Colour and Texture Image Retrieval using Chromaticity Histograms and Wavelet Frames

S. Liapis and G. Tziritas

Department of Computer Science, University of Crete,

P.O. Box 2208, Heraklion, Greece

E-mails: {*liapis, tziritas*}@*csd.uoc.gr*

Tel. +30-81-393517, Fax +30-81-393501

Abstract

This paper explores image retrieval mechanisms based on a combination of texture and colour features. Texture features are extracted using the Discrete Wavelet Frame analysis, which is an over-complete decomposition in scale and orientation. 2-D or 1-D histograms of the *CIE Lab* chromaticity coordinates are used as colour features. The 1-D histograms of the a, b coordinates were also modeled according to the generalized Gaussian distribution. The similarity measure is defined on the features distribution and 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 colour separately and in combination.

Index terms – Discrete Wavelet Frames, texture retrieval, colour image similarity, Bhattacharya distance

EDICS: 4-KEEP

I. Introduction

The current explosion in the generation rate of image archives necessitates the development of effective ways of managing (describing, indexing and retrieving) the 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 the image content, low level arithmetic features must be extracted that will be quantitatively comparable. The MPEG-7 working groups are seeking to define and standardize the image content description for automatic indexing.

Numerous features have been proposed and used to describe quantitatively the visual information, such as shape, colour, texture, motion, etc... [5]. Many image retrieval systems have indeed been developed using all or some of these features, including QBIC [4], Photobook [19], Chabot [18], Virage [6].

In this work texture and colour features are explored as a means of image content description. Image classification is performed according to global features for describing the texture and colour content for the whole image. It could be also possible to extract the same features for previously segmented objects.

Most approaches to classiffying textured images proposed during the past decades fall into three broad categories: statistical, model-based, and signal processing techniques [23], [22]. The statistical approaches, such as [7], were primarily investigated in the 80's. Model-based methods include among others Markov random field [9], simultaneous autoregressive [17], 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. Common to most signal processing methods is that the image is submitted to a filter bank, followng which local energy measurements are made. In this context the Gabor filter analysis could be used [20], [1], [8]. Having similar properties, but admitting a simpler implementation, wavelet transform representations could also be used. Pioneering work on texture classification using the wavelet packet representation is documented [3], [12]. Related to this approach is also the quadrature mirror (QMF) filter bank decomposition presented in [10].

Manjunath and Ma [16] studied certain aspects of image processing important for extracting texture features relevant to browsing and retrieval. They propose the use of Gabor wavelet features, where an adaptive filter selection algorithm is also proposed. The 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 more efficient in terms of computational complexity was the pyramid-structured wavelet transform.

Liu and Picard [14] proposed the use of a Wold decomposition for extracting perceptual features, described as periodicity, directionality and randomness. The performance of the Wold model is compared with other methods: the multiresolution simultaneous autoregressive model, the principal component analysis, the tree-structured wavelet transform, and the Tamura features. The experiments were performed over the Brodatz database, and the Wold model was shown to be the best, 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 aim of the analysis is to determine characteristics corresponding to each texture type, so that each texture pattern is uniquely defined. Such a distinction takes place in the frequency domain, where the input image is equivalently decomposed to different scale levels using the Discrete Wavelet Frames (DWF). Once these characteristics are deduced, statistical methods are applied to extract the features which are necessary for the description and the classification of the texture pattern. Although the philosophy of this approach has been introduced in the past [24], 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 colour features the *CIE Lab* colour system was chosen, which is designed to be perceptually uniform. Only the chromaticity coordinates (a, b) are used to describe colour. In general, colour content is best described by the chromaticity distribution which is given by 1-D or 2-D histograms. The computational complexity is reduced if Gaussian or Laplacian models are assumed for these distributions.

In order to compare texture and colour 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] which is a collection of various natural grayscale textures. The second data set is obtained from the MIT Media Laboratory VisTex database [11], which contains images of scenes of physical colour textures. The third is obtained from the Corel Photo Gallery.

