Journal of Electronic Imaging 10(4), 1–0 (October 2001).

# Image retrieval using wavelet-based salient points

**Q. Tian**
Beckman Institute
405 N. Mathews
Urbana, Illinois 61801
E-mail: qitian@ifp.uiuc.edu

**N. Sebe**
**M. S. Lew**
Leiden Institute of Advanced Computer Science
Leiden, The Netherlands

**E. Loupias**
Laboratories Reconnaissance, de Formes et Vision
INSA-Lyon
Lyon, France

**T. S. Huang**
Beckman Institute
405 N. Mathews
Urbana, Illinois 61801



**Abstract.** *Content-based image retrieval (CBIR) has become one of the most active research areas in the past few years. Most of the attention from the research has been focused on indexing techniques based on global feature distributions. However, these global distributions have limited discriminating power because they are unable to capture local image information. The use of interest points in content-based image retrieval allow image index to represent local properties of the image. Classic corner detectors can be used for this purpose. However, they have drawbacks when applied to various natural images for image retrieval, because visual features need not be corners and corners may gather in small regions. In this paper, we present a salient point detector. The detector is based on wavelet transform to detect global variations as well as local ones. The wavelet-based salient points are evaluated for image retrieval with a retrieval system using color and texture features. The results show that salient points with Gabor feature perform better than the other point detectors from the literature and the randomly chosen points. Significant improvements are achieved in terms of retrieval accuracy, computational complexity when compared to the global feature approaches.* © *2001 SPIE and IS&T.* [DOI: 10.1117/1.1406945]

## 1 Introduction

Recent years have witnessed a rapid increase of the volume of digital image collections, which motivates the research of image retrieval.[1–3] Early research in image retrieval proposed manually annotated images for their retrieval. However, these text-based techniques are impractical for two reasons: large size of image databases and subjective meaning of images. To avoid manual annotation, an alternative approach is content-based image retrieval (CBIR), by which images would be indexed by their visual contents such as color, texture, shape, etc. Many research efforts have been made to extract these low-level image features,[4,5] evaluate distance metrics,[6,7] look for efficient searching schemes,[8,9] and more recently propose statistically learning approaches.[10–12]

In a typical content-based image database retrieval application, the user has an image he or she is interested in and wants to find similar images from the entire database. A two-step approach to search the image database is adopted. First, for each image in the database, a feature vector characterizing some image properties is computed and stored in a feature database. Second, given a query image, its feature vector is computed, compared to the feature vectors in the feature database, and images most similar to the query images are returned to the user. The features and the similarity measure used to compare two feature vectors should be efficient enough to match similar images as well as being able to discriminate dissimilar ones.

In this context, an image index is a set of features, often computed from the entire image. However natural images are mainly heterogeneous, with different parts of the image with different characteristics, which cannot be handled by these *global* features.

Paper SPR-08 received Mar. 15, 2001; revised manuscript received June 27, 2001; accepted for publication July 3, 2001.
1017-9909/2001/$15.00 © 2001 SPIE and IS&T.

Local features can be computed to obtain an image index based on local properties of the image. These local features, which need to be discriminant enough to ''summarize'' the local image information, are mainly based on filtering, sometimes at different image scales. These kinds of features are too time consuming to be computed for each pixel of the image. Therefore the feature extraction is limited to a subset of the image pixels, the *interest points*,[13–15] where the image information is supposed to be the most important.

Besides saving time in the indexing process, these points may lead to a more discriminant index because they are related to the visually most important parts of the image. Schmid and Mohr introduced the notion of *interest point* in image retrieval.[13] To detect these points, they compute local invariants. They use the Harris' detector, one of the most popular corner detectors. This detector, as many others, was initially designed for robotics, and it is based on a mathematical model for corners. The original goal was to match same corners from a pair of stereo images, to obtain a representation of the three-dimensional (3D) scene. Since corner detectors were not designed to give a ''summary'' as comprehensive as possible of an image, they have drawbacks when applied to various natural images for image retrieval:

1. Visual focus points need not to be corners: When looking at a picture, we are attracted by some parts of the image, which are the most meaningful for us. We cannot assume them to be located in corner points, as mathematically defined in most corner detectors. For instance, smoothed edges can also be visual focus points, and they are usually not detected by a corner detector. The image index we want to compute should describe them as well.
2. Corners may gather in small regions: In various natural images, regions may well contain textures (trees, shirt patterns, etc). Many gathered corners are detected in these regions by a corner detector. However, a preset number of points per image are used in the indexing process, to limit the indexing computation time. With this kind of detector, most of the points are in a small region, and the local features are computed from the same texture region, while other parts of the image will not be described in the index at all.

For these reasons, corner points, as designed in robotics, may not represent the most interesting subset of pixels for image indexing. Indexing points should be related to any visual ''interesting'' part of the image, whether it is smooth or corner-like. To describe different parts of the image, the set of interesting points should not be clustered in few regions. From now on, we will refer to these points as *salient points*, which are not necessarily corners. We will avoid the term *interest points*, which is ambiguous, since it was previously used in the literature as *corner*.

The aim of the following sections is to present a CBIR system using wavelet-based salient points for image indexing, based on wavelet transform.[16–18] In Sec. 2, we will briefly review previous point detectors, especially the ones that also use wavelets and multiresolution. Section 3 describes how we extract points from a wavelet representation and gives some examples. It illustrates the different behavior of our detector compared to the known corner detectors. In Sec. 4, the wavelet-based salient points are evaluated for image retrieval with the Multimedia Analysis and Retrieval System (MARS) at the University of Illinois[19,20] using color and texture features. Conclusions will be given in Sec. 5.

## 2 Point Detectors

The need for a corner detector first appeared in applications such as shape recognition or 3D reconstruction. Corners are interesting points to accurately describe a shape or represent a 3D scene for image matching. A number of corner detectors are briefly reviewed in Sec. 2.1.[18]

The points to use in image indexing are not necessarily corners. We believe multiresolution representations are interesting to detect this kind of point. In Sec. 2.2, we show how multiresolution was previously used for point detection, either in retrieval context or not.

### 2.1 *Corner Detectors*

Corners are usually defined as points where gradient is high in multiple orientations. This definition leads to corner detectors based on local derivatives (usually first or second order). Several of these techniques are described in Ref. 21.

Harris' detector is also based on local derivatives, using autocorrelation of the image.[22] This detector is the most used corner detector in many applications (include image retrieval),[13,15] and is often evaluated as the best detector toward different criteria.[23,24] However, there are a lot of different implementations of it, because five parameters need to be set: the derivative kernel, the smoothing kernel ($\sigma$), $k$, the local maximum neighborhood, and the final threshold. Recently, Zheng adapted the Harris' detector to improve the corner localization.[25]

Other corner detectors use a local neighborhood in each pixel to evaluate whether it is a corner. Zitova *et al.* uses the difference between the neighbors and the local mean value.[26] For Smith and Brady (SUSAN detector), corners are points that have few neighbors with similar value.[27] These methods, even though various, all focus on local information because their purpose is to extract points where the image is locally corner-like. In retrieval context, images must be studied as a whole: images can be very heterogeneous, with some textured parts and others smoothed. Since the process should be efficient, the number of points used must be as small as possible to represent the whole image. With this kind of corner detector, most of the points are in the textured part, passing up other parts that can be meaningful for the user (Fig. 1). The resulting index does not fully describe the image. The points to use in a retrieval process should be computed with more global information, to take into account the entire image. An attractive way to do this is through the use of multiresolution information, to extract meaningful points from different resolutions.

### 2.2 *Multiresolution Point Detectors*

We review here previous point detectors using multiresolution representation such as wavelets. For wavelet theory, see Refs. 28–30. In Ref. 31, corners are detected from the one-dimensional wavelet transform of the contour orientation function. However, curve function cannot be automati-

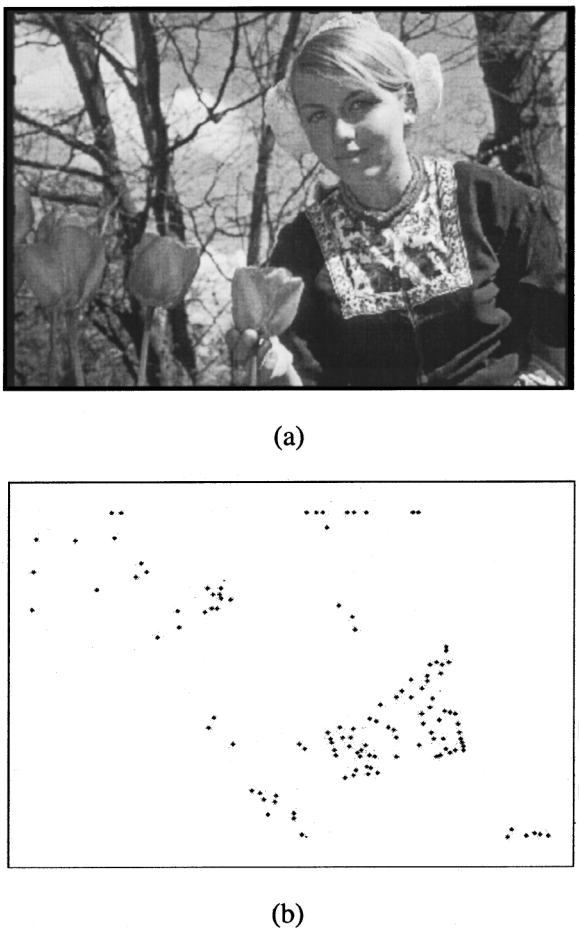


**Fig. 1** (a) A natural image with texture in the dress and (b) Zheng corners. Most points gathered in the textured region (Dutch dress).

cally extracted from natural images. In Ref. 32 a wavelet-based corner detector for gray-level images is presented. The authors compute magnitude information from the wavelet transform, which is proportional to the scale for the corner position and some edge points. Points are extracted using this property at two different scales, according to the corner model. Therefore it does not allow detection of salient features from different resolutions.

