieval benchmark proposed in [14]. Three data sets are considered. The first database is the whole Brodatz album [2], a collection of various natural grayscale textures. The second is obtained from the MIT Media Laboratory VisTex database [11], which contains images of scenes of physical color textures. The third is obtained from the Corel Photo Gallery.

This work is organized as follows. In Section II, we describe the underlying theory of the basic filters, the necessary decomposition by upsampling, and the use of DWFs, as applied to 2-D signals. Next, in Section III, the texture- and color-describing features are introduced, as well as the dissimilarity measure, which is based on the Bhattacharya distance. In Section IV, the benchmarking results are presented, compared with other methods and discussed.

II. Texture Analysis and Characterization

The basic tools used for processing the textured images are a filter bank and the concept of wavelet frames, ensuring the translation invariance of the features. The filter bank decomposes the image into orthogonal components, simplifying the classification problem. A low-pass filter $H(z)$ and its conjugate quadrature high-pass $G(z)$ form the pair of prototype filters; from these the entire filter bank is generated by upsampling, so that the whole range of bands is covered. Cubic splines having

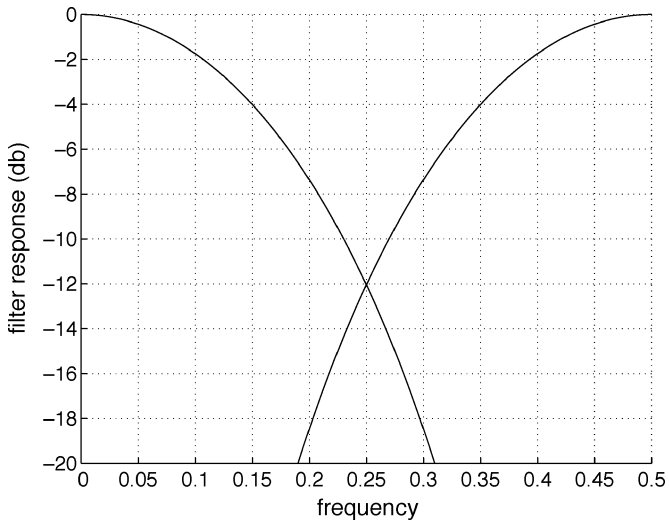


Fig. 1. Frequency responses of the pair of conjugate filters.

appropriate scaling properties could be used for designing the pair of prototype filters [27]. Here we use the fourth-order binomial filter and its conjugate quadrature filter:

$$H(z) = \frac{3}{8} + \frac{1}{4}(z + z^{-1}) + \frac{1}{16}(z^2 + z^{-2}) \tag{1}$$

$$G(z) = zH(-z^{-1}) \tag{2}$$

in the frequency domain, and

$$h(n) = \begin{cases} \frac{6}{16}, & n = 0 \\ \frac{4}{16}, & |n| = 1 \\ \frac{1}{16}, & |n| = 2 \\ 0, & |n| > 2 \end{cases} \quad g(n) = (-1)^{1-n} h(1-n) \tag{3}$$

in the spatial domain. The two-channel filter bank amplitude responses in *db* are illustrated in Fig. 1. The filter bank is obtained iteratively, indexed by the scale factor $i = 0, \ldots, I$:

$$H_{i+1}(z) = H\left(z^{2^i}\right) H_i(z) \tag{4}$$

$$G_{i+1}(z) = G\left(z^{2^i}\right) H_i(z) \tag{5}$$

where $H_0(z) = 1$. Such filters can form orthogonal wavelet basis functions of the form

$$\phi_{i,t}(k) = 2^{\frac{i}{2}} h_i(k - 2^i t)$$

$$\psi_{i,t}(k) = 2^{\frac{i}{2}} g_i(k - 2^i t) \tag{6}$$

where ϕ, ψ are the wavelet basis functions, i is the scale index, and t is the translation index.

At each iteration the filters are dilated by a factor of two and, therefore, the signal is decomposed in subbands of approximately one octave each. The filter bank is iteratively implemented. Each step involves a convolution with the prototype filters, which are expanded by insertion of an appropriate number of zeros between taps [27], as shown in Fig. 2.

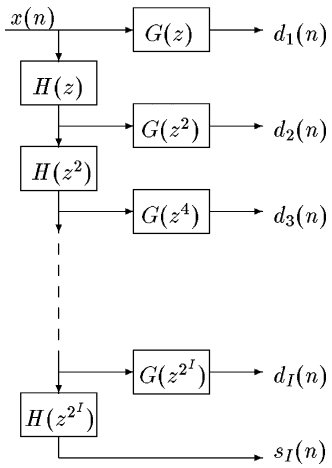


Fig. 2. Iterative implementation of the DWF decomposition.

Thus, the input signal is decomposed into wavelet coefficients corresponding to different scales. The DWFs can handle important texture characteristics, such as periodicity and translational invariance. The representation is overcomplete because the filtered images are not subsampled. This leads to fewer constraints on the choice of the prototype filters. Even more importantly, it facilitates the selection of orthonormal basis functions. The filter pair in (3) is selected because of its good orthogonalization properties.

Next, we present two indicators for band separation by the filter pair. The antialiasing coefficient η measures the energy concentration in the low-frequency band by the low-pass filter:

$$\eta = \frac{\int_0^{\frac{1}{4}} |\mathcal{H}(f)|^2 \, df}{\int_0^{\frac{1}{2}} |\mathcal{H}(f)|^2 \, df}$$

where $\mathcal{H}(f)$ is the Fourier transform of the low-pass filter H. For the above fourth-order binomial filter we have $\eta = 0.985$, close to the ideal value of 1. In addition we use the correlation coefficient between the approximation signal and the details signal as follows:

$$\gamma_{LH}(m) = \frac{\sum_k \sum_l h(k) g(l) \gamma_s(m+k-l)}{\sqrt{\sum_k \sum_l h(k) h(l) \gamma_s(k-l)} \sqrt{\sum_k \sum_l g(k) g(l) \gamma_s(k-l)}}$$

where $\gamma_s(\cdot)$ is the autocovariance function of the input signal. For a fully correlated input signal, $\gamma_{LH}(m)$ is zero-valued for any m, because the sum of all $g(\cdot)$ coefficients is zero. An interesting and important indication of the output correlation is given by the maximum value of $\gamma_{LH}(m)$ under the assumption of a first-order Markov process for the input signal. Since at each layer the input signal processed is an approximation signal with strong correlation, we provide here the maximum $\gamma_{LH}(m)$ value for $\rho = 0.9$:

$$\max_m \gamma_{LH}(m) = 0.027,$$

which is small enough to consider the interband decorrelation hypothesis to be valid in practice.