The presented work is organized as follows. In Section II, the underlying theory of the basic filters, the necessary decomposition by upsampling and the use of Discrete Wavelet Frames, as applied to 2-D signals, are described. Next, in Section III, the texture and colour 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 fundamental 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, hence simplifying the classification problem. A low-pass filter $H(z)$ and its conjugate quadrature high-pass $G(z)$ form the pair of prototype filters for generating the whole filter bank by upsampling, so that the whole range of bands is covered. Cubic splines having appropriate scaling properties could be used for designing the pair of prototype filters [24]. Here the fourth-order binomial filter and its conjugate quadrature filter are used,

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

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

in the frequency domain, and

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

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 $i = 0, \ldots, I$,

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

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

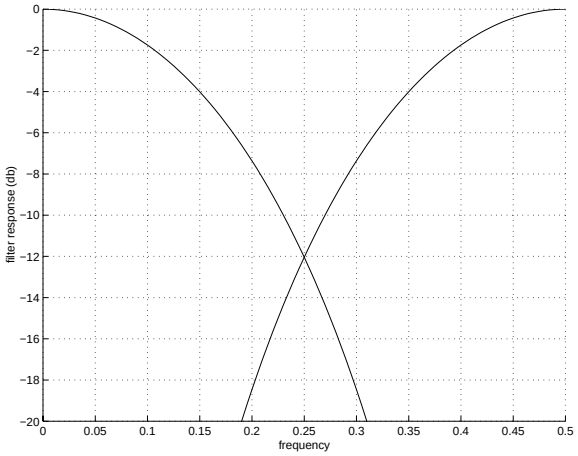

Fig. 1. The frequency responses of the pair of conjugate filters

where $H_0(z) = 1$.

Therefore, the input signal can be decomposed into wavelet coefficients corresponding to different scales. Thus the Discrete Wavelet Frames (DWF) could 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 of Equation (3) is selected because of its good orthogonalization properties.

We present here two criteria for band separation by the filter pair. The antialiasing coefficient measures the energy concentration in the low-frequency band by the low-pass filter

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

where $\mathcal{H}(f)$ is the Fourier transform of the low-pass filter H. For the above filter we have $\eta = 0.985$. In addition we use the correlation coefficient between the approximation and the details signals 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 sufficiently weak for considering 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, 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. 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 [22] 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 evidence 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 [22].

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 Discrete Wavelet Frames analysis because of the above separability property, which parmits 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:

$$y(m,n) = [y_1(m,n), \ldots, y_{N-1}(m,n), y_N(m,n)] \tag{6}$$

where each element of $y(m,n)$ has been determined according to 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 in each resolution level three new feature channels are obtained, which characterize for the given scale depth the horizontal, the vertical and

the diagonal directions. The first analysis layer corresponds to high frequencies, while the higher-order layers represent represent successively lower 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 (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 a texture is better described through the higher frequency channels than through the approximation component. The first reason for this is that a mean-invariant classifier is required, and secondly, the approximation component might not be stationary, with a shift-variant variance. In other words 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 σ_i^2 $(i = 1, \ldots, N-1)$. Indeed, 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. Colour features

In order to characterize the colour content of an image the *CIE Lab* colour space is used. The *Lab* colour coordinate system has the advantage that it is designed to be perceptually uniform, meaning that the same distance in the colour space leads to equal human colour difference perception, *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 colour 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 ($[-137, 96], [-99, 133]$ for (a, b) which yields 54056 bins). The number of chromaticities is so large because 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 colours 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 colour feature could use the 232 and 233 bins of the (a, b) histograms respectively.

In order to reduce the number of the colour features we could assume a model for each coordinate distribution. In our case the Gaussian and Laplacian distribution are used as models. This model requires only the mean value and the variance of the image's colour coordinates. The storage demands are minimized and the comparison of colour 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 colour and texture is given in [21]. 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 of 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 Bhattacharya distance for retrieving images by colour and texture. In this paragraph we present both, and in benchmarking we use the latter.

