Image retrieval by colour and texture using chromaticity histograms and wavelet frames

Spyros Liapis
Georgios Tziritas
Computer Science Department, University of Crete, Heraklion, Greece,
{liapis, tziritas}@csd.uch.gr

Abstract

In this paper the combination of texture and colour features is used for image classification. Texture features are extracted using the Discrete Wavelet Frame analysis. 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 defined on the features distribution is based on the Bhattacharya distance. Retrieval benchmarking is performed on textured colour images from natural scenes, obtained from the VisTex database of MIT Media Laboratory and from the Corel Photo Gallery.

1 Introduction

The current explosion in the generation rate of image archives necessitates the development of effective ways of managing (describing, indexing and retrieving) visual information by its content [2], 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 aimed to define and standardize the image content description for automatic indexing.

Numerous features were proposed and used to describe quantitavily the visual information, like shape, colour, texture, motion etc... [2]. Also a lot of image retrieval systems were developed using all or some of these features, like QBIC [1], Photobook [9], Chabot [8], Virage [3].

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

In this paper, for texture feature extraction the Discrete Wavelet Frames (DWF) analysis is used [11] [5]. Texture characterization is obtained from spatial frequency decomposition into distinctive bands that differ in scale and orientation.

For colour features the *CIE Lab* colour system is 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 could be 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 [6]. Two data sets are considered. The first data set is obtained from the MIT Media Laboratory VisTex database [4], which contains images of scenes of physical colour textures. The second one is obtained from the Corel Photo Gallery.

2 Texture feature extraction

Texture analysis is performed with the use of Discrete Wavelet Frames. The aim of the analysis is to determine characteristics corresponding to each texture pattern, 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. The decomposition is performed with multichannel filtering. For this purpose 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 with a factor of 2, so that the whole range of bands is covered. The fourth-order binomial filter and its conjugate quadrature filter are used,

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

in the frequency domain. In addition, the generated filters can form orthogonal wavelet base functions [7], so the input signal can be decomposed into discrete wavelet frame coefficients, each corresponding to a different frequency layer. The previous decomposition can be extended to 2-D signals (images), by forming wavelet bases which result from the cross product of separable bases in each direction. These four base functions deduce the following decomposition algorithm:

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

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

The previous analysis can be applied to texture images, yielding the following representative vector:

$$y(k,l) =< y_1(k,l), \ldots, y_{N-1}(k,l), y_N(k,l) > \tag{3}$$

where each element of $y(k,l)$ has been determined according to the analysis in (2) and the dimension of the vector is $N = 3I+1$, composed of $3I$ detail components and the approximation at level I component.

The texture content is then characterized by the variances σ_i^2 of the $N-1$ detail components of the representative vector $(i = 1, \ldots, N-1)$. This characterization is based on the fact that the mean value of the details is zero, because $\mathrm{G(z)}|_{z=1} = 0$, and the different components are uncorrelated, because the values of the covariance matrix except the diagonals are practically zero. In addition, the components of vector $y(k,l)$ could be assumed according to the generalized Gaussian distribution.

The main advantage of this analysis is that the coefficients are computed in a separable way, which makes it no computational expensive. Also DWF decomposition provides good feature localization. Each point has a representative vector of DWF coefficients, because the scale of input signal does not change, in contrast with Discrete Wavelet Transform [7].

3 Color 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. It also has the advantage that lightness L is distinct and independent from the other coordinates (a, b), which are the chromaticity coordinates. For colour image classification and retrieval it is more relevant to compare the chromaticity distribution of an image, disregarding the lightness component, *i.e.*, images which are perceptually similar have the same chromaticity components. 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 is 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 has the disadvantage that needs 1024 floating point numbers for storage for 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, which require 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.