The approach described in Ref. 14 considered energy-based points for image retrieval instead of usual corners. They use a multiresolution contrast pyramid to extract them. However, a lot of points are also extracted in textured regions because these regions are very contrasted.

In another image retrieval paper,[33] the author is also looking for visually meaningful points, which are not necessarily corners. A specific wavelet is used to detect points. But since only a given scale is used, different resolution features cannot be detected.

## 3 Wavelet-Based Salient Points

### 3.1 *Salient Point Extraction*

The wavelet representation gives information about the variations in the image at different scales. In our retrieval context, we would like to extract salient points from any part of the image where ''something'' happens in the image at any resolution. A high wavelet coefficient (in absolute value) at a coarse resolution corresponds to a region with high global variations. The idea is to find a relevant point to represent this global variation by looking at wavelet coefficients at finer resolutions.

We first consider the wavelet transform for one dimension. Most notations and equations here can be found in Ref. 29. A wavelet is an oscillating and attenuating function with zero integral. We study the image $f$ at the scales (or resolutions) 1/2, 1/4,...,$2^j$, $j \in \mathbf{Z}$, and $j \leqslant -1$. The wavelet detail image $W_{2^j}f$ is obtained as the convolution of the image with the wavelet function dilated at different scales. We considered orthogonal wavelets with compact support. First, this assures that we have a complete and nonredundant representation of the image. Second, since the wavelets have a complete support, we know from which signal points each wavelet coefficient at the scale $2^j$ was computed. We can further study the wavelet coefficients for the same points at the finer scale $2^{j+1}$. There is a set of coefficients at the scale $2^{j+1}$ computed with the same points as a coefficient $W_{2^j}f(n)$ at the scale $2^j$. We call this set of coefficients the children $C(W_{2^j}f(n))$ of the coefficient $W_{2^j}f(n)$. The children set in one dimension is:

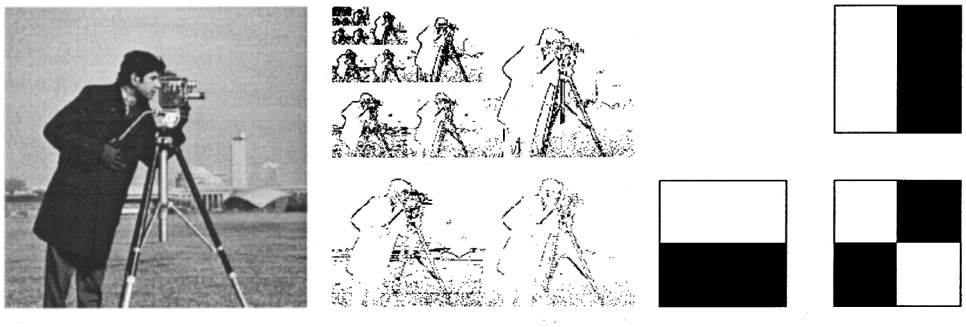
$$C(W_{2^j}f(n))=\{W_{2^{j+1}}f(k), 2n \leqslant k \leqslant 2n+ap-1\}, \qquad (1)$$

where $p$ is the wavelet regularity ($p=1$ for *Haar* wavelet, $p=2$ for *Daubechies 4* wavelet) and $0 \leqslant n \leqslant 2^j N$, with $N$ the length of the signal. The simplest orthogonal compactly supported wavelet is the *Haar* wavelet, which is the discontinuous crenel function. Other orthogonal and compactly supported wavelets were found by Daubechies.[30]

Each wavelet coefficient $W_{2^j}f(n)$ is computed with $2^{-j}p$ signal points. It represents their variation at the scale $2^j$. Its children coefficients give the variations of some particular subsets of these points (with the number of subsets depending on the wavelet). The most salient subset is the one with the highest wavelet coefficient at the scale $2^{j+1}$, that is the maximum in absolute value of $C(W_{2^j}f(n))$. In our salient point extraction algorithm, we consider this maximum, and look at his highest child. Recursively applying this process, we select a coefficient $W_{2^{-1}}f(n)$ at the finer resolution 1/2 [Figs. 2(b) and 5]. Hence, this coefficient represents $2p$ signal points. To select a salient point from this tracking, we choose among these $2p$ points the one with the highest gradient. We set its saliency value as the sum of the absolute value of the wavelet coefficients in the track:

$$\text{saliency}=\sum_{k=1}^{-j} |C^{(k)}(W_{2^j}f(n))|, \quad -\log_2 N \leqslant j \leqslant -1. \qquad (2)$$

The tracked point and its saliency value are computed for every wavelet coefficient. A point related to a global variation has a high saliency value, since the coarse wavelet coefficients contribute to it. A finer variation also leads to an extracted point, but with a lower saliency value. We then need to threshold the saliency value in relation to the de-



(a) (b) (c)

**Fig. 2** (a) Cameraman image, (b) *Haar* transform, and (c) *Haar* derivative filters used to compute the transform.



sired number of salient points. We first obtain the points related to global variations; local variations also appear if enough salient points are requested.

The salient points extracted by this process depend on the wavelet we use. *Haar* is the simplest orthogonal wavelet with compact support, so is the fastest for execution. The larger the spatial support of the wavelet, the more the number of computations. Nevertheless, some localization drawbacks can appear with *Haar* due to its nonoverlapping wavelets at a given scale. This drawback can be avoided with the simplest overlapping wavelet, *Daubechies 4*. However, this kind of drawback is not likely in natural images and therefore, we used *Haar* transform in our experiments.

### 3.2 *Extension to Images*

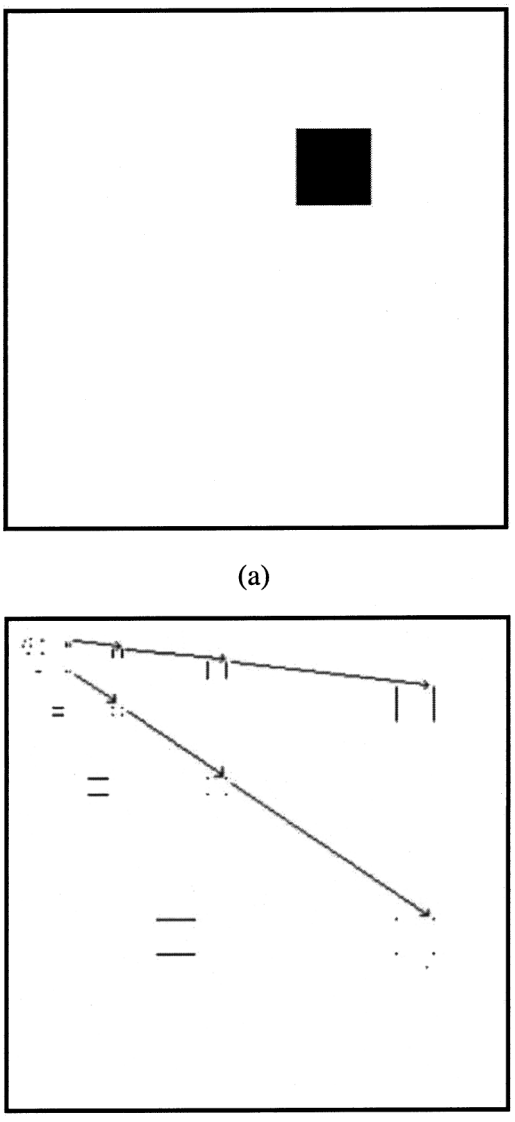
We presented the salient point extraction in one-dimensional signal for simplicity. We now extend it to images. We first need to perform the image wavelet transform. There are two ways to do it. A simple approach is to compute the one-dimensional wavelet transform to each row of the image, then compute the one-dimensional wavelet transform to each line of the result (to compute first lines then rows gives the same result). However, the resulting wavelets have supports with different shapes at different scales, which make them difficult to interpret.

The extension of the wavelet model to two dimensions leads to three different wavelet functions ($\psi^1$, $\psi^2$, and $\psi^3$), related to three different *spatial orientations* (horizontal, diagonal, and vertical).[34] Then the wavelets all have square supports (see Fig. 2). In this framework, the extension of our salient point extraction is straightforward.