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,

$$d_j^L(y) = \sum_{(m,n)} \sum_{i=1}^{N-1} \left(\frac{y_i^2(m,n)}{\sigma_{i,j}^2} + \ln \sigma_{i,j}^2 \right) \propto \sum_{i=1}^{N-1} \left(\frac{S_i^2}{\sigma_{i,j}^2} + \ln \sigma_{i,j}^2 \right) \tag{7}$$

where the first sum is taken over all image points, S_i^2 is the variance of the i component of 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 two-classes case, and because the two texture patterns, the

one of the data set and the one tested, are considered in an symmetric way. The Bhattacharya distance is in fact a special case of the Chernoff bound of the Bayes error [25]. It is well known that the Chernoff information gives the highest achievable exponent for the error probability. The definition of the Bhattacharya distance is

$$d_B(p_1, p_2) = -\ln\left(\int_x \sqrt{p_1(x)p_2(x)}dx\right) \tag{8}$$

where p_1, p_2 are probability density functions of vector x of any dimension. The Bhattacharya distance satisfies the symmetric property, ($\mathrm{d}(p_1,\ p_2) = \mathrm{d}(p_2,\ p_1)$). The triangular property is satisfied only for specific configurations.

In our case this distance could be defined on empirical probability distributions. The discrete expression is

$$d_B(h_1, h_2) = -\ln\left(\sum_i \sqrt{h_1(i)h_2(i)}\right) \tag{9}$$

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 feature's 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 the generalized Gaussian [15]

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

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

$$d_j^B(y) = \frac{1}{c}\sum_{i=1}^{N-1} \ln \frac{S_i^c + \sigma_{i,j}^c}{2\sqrt{S_i^c \sigma_{i,j}^c}} \ . \tag{10}$$

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

When texture features and colour 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 of the distance, because the decomposition is approximately orthogonal.

IV. Retrieval benchmark

In order to assess the capabilities of the texture and colour 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 database of icons is thus obtained with a large number of items. Each small icon in the database is used to retrieve from the database the nearest (more similar) icons, except itself. The similarity between two icons is determined using as measure the Bhattacharya distance described in the previous section. For each number of retrieved icons, we record the *recall*, that is the number of relevant images relative to the total number of relevant images in the database. This result is presented graphically in a hit rate curve versus the number of retrieved images. It is obvious that 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 Figure 2 for the relatively heterogenous *D13* Brodatz image. In this example for 8 retrieced icons the *recall* is 50 %, and for 17 it is 75 %.

The first database used is the entire Brodatz collection containing 112 grayscale texture pictures. Each of them is subdivided into nine non-overlapping subimages, resulting in 1008 texture patches. The DWF analysis is performed up to five scale levels. In Figure 3 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 (Tab. I). The number of the retrieved images is equal to the number of expected relevant subimages. As in our experiments the number of subimages per texture picture is equal to nine, the number of retrieved subimages is eight. We should note that there may be some differences in the data set, as for example a different scanning procedure, or in the implementation of the various methods. This global result shows that the proposed method yields very good results, while being much simpler in computations in comparison with the Wold model, the Gabor analysis and the MRSAR model.

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 Figure 4. 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