4 Dissimilarity measure

Measuring the dissimilarity between images is of central importance for retrieving images by content. Some different dissimilarity measures for colour and texture were empirically evaluated in [10]. In our work another dissimilarity measure, the Bhattacharya distance, was used in order to compare the extracted features and measure their dissimilarity. The definition of the Bhattacharya distance is

$$d_B(p_1, p_2) = -\ln\left(\int_x \sqrt{p_1(x)p_2(x)}dx\right) \tag{4}$$

where p_1, p_2 probability density functions of vector x of any dimension. This measure has the advantage that is designed to compare features for the two classes case. It is a special case of the Chernoff bound of the error probability in binary classification [12]. It is well known that the Chernoff information gives the highest achievable exponent for the error probability. The Bhattacharya distance has the symmetric property, ($\mathrm{d}(p_1,\, p_2) = \mathrm{d}(p_2,\, p_1)$). The triangle property is only satisfied for specific configurations.

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

$$d_B(h_1, h_2) = -\ln\left(\sum_i \sqrt{h_1(i)h_2(i)}\right) \tag{5}$$

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

In the case that we have a model for the histogram's distribution, a simpler expression of the Bhattacharya distance can be deduced. In this work we assume that some features might follow the generalized Gaussian distribution

$$p(y) = \frac{c}{2\sigma\Gamma(\frac{1}{c})}e^{-(\frac{|y-\mu|}{\sigma})^c} \tag{6}$$

where the parameter σ is directly related to the variance, and c with the sharpness of the probability density function. For $c = 2$ we have the Gaussian and for $c = 1$ the Laplacian distribution.

For example generalized Gaussian distribution is suitable for DWF coefficients [7]. Also we assume that each feature is uncorrelated to each other (*e.g.* for DWF coefficient which is practically true). The simplified expression assuming generalized Gaussian distribution and uncorrelated features is

$$d_B^{1,2} = \frac{1}{c}\sum_{i=1}^{N} \ln \frac{\sigma_{i,1}^c + \sigma_{i,2}^c}{2\sqrt{\sigma_{i,1}^c \sigma_{i,2}^c}} + \frac{1}{2c}\sum_{i=1}^{N} \frac{|\mu_{i,1} - \mu_{i,2}|^c}{\sigma_{i,1}^c + \sigma_{i,2}^c} \tag{7}$$

where N is the dimension of the feature vector and the parameters σ_1^c and σ_2^c are estimated from the data. In this work values $c = 2$ (Gauss) or $c = 1$ (Laplace) are used. For the texture features mean values are zero because the high-pass filters have coefficients with zero sum, which results in omitting the second term in formula (7). On the other hand for colour features both terms are used, because mean colour values, obviously, are not zero.

When texture features (variances) and colour histogram features (1-D or 2-D a, b histograms) need to be combined, the simpler expression (7) is used for texture features and the initial discrete expression (5) is used for histograms. The combined distance formula is formed by the independent sumation of the distance expression for each feature. This holds because all terms are depicted from the same initial expression and because features are assumed uncorrelated.

5 Benchmark

In order to exploit the capabilities of the texture and colour features a retrieval benchmark was performed [6]. The purpose of this classification experiment is to find out if the image features overcome the images inhomogeneities.

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 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 with the distance measure 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. 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 *recall* rate is increasing.

We performed this experiment on a data set obtained from VisTex database of the Media Laboratory in MIT. From this database of homogeneous colour textures from natural scenes were chosen 55 images (512×512). These images contain wood, bark, food, sand, flowers, trees, tiles, fabric and other. In order to perform the retrieval experiment they were cut to 16 icons 128×128 each, yielding 880 icons.

The benchmarking experiment with this data set was performed with all the texture and colour features. Figure 1 shows the classification curve for all the combinations. For texture the DWF features are used. For colour are used 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 the levels of decomposition were 5, yielding 15 dimension feature vector. Also Laplace distribution modeling was used for texture features, because after experimental results has better performance than assuming Gauss.