The wavelet representation of an image $I$ is the set of coefficients for all orientations and all scales:

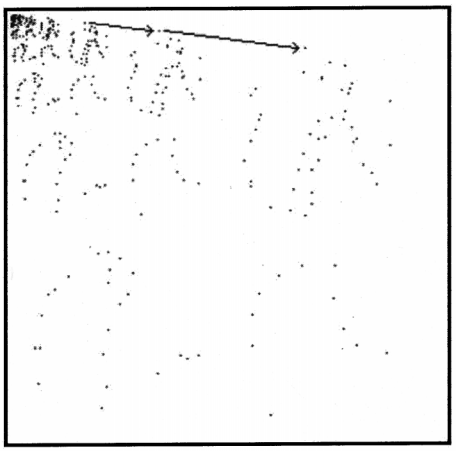
$$WI=(W^1_{2^j}I, W^2_{2^j}I, W^3_{2^j}) - J_{\max} \leqslant j \leqslant -1. \quad (3)$$

Each orientation gives information about the horizontal, vertical, or ''diagonal'' variations in the signal at different scales. Hence, we apply the process described in one dimension independently for each direction (Fig. 3).



(a) (b)

**Fig. 3** (a) An image. (b) The wavelet transform and the track of the wavelet coefficients for each orientation.



(a) The wavelet coefficients tracked

(b) 100 salient points on the original image

**Fig. 4** The cameraman image [Fig. 2(a)]. (a) The wavelet coefficients tracked (b) 100 salient points on the original image.

The spatial support of the wavelet discrete filter is $S(g(x,y))=[0,2p-1]\times[0,2p-1]$, where $g$ is the wavelet discrete filter[29] ($p=1$ for *Haar* wavelet, $p=2$ for *Daubechies 4* wavelet), and the children set for a given wavelet coefficient:

$$C(W_{2^j}^d f(x,y))=\{W_{2^{j+2}}^d f(k,l),2x\leqslant k\leqslant 2x+2p-1,2y\leqslant l\leqslant 2y+2p-1\}$$
$$0\leqslant x\leqslant 2^j N,\quad 0\leqslant y\leqslant 2^j M,\quad 1\leqslant d\leqslant 3, \tag{4}$$

where $N\times M$ is the number of pixels of the image $I$.

For each orientation, we track the wavelet coefficients for this orientation until we find a pixel, to which we give the saliency value. If different wavelet coefficients from different orientations lead to the same pixel, then we keep the highest saliency value, as we did for overlapping wavelets.

An example of tracked coefficients and the extracted points are given for the cameraman image in Figs. 4 and 5. Note that our method extracts salient points not only in the foreground but also in the background where some smooth details are present.

### 3.3 *Implementation*

The main steps to implement the salient points extraction described before are:

- For each wavelet coefficient, find the maximum child coefficient.
- Track it recursively in finer resolutions.
- At the finer resolution (1/2), set the saliency value of the tracked pixel: the sum of the wavelet coefficients tracked.
- Threshold to extract the most prominent points.

Interested readers should see Ref. 18 for details.

### 3.4 *Examples*

In this section, we show the different behavior of the salient point detector comparing it with the corner detector. The wavelets *Haar* and *Daubechies 4* are used for our process.

Two different versions of Harris' detector were implemented: *Harris1* uses a standard derivative, and a local maximum neighborhood of three pixels; *Harris2* uses recursive derivative,[35] and a local maximum neighborhood of five pixels; both use $k=0.04$.

Natural images may contain features that are visually meaningful, and which are not necessarily corners. In Fig. 6 the image contains smoothed edges (the fox fur), which are

**Fig. 5** Spatial support of tracked coefficients.

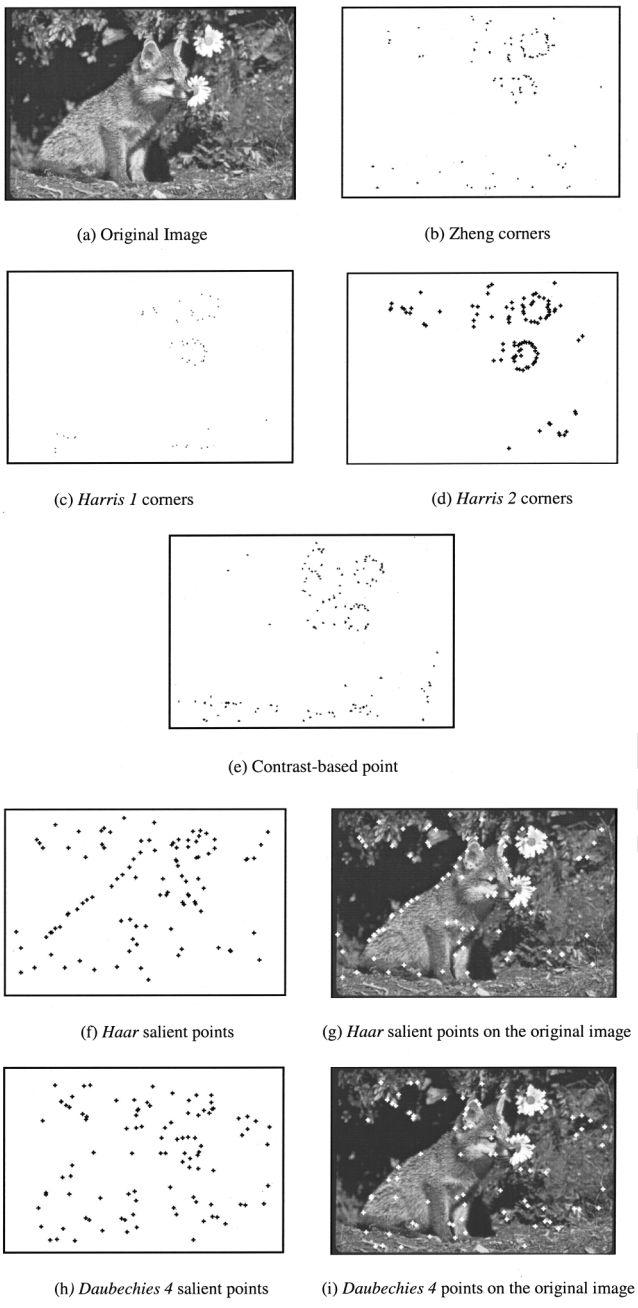



**Fig. 6** (a) A natural image with smoothed edges, and 100 points extracted with various detectors (b)–(i). Points superimposed on the original image are given to evaluate salient points location (g), (i).

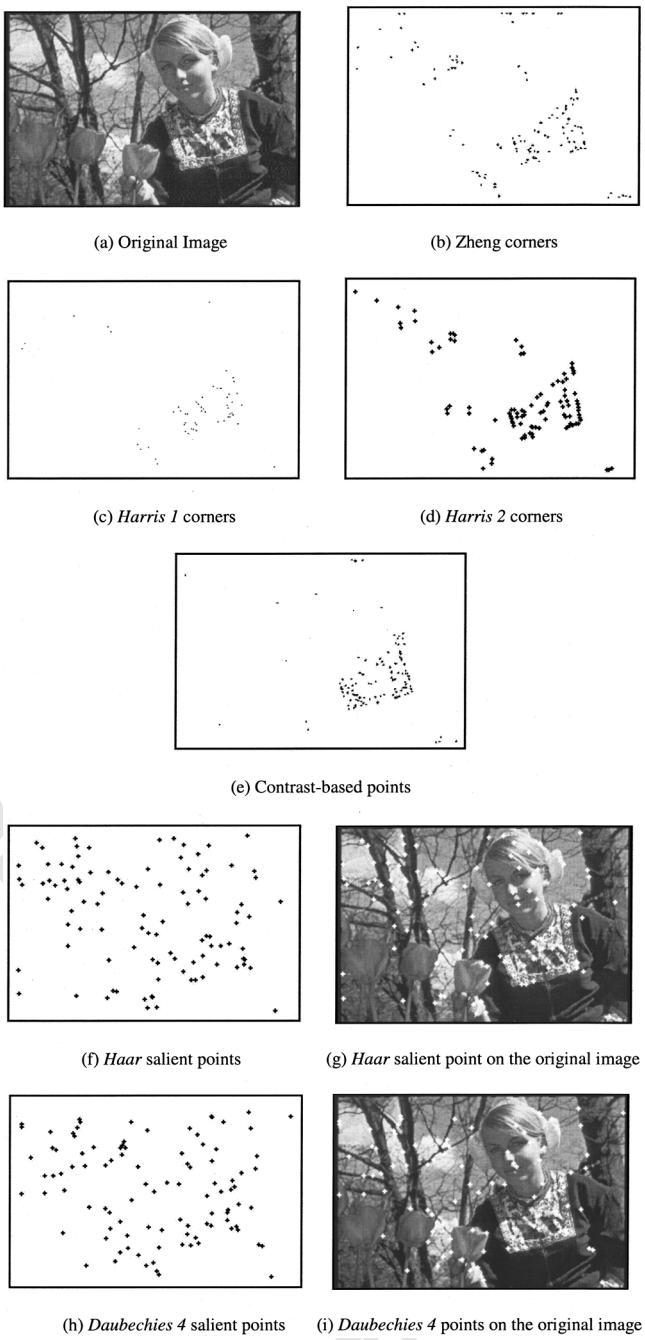
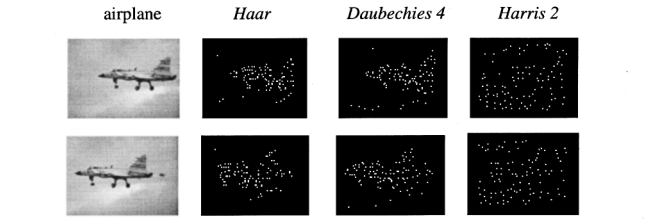



**Fig. 7** (a) A natural image with texture, and 100 points extracted with various detectors (b)–(i). Points superimposed on the original image are given to evaluate salient points location (g), (i).