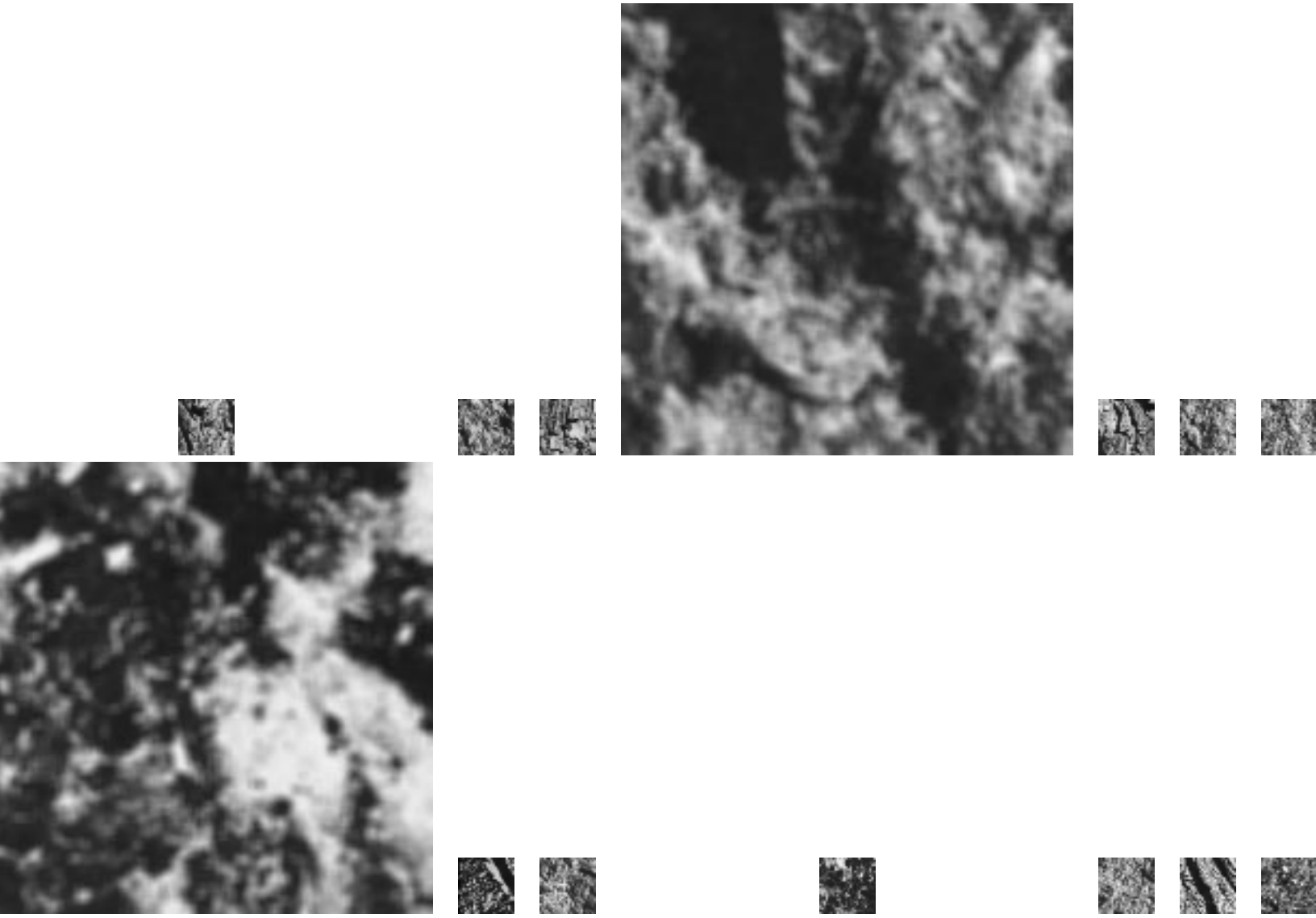

Fig. 2. A retrieval example: the top-left icon is the query picture

Method	Retrieval performance
Wold [14]	75%
DWF (our result)	73.4%
Gabor [16]	73.1%
MRSAR [16]	72.0%
MRSAR [14]	70 %
TWT [16]	68.0%
PWT [16]	67.4%

TABLE I

The average recognition rate for the Brodatz collection and for different methods and implementations

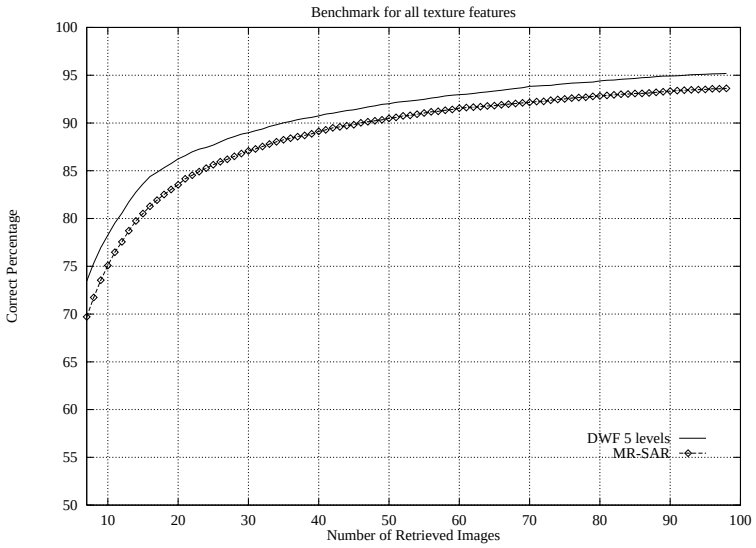

Fig. 3. Correct percentage curves over the Brodatz database

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 sixteen 128 $\times$ 128 subimages, 880 subimages in total. The average recognition rate is computed for different numbers of the top retrieved images. Each image in the database is used once as query prototype. 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 behaviour 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: in the first 15 retrieved images 77.2% were correct, in comparison with 76.9% of the Wavelet analysis.