All of the above are extensible into 2-D, thus becoming applicable to images with texture whose features must be extracted. This can be accomplished by forming wavelet bases which result from the cross-product of separable bases in each direction. Thus, the analysis is computationally straightforward, since rows and columns of the image are processed separately, as if they were 1-D signals.

Of course, other filter prototypes could be used in the same processing scheme. Randen and Husoy [24] compared the most frequently used filtering approaches, including Laws masks, Gabor filter banks, wavelet transforms, wavelet packets, wavelet frames, quadrature mirror filters, discrete cosine transform, eigenfilters, and linear predictors. They do not give conclusive reasons for preferring a particular filter, because often the effectiveness of the various filters depends on the image in question. Nonetheless, the QMF and wavelet frames approaches were among the best in the comparisons performed in [24].

In addition to wavelet frames decomposition, other filtering analysis approaches could also be used in the following section considering the feature extraction, classification, and retrieval problems. We adopted the DWF analysis because of the above separability property, which permits a simpler implementation.

III. Feature Extraction and Dissimilarity Measure

A. Texture Feature Extraction

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

$$y(m,n) = [y_1(m,n), \ldots, y_{N-1}(m,n), y_N(m,n)] \quad (7)$$

where each element of $y(m,n)$ has been determined by the wavelet frame analysis and the dimension of the vector is $N = 3I + 1$, composed of $3I$ detail components and the approximation at level I component. It is evident that at each resolution level three new feature channels are obtained, which characterize for the given scale depth the horizontal, the vertical, and the diagonal components. The first analysis layer corresponds to high frequencies, while the higher-order layers represent successively lower frequencies. Thus, depending on the value of the corresponding vector coefficient, the direction and the contribution by a certain frequency are deduced at a given image point (m,n). Different textures are distinguished based on these last two characteristics.

In this work, the discrimination of different textures is based only on the $N-1$ high-frequency components. The low-pass component is not used because textures are better described through the higher frequency channels than through the approximation component. One reason for this is that a mean-invariant classifier is required; secondly, the approximation component might not be stationary, with a shift-variant variance, i.e., the approximation component is more sensitive to local inhomogeneities. Each texture class is then characterized by the variances of the high-frequency components $y_i(m,n)$, say $\sigma_i^2 (i = 1, \ldots, N-1)$. The mean value of each high-frequency component is zero, because the sum of the corresponding filter coefficients is zero. In addition, the correlation coefficients between different components could be assumed to be zero, as explained in the previous section, dealing with the filter properties.

B. Color Features

In order to characterize the color content of an image, the *CIE Lab* color space is used. This color coordinate system has the advantage that it is designed to be perceptually uniform, meaning that equal distances in the color space correspond to equal color differences, as perceived by humans, i.e., images which are perceptually similar have the same chromaticity components. It also has the advantage that lightness L is distinct and independent from the chromaticity coordinates (a, b). For color image classification and retrieval it is more relevant to compare the chromaticity distribution of an image, disregarding the lightness component. This exclusion of lightness is enforced in our case by the fact that lightness is used to extract texture features.

In order to characterize the chromaticity content of an image, the 2-D histogram of the (a, b) coordinates are used. A uniform quantization of the 2-D histogram down to 1024 chromaticity bins is performed, because otherwise it would be very large and very sparse (a, b range over $[-137, 96]$, $[-99\,133]$, yielding 54 056 bins). Most of the values of these coordinates are very dense in a small region around zero; higher absolute values are found only when the image contains pure colors such as high saturated red or blue. Empirically the values of (a, b) found in natural images are compact and occupy a small portion of the whole range of values.

This method has the advantage of describing exactly the 2-D distribution of the chromaticity coordinates. However it has the disadvantage that 1024 floating point numbers are needed for storage for the empirical probability density function of each image. This size could be reduced if the coordinates are uncorrelated, in which case the 1-D histograms of each coordinate could be used. Thus color feature could use the 232 and 233 bins of the (a, b) histograms, respectively.

In order to reduce the number of color features, we can assume a model for each coordinate distribution. In our case, the Gaussian and Laplacian distribution are used as models, which require only the mean and variance of the image's color coordinates. The storage demands are minimized and the comparison of color features is accelerated. Detracting from this model's usefulness is that its assumptions are not always valid. This fact leads us to a constrained data set in which each image will contain chromaticities concentrated around a concrete value at each coordinate.

C. Dissimilarity Measures

Measuring the dissimilarity between images is of central importance for retrieving images by content. A thorough empirical evaluation of many different dissimilarity measures for color and texture is given in [23], where for large histograms it is concluded that of these the χ^2 distance, the Kullback divergence and the Jensen divergence perform best. In our work, we considered two dissimilarity measures in the same category: the likelihood ratio and the Bhattacharya distance. A preliminary study [13] and empirical evaluation of these two measures lead us to select the latter for retrieving images by color and texture. In this paragraph, we present both, and in benchmarking, we use the Bhattacharya distance.

Assuming Gaussian probability density functions with the previous statistics, the logarithmic likelihood criterion gives the distance of a test texture y from a class j:

$$d_j^L(y) = \sum_{(m,n)} \sum_{i=1}^{N-1} \left(\frac{y_i^2(m,n)}{\sigma_{i,j}^2} + \ln \sigma_{i,j}^2 \right) \propto \sum_{i=1}^{N-1} \left(\frac{S_i^2}{\sigma_{i,j}^2} + \ln \sigma_{i,j}^2 \right) \tag{8}$$

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

A more interesting criterion is the Bhattacharya distance, because it is strongly linked to the minimum classification error for the case of two classes, and because the two texture patterns, the that of the data set and the one being tested, are treated symmetrically. The Bhattacharya distance is in fact a special case of the Chernoff bound of the Bayes error [28]. It is well known that the Chernoff information gives the highest achievable exponent for the error probability. The Bhattacharya distance is defined by

$$d_B(p_1, p_2) = -\ln \left(\int_x \sqrt{p_1(x) p_2(x)} dx \right) \tag{9}$$

where p_1, p_2 are probability density functions of vector x of any dimension under two hypotheses indexed by 1 or 2. The Bhattacharya distance satisfies the symmetric property, $d(p_1, p_2) = d(p_2, p_1)$; the triangular property is satisfied only for specific configurations.

In our case this distance could be defined on empirical probability distributions. The discrete form of the Bhattacharya distance is

$$d_B(h_1, h_2) = -\ln \left(\sum_i \sqrt{h_1(i) h_2(i)} \right) \tag{10}$$

where i is an index into the bins of the normalized histograms h_1, h_2.