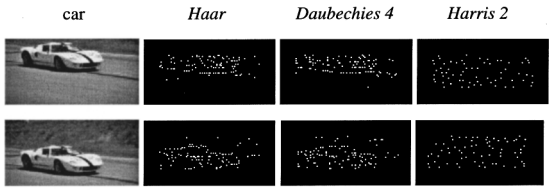
rarely selected by corner detectors, because they are not corners. However, we need to detect points there to include these edges in the image description. Wavelet-based salient points are extracted both in these smoothed edges and finer edges (like the flowers).

In natural images, regions may well contain textures (trees, shirt patterns, etc.). In Fig. 7, many gathered corners are detected in the Dutch dress by corner detectors. Using these points does not lead to a complete description of the image. Wavelet-based salient points are extracted in the dress as well as other parts of the image (face, background).

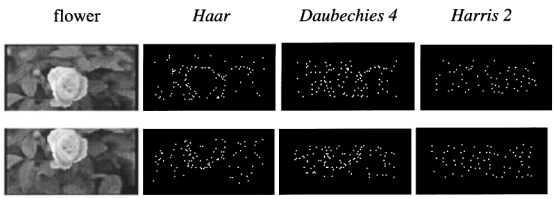
Figure 8 shows examples of the salient points (*Haar* and *Daubechies 4*) and *Harris 2* points under different spatial translations for airplane (left-shifted), car (down-shifted), flower (up-shifted), tiger (upper-left-shifted), and eagle (right-shifted). It is not difficult to find out that most of the salient points still capture the objects (airplane, car, flower, tiger, and eagle) under different spatial translation while the *Harris 2* points do not. This fact experimentally demonstrates the spatial translation invariant property of the salient point approach. It is not hard to understand it theoretically because the salient points are taken at the points where the ''multiresolution'' gradient is high and it is spatial translation invariant.



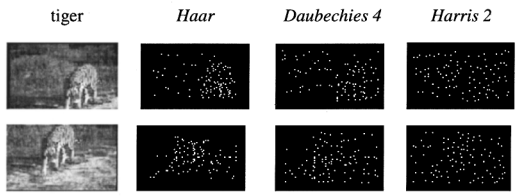

(a) The *Haar*, *Daubechies 4* salient points and *Harris 2* points for the airplane (first row) and its left-shifted version (second row)



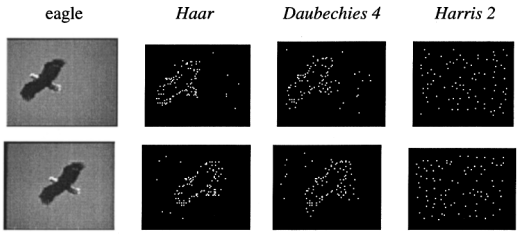

(b) The *Haar*, *Daubechies 4* salient points and *Harris 2* points for the car (first row) and its downshifted version (second row)




(c ) The *Haar*, *Daubechies 4* salient points and *Harris 2* points for the flower (first row) and its up-shifted version (second row)




(d)The *Haar*, *Daubechies 4* salient points and *Harris 2* points for the tiger (first row) and its upper-left-shifted version (second row)




(e) The *Haar*, *Daubechies 4* salient points and *Harris 2* points for the airplane (first row) and its right-shift version (second row)

**Fig. 8** Examples of salient points (*Haar* and *Daubechies 4*) and *Harris 2* points under different spatial translations. (a) The *Haar, Daubechies 4* salient points and *Harris 2* points for the airplane (first row) and its left-shifted version (second row). (b) The *Haar, Daubechies 4* salient points and *Harris 2* points for the car (first row) and its downshifted version (second row). (c) The *Haar, Daubechies 4* salient points and *Harris 2* points for the flower (first row) and its upshifted version (second row). (d) The *Haar, Daubechies 4* salient points and *Harris 2* points for the tiger (first row) and its upper-left-shifted version (second row). (e) The *Haar, Daubechies 4* salient points and *Harris 2* points for the airplane (first row) and its right-shift version (second row).

## 4 Image Retrieval Using Salient Points

### 4.1 *Color Features*

Of the visual media retrieval methods, color indexing is one of the dominant methods because it has been shown to be effective in both the academic and commercial arenas. In color indexing, given a query image, the goal is to retrieve all the images whose color compositions are similar to the color composition of the query image. In color indexing, color histograms are often used because they have sufficient accuracy.[36] While histograms are useful because they are relatively insensitive to position and orientation changes, they do not capture the spatial relationship of color regions, and thus they have limited discriminating power. Stricker and Prengo[37] showed that characterizing one dimensional color distributions with the first three moments is more robust and more efficient than working with color histograms.

The idea of using color distribution features for color indexing is simple. In the index we store dominant features of the color distributions. The retrieval progress is based on similarity function of color distributions. The mathematical foundation of this approach is that any probability distribution is uniquely characterized by its moments. Thus, if we interpret the color distribution of an image as a probability distribution, then the color distribution can be characterized by its moments.[37] Furthermore, because most of the information is concentrated on the low-order moments, only the first moment (mean), the second and the third central moments (variance and skewness) were used. If the value of the $i$th color channel at the $j$th image pixel is $I_{ij}$ and the number of image pixels is $N$, then the index entries related to this color channel are

$$\mu_i=\frac{1}{N}\sum_{j=1}^{N} I_{ij}, \quad \sigma_i=\left(\frac{1}{N}\sum_{j=1}^{N}(I_{ij}-\mu_i)^2\right)^{1/2}, \qquad s_i=\left(\frac{1}{N}\sum_{j=1}^{N}(I_{ij}-\mu_i)^3\right)^{1/3}. \tag{5}$$

We were working with the HSV color space so, for each image in the database a nine-dimensional color feature vector was extracted and stored off-line.

### 4.2 *Texture Features*

Color indexing is based on the observation that often color is used to encode functionality (sky is blue, forests are green) and in general will not allow us to determine an object's identity.[38] Therefore, texture or geometric properties are needed to identify the object.[39] Consequently, color indexing methods are bound to retrieve false positives, i.e., images, which have a similar color composition as the query image but with a completely different content. Therefore, in practice, it is necessary to combine color indexing with texture and/or shape indexing techniques.

Texture analysis is important in many applications of computer image analysis for classification, detection, or

segmentation of images based on local spatial patterns of intensity or color. Textures are replications, symmetries, and combinations of various basic patterns or local functions, usually with some random variation. Textures have the implicit strength that they are based on intuitive notions of visual similarity. This means that they are particularly useful for searching visual databases and other human computer interaction applications. However, since the notion of texture is tied to the human semantic meaning, computational descriptions have been broad, vague, and somewhat conflicting.

The method of texture analysis chosen for feature extraction is critical to the success of texture classification. Many methods have been proposed to extract texture features either directly from the image statistics, e.g., cooccurrence matrix, or from the spatial frequency domain.[40] Ohanian and Dubes[41] studied the performance of four types of features: Markov random fields parameters, Gabor multichannel features, fractal-based features, and cooccurrence features. Comparative studies to evaluate the performance of some texture measures were made.[42,43]

Recently there was a strong push to develop multiscale approaches to the texture problem. Smith and Chang[44] used the statistics (mean and variance) extracted from the wavelet subbands as the texture representation. To explore the middle-band characteristics, tree-structured wavelet transform was studied by Chang and Kuo.[45] Ma and Manjunath[46] evaluated the texture image annotations by various wavelet transform representations, including orthogonal and biorthogonal, tree-structured wavelet transforms, and Gabor wavelet transform. They found out that Gabor transform was the best among the tested candidates, which matched the human vision study results.[47]

Gabor filters produce spatial-frequency decompositions that achieve the theoretical lower bound of the uncertainty principle. They attain maximum joint resolution in space and spatial frequency bounded by the relations: $\Delta_x^2 \cdot \Delta_u^2 \geqslant 1/4\pi$ and $\Delta_y^2 \cdot \Delta_v^2 \geqslant 1/4\pi$, where $[\Delta_x^2, \Delta_y^2]$ gives resolution in space and $[\Delta_u^2, \Delta_v^2]$ gives resolution in spatial frequency. In addition to good performances in texture discrimination and segmentation, the justification for Gabor filters is also supported through psychophysical experiments. Texture analyzers implemented using 2D Gabor functions produce a strong correlation with actual human segmentation.[48] Furthermore, the receptive visual field profiles are adequately modeled by 2D Gabor filters.[49] The Gabor filter masks can be considered as orientation and scale tunable edge and line detectors.

Another texture representation is wavelet-based texture.[44] The original image is fed into a wavelet filter bank and is decomposed into ten decorrelated subbands. Each subband captures the characteristics of a certain scale and orientation of the original image. For each subband, we extract the standard deviation of the wavelet coefficients and therefore have a texture feature vector of length 10.

### 4.3 *Similarity Measurement*

The image similarity is a fuzzy concept, which must be clarified. For the user, the implicit image similarity is usually based on the human perceptual similarity. However, this kind of descriptions cannot be extracted automatically from the image without specific knowledge. Image similarity is therefore mainly based on low-level features such as color and texture.