We have also compared our method to the MR-SAR method [17] which could be regarded as representative of the state-of-the-art modeling for texture classification. Two types of MR-SAR are used. The first is the one proposed by Mao for classification, and uses a Gaussian pyramid of the textured image in order to extract SAR model parameters at different resolutions. The second type is a "pseudo" MR-SAR 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 [17]

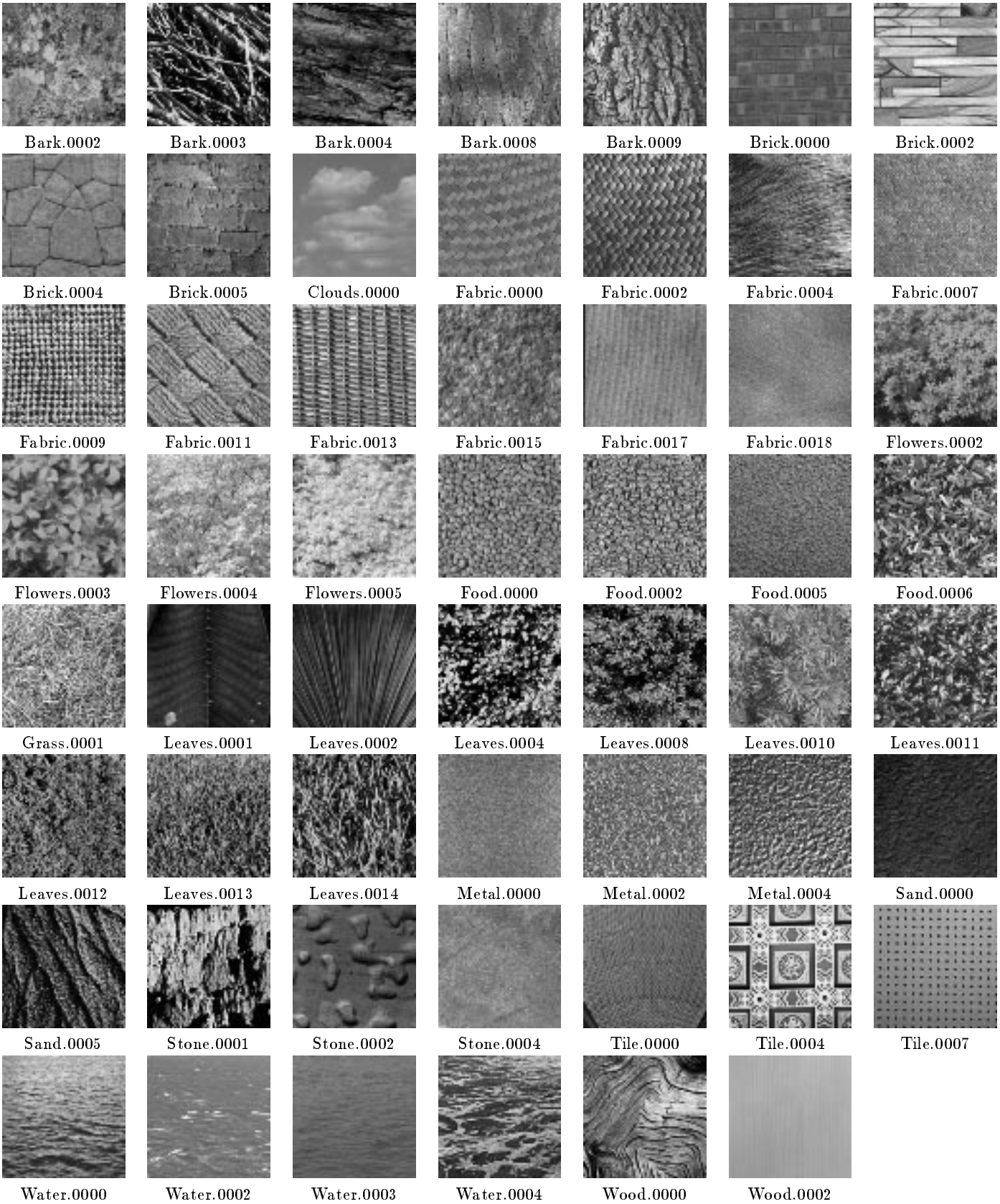

Fig. 4. The VisTex texture image database used for retrieval experiments

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 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
Retrieval performance	76.9	82.6	86.0	88.5	89.9	91.1	92.1