In the case that we have a model for the features' distribution, a simpler expression for the Bhattacharya distance can be deduced. Statistical analysis of experimental results have shown that the probability distribution of the high-frequency components could be taken to be the generalized Gaussian [15]

$$p(x) = \frac{c}{2\sigma\Gamma\left(\frac{1}{c}\right)} e^{-\left(\frac{|x|}{\sigma}\right)^c}$$

where the parameter σ is related to the variance and c reflects the sharpness of the probability density function. For $c = 2$, we obtain the Gaussian function, and, for $c = 1$, the Laplacian function. Taking into account that we have in practice an interband decorrelation, the Bhattacharya distance will be the sum of the corresponding distances on all the components

$$d_j^B(y) = \frac{1}{c} \sum_{i=1}^{N-1} \ln \frac{S_i^c + \sigma_{i,j}^c}{2\sqrt{S_i^c \sigma_{i,j}^c}}. \tag{11}$$

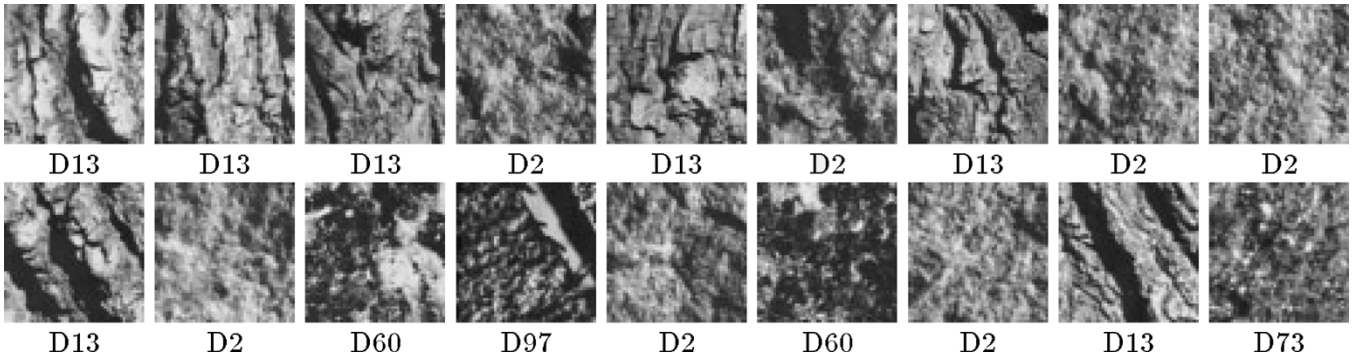


Fig. 3. Retrieval example: the top-left icon is the query picture; the remaining 17 icons are shown in order of decreasing similarity w.r.t the first icon.

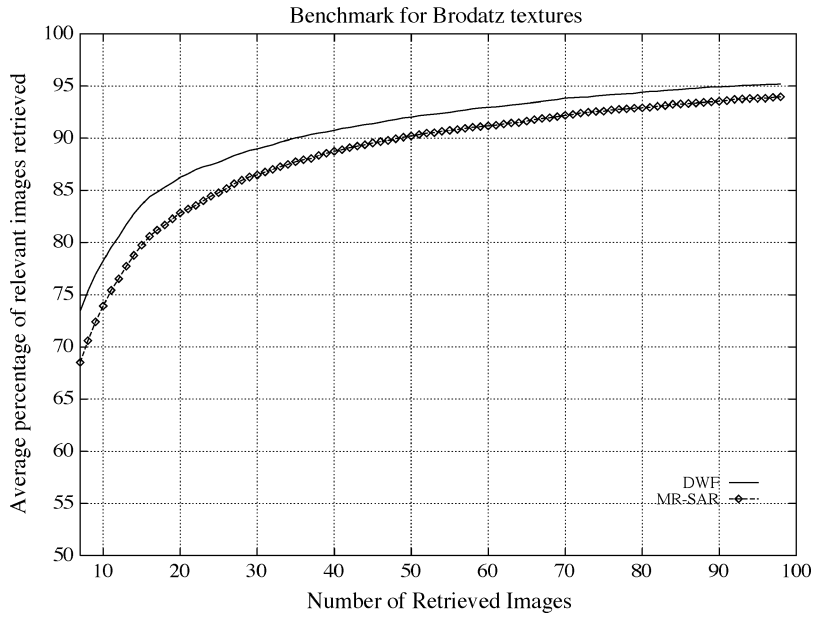


Fig. 4. Average percentage curves over the Brodatz database.

In our work, we have considered and give results for both $c = 2$ and $c = 1$.

When texture features and color histogram features need to be combined, the last model-based expression is used for texture features and the initial expression is used for histograms. Each term for each feature is summed independently of the expression for the distance, because the decomposition is approximately orthogonal.

IV. Retrieval Benchmark

In order to assess the capabilities of the texture and color features alone or in combination, a retrieval benchmark comparison was performed. For this purpose, all the images in the database are sectioned into an equal number of icons, all of the same size, provided that all the images in the database have the same size. A large database of icons is thus obtained. Each small icon in the database is used to retrieve from the database the nearest (most similar) icons, except itself. The similarity between two icons is determined using the Bhattacharya distance as described in the previous section. For each number of retrieved icons, we record the *recall*, i.e., the number of relevant images retrieved relative to the total number of relevant images in the database.

TABLE I
Average Recognition Rate for the Brodatz Collection and for Different Methods and Implementations

Method	Retrieval performance
Wold [14]	75%
DWF (our result)	73.4%
Gabor [17]	73.1%
MRSAR [17]	72.0%
MRSAR [14]	70 %
TWT [17]	68.0%
PWT [17]	67.4%

This result is presented graphically in Fig. 4 as a hit-rate curve versus the number of retrieved images. Clearly, this curve will be increasing, because as the number of the retrieved icons is increasing the classification rate is increasing. A retrieval example is shown in Fig. 3 for the relatively heterogenous *D13* Brodatz image. In this example for eight retrieved icons, the *recall* is 50%, and for 17, it is 75%. We note that, even if this example is chosen for its poor result, the retrieved images are perceptually similar.

