

Windsurf: Region-Based Image Retrieval Using Wavelets

Stefania Ardizzoni
Computer Science Lab.
University of Bologna, Italy
stefyardy@iol.it

Ilaria Bartolini
DEIS - CSITE-CNR
University of Bologna, Italy
ibartolini@deis.unibo.it

Marco Patella
DEIS - CSITE-CNR
University of Bologna, Italy
mpatella@deis.unibo.it

Abstract

In this paper we present WINDSURF (Wavelet-Based Indexing of Images Using Region Fragmentation), a new approach to content-based image retrieval. The method uses the wavelet transform to extract color and texture features from an image and applies a clustering technique to partition the image into a set of “homogeneous” regions. Similarity between images is assessed by using the Bhattacharyya distance to compare region descriptors, and then combining the results at image level. Experimental results on a testbed of 10,000 general-purpose images show that our approach is very effective in retrieving images that are “semantically” similar to the query image. In particular, we compared results of WINDSURF with the approach by Stricker and Orengo [11], showing that a significant improvement is obtained in the quality of the result.

1. Introduction

The goal of content-based image retrieval (CBIR) systems is to define a set of properties (*similarity features*) able to effectively characterize the content of images and then to use such features during the retrieval. The features should be simple enough to allow their automatic extraction but meaningful enough to capture the image content.

Recent studies have highlighted the fact that features like color, texture, shape, and spatial position indeed possess a very high semantic value, and are effectively used in several CBIR systems [6, 11, 8, 9]. The usual approach of such systems is to represent each image as a feature vector and to assess similarity between images by way of a metric function. The choice on the number and the type of such features is peculiar to each method. As an example, in the IBM QBIC System [6] a number of features is extracted from each image, representing its color, texture and shape. Stricker and Orengo [11] propose a different approach to color similarity, where each image is represented by way of the first three moments of the distribution of each color channel. The Photobook systems [8] uses a stochas-

tic model (*Wold-decomposition*) to assess the similarity between images based on texture.

All of the previously described approaches use *global features* to represent image semantics. This is due to the fact that all these methods consider color, texture and shape as separate features, whereas, in order to fragment an image in regions, it would be desirable to operate on a combined space (e.g. on a *color-texture* space). This issue has only recently emerged as a need for CBIR in order to support more specific queries (e.g. find all those images containing a small grass region under a big blue region). As an example, in [10] a system is proposed considering information on both the spatial and the frequency domain in order to decompose each image into regions. The similarity between two images is computed by taking into account the location, the dimension, and the relative positioning of regions. The major drawbacks of this approach are the quality of the segmentation and the limited number of image regions. In [1] the authors present the Blobworld system, where images are represented as sets of *blob*, i.e. ellipses in the 2-*D* space, corresponding to regions of the image which are homogeneous for color and texture. Each blob is represented by a 6-*D* vector (HSV descriptors for color, *anisotropy*, *orientation*, and *contrast* for texture).

In recent years a number of new techniques has been developed for signal analysis, that can be applied to the compression, the edge detection, and the texture analysis of images. Such techniques operate in the time-frequency domain by using the wavelet transform (WT, [5]) to obtain a multi-resolution analysis of the space. As an example, the WBIIS system [14] uses Daubechies' wavelets [5] to build the vector representing each image. This approach offers a better frequency location with respect to other classic algorithms, and the quality of its results is high when the query image contains high-frequency information due to sharp color variations. However, the WBIIS system assess the similarity between images by using whole pixel matrices, therefore operating in a very high-dimensional space.

In WINDSURF we use the Discrete WT (DWT) to extract a set of features representing each image in the color-

texture space. Such features are then used to fragment the image into homogeneous regions by using a clustering algorithm. Finally, similarity between images is computed by comparing wavelet coefficients for the extracted regions and combining the results.

1.1. The Wavelet Transform

The basic idea of the wavelet transform is similar to that of Fourier transform: Approximate a signal through a set of basic mathematical functions. However, wavelet functions are able to give a multi-resolution representation of the signal, since each frequency component can be analyzed with a different resolution and scale, whereas the Fourier transform divides the time-frequency domain in an homogeneous way. This allows the WT to represent discontinuities in the signal by using “short” functions and, at the same time, to emphasize low frequency components using “wide” functions.