TABLE III

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

for segmentation, and for classification in [14], [16]. The implementation is like the one described by Picard, using three resolutions and second order SAR mask. A multiresolution SAR model with 3 resolution levels is tested, as well as a pseudo MR-SAR model. The retrieval performance operating characteristics are given in Figure 5. Our method outperforms the MR-SAR model-based method of about 5% for any number of retrieved images. However, this comparison indicates an average performance over the whole image database. There exists an important dispersion on the differential performance on the individual texture images. The multichannel approaches perform better on structured, quasi-periodic textures, while MR-SAR modeling could be much better on textures without specific structure. In order to illustrate the influence of the texture pattern we give the list of 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 colour texture retrieval benchmark we use the same images. Figure 6 shows the retrieval performance curve for all the combinations. For texture the DWF features are used. For colour we use the 2-D histogram of (a, b), the two 1-D histograms of a, b respectively, the parameters of a Gaussian and a Laplacian model. For the DWF analysis 5 levels of decomposition were used, yielding a 15-dimensional feature vector. Also Laplace distribution modeling was used for texture features, for which experimental results show to perform better than the Gaussian assumption.

As expected the 2-D histogram has the best performance, even with small difference from the performance using 1-D histograms (91.3% against 90.6%). The modeling of the histograms distribution with Gauss and Laplace distribution provide good performance when combined with texture features yielding 88.5% and 85.3% of correct classification respectively. In practice a, b

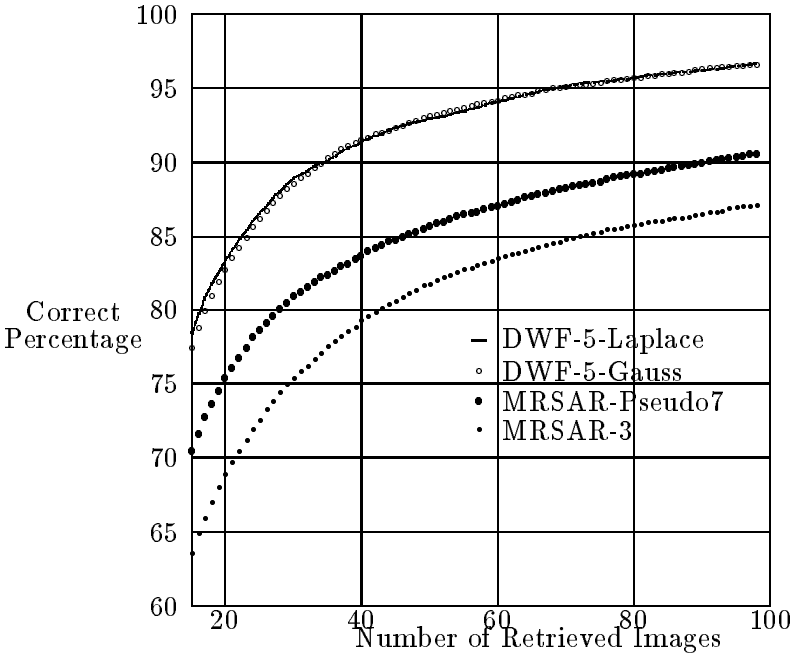

Fig. 5. Benchmark comparison of texture retrieval (VisTex images)

1-D histograms are close to the Gauss distribution in most cases. This is because most of the images are homogeneous which yields homogeneous chromaticities. Also Gauss modeling is enforced from *Lab* colour system which has chromaticity a, b coordinates very compact in a small range of all possible values.

Figure 7 shows the performance according to the benchmark using each texture or colour feature alone. Texture features have the best performance with 76.8%. Then the colour features follow, 2-D histograms, 1-D histograms, Gauss and Laplace modeling with 71.1%, 70.2%, 62.8%, 49.8% respectively. Texture features result in the best performance because the data set is texture-oriented.