As expected the 2-D histogram has the best performance, even with small difference from 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 1-D histograms are close to Gauss distribution in most

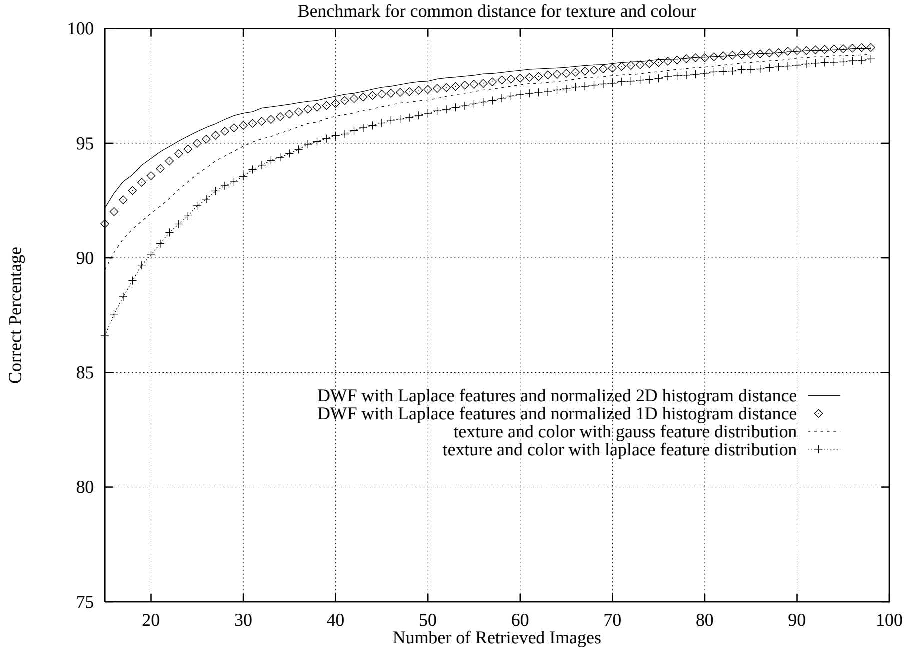

Figure 1: Correct percentage curves using the combination of texture and colour features, for VisTex database

of the 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 2 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 3 are presented the results for the Corel Photo Gallery data set. The tested data set contains 350 images of 384×256 pixels. As for the VisTex data set, 128×128 subimages are considered. The total number of subimages is therefore 2100 belonging to 350 classes. Among these classes there are some similar in colour or in texture. Retrieving by only colour or only texture might give ambiguous results. The combination of both colour and texture gives much better classification rates. The benchmark is defined in the same way, as for the VisTex data set, and the results show that the combination of texture and colour features gives a percentage of correct classification of the five first retrieved subimages equal to 93.6%. If only texture features are used the performance becomes 52.2%, and in the case of only colour features 83.5%. In Figure 4 are given the more inhomogeneous, in either colour or texture, images, for which the retrieval is less performant.

6 Summary

In this paper we presented texture and colour feature extraction methods. The Discrete Wavelet Frames analysis provides the texture features, which are the variances of the sub-bands. Color was described by the chromaticity distribution. These features were combined using a common distance measure, the Bhattacharya distance. The performance of the proposed image classification method was tested using