The Continuous WT decomposes a signal f into a set of *scaling functions* by using a wavelet functions basis:

$$(W_a f)(b) = \int f(x) \psi_{a,b}^*(x) dx \quad (1)$$

The basis of wavelet functions is obtained by scaling and shifting a single *mother wavelet* function $\psi(x)$:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \quad (2)$$

The mother wavelet should only satisfy the zero-average condition, i.e. $\int \psi(x) dx = 0$. The Discrete WT is obtained by taking $a = 2^n$ and $b \in \mathbb{Z}$. The oldest, and simplest, example of a mother wavelet is the Haar function, which was first introduced in 1910, and is composed by a pair of rectangular pulses:

$$\psi(x) = \begin{cases} 1 & 0 \leq x < 1/2 \\ -1 & 1/2 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

The following example shows how a discrete signal is decomposed using the Haar wavelet.

Example 1 Consider a discrete signal $x = (x_0, x_1, \dots, x_{2^L-1})$ having length 2^L . The DWT is computed through the following steps:

1. For each pair of consecutive samples (x_{2i}, x_{2i+1}) , $(0 \leq i < 2^{L-1})$, compute $a_i^1 = \frac{1}{\sqrt{2}}(x_{2i} + x_{2i+1})$ and $d_i^1 = \frac{1}{\sqrt{2}}(x_{2i} - x_{2i+1})$.
2. Consider the new signal $(a_0^1, \dots, a_{2^{L-1}-1}^1)$ and proceed as in step 1., obtaining a_i^2 and d_i^2 $(0 \leq i < 2^{L-2})$.

3. Continue until a single value of a_0^L is obtained.

The Haar transform of x is given by the set of “difference” values d_i^l $(0 < l \leq L, 0 \leq i < 2^{L-l})$, and the “average” value for the last level a_0^L . In the frequency domain, the values a_i^l correspond to the output of a low pass filter, thus representing low-frequency information, whereas the d_i^l values correspond to the output of a high pass filter, thus representing high-frequency information. \square

In our case, the signal is a 2-D color image, where the “time” domain is the spatial location of pixels and the “frequency” domain is the color variation between adjacent pixels. In order to build a wavelet basis for the 2-D space, one can start from the 1-D domain and define the prototype function as the product of two 1-D basis functions: $\Psi_{j_1, k_1; j_2, k_2}(x_1, x_2) = \psi_{j_1, k_1}(x_1) \cdot \psi_{j_2, k_2}(x_2)$. For an $N \times M$ image, the first transformation step decomposes the signal into four sub-images of size $N/2 \times M/2$, representing the sub-bands in the frequency domain. The obtained sub-images are labelled as LL, LH, HL, HH , where L and H represent low- and high-frequency information, respectively, and the first and the second position refer to the horizontal and the vertical direction, respectively. The second transformation level decomposes the LL sub-image, obtaining four images of size $N/4 \times M/4$, and so on. Figure 1 shows the decomposition of the frequency domain at different scale levels: $A_{2^{-L}}^d f$ contains low-frequency information, whereas $D_{2^j}^1 f$, $D_{2^j}^2 f$, and $D_{2^j}^3 f$ contain horizontal, vertical and diagonal information, respectively.

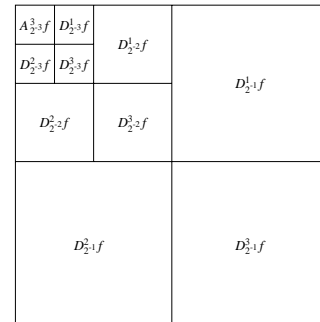


Figure 1. The sub-images $D_{2^j}^k f, A_{2^{-L}}^d f$ in the “wavelet” image representation.

2. The Windsurf Approach

In WINDSURF the similarity between a pair of images is assessed by way of the similarity between their regions. Each region is represented by means of a (restricted) domain of features and the match between regions is measured using a specific metric function on such a domain. The global architecture of the system is sketched in Figure

2. Each image is processed through a number of steps described in the following:

DWT The image is analyzed in the time-frequency domain using a 2-*D* discrete wavelet transform.

Clustering The image is fragmented into a set of regions using the wavelet coefficients (*clustering features*).

Feature Indexing Regions so obtained are stored using a set of similarity features.