In Figure 8 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 128×128 subimages are considered. The benchmark is defined in the same way, and the results show that the combination of texture and colour 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 80%, and in the case of colour features alone 75%. In Figure 9 we give the more heterogeneous, in either colour or texture, images, for which the retrieval is less performant. Note that scale- or rotation-invariance are not

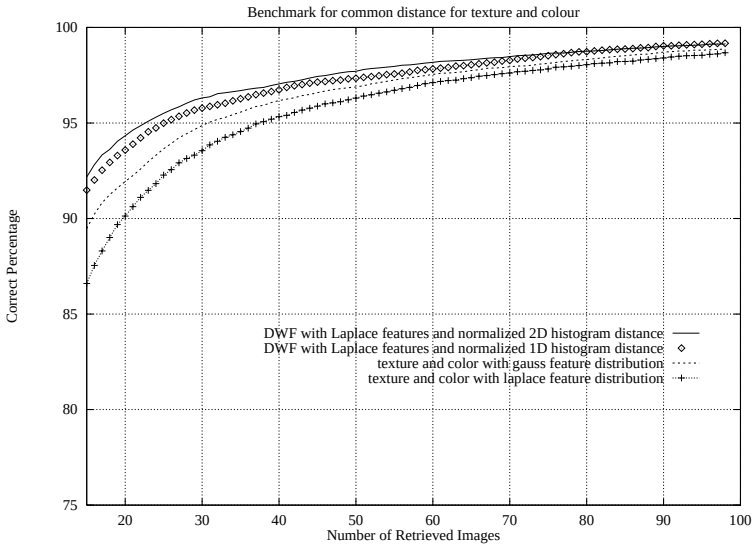

Fig. 6. Correct percentage curves using the combination of texture and colour features

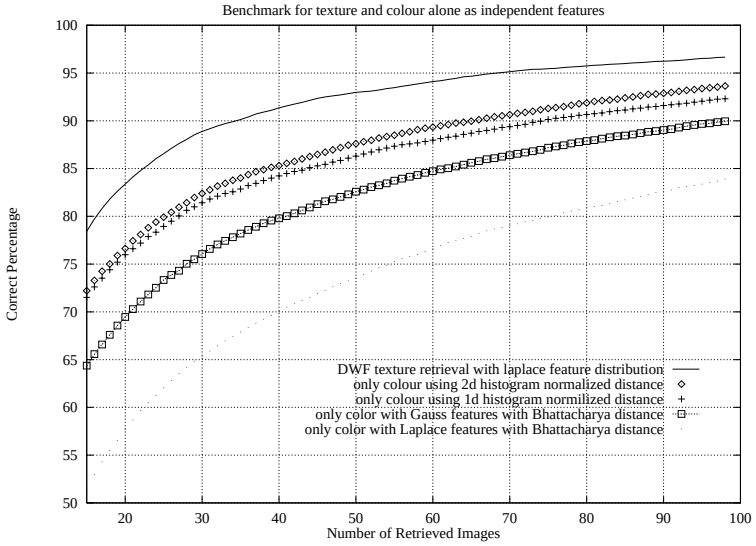

Fig. 7. Correct percentage curves using only texture or colour features

addressed in this paper.

V. Conclusion

The Wavelet Frame Analysis decomposes images into different frequency bands, expressing them into 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 our work is extremely simple, implemented using 5-tap 1-D filters. In