The first database used is the entire Brodatz collection containing 112 grayscale texture pictures. Each of them is subdivided into nine nonoverlapping subimages, resulting in 1008

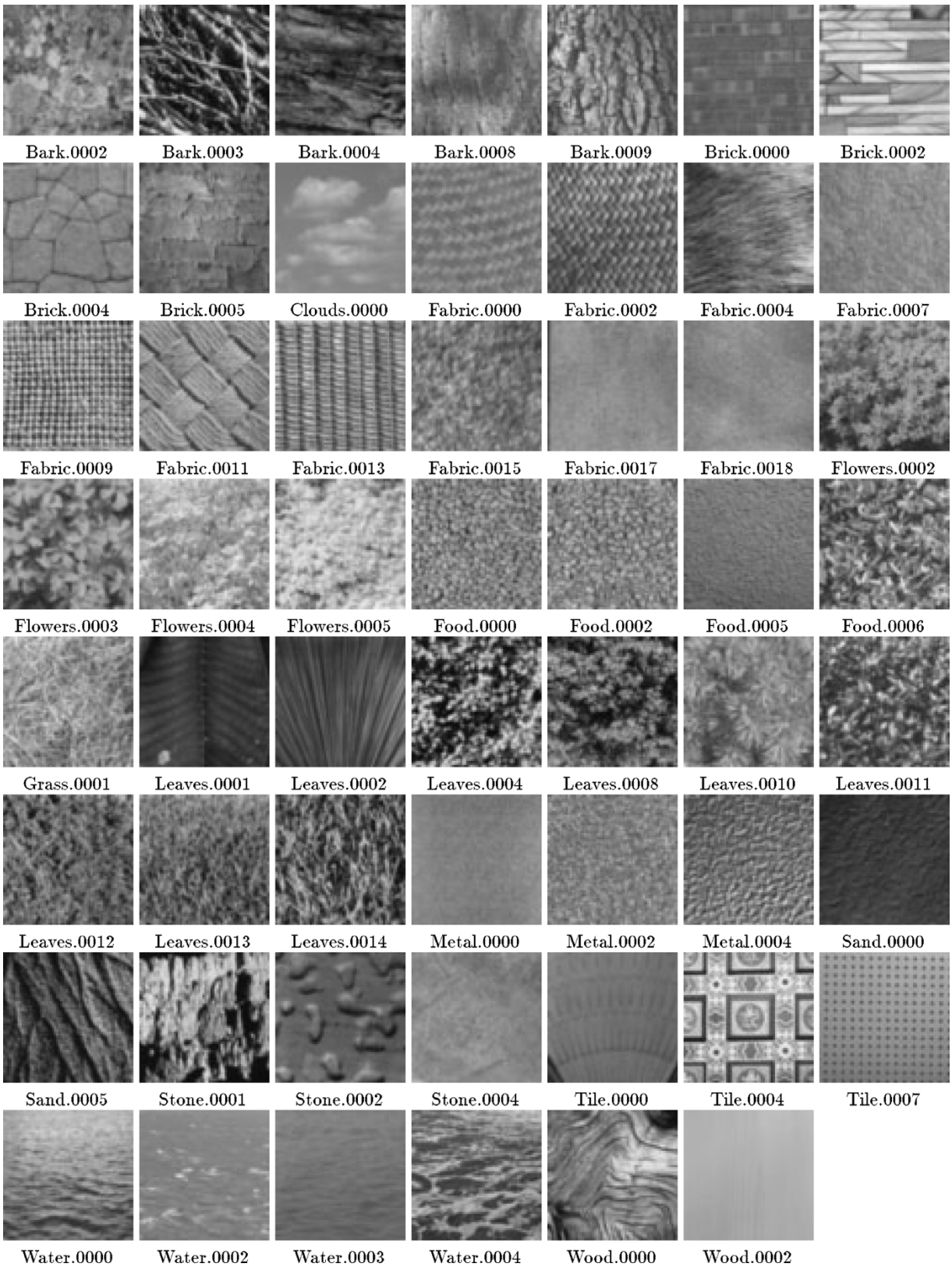


Fig. 5. VisTex texture image database used for retrieval experiments.

texture patches. The DWF analysis is performed up to five scale levels and the Laplacian distribution is assumed. In Fig. 4, the average percentage of relevant icons retrieved versus the number of top retrieved icons is graphically presented. In the same figure a comparison with the MRSAR method is given. As the Brodatz album is used in other published work for benchmarking, we present the *recall* measure for different methods as reported in the literature (see Table I). In this comparison, we retrieve a

number of images equal to the number of relevant subimages in the database. In our experiment this is eight because each image was divided into nine patches. We should note that there may be some differences in the data set, stemming from, for example, a different scanning procedure, or the implementation of the various methods. This global result shows that the proposed method yields very good results, while being more computationally tractable than the Wold model, the Gabor analysis, or the MRSAR model. Indeed, the Wold model needs at least and in average as many computations as the MRSAR model [14]. In [17], the Gabor analysis, the MRSAR model and wavelet transform methods are compared. They conclude that the more expensive for feature extraction is the MRSAR model, then comes the Gabor filter bank, and the less expensive is the wavelet analysis. As the wavelet frame analysis proposed here uses five-tag symmetric or antisymmetric filters for computing only 15 features per texture, we can claim that our method is less expensive.

The database used for the second retrieval experiment consists of 55 images of different texture classes derived from the MIT Media Laboratory VisTex texture database, shown in Fig. 5. The DWF algorithm is applied to analyze the images from the database in five scale levels. The variances σ_i^2 are then calculated in order to characterize each textured image, based on the previously described feature extraction algorithm.

Tables II and III present results on the retrieval accuracy of our method for the same database. Each of the 55 texture images provides 16 128×128 subimages, 880 subimages in total. Each image in the database is used once as a query prototype and the average recognition rate is computed for different numbers of the best retrieved images. The Bhattacharya distance is used for classifying and retrieving the subimages. Two hypotheses are considered for the probability distribution of the wavelet frame components: Gaussian and Laplacian. Table II shows detailed results on the recognition rate for assessing the efficiency of the method with respect to the image content. Table III shows global rates under the Laplacian assumption, knowing that the behavior on the individual images is very similar to that of the Gaussian distribution assumption. In addition, we have implemented a Gabor filter for the diagonal components for separating the two diagonal directions. A small improvement is achieved: for 15 retrieved images using the "diagonal" Gabor filter 77.2% were correct, in comparison with 76.9% with the Wavelet analysis.