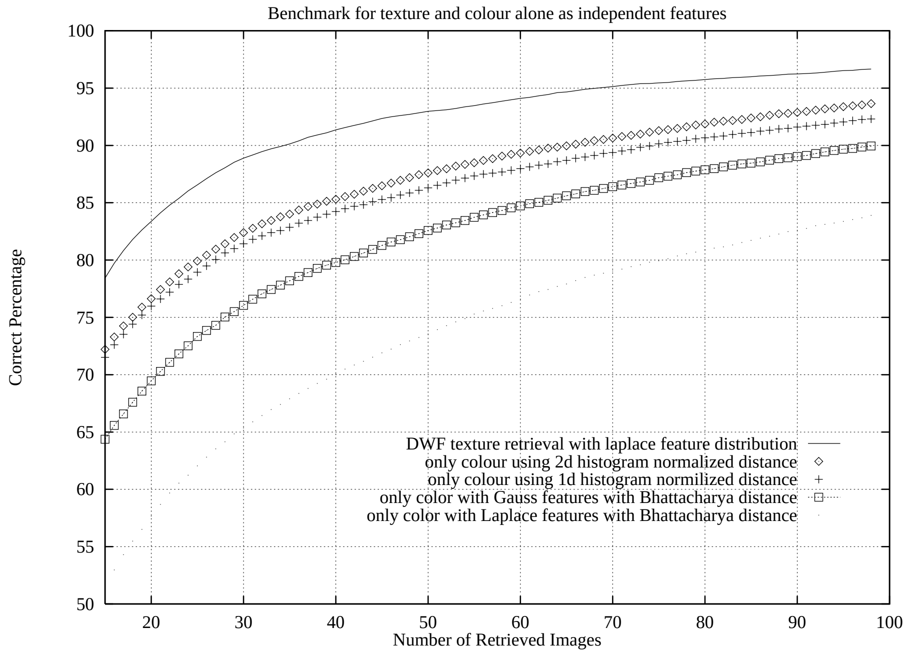

Figure 2: Correct percentage curves using only texture or colour features, for VisTex database

a retrieval benchmark, where the performance is defined by the percentage of the correct for a given number of retrieved images. The data set was from nearly homogeneous natural colour textures from VisTex database and from Corel photo gallery.

7 Aknowledgments

This work was funded in part by the PENED99 project, of the Greek General Secretariat of Research and Technology.

References

[1] 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.

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

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

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

[5] S. Liapis, N. Alvertos, and G. Tziritas. Maximum likelihood texture classification and bayesian texture segmentation using discrete wavelet frames. *Intern. Conf. in Digital Signal Processing DSP97*, 2:1107–1110, July 1997.

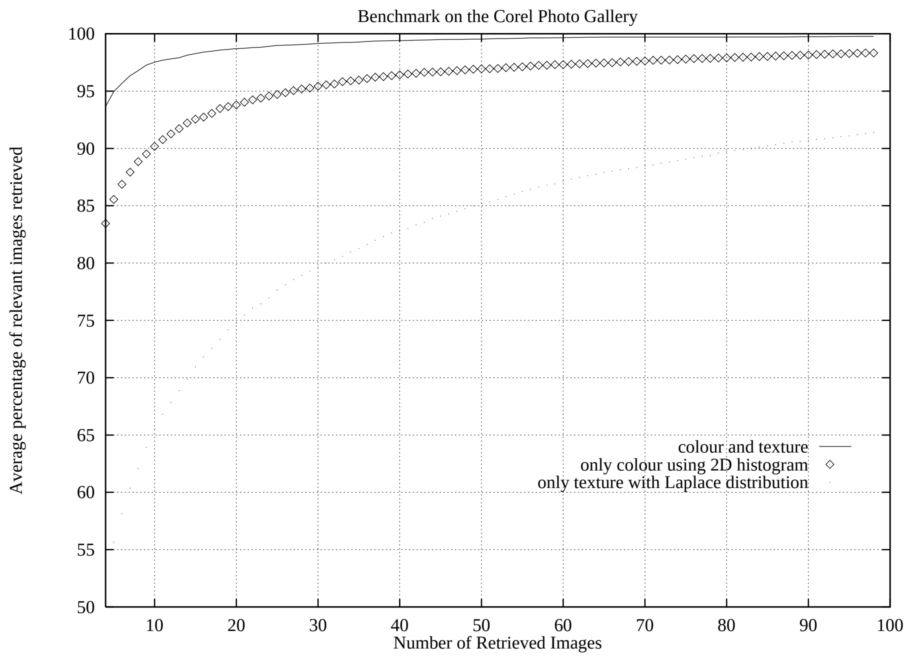

Figure 3: Correct percentage curves for the benchmark on the Corel Photo Gallery data set

[6] F. Liu and R. W. Picard. Periodicity directionality and randomness wold features for image modeling retrieval. *Pattern Recognition*, 18(7):722–733, Jul 1996.

[7] S. G. Malat. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11:674–693, January 1989.

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

[9] 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.

[10] 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.

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

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

Figure 4: Images with inhomogeneities from the Corel data set