In WINDSURF each image is divided into the correspondent color channels and the DWT is applied to each channel. It is well known [9, 14] that the RGB color space is not suitable to reflect human perception of color. To overcome this problem, we considered the transformation of each image in a color space where each color component is perceptually independent and uniform. In particular, we investigated the HSV and the OPP color spaces [9]. Experimental results, however, showed that the HSV color space is most appropriate for our approach, since in the HSV space the WINDSURF fragmentation technique generates perceptually better regions with respect to those generated in the OPP space. In the following, the j -th wavelet coefficient in the S sub-band ($S \in \{LL, LH, HL, HH\}$) of the l -th DWT level will be indicated as $w_j^{l;S} = (w_{0_j}^{l;S}, w_{1_j}^{l;S}, w_{2_j}^{l;S})$, where $c \in \{0, 1, 2\}$ denotes a color channel.

In the WINDSURF system we used the Haar wavelet as a special case of the biorthogonal wavelet of Cohen-Daubechies-Feauveau (CDF), as provided by the WAILL software library [12].

2.1. Region Fragmentation

To fragment each image into regions, we use a clustering algorithm on the coefficients obtained through the DWT. We use a simple *k-means* algorithm with a validity function which is slightly different from that proposed for the *fuzzy k-means* [15].

Given the set $X = \{x_1, \dots, x_N\}$ of N points to be clustered, let $V = \{\mu_1, \dots, \mu_k\}$ be the set of cluster centroids. The goal of the *k-means* algorithm is to minimize the function

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} \delta(x_j, \mu_i)^2 \quad (4)$$

where $\delta(x_j, \mu_i)$ is the distance between x_j and the centroid μ_i of the i -th cluster C_i . Obviously, the value of J depends both on the value of k and on the choice of the distance function $\delta()$, i.e. different choices of $\delta()$ lead to different results. In order to provide semantically meaningful regions, we ignore clusters having size smaller than a minimum value.

For $\delta()$ we adopted the *Mahalanobis distance* applied on the 3-*D* wavelet coefficients of the *LL* sub-band of the 3-rd level, i.e. $x_j = w_j^{3;LL}$. Since we only use wavelet coefficients to divide the image into regions, *no spatial information is taken into account*, i.e. image pixels are grouped together only using information on color and texture. This could be viewed as a limit of the system, since spatial queries, where the user requests for images having a spatial arrangement of regions similar to that of the query image, cannot be issued. However, by ignoring spatiality of regions, the WINDSURF system is able to consider as similar images having “similar” regions with different positioning (and, by the way, also different orientation, rotation, etc.).

The Mahalanobis distance between points x_i and x_j is given by:

$$\delta(x_i, x_j) = (x_i - x_j)^T \cdot (\mathcal{C}^{l;S})^{-1} \cdot (x_i - x_j) \quad (5)$$

where $\mathcal{C}^{l;S} = \{cov_{c,d}^{l;S}\}$ is the covariance matrix of the points, where diagonal values represent energies and non-diagonal elements represent correlations. By using the Mahalanobis distance we obtain both a normalization of vectors (given by diagonal elements of the covariance matrix), and the fact that texture characteristics are also taken into account. Indeed, as recent studies demonstrate [13], texture information is captured by the cross-correlation between color channels, i.e. by the off-diagonal coefficients of matrix $\mathcal{C}^{l;S}$. In our case, $\mathcal{C}^{l;S}$ is a 3×3 matrix whose coefficients $cov_{c,d}^{l;S}$ are given as:

$$cov_{c,d}^{l;S} = \frac{1}{N} \left(\sum_{j=1}^N w_{c_j}^{l;S} w_{d_j}^{l;S} - \sum_{j=1}^N w_{c_j}^{l;S} \cdot \sum_{j=1}^N w_{d_j}^{l;S} \right) \quad c, d \in \{0, 1, 2\}. \quad (6)$$