We have also compared our method to the MRSAR method [18] which could be regarded as representing state-of-the-art modeling for texture classification. Two types of MRSAR are possible. The first is that proposed by Mao and Jain [18] for classification, using a Gaussian pyramid of the textured image in order to extract SAR model parameters at different resolutions. The second is a "pseudo" MRSAR that does not change the image scale, but enlarges the SAR model mask, capturing lower resolution characteristics. This type is used by Mao and Jain [18] for segmentation, and for classification in [14], [17]. A multiresolution SAR model with three resolution levels is tested, as well as a "pseudo"-MRSAR model implemented similarly to that in [14], using three resolutions and a second-order SAR mask. The retrieval performance operating characteristics are given in Fig. 6. Our method outperforms the MRSAR model-based method by roughly 5% for any number of retrieved images. However, this comparison indicates an average performance over the whole image database.

TABLE II
AVERAGE PERCENTAGE RECOGNITION RATE FOR ALL TEXTURE CLASSES VERSUS THE NUMBER OF THE TOP RETRIEVED SUBIMAGES UNDER THE GAUSSIAN HYPOTHESIS

Number of retrieved images	15	20	25	30	35	40	45
Bark.0002	68.8	80.0	85.4	91.2	96.7	97.9	99.2
Bark.0003	52.1	59.2	67.5	74.2	77.1	81.7	83.3
Bark.0004	77.5	85.4	87.9	90.0	92.1	92.9	94.2
Bark.0008	57.1	61.2	65.4	68.3	70.0	72.1	74.2
Bark.0009	37.5	40.8	42.9	45.8	47.1	49.2	51.2
Brick.0000	89.2	93.3	96.2	97.9	99.6	100.0	100.0
Brick.0002	74.2	77.9	81.2	82.1	83.3	85.8	87.5
Brick.0004	75.4	82.1	82.9	85.8	87.5	89.2	89.6
Brick.0005	91.7	97.5	97.9	99.2	99.2	99.6	100.0
Clouds.0000	92.5	96.2	97.9	98.3	99.2	99.6	99.6
Fabric.0000	78.3	81.7	84.2	87.9	90.4	91.7	92.9
Fabric.0002	85.0	87.9	89.2	90.0	91.2	92.1	93.3
Fabric.0004	75.4	78.8	80.4	84.2	85.8	86.7	88.8
Fabric.0007	97.1	97.9	99.2	100.0	100.0	100.0	100.0
Fabric.0009	98.3	99.2	100.0	100.0	100.0	100.0	100.0
Fabric.0011	75.0	84.6	90.8	93.8	95.4	97.9	98.8
Fabric.0013	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Fabric.0015	81.2	91.2	95.0	96.7	97.5	98.8	99.2
Fabric.0017	93.8	96.7	97.1	97.9	99.2	99.2	100.0
Fabric.0018	93.3	99.2	100.0	100.0	100.0	100.0	100.0
Flowers.0002	57.9	64.2	73.8	80.4	86.2	92.1	93.8
Flowers.0003	81.2	91.2	95.0	98.3	100.0	100.0	100.0
Flowers.0004	63.8	72.9	77.1	80.8	84.2	87.5	87.5
Flowers.0005	56.7	73.3	86.2	92.5	94.6	95.0	95.8
Food.0000	66.2	77.9	84.6	88.3	90.0	92.9	95.0
Food.0002	53.3	63.3	75.8	84.2	88.8	92.1	92.5
Food.0005	81.7	90.8	94.2	95.0	97.1	98.3	98.3
Food.0006	82.5	92.5	96.2	97.5	99.6	100.0	100.0
Grass.0001	97.1	99.6	100.0	100.0	100.0	100.0	100.0
Leaves.0001	66.7	74.6	81.2	85.4	88.3	88.8	90.4
Leaves.0002	87.9	89.6	89.6	90.8	92.1	92.9	93.3
Leaves.0004	82.9	94.6	97.9	99.6	99.6	100.0	100.0
Leaves.0008	70.4	84.2	90.8	94.6	96.7	97.9	98.3
Leaves.0010	55.0	61.7	68.8	74.2	76.7	79.2	81.2
Leaves.0011	53.8	60.0	64.6	69.2	72.1	76.2	80.4
Leaves.0012	67.5	80.0	90.4	92.9	95.8	97.1	97.9
Leaves.0013	59.6	62.5	65.4	66.2	68.8	70.4	72.9
Leaves.0014	72.9	81.7	85.4	89.2	90.8	92.1	92.5
Metal.0000	81.2	88.3	93.3	98.3	100.0	100.0	100.0
Metal.0002	98.3	100.0	100.0	100.0	100.0	100.0	100.0
Metal.0004	79.2	87.5	93.3	96.7	97.9	98.3	99.6
Sand.0000	97.5	99.2	99.6	100.0	100.0	100.0	100.0
Sand.0005	61.7	67.9	71.7	74.2	76.7	79.2	79.6
Stone.0001	62.5	73.3	81.2	88.3	92.9	95.4	96.7
Stone.0002	85.0	88.3	91.2	95.0	96.2	97.1	97.1
Stone.0004	86.2	92.1	93.3	95.0	95.8	97.5	97.5
Tile.0000	77.5	78.3	80.0	80.4	82.5	84.2	84.2
Tile.0004	99.2	100.0	100.0	100.0	100.0	100.0	100.0
Tile.0007	77.9	80.8	84.2	85.8	88.3	89.2	89.6
Water.0000	61.2	67.1	70.0	71.7	74.6	77.5	80.8
Water.0002	62.1	65.4	68.3	70.4	72.5	75.8	77.9
Water.0003	99.2	100.0	100.0	100.0	100.0	100.0	100.0
Water.0004	64.6	72.1	77.1	82.9	85.8	87.9	90.8
Wood.0000	32.5	40.4	45.8	47.9	50.0	52.9	55.4
Wood.0002	100.0	100.0	100.0	100.0	100.0	100.0	100.0
All	75.9	81.9	85.6	88.2	89.9	91.3	92.2

TABLE III
AVERAGE PERCENTAGE RECOGNITION RATE FOR ALL TEXTURE CLASSES VERSUS THE NUMBER OF THE TOP RETRIEVED SUBIMAGES UNDER THE LAPLACIAN HYPOTHESIS

Number of retrieved images	15	20	25	30	35	40	45
Retrieval performance	76.9	82.6	86.0	88.5	89.9	91.1	92.1