Let a distance measure between two images $g$ and $h$ written as $D(g,h)$. Let $g_i$ denote the $i$th feature vector of image $g$. The features similarity distance can be compactly written as

$$S_i(g,h) = \|g_i - h_i\|^2 \quad i = 1,\cdots,L, \tag{6}$$

where $L$ is the total number of features, e.g., $L=2$ and $i=1$ for color and $i=2$ for texture. The overall similarity distance between images $g$ and $h$ is

$$D(g,h) = \sum_{i=1}^{L} W_i \cdot S_i(g,h). \tag{7}$$

The low-level feature weights $W_i$ for color and texture in Eq. (7) are set to be equal, e.g., $W_i=1$, $i=1,...,L$.

If the number of the query images is greater than one and a preference weight for each query image is assigned, (e.g., how much the user likes each query image), the weights in Eq. (7) will be automatically adjusted during the relevance feedback.[50]

### 4.4 *Experiments*

The setup of our experiments is as follows. First we extracted a fixed number of salient points for each image in the database using *Haar* wavelet transform and the algorithm described in Sec. 3. The number of salient points cannot be too small or too large. According to our experiments, 50–100 is a reasonable range for the number of salient points. The number of the extracted salient points is 50 in the following experiments. Figure 8 shows some salient point examples for COREL images. The COREL image database contains more than 17 000 images. It covers a wide range of more than 500 categories ranging from animals and birds to Tibet and the Czech Republic. The original images are shown in Fig. 9(a). Their salient point maps are shown in Fig. 9(b). It is not hard to find out that most salient points are located at the boundary and inside of the objects (e.g., bird, airplane, flower, tiger, and car). Fewer points are located at the background, e.g., sky, cloud, green leaves, rocks, etc.

For feature extraction, we considered the pixels in a small neighborhood around each salient point that form the image signature. For each image signature in the database we computed the color moments for color and the Gabor moments for texture. In this paper, the $3\times3$ neighborhood for color features extraction and the $9\times9$ neighborhood for texture features extraction was used. For convenience, this approach is denoted as the *salient approach*.

When the user selects one query image, the system computes the corresponding feature vector from the query image signature and compares it with the feature vector of every image in the database. In the MARS[19,20] system, the color moments[37] and wavelet moments[44] were extracted from the entire images to form the feature vectors. For convenience, this approach is denoted as the *global CW approach*. Since the local color and Gabor feature were extracted in the salient approach, as a benchmark, we also

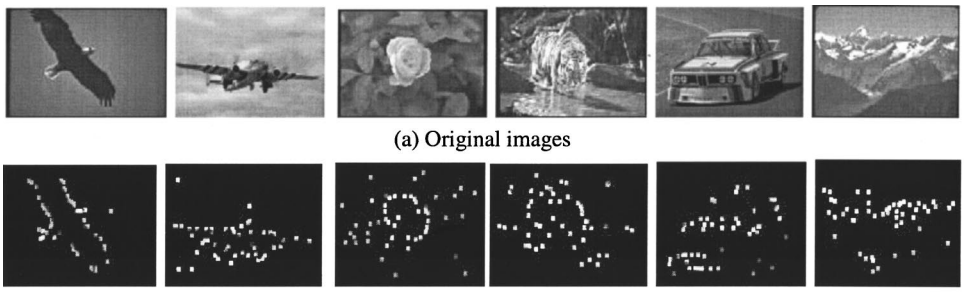



**Fig. 9** Examples of salient points for COREL images: (a) original images, and (b) salient points and their $3\times3$ neighborhood.

considered the results obtained using the color moments and Gabor texture features extracted from the entire image. This approach is denoted as the *global CG approach*. The above-mentioned three approaches will be compared in the following experiments.

In the first experiment we considered a subset of COREL database consisting of 142 images of seven classes such as airplane (21 images), bird (27 images), car (18 images), tiger (18 images), flower (19 images), mountain (19 images), and church paintings (20 images). All images in the database have been labeled as one of these classes and this serves as the ground truth.

Figure 10 shows two examples of the retrieved images in MARS using the salient points. The top left image is the query. The query images are airplane and tiger, respectively. The top 20 retrieved images are shown in Figs. 10(a) and 10(c). The corresponding salient points together with their $3\times3$ neighborhood are shown in Figs. 10(b) and 10(d), respectively. In the example of an airplane, its background is relatively simpler (smaller variations) than the background of the tiger and thus fewer points are located at the background compared to the tiger example. Therefore, satisfied retrieval results are obtained [Figs. 10(a) and 10(b)] while in the example of the tiger, the flowers (wrong classifications) are retrieved, i.e., the 16th and 17th retrieved images. The reason is that the flowers and tiger have similar color information, i.e., yellow, in the neighborhood of salient points.

For retrieval performance evaluation, we randomly picked five images from each class and used them as queries. For each individual class we computed the retrieval accuracy as the average percentage of images from the same class as the query image that were retrieved in the top 15 images. The results are given in Table 1. Since our first testing database is relatively small and the images are clustered (seven classes), only color indexing was used. It should be noted that with more complex features used, the retrieval performance would be improved for both salient approach and the global approach.

Note that for the classes where the background was complex (car, flower), the retrieval accuracy was worse than the other classes. However, in general the salient points capture the details of the foreground objects and therefore the results were better than or comparable to that of using global color moments.

In our second experiment we considered a database of 479 images (size $256\times256$) of color objects such as domestic objects, tools, toys, food cans, etc. As ground truth we used 48 images of eight objects taken from different camera viewpoints (six images for a single object). The problem is formulated as follows:

Let $Q_1,\cdots,Q_n$ be the query images and for the $i$th query $Q_i$, $I_1^{(i)},\cdots,I_m^{(i)}$ be the images similar with $Q_i$ according to the ground truth. The retrieval method will return this set of answers with various ranks. In this experiment both color and texture features were used. Three approaches, the salient approach, the global CW approach and the global CG approach, were compared. Color features (color moments) were extracted from the entire image for the global CW and global CG approaches and from the $3\times3$ neighborhood of each salient point for the salient approach. For texture features, wavelet moments were used for the global CW approach. The wavelet texture feature length was 10. For the salient approach, we extracted the Gabor texture feature from the $9\times9$ neighborhood of each salient point. The dimension of the Gabor filter was $7\times7$. 24 Gabor features were extracted from each neighborhood of the salient points using two scales and six orientations. The first 12 features represented the averages over the filter outputs obtained in order for: scale 1 and orientation 1,$\cdots$, scale 1 and orientation 6, scale 2 and orientation 1,$\cdots$, scale 2 and orientation 6. The last 12 features were the corresponding variances. For the global CG approach, the global Gabor texture features were extracted. The dimension of the global Gabor filter was $61\times61$. 36 Gabor features were extracted using 3 scales and 6 orientations. The first 18 features were the averages over the filter outputs and the last 18 features were the corresponding variances. Note that these color and texture features were independent so that they had different ranges. Therefore each feature was then Gaussian normalized over the entire image database.

We expect the salient point method to be more robust to the viewpoint change because the salient points are located around the object boundary and capture the details inside the object, neglecting the noisy background. In Fig. 11 we represented an example of a query image and the similar images from the database. The salient approach outperforms the global CW approach in capturing the last two images. Even when the image was taken from a very dif-