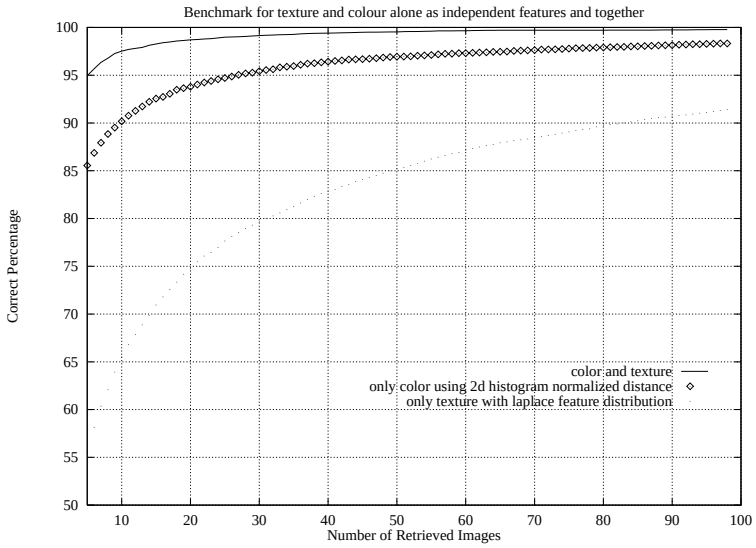

Fig. 8. Correct percentage curves for the benchmark on the Corel Photo Gallery data set

the future it will be interesting to find the best filter-bank for a given retrieval set. Colour was described by the chromaticity distribution in the *CIE Lab* colour space.

The proposed classification and retrieval method, based on the Bhattacharya distance as a dissimilarity measure between texture patterns and colour 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. These data sets, consisting of pure texture and coloured textures, are essentially homogeneous in texture and more spatially variant in colour. As expected texture features were shown to be more relevant for image classification and retrieval. 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 quasi-periodic texture patterns, and poorer results for "random" or "chaotic" patterns. Other comparisons with other known methods show that our method is either equivalent or better in performance, and computationally less time-consuming. In particular in comparison with other wavelet-based methods we obtain clearly better results. In addition, our method easily combines different features, as shown with texture and colour.

References

[1] A.C. Bovic, M. Clark, and W.S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Trans. Patt. Anal. Machine Intell.*, 12:55–73, January 1990.

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

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

[4] M. Flickner, H. Sawhney, W. Niblack, and J Ashley. Query by image and video content: the qbic system. *IEEE Computer*, 28(9):23–32, Sep. 1995.

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

[6] A. Gupta and R. Jain. Visual information retrieval. *Communicastions of the A.C.M.*, 40(5):70–79, May 1997.

[7] R. M. Haralick, K. Shanmugan, and I. Dinstein. Textural features for image classification. *IEEE Trans. Syst., Man, Cybern.*, 8:610–621, Nov. 1973.

[8] A. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24:1167–1186, 1991.

[9] R.L. Kashyap, R. Chellappa, and A. Khotanzad. Texture classification using features derived form random field models. *Pattern Recognition Letters*, 1:43–50, 1982.

[10] A. Kundu and J.-L. Chen. Texture classification using qmf bank-based subband decomposition. *CVGIP: Graphical Models and Image Processing*, 54:369–384, Sept. 1992.

[11] MIT Media Laboratory. Vistex: Texture image database. *http://wwwwhite.media.mit.edu/vismod/imagery/-VisionTexture/vistex.html.*

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

[13] S. Liapis, N. Alvertos, and G. Tziritas. Maximum likelihood texture classification and bayesian texture segmentation using discrete wavelet frames. *Intern. Conference on Digital Signal Processing*, pages 1107–1110, 1997.

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

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

[16] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. Patt. Anal. Machine Intell.*, 18:837–842, Aug. 1996.

[17] J. Mao and A. Jain. Texture classification and segmentation using multiresolution simulttaneous autoregressive models. *Pattern Recognition*, 25:173–188, 1992.

[18] V. Ogle and M. Stonebraker. Chabot: Retrieval from a relational database of images. *IEEE Computer*, 28(9):40–48, Sep. 1995.

[19] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *M.I.T. Media Lab. Tech. Rep. No. 255*, Nov. 1993.

[20] 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.

[21] J. Puzicha, J. Buhmann, Y. Rubner, and C. Tomasi. Empirical evaluation of dissimilarity measures for color and texture. *Intern. Conf. on Computer Vision*, Sep. 1999.

[22] T. Randen and J. H. Husoy. Filtering for texture classification: a comparative study. *IEEE Trans. Patt. Anal. Machine Intell.*, 21:291–310, April 1999.

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

[24] M. Unser. Texture classification and segmentation using wavelet frames. *IEEE Trans. on Image Processing*, 4:1549–1560, November 1995.

[25] T. Young and K.-S. Fu eds. *Handbook of pattern recognition and image processing*. Academic Press, 1986.

Fig. 9. Images with inhomogeneities from the Corel data set