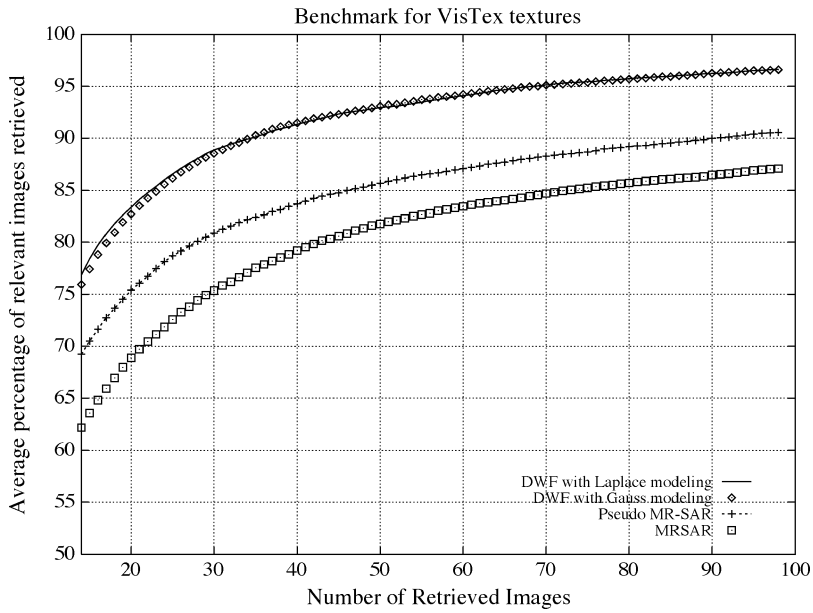


Fig. 6. Benchmark comparison of texture retrieval (VisTex images).

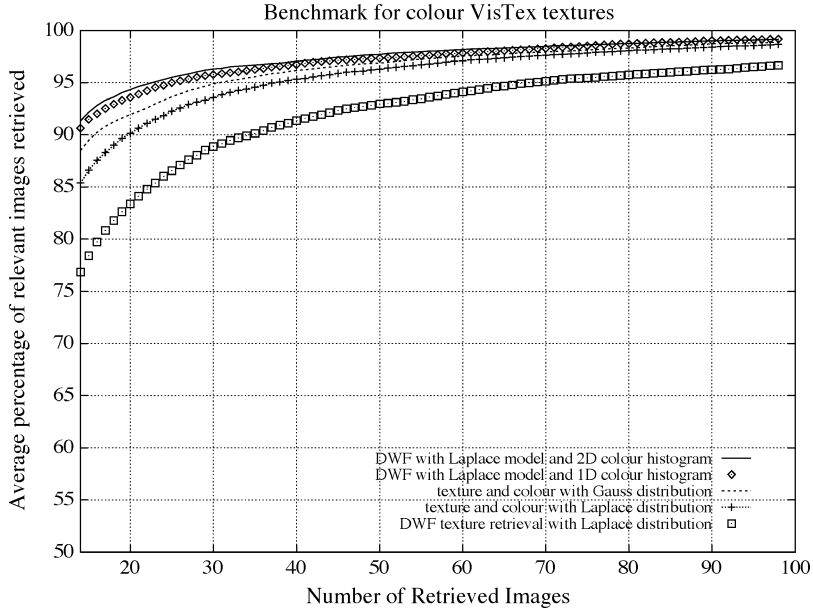


Fig. 7. Average percentage curves using the combination of texture and color features (VisTex data set).

There are important differences in performance on individual texture images. The multichannel approaches perform better on structured, quasiperiodic textures, while MRSAR modeling is in general much better on textures without a specific structure. In order to illustrate the influence of the texture pattern, we list the images for which the Wavelet approach performs much better: *Brick.0002, Brick.0004, Fabric.0000, Fabric.0002, Fabric.0011, Flowers.0003, Food.0006, Grass.0001, Leaves.0001, Leaves.0002, Leaves.0010, Leaves.0014, Tile.0000, Tile.0004, Wood.0002*; and the list of images where the auto-regressive modeling is much more efficient: *Bark.0009, Fabric.0015, Fabric.0018, Food.0005, Leaves.0004, Leaves.0011, Metal.0000, Metal.0004, Tile.0007, Water.0000, Water.0004*.

For the color texture retrieval benchmark, we use the same images. Fig. 7 shows the retrieval performance curve for all the combinations. For texture the DWF features are used; for color we use the 2-D histogram of (a, b), the two 1-D histograms of a, b, the parameters of a Gaussian and a Laplacian model. For the DWF analysis, five levels of decomposition were used, yielding a 15-dimensional feature vector. For texture features we used Laplace distribution modeling, which which experimental results show to perform better than the Gaussian assumption.

As expected, the 2-D histogram has the best performance, marginally outperforming even the 1-D histograms (91.3% versus 90.6%). The modeling of the histograms' distribution with Gauss and Laplace distributions provides good performance when combined with texture features, yielding 88.5% and 85.3% of correct classification respectively. In practice a, b 1-D histograms are close to the Gauss distribution in most cases, because most of the images are homogeneous which yields homogeneous chromaticities. Also, Gauss modeling is

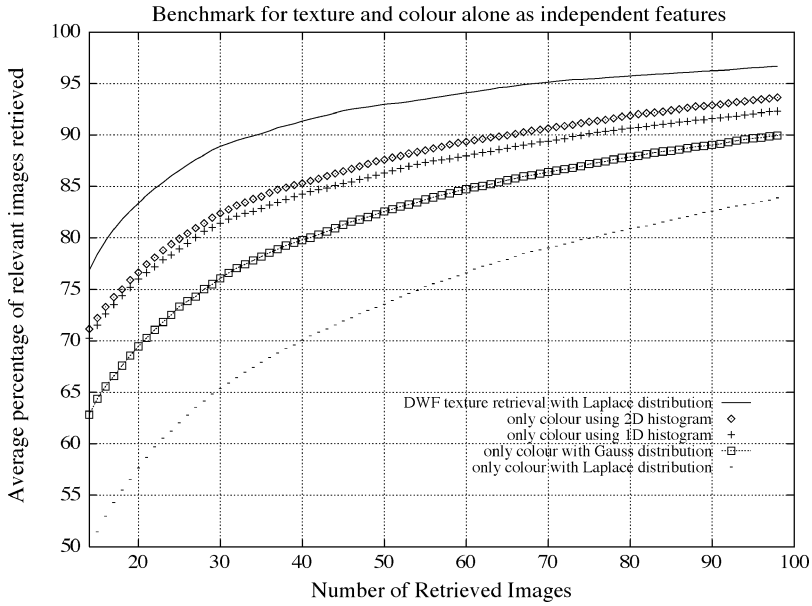


Fig. 8. Average percentage curves using only texture or color features (VisTex data set).