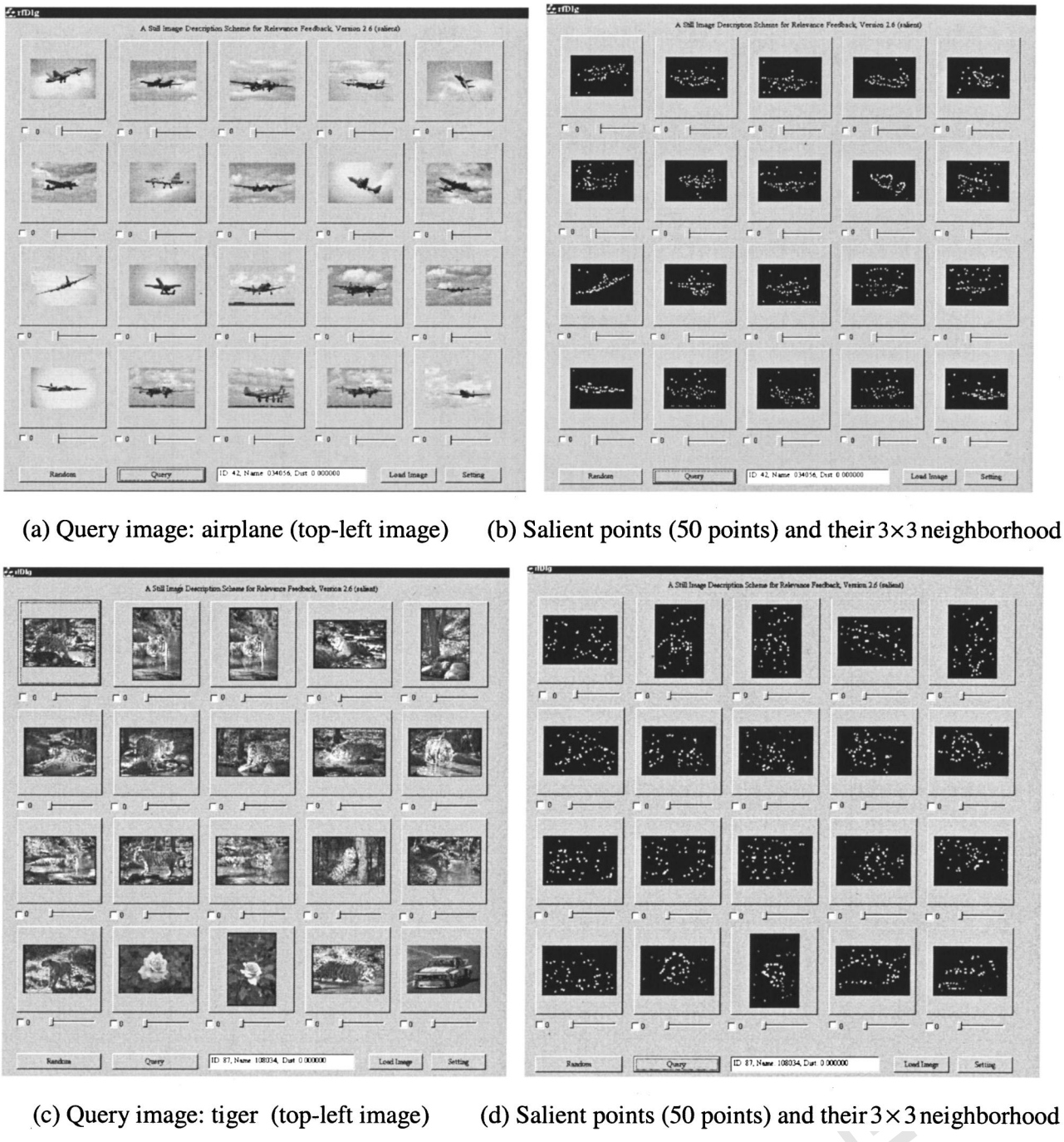



**Fig. 10** User interface: experimental results using the color moments extracted from the 3×3 neighborhood of the salient points (rank from left to right and from top to bottom, the top left is the query image): (a) query image: airplane (top-left image), (b) salient points (50 points) and their 3×3 neighborhood, (c) query image: tiger (top-left image), and (d) salient points (50 points) and their 3×3 neighborhood.

ferent viewpoint, the salient points captured the object details enough so the similar image was retrieved with a good rank. The global CG approach shows better performance than the global CW approach and comparable performance to the salient point approach. This fact demonstrates that the Gabor feature could be a powerful feature for texture classification. However, it should be noted that: (1) the salient point approach only uses the information from a very small part of the image, but still achieves a very good representation of the image, which shows its representing power. For example, in our object database at most 9×9 ×50 pixels were used to represent the image. Compared to the Global approach (all 256×256 pixels were used), only 1/16 of the whole image pixels were used. (2) Compared to the global CG approach, the salient approach is computation efficient. Table 2 shows the computation complexity for the three image databases using the salient point ap-

**Table 1** Retrieval accuracy (%) for each individual class using five randomly chosen images from each class as queries.

| Class | Salient | Global moments |
|---|---|---|
| Airplane | 88 | 86 |
| Bird | 97 | 97 |
| Tiger | 89 | 81 |
| Car | 63 | 49 |
| Flower | 58 | 60 |
| Church painting | 93 | 98 |
| Mountain | 97 | 100 |

**Table 2** Comparison of the computational complexity between the salient approach and the global approach for extracting texture feature using Gabor filters.

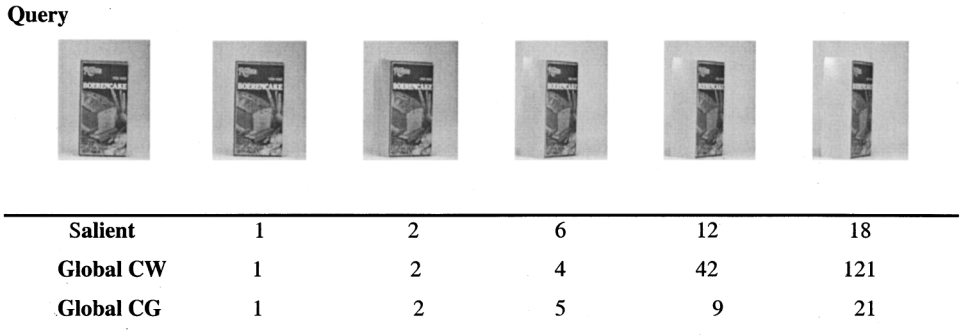
| Database | 1 | 2 | 3 |
|---|---|---|---|
| Description | Object | Natural images | Scenery images |
| Number of images | 479 | 1505 | 4013 |
| Resolution | 256×256 | 384×256 | 360×360 |
| Salient points extraction (min) | 23.9 | 77.5 | 225 |
| Salient Gabor feature extraction (min) | 7.98 | 37.6 | 108 |
| Salient total time (min) | 31.88 | 115.1 | 333 |
| Global Gabor feature extraction (min) | 149 | 928 | 2340 |
| Cost ratio (global/salient) | 4.67 | 8.06 | 7.02 |

proach and the global approach for extracting Gabor texture features. The computation is done on the same SGI O2 R10000 workstation. The total computational cost for the salient approach comes from the two sources. The first source is the time spent on the salient point extraction. The second source is the Gabor texture feature extraction from the neighborhood of the salient points. From Table 2, the average computational complexity ratio of the global approach to the salient approach is about 6.58 (average of 4.67, 8.06, and 7.02) for the three listed image databases. It implies that the computational gain would be huge for the very large database, e.g., millions of images using the salient approach. (3) When the computational complexity of the color feature extraction is compared, the salient points' approach is much faster than the global approach. Color moments were extracted only from the 3×3 neighborhood of the salient points and a small number of salient points were used. Table 3 summarizes the computational gain of color feature extraction for the same three databases used in Table 2. As one can see, the computational gain is very large for the salient approach compared to the global approach.

Table 4 shows the retrieval accuracy of the retrieved images in the top 6, 10, and 20 returned images for the object database mentioned before. Each of the six images from the eight classes was considered as the query image and the average retrieval accuracy was calculated.

Results in Table 4 show that by using the salient point information the retrieval results are significantly improved (>10%) compared to the global CW approach implemented in the MARS.[19,20] When compared to the global CG approach, the retrieval accuracy of the salient approach is slightly (1.9%, 0.4% and 1.5%) lower in the top 6, 10 and 20 returned images, respectively. Although the salient approach is not the best in terms of the retrieval accuracy among the three approaches, it has the best representing power [it only uses a very small part of the whole image pixels but achieves comparable performance to the global CG approach with much less computational complexity (Tables 2 and 3)]. The global wavelet texture features are fast to compute, but their retrieval performance is much worse than the other two methods. Therefore, in terms of overall retrieval accuracy, computational complexity, the salient approach, is considered the best among the three approaches.

In our third experiment, two databases were evaluated. The first database consists of 1505 various natural images. They cover a wide range of natural scenes, animals, buildings, construction sites, textures, and paintings. The second database consists of 4013 various scenes. Most of them are outdoor images like mountains, lakes, buildings, and roads, etc. For the purpose of quantitative analysis, we randomly chose five images from some categories, e.g., building, flower, tiger, road, mountains, forest, and sunset, and used each of the five randomly chosen images as a query. The




| | | | | | |
|---|---|---|---|---|---|
| **Salient** | 1 | 2 | 6 | 12 | 18 |
| **Global CW** | 1 | 2 | 4 | 42 | 121 |
| **Global CG** | 1 | 2 | 5 | 9 | 21 |

**Fig. 11** Example of images of one object taken from different camera viewpoints. The number means the rank of the returned image in order of the decreasing similarities to the query.

**Table 3** The computation complexity of color feature extraction.

| Database | 1 | 2 | 3 |
|---|---|---|---|
| Computational gain (salient/global) | 145 | 218 | 288 |

**Table 4** Retrieval accuracy (%) using 48 images from eight classes for object database.

| Top | 6 | 10 | 20 |
|---|---|---|---|
| Global CW | 47.3 | 62.4 | 71.7 |
| Global CG | 61.2 | 74.2 | 84.7 |
| Salient | 59.3 | 73.8 | 83.2 |

retrieval accuracy was calculated in terms of number of hits, i.e., how many images are similar to the query in the top 20 returned images.

Figure 12 shows the average number of hits for each category using the global CW approach, the global CG approach, and the salient approach. Clearly the salient approach has a similar performance as the global CG approach and outperforms the global CW approach for the first five categories, which are building, flower, tiger, lion, and road. For the last three categories, which are forest, mountain, and sunset, the global approaches (both global CW and global CG) perform better than the salient approach. This is reasonable because the image contents show more global property in the last three categories than the first five categories. Therefore the global approach will result in better performance for these categories. This shows that the salient point performance for image indexing also depends on the image database. The detector choice for a specific database should be investigated in future work.

In our last experiment, we compared the salient point detector with several other point detectors, e.g., the Harris detector,[22] the contrast-based detector,[14] and randomly chosen points. The randomly chosen points were recalculated for each query and were assumed to have a uniform distribution. The choice of random points of different distribution other than uniform distribution has not been investigated yet but will be considered in future work. Figure 12 shows the precision-recall graph, computed from different numbers of returned image $n$ using the Gabor feature for the different points' detector.[51] The system retrieves $r$ images that belong to the same class $C$ as the query ($r \leqslant n$). There are $N_c$ images in the class $C$ of the query. $P=r/n$ is the precision and $R=r/N_c$ the recall for this query. The database consists of 577 various natural images of nine classes (animals, flowers, landscapes, building, cities, etc.). We used each image in the database as a query, and calculated the average recall and precision for the graph (Fig. 13).