The *k-means* algorithm provides, as the output, the N points divided in k clusters. However, different values of the parameter k lead, as intuition may suggest, to different results. To compute the “optimal” value for k , we iterate it between a minimum ($k_{\min} = 2$) and a maximum ($k_{\max} = 10$) value, choosing the best solution which minimizes the *validity* V defined as:

$$V = \frac{J}{N \cdot d_{\min}^2} + \sum_{i=1}^k \frac{1}{1 + \|C_i\|} \quad (7)$$

where J is given by Equation 4, and $d_{\min} = \min_{i \neq j} \{d(\mu_i, \mu_j)\}$ is the minimum distance between cluster centroids. The first term of Equation 7 represents the goal function J divided by d_{\min} , i.e. clusters well separated provide better solutions, whereas the second term represents a penalty factor for small clusters. As an example, Figure 3

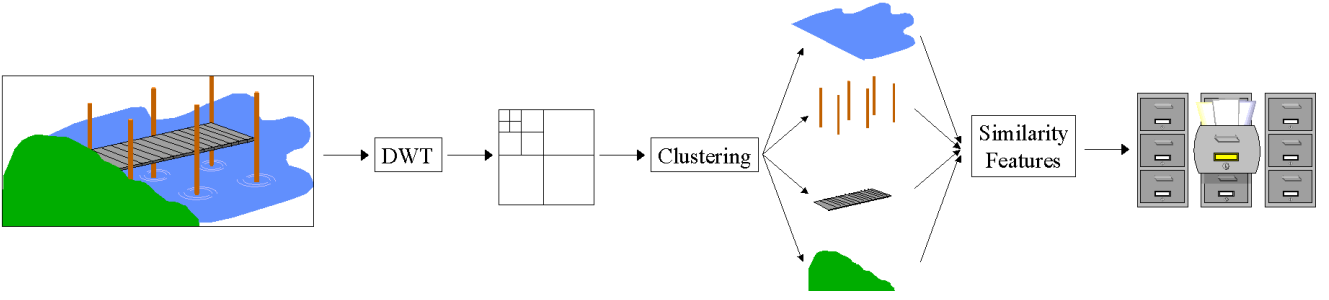


Figure 2. Steps for image features indexing in the WINDSURF system.

shows the result of the k -means algorithm applied to the image on the left, when $k = 2$, $k = 10$, and $k = 4$, the latter being the optimal solution according to the validity function of Equation 7. In the clustered images, points having the same color belong to the same cluster.

The validity criterion expressed by Equation 7 is not applicable when $k = 1$. However, our approach should deal with general-purpose images which can also present an uniform pattern, i.e. that are perceptually homogeneous. In such cases, we do not apply the clustering and represent images globally as a single region. We choose if an image should be clustered by looking at the trace $\mathcal{T}_{\mathcal{C}^{3;LL}}$ of the covariance matrix $\mathcal{C}^{3;LL}$ used for clustering.¹ Intuitively, if the eigenvalues have all small values, then the coefficients have a small variability, and the image can be considered as an homogeneous pattern.

3. Similarity Assessment

Our clustering algorithm takes into account only the low-frequency information to fragment the image (this is intuitive, since regions are low-frequency information). However, when comparing two images, our goal is to use the information provided by all the frequency sub-bands. To this end, regions are represented by way of the coefficients of the covariance matrices in all the sub-bands. In short, the similarity features for each region R_i are defined as:

Size The number of points in the region, $size(R_i)$.

Centroid The centroid of R_i is defined through a 12- D vector $V_{R_i} = (\mu_i^{LL}, \mu_i^{LH}, \mu_i^{HL}, \mu_i^{HH})$, where, for each sub-band S , μ_i^S is a 3- D point representing the centroid coefficient for each color channel.

Features These correspond to the coefficients of the 3×3 covariance matrices, $\mathcal{C}_{R_i}^{3;S}$, of the points contained in R_i . Since the covariance matrices are symmetrical, we only store 6 values for each matrix $\mathcal{C}_{R_i}^{3;S}$, obtaining a 24- D vector \mathcal{C}_{R_i} .

¹Indeed, the trace of a matrix is equal to the sum of its eigenvalues.

In the WINDSURF system, the similarity between two regions, R_i and R_j , is computed as follows:

$$sim(R_i, R_j) = \alpha_{ij} \cdot \beta_{ij} \cdot h(d(R_i, R_j)) \quad (8)$$

where $d()$ is a distance function, $h()$ is a so-called *correspondence function* [4] relating distance values to similarity scores, α_{ij} and β_{ij} are similarity coefficients.