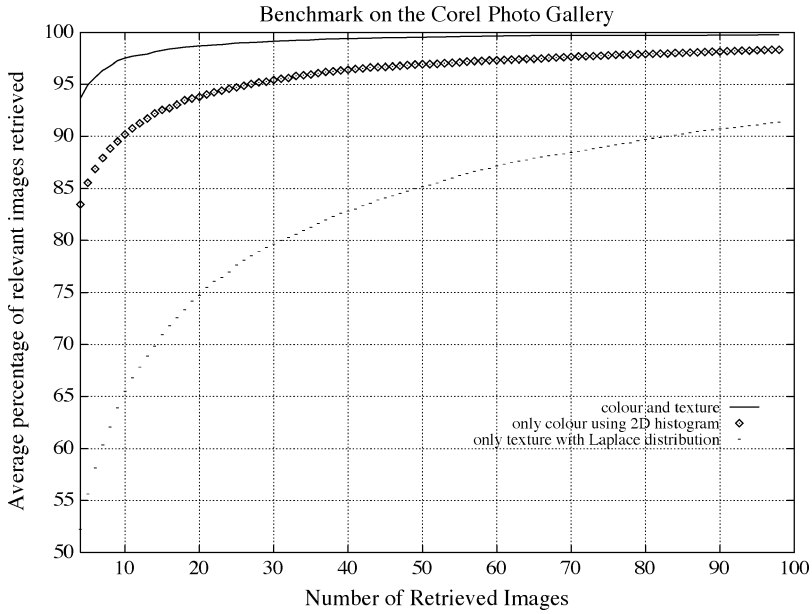


Fig. 9. Average percentage curves for the benchmark on the Corel Photo Gallery data set.

enforced by the *Lab* color system in which chromaticity a, b coordinates are concentrated in a small subrange of all possible values.

Fig. 8 shows the performance according to the benchmark using each texture or color feature alone. Texture features have the best performance with 76.8%. The color features follow with 2-D histograms, 1-D histograms, Gauss, and Laplace modeling giving 71.1%, 70.2%, 62.8%, and 49.8% respectively. Texture features result in the best performance because the data set used is intrinsically texture-oriented.

In Fig. 9, we present the results for the Corel Photo Gallery data set, comprised of 210 images of 384×256 pixels. As for the VisTex data set, subimages of size 128×128 are considered. The benchmark is defined in the same way, and the results show that the combination of texture and color features gives a percentage of correct classification of the first retrieved subimage equal to 93%. If texture features are used alone, the performance drops to 52%, and in the case of color features alone, 83%. In Fig. 10, we show the more heterogeneous images, with respect to either color or texture, for which the retrieval is less performant. Note that neither scale- nor rotation-invariance are addressed in this paper.

V. Conclusions

The Wavelet Frame Analysis decomposes images into different frequency bands, expressing them in various scales and orientations. For this reason, the components of the wavelet frames are particularly effective in exploiting periodicities in texture patterns. The Wavelet Frames Decomposition used in

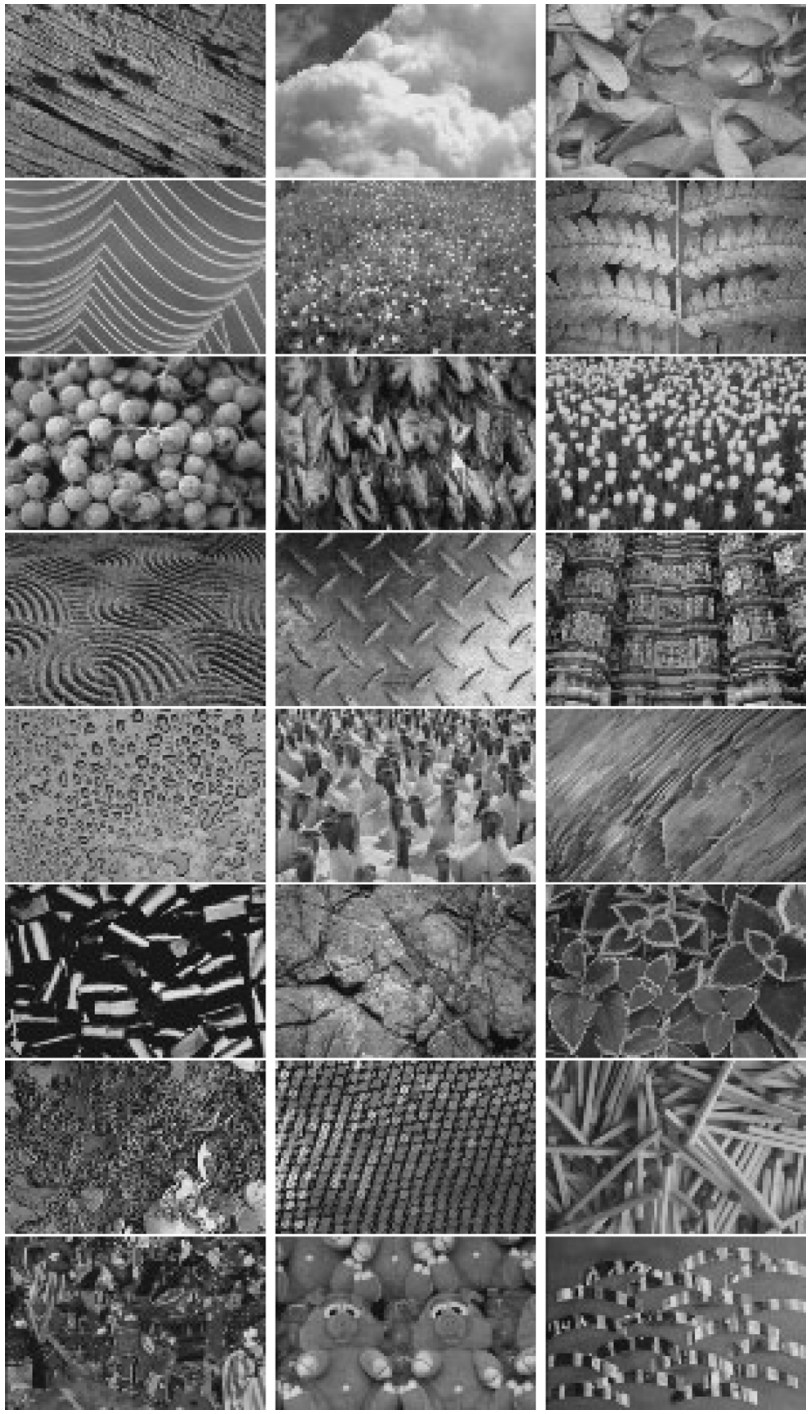


Fig. 10. Images with inhomogeneities from the Corel data set.

our work is extremely simple, implemented using five-tap 1-D filters. In the future, it will be interesting to find the best filter-bank for a given retrieval set. Color was described by the chromaticity distribution in the *CIE Lab* color space.