The retrieval results in Fig. 13 show that the salient point approach (*Daubechies 4* and *Haar* salient points) performs better than the other point detectors from the literature and the randomly chosen points.

## 5 Conclusions

In this paper, we presented a CBIR system using wavelet-based salient points. The wavelet-based salient points are interesting for image retrieval because they are located in visual focus points, whether they are corner-like or not, without gathering in the textured region, and therefore they can capture the local image information. Two demands were imposed for the salient points extraction algorithm. First, the salient points should be located in any visually interesting part of the image. Second, they should not be clustered in a few regions.

To accomplish these demands we used a *Haar-based* wavelet salient point extraction algorithm that is fast and captures the image details at different resolutions. A fixed number of salient points were extracted for each image. Color moments for color feature and Gabor moments for texture feature were extracted from $3\times3$ and $9\times9$ neighborhood of the salient points, respectively. For the bench-

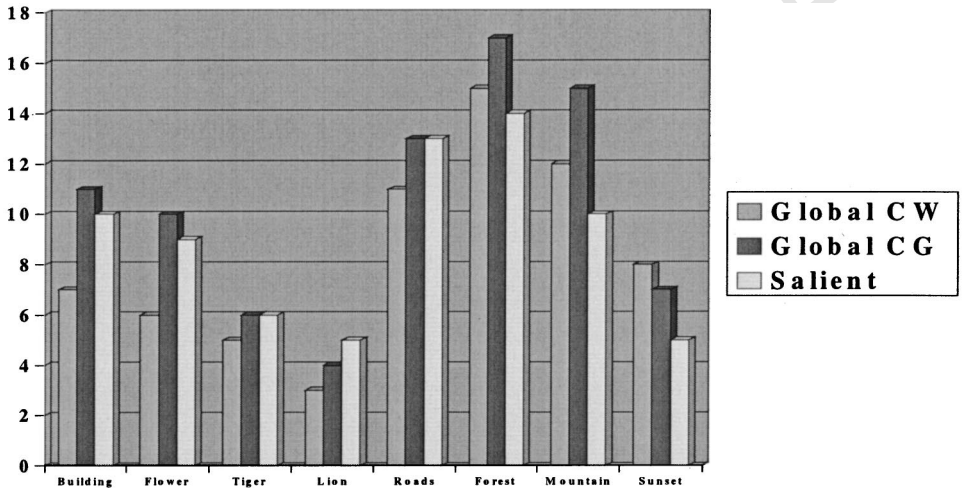



**Fig. 12** The average number of hits for each category using the global color and wavelet moments (global CW), the global color and Gabor moments (global CG), and the salient point approach (salient).

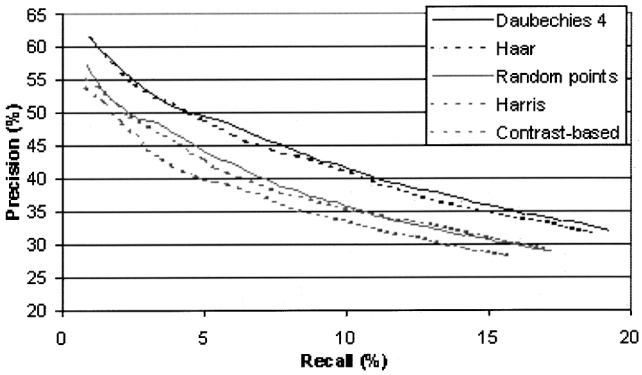



**Fig. 13** Retrieval results.

mark purpose, the salient point approach was compared to the global color and wavelet moments (global CW) approach and the global color and Gabor moments (global CG) approach.

Several experiments were conducted and the results show that: (1) the salient point approach has significantly improved performance compared to the global CW approach. The salient point approach proved to be robust to the viewpoint change because the salient points were located around the object boundaries and captured the details inside the objects, neglecting the background influence. (2) The salient points with Gabor features perform better than the other point detectors (e.g., the Harris detector[22] and contrast-based detector)[14] from the literature and the randomly chosen points. (3) The salient point approach has a comparable performance compared to the global CG approach in terms of the retrieval accuracy but with much less computational cost and using much less image pixels information, which shows the representing power of salient points. In the overall considerations of retrieval accuracy and computational complexity, the salient point approach is considered the best, especially useful for the very large image database.

Our experimental results also show that the Gabor texture features perform much better than the wavelet texture features and show that the Gabor feature is a very powerful candidate for texture classification. This fact is consistent with the results claimed by the other researchers in the field.[52]

In conclusion, the content-based image retrieval can be significantly improved by using the local information provided by the wavelet-based salient points. The salient points are able to capture the local feature information and therefore they can provide a better characterization for object recognition.

There are a number of issues to be addressed in our future work. First, we are going to explore the robustness of the salient points to the white noise, different compression format, changing background and multiple objects in the image. Second, spatial information of the salient points will be integrated into the current system. Grouping salient points will be investigated to allow more advanced spatial and visual similarity retrieval. The number of groups of the salient points could be a good indicator whether an image exhibits more global characteristics. If all or most of the salient points have similar properties, it is very likely that they belong to the same one group. In other words, the image exhibits more global features and it might be more appropriate to consider a global approach rather than the salient point approach. Third, we are going to extract the local shape information from the neighborhood of the salient points and combine the use of the multiple features, e.g., color, texture, and shape to improve the CBIR system performance. Finally, implementation of fast 2D Gabor transform will also be interesting to explore in our future work.

## Acknowledgments

This work was supported in part by National Science Foundation Grant Nos. CDA-96-24396 and EIA-99-75019.

## References

1. S. Chang, J. Smith, M. Beigi, and A. Benitez, ''Visual information retrieval from large distributed online repositories,'' *Commun. ACM* ■, 12–20 (1997).
2. A. Pentland, R. Picard, and S. Sclaroff, ''Photobook: Content-based manipulation of image database,'' *Int. J. Comput. Vis.* ■■, ■■■–■■■ (1996).
3. Y. Rui, T. Huang, and S. Chang, ''Image retrieval: Current techniques, promising directions and open issues,'' *J. Visual Commun. Image Represent* **10**, 1–23 (1999).
4. B. Manjunath and W. Ma, ''Texture features for browsing and retrieval of image data,'' *IEEE Trans. Pattern Anal. Mach. Intell.* ■■, ■■■–■■■ (1996).
5. S. X. Zhou, Y. Rui, and T. S. Huang, ''Water-filling algorithm: A novel way for image feature extraction based on edge maps,'' in *Proc. IEEE Intl. Conf. On Image Proc.*, Japan (1999).
6. M. Popescu and P. Gader, ''Image content retrieval from image database using feature integration by Choquet integral,'' *Storage and Retrieval for Image and Video Databases, Proc. SPIE* ■■, ■■■–■■■ (1998).
7. S. Santini and R. Jain, ''Similarity measures,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(9), ■■■–■■■ (1999).
8. D. Swets and J. Weng, ''Hierarchical discriminant analysis for image retrieval,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(5), 386–400 (1999).
9. H. Zhang and D. Zhong, ''A scheme for visual feature based image retrieval,'' *Storage and Retrieval for Image and Video Databases, Proc. SPIE* ■■, ■■■–■■■ (1995).
10. Y. Rui and T. S. Huang, ''Optimizing learning in image retrieval,'' in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. ■■■–■■■ Hilton Head, SC (June 2000).
11. Q. Tian, Y. Wu, and T. S. Huang, ''Incorporate discriminant analysis with EM algorithm in image retrieval,'' in *Proc. IEEE 2000 International Conference on Multimedia and Expo (ICME'2000)*, pp. ■■■–■■■, New York, NY (July–Aug. 2000).
12. Q. Tian, P. Hong, and T. S. Huang, ''Update relevant image weights for content-based image retrieval using support vector machines,'' in *Proc. IEEE International Conference on Multimedia and Expo (ICME'2000)*, pp. ■■■–■■■, New York, NY (July–Aug. 2000).
13. C. Schmid and R. Mohr, ''Local grayvalue invariants for image retrieval,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(5), 530–535 (May 1997).
14. S. Bres and J. M. Jolion, ''Detection of interest points for image indexation,'' in *Proc 3rd Intl. Conf. on Visual Information Systems, Visual'99*, pp. 427–434, Amsterdam, The Netherlands (June, 1999).
15. T. Tuytelaars and L. Van Gool, ''Content-based image retrieval based on local affinely invariant regions,'' in *Proc. 3rd Intl. Conf. on Visual Information Systems, Visual'99*, pp. 493–500, Amsterdam, The Netherlands (June, 1999).
16. Q. Tian, N. Sebe, M. S. Lew, E. Loupias, and T. S. Huang, ''Content-based image retrieval using wavelet-based salient points,'' *Storage and Retrieval for Media Databases, Proc. SPIE* ■, ■■■–■■■, San Jose, CA (January 2001).
17. N. Sebe, Q. Tian, E. Loupias, M. S. Lew, and T. S. Huang, ''Color indexing using wavelet-based salient points,'' in *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries, IEEE CVPR'2000*, pp. ■■■–■■■, Hilton Head Island, SC (June 2000).
18. E. Loupias and N. Sebe, ''Wavelet-based salient points for image retrieval,'' Report No. RR 99-11, Laboratoire Reconnaissance de Formes et Vision, INSA Lyon (1999).
19. T. S. Huang, S. Mehrotra, and K. Ramchandran, ''Multimedia analysis and retrieval system (MARS) project,'' in *Proc. 33rd Annual Clinic on Library Application of Data Processing—Digital Image Access and Retrieval* (1996).