The α_{ij} coefficient, used to favor match between large regions, takes into account the relative size of the regions with respect to those of the images they were extracted from. The β_{ij} coefficient takes into account the similarity in size between the two regions. If R_i is extracted from image I and R_j from image J , α_{ij} and β_{ij} can be defined as:

$$\alpha_{ij} = \frac{size(R_i) + size(R_j)}{size(I) + size(J)} \quad (9)$$

$$\beta_{ij} = 1 - \frac{|size(R_i) - size(R_j)|}{size(R_i) + size(R_j)}$$

The correspondence function $h()$ is used to transform distance values into similarity scores. The function $h : \mathbb{R}_0^+ \rightarrow [0, 1]$ has to satisfy the following properties: $h(0) = 1$ and $d_1 \leq d_2 \Rightarrow h(d_1) \geq h(d_2), \forall d_1, d_2 \in \mathbb{R}_0^+$. In all the experiments, we used $h(d) = e^{-d/\sigma_d}$, where σ_d^2 is the distance variance computed over a sample of regions.

Finally, distance between two regions is computed by way of the *Bhattacharyya metric* [2], used to compare ellipsoidal clusters:

$$d_S(R_i, R_j)^2 = \frac{1}{2} \ln \left(\frac{\left| \frac{\mathcal{C}_{R_i}^{3;S} + \mathcal{C}_{R_j}^{3;S}}{2} \right|}{\left| \mathcal{C}_{R_i}^{3;S} \right|^{\frac{1}{2}} \cdot \left| \mathcal{C}_{R_j}^{3;S} \right|^{\frac{1}{2}}} \right) + \frac{1}{8} \left[(V_{R_i} - V_{R_j})^T \cdot \left(\frac{\mathcal{C}_{R_i}^{3;S} + \mathcal{C}_{R_j}^{3;S}}{2} \right)^{-1} \cdot (V_{R_i} - V_{R_j}) \right] \quad (10)$$

where $|A|$ is the determinant of matrix A . Note that Equation 10 is composed of two terms, the second one being the

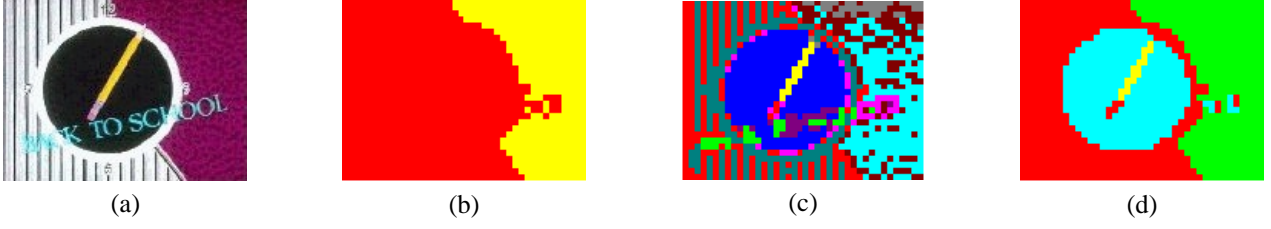


Figure 3. (a) The input image. (b) Clusters obtained for $k = 2$. (c) Clusters obtained for $k = 10$. (d) Clusters obtained for $k = 4$ (optimal solution).

Mahalanobis distance (Equation 5) between regions centroids, using the average covariance matrix. The first term is used to compare the covariance matrices of the two regions. Indeed, if the two regions have the same centroid, the second term of Equation 10 has a value of zero, and the first term is used to distinguish between the two regions (see Figure 4).

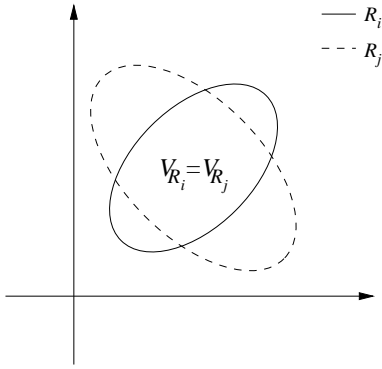


Figure 4. Two regions with different shape and coincident centroids.