The proposed classification and retrieval method, based on the Bhattacharya distance as a dissimilarity measure between texture patterns and color distributions, was tested according to a standard retrieval benchmark on three relatively large sets of natural textures from the Brodatz collection, the VisTex database, and the Corel Photo Gallery. The results were very satisfactory for such a rich database, and in comparison with another well-known and effective method of random field modeling, the multiresolution autoregressive method, our method exhibits much better results for quasiperiodic texture patterns, while poorer results for "random" or "chaotic" patterns. One reason for this is that the MRSAR model is more appropriate for modeling random fields with continuous spectra. Other comparisons with other known methods show that our method is either equivalent or better in performance, and computationally less time-consuming. In particular, we obtain clearly better results than other methods. In addition, our method easily combines different features, wavelet-based as shown with texture and color.

REFERENCES

[1] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 55–73, Jan. 1990.

[2] P. Brodatz, *A Photographic Album for Artists and Designers*. New York: Dover, 1966.

[3] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. Image Processing*, vol. 2, pp. 429–441, Oct. 1993.

[4] M. Flickner, H. Sawhney, W. Niblack, and J. Ashley, "Query by image and video content: the QBIC system," *IEEE Computer*, vol. 28, no. 9, pp. 23–32, Sept. 1995.

[5] V. N. Gudivada and V. V. Raghavan, "Content based image retrieval systems," *IEEE Computer*, vol. 28, Sept. 1995.

[6] A. Gupta and R. Jain, "Visual information retrieval," *Commun. ACM*, vol. 40, no. 5, pp. 70–79, May 1997.

[7] R. M. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, pp. 610–621, Nov. 1973.

[8] A. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognit.*, vol. 24, pp. 1167–1186, 1991.

[9] R. L. Kashyap, R. Chellappa, and A. Khotanzad, "Texture classification using features derived form random field models," *Pattern Recognit. Lett.*, vol. 1, pp. 43–50, 1982.

[10] A. Kundu and J.-L. Chen, "Texture classification using qmf bank-based subband decomposition," *CVGIP: Graph. Model Image Process.*, vol. 54, pp. 369–384, Sept. 1992.

[11] MIT Media Laboratory. Vistex: Texture Image Database, Cambridge, MA. [Online] Available http://www.white.media.mit.edu/vismod/imagery/-VisionTexture/vistex.html

[12] A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 1186–1191, Nov. 1993.

[13] S. Liapis, N. Alvertos, and G. Tziritas, "Maximum likelihood texture classification and Bayesian texture segmentation using discrete wavelet frames," in *Proc. Int. Conf. Digital Signal Processing*, 1997, pp. 1107–1110.

[14] F. Liu and R. W. Picard, "Periodicity, directionality, and randomness: wold features for image modeling and retrieval," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 722–733, July 1996.

[15] S. G. Mallat, "A theory of multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, Jan. 1989.

[16] B. Manjunath, J.-R. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 703–715, June 2001.

[17] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 837–842, Aug. 1996.

[18] J. Mao and A. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Pattern Recognit.*, vol. 25, pp. 173–188, 1992.

[19] V. Ogle and M. Stonebraker, "Chabot: retrieval from a relational database of images," *IEEE Computer*, vol. 28, no. 9, pp. 40–48, Sep. 1995.

[20] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: Content-Based Manipulation of Image Databases,", Cambridge, MA, MIT Media Lab., Tech. Rep. 255, Nov. 1993.

[21] *Signal Processing: Image Communication, Special Issue on MPEG-7 Technology*, F. Pereira and P. Salembier, Eds., vol. 16, pp. 1–294, Sept. 2000.

[22] M. Porat and Y. Y. Zeevi, "Localized texture processing in vision: analysis and synthesis in Gaborian space," *IEEE Trans. Biomed. Eng.*, vol. 36, pp. 115–129, Jan. 1989.

[23] J. Puzicha, J. Buhmann, Y. Rubner, and C. Tomasi, "Empirical evaluation of dissimilarity measures for color and texture," in *Int. Conf. Computer Vision*, Sept. 1999 PAGES???.

[24] T. Randen and J. H. Husoy, "Filtering for texture classification: a comparative study," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 291–310, Apr. 1999.

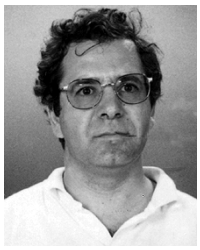
[25] T. Reed and J. M. H. du Buf, "A review of recent texture segmentation and feature extraction techniques," *CVGIP: Image Understand.*, vol. 57, pp. 359–372, May 1993.

[26] T. Sikora, "The MPEG-7 visual standard for content description—an overview," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 696–702, June 2001.

[27] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Processing*, vol. 4, pp. 1549–1560, Nov. 1995.

[28] *Handbook of Pattern Recognition and Image Processing*, T. Young and K.-S. Fu, Eds., Academic, New York, 1986.

Spyros Liapis received the B.Sc. degree in 1997 and the M.Sc. degree 1999, both in computer science, from the University of Crete, Heraklion, Greece. His research interests are in image analysis, indexing and retrieval.



Georgios Tziritas was born in Heraklion, Crete, Greece, on January 7, 1954. He received the Diploma of Electrical Engineering in 1977 from the Technical University of Athens, Greece, the Diplome d'Etudes Approfondies (DEA) in 1978, the Diplome de Docteur Ingenieur in 1981, and the Diplome de Docteur d'Etat inn 1985, all from the Institut Polytechnique de Grenoble, France.

From 1982 to 1992, he was a Researcher with the Centre National de la Recherche Scientifique, with the Centre d'Etudes des Phenomenes Aleatoires (CEPHAG) from September 1992 until August 1985, with the Institut National de Recherche en Informatique et Automatique (INRIA) from September 1985 until January 1987, and with the Laboratoire des Signaux et Systemes (LSS) from February 1987 until August 1992. From September 1992 until December 2002, he was Associate Professor at the Department of Computer Science, University of Crete, teaching digital signal processing, digital image processing, digital video processing, and information and coding theory. Since January 2003, he is Full Professor in the same department. He is coauthor, with C. Labit, of the book *Motion Analysis for Image Sequence Coding* (Amsterdam, The Netherlands: Elsevier, 1994), and of more than 70 journal and conference papers on signal and image processing, and image and video analysis. His research interests are in the areas of multimedia signal processing, image processing and analysis, computer vision, motion analysis, image and video indexing, and image and video communication.