The overall distance between regions R_i and R_j is assessed by computing Equation 10 over all the frequency sub-bands: $d(R_i, R_j)^2 = \sum_S \gamma_S \cdot d_S(R_i, R_j)^2$, where $d_S(R_i, R_j)^2$ is computed by way of Equation 10 and coefficients γ_S are used to give different weights to frequency sub-bands. In our experiments we equally weigh the frequency coefficients, therefore it is $\gamma_{LL} = \gamma_{LH} = \gamma_{HL} = \gamma_{HH} = 1$.

When computing Equation 10, we have to take into account particular cases arising when dealing with singular covariance matrices. Such situations arise when dealing with uniform images (e.g. a totally black image), where the covariance between coefficients is null.

3.1. Image Similarity

Having defined how the similarity between regions is computed, we now need to assess the overall similarity between two images (e.g. the query image Q and a DB image

T). When matching regions in Q with regions in T , we have to satisfy two basic constraints:

1. A region of Q cannot match with two different regions in T (Figure 5 (a)).
2. Two different regions of Q cannot match with a single region of T (Figure 5 (b)).

To this end, each region q_i of Q is associated to its “best match” region t_j in T by only considering $d(q_i, t_j)$, i.e. regions size is not taken into account. If, however, two regions q_i and q_l of Q are associated to the same region t_j of T , only the best match is kept, e.g. if $d(q_i, t_j) < d(q_l, t_j)$, then the match between q_l and t_j is removed. The best match of region q_i in image T is indicated as $t_{j(i)}$. Note that the best match for a certain region q_i can be undefined.

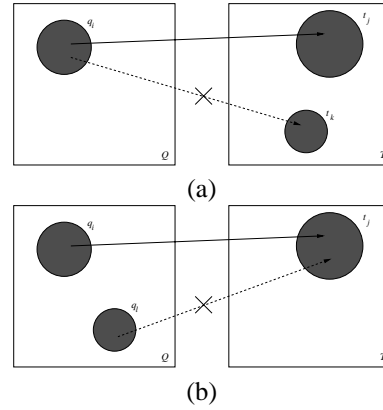


Figure 5. A region of Q cannot match with two regions of T (a) and two regions of Q cannot match with the same region of T (b).

We are now ready to define the overall similarity between two images Q and T as:

$$\begin{aligned}
 SIM(Q, T) &= \sum_i sim(q_i, t_{j(i)}) = & (11) \\
 &= \sum_i \alpha_{ij(i)} \cdot \beta_{ij(i)} \cdot h(d(q_i, t_{j(i)}))
 \end{aligned}$$

Of course, $sim(q_i, t_{j(i)}) = 0$ if $t_{j(i)}$ is undefined.

The user can now express a query by giving an input image and an integer value n . The WINDSURF system will retrieve the n DB images most similar to the query image, with ties arbitrarily broken (n -nearest neighbor query). At the moment, the system performs a sequential scan of the regions in the DB, but we are considering a speed-up of the retrieval phase by exploiting a metric tree (like M-tree [3]) to index regions features.

4. Experimental Results

In order to evaluate the effectiveness of our approach, we compare the results of a number of queries in the WINDSURF system with those obtained by applying the (SO) method proposed in [11]. For our experiments, we considered a dataset composed by 10,000 images extracted from a CD-ROM of *IMSI-PHOTOS*.² The regions database is composed of over 40,000 regions. These, of course, are only preliminary results comparing performance of the WINDSURF system with those of a simple method which was available to us at the time of this work. In the future, see Section 5, we plan to perform an exhaustive comparison of results between WINDSURF and other CBIR systems.

From a semantic point of view, results obtained by the WINDSURF system are considerably better with respect to those obtained by the SO method, as it is clear by observing the following results, which are only a minor part of those obtained during experimentation. As an example, consider Figure 6: Results for SO (SO1 - SO5) contain images semantically uncorrelated to the query image (e.g. image (SO3), a house, and image (SO5), a harbour). As for the results for the WINDSURF system (WS1 - WS5), all present a “sky” region and a darker area.

The superior effectiveness of our approach is confirmed when considering “difficult” queries, i.e. queries having a low number of similar images in the DB. In Figure 7 we show the results for a query having only two similar images: For SO, none of the two images is included in the result. The WINDSURF system, on the other hand, retrieves both images.

Finally, we compared the two approaches when dealing with “partial match” queries, i.e. queries specifying only a part of the image. As an example, consider Figure 8, where we query the system giving, as the query image, an image obtained by “cropping” a DB image, in this case, the dome of St. Peter in Rome. For WINDSURF all the retrieved images refer to St. Peter, with the exception of image (WS3), representing the dome of St. Marcus in Venice. Indeed, the query image was extracted from image (WS1). When we analyze the result obtained by SO, we see that only an image referring to the query image is retrieved in third position, whereas other images, with the exception of image (SO2)

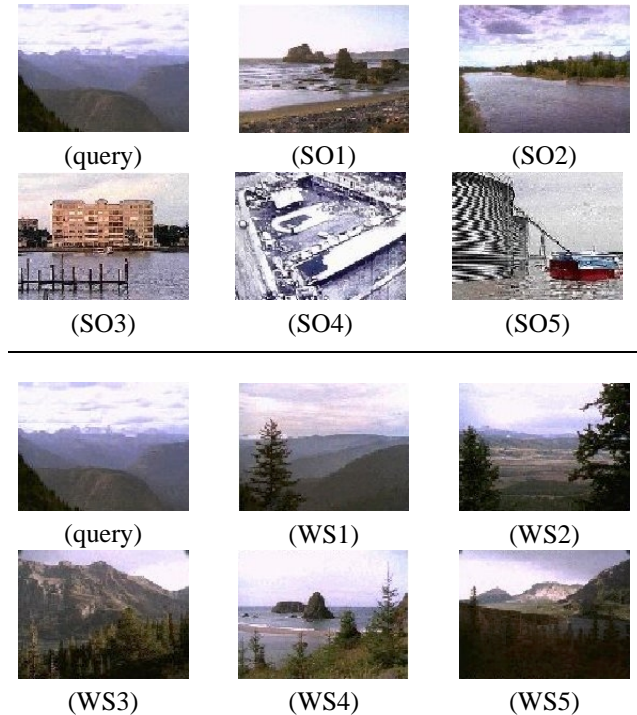


Figure 6. Results for the “mountains” query.

(again the dome of St. Marcus), are totally uncorrelated to the query image.

5. Conclusions

In this work we introduced WINDSURF, a new approach to content-based image retrieval. The novelty of our approach is given by the use of a combined transform-clustering approach. In particular, each image is analyzed in the spatial-frequency domain by applying a discrete wavelet transform on the color channels. This corresponds to consider combined color *and* texture information to represent the image content, as opposed to previous approaches which only consider color and texture as separate features. Such information is then used to divide each image into a number of regions, which are therefore homogeneous in color and texture. Similarity between images is then assessed by means of a distance function comparing region features and combining the results at a global level. Finally, we have experimentally demonstrated the superior retrieval effectiveness of the WINDSURF approach over an established color retrieval method [11].

The major current limit of the presented approach is its low speed during the retrieval phase, since the regions DB is sequentially scanned. To overcome this problem, we are considering to index region features using an M-tree [3]. Other planned work includes the comparison of the WINDSURF approach to other CBIR systems, like QBIC [6], WALRUS [7], and WBIIS [14].

²IMSI MasterPhotos 50,000: <http://www.imsisoft.com>

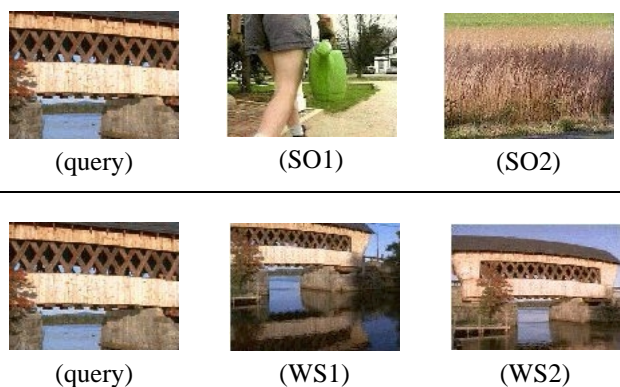


Figure 7. Results for the “bridge” query.

References

- [1] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color and texture-based image segmentation using EM and its application to content-based image retrieval. In *Proceedings of the 6th International Conference on Computer Vision ICCV'98*, Mumbai, India, Jan. 1998.
- [2] J. P. Campbell. Speaker recognition: A tutorial. *Proceedings of the IEEE*, 85(9):1437–1461, Sept. 1997.
- [3] P. Ciaccia, M. Patella, and P. Zezula. M-tree: An efficient access method for similarity search in metric spaces. In *Proceedings of the 23rd International Conference on Very Large Data Bases (VLDB'97)*, pages 426–435, Athens, Greece, Aug. 1997.
- [4] P. Ciaccia, M. Patella, and P. Zezula. Processing complex similarity queries with distance-based access methods. In *Proceedings of the 6th International Conference on Extending Database Technology (EDBT'98)*, pages 9–23, Valencia, Spain, Mar. 1998.
- [5] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.
- [6] C. Faloutsos, W. Equitz, M. Flickner, W. Niblack, D. Petkovic, and R. Barber. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 3(3/4):231–262, July 1994.
- [7] A. Natsev, R. Rastogi, and K. Shim. WALRUS: A similarity retrieval algorithm for image databases. In *Proceedings 1999 ACM SIGMOD International Conference on Management of Data*, Philadelphia, PA, June 1999.
- [8] A. Pentland, R. W. Picard, and S. Sclaroff. Photo-book: Content-based manipulation of image databases. In B. Furht, editor, *Multimedia Tools and Applications*, chapter 2, pages 43–80. Kluwer Academic Publishers, 1996.
- [9] J. R. Smith. *Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression*. PhD thesis, Columbia University, 1997.
- [10] J. R. Smith and S.-F. Chang. VisualSEEK: A fully automated content-based image query system. In *Proceedings of the 4th ACM International Conference on Multimedia*, pages 87–98, Boston, MA, Nov. 1996. <http://www.ctr.columbia.edu/visualeek/>.

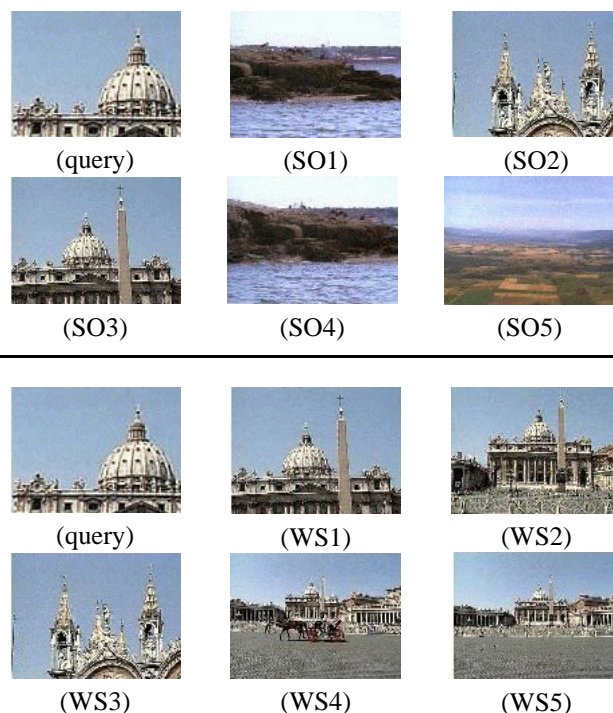


Figure 8. Results for the “dome” query.

- [11] M. Stricker and M. Orengo. Similarity of color images. In *Storage and Retrieval for Image and Video Databases SPIE*, volume 2420, pages 381–392, San Jose, CA, Feb. 1995.
- [12] G. Uytterhoeven, F. Van Wulpen, M. Jansen, D. Roose, and A. Bultheel. WAILI: Wavelets with integer lifting. Technical Report 262, Department of Computer Science, Katholieke Universiteit Leuven, Heverlee, Belgium, July 1997.
- [13] G. Van de Wouwer, P. Scheunders, S. Livens, and D. Van Dyck. Color texture classification by wavelet energy-correlation signatures. To appear on *Pattern Recognition*.
- [14] J. Z. Wang, G. Wiederhold, O. Firschein, and S. X. Wei. Wavelet-based image indexing techniques with partial sketch retrieval capability. In *Proceedings of the 4th IEEE Forum on Research and Technology Advances in Digital Libraries (ADL'97)*, Washington, DC, May 1997.
- [15] X. L. Xie and G. Beni. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):841–847, Aug. 1991.