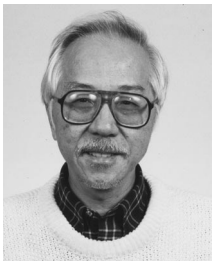
20. Q. Tian, Y. Wu, and T. S. Huang, ''Combine User Defined Region-of-Interest and Spatial Layout for Image Retrieval,'' in *Proc. IEEE 2000 International Conference on Image Processing (ICIP'2000)*, pp. ■■■–■■■, Vancouver, BC, Canada (Sep. 2000).
21. L. Kitchen and A. Rosenfield, ''Gray-level corner detection,'' *Pattern Recogn. Lett.* **1**(2), 95–102 (1982).
22. C. Harris and M. Stephens, ''A combined corner and edge detector,'' in *Proc. of 4th Alvey Vision Conference*, pp. 147–151 (1988).
23. C. Schmid, R. Mohr, and C. Bauckhage, ''Comparing and evaluating interest points,'' in *Proc. International Conference on Computer Vision*, pp. ■■■–■■■ (January, 1998).
24. M. Trajkovic and M. Hedley, ''Fast corner detection,'' *Image Vis. Comput.* **16**(2), 75–87 (1998).
25. Z. Zheng, H. Wang, and E. Teoh, ''Analysis of gray level corner detection,'' *Pattern Recogn. Lett.* **20**, 149–162 (1999).
26. B. Zitova, J. Kautsky, G. Peters, and J. Flusser, ''Robust detection of significant points in multi-frame images,'' *Pattern Recogn. Lett.* **20**, 199–206 (1999).
27. S. Smith and J. Brady, ''SUSAN—A new approach to low level image processing,'' *Int. J. Comput. Vis.* **23**(1), 45–78 (1997).
28. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic, New York (1998).
29. S. Mallat, ''A theory for multiresolution signal decomposition: the wavelet representation,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(7), 674–693 (1989).
30. I. Daubechies, ''Orthonormal bases of compactly supported wavelets,'' *Commun. Pure Appl. Math.* **41**, 909–996 (1988).
31. J.-S. Lee, Y.-N. Sun, and C. H. Chen, ''Multiscale corner detection by using wavelet transform,'' *IEEE Trans. Image Process.* **4**(1), 100–104 (1995).
32. C.-H. Chen, J.-S. Lee, and Y.-N. Sun, ''Wavelet transformation for gray-level corner detection,'' *Pattern Recogn.* **28**(6), 853–861 (1995).
33. S. Bhattacharjee and T. Ebrahimi, ''Image retrieval based on structural content,'' in *Proc. Workshop on Image Analysis for Multimedia Interactive Services*, pp. ■■■–■■■, Heinrich-Hertz-Institut (HHI), Berlin, Germany (May–June, 1999).
34. S. Mallat and S. Zhong, ''Characterization of signals from multiscale edges,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(7), 710–732 (1992).
35. R. Deriche, ''Optimal edge detection using recursive filtering,'' in *Proc. of 1st International Conference on Computer Vision*, pp. 501–505, London (June 1987).
36. M. J. Swain and D. H. Ballard, ''Color indexing,'' *Int. J. Comput. Vis.* **7**(1), 11–32 (1991).
37. M. Stricker and M. Prengo, ''Similarity of color images,'' in *Storage and Retrieval for Image and Video Databases*, *Proc. SPIE* ■■, ■■■–■■■ (1995).
38. M. Stricker and A. Dimai, ''Spectral covariance and fuzzy regions for image indexing,'' in *Storage and Retrieval for Image and Video Databases*, *Proc. SPIE* ■■, ■■■–■■■ (1997).
39. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, ''Query by image and video content: The QBIC system,'' *Computer* **28**(9), 23–32 (1995).
40. L. Van Gool, P. Dewaele, and A. Oosterlinck, ''Texture analysis,'' *Comput. Vis. Graph. Image Process.* **29**(3), 336–357 (1985).
41. P. Ohanian and R. Dubes, ''Performances evaluation for four classes of textural features,'' *Pattern Recogn.* **25**, 819–833 (1992).
42. T. Ojala, M. Pietikainen, and D. Harwood, ''A comparative study of texture measures with classification based on feature distribution,'' *Pattern Recogn.* **29**, 51–59 (1996).
43. T. Reed and H. Wechsler, ''A review of recent texture segmentation and feature extraction techniques,'' *Comput. Vis. Graph. Image Process.* **57**(3), 359–373 (1993).
44. J. Smith and S.-F. Chang, ''Transform features for texture classification and discrimination in large image database,'' in *Proc. IEEE Intl. Conf. on Image Proc.* (1994).
45. T. Chang and C. Kuo, ''Texture analysis and classification with tree-structured wavelet transform,'' *IEEE Trans. Image Process.* **2**(4), 429–441 (1993).
46. Y. Ma and B. Manjunath, ''Texture features of wavelet transform features for texture image annotation,'' in *Proc. IEEE Intl. Conf. on Image Proc.* (1995).
47. J. Beck, A. Sutter, and R. Ivry, ''Spatial frequency channels and perceptual grouping in texture segregation,'' *Comput. Vis. Graph. Image Process.* **37**, ■■■–■■■ (1987).
48. T. Reed and H. Wechsler, ''Segmentation of textured images and gestalt organization using spatial/spatial-frequency representation,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(1), 1–12 (1990).
49. J. Daugman, ''Entropy reduction and decorrelation in visual coding by oriented neural receptive field,'' *IEEE Trans. Biomed. Eng.* **36**(1), ■■■–■■■ (1989).
50. Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, ''Relevance feedback: a power tool for interactive content-based image retrieval,'' *IEEE Trans. Circuits Syst. Video Technol.* **8**, 644–655 (1998).
51. E. Loupias, N. Sebe, S. Bres, and J.-M. Jolion, ''Wavelet-based salient points for image retrieval,'' in *Proc. IEEE Intl. Conf. on Image Proc.* (2000).
52. W. Y. Ma and B. S. Manjunath, ''Texture features and learning similarity,'' in *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)* (1996).


**Qi Tian** received his BE degree in electronic engineering from Tsinghua University, China and his MS degree in electrical engineering from Drexel University in 1992 and 1996, respectively. He was a graduate research and teaching assistant in the Department of Electrical and Computer Engineering at Drexel University from 1994 to 1997 and a graduate research assistant in the Bioaccoustics Research Laboratory of the Beckman Institute for Advanced Science and Technology at University of Illinois at Urbana-Champaign (UIUC) from August 1997 to January 1999. Since February 1999 he has been a graduate research assistant at the Image Formation and Processing Group of the Beckman Institute for Advanced Science and Technology at UIUC. During the summer of 2000 and 2001, he was a summer research intern at the Mitsubishi Electric Research Laboratory (MERL), Cambridge, MA and has one patent filed for summer work. He was a visiting researcher at MERL from February 2001 to May 2001. His current research interests include multimedia information retrieval, digital image/signal processing, and computer vision. He has published over 25 journal and conference papers in the above areas.

**Nicu Sebe** received his PhD degree in computer science from Leiden University in 2001. Since 1997, he is with the Leiden Institute of Advanced Computer Science, The Netherlands, and currently he is doing postdoctoral research. His main interest is in the fields of computer vision and pattern recognition, in particular content-based retrieval and maximum likelihood analysis. He is a member of the IEEE.

**Michael S. Lew** received his PhD in electrical engineering from the University of Illinois at Urbana-Champaign in 1995. Currently, he is an associate professor and academic fellow of the Leiden Institute of Advanced Computer Science (LIACS) at Leiden University in The Netherlands. His distinctions include best research proposal (Netherlands's National Science Foundation, 1996), best presentation (Leiden University, 1998), and he was ranked as one of the top young scientists at Leiden University (2000, Faculty of Science at Leiden University). He has published more than 70 articles in refereed journals and conferences in the fields of multimedia search, computer vision, human computer interaction (HCI), and virtual communities. In addition to research and teaching, he is the co-director of the LIACS Media Lab. He is a member of the IEEE.

**Etienne Loupias** received his PhD degree in computer science from the Institute National des Sciences Appliquees (INSA) Lyon in 2000. During 1999 he was a visiting researcher at the Leiden Institute of Advanced Computer Science. His main interest is in the field of computer vision and pattern recognition, in particular content-based retrieval.



**Thomas S. Huang** (F'79) received his BS degree in electrical engineering from National Taiwan University, Taipei, and his MS and ScD degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge. He was on the Faculty of the Department of Electrical Engineering at M.I.T. from 1963 to 1973, and on the Faculty of the School of Electrical Engineering and Director of the Laboratory for Information and Signal Processing at Purdue University from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, where he is now a William L. Everitt distinguished professor of electrical and computer engineering, and research professor at the Coordinated Science Laboratory, and the head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology. During his sabbatical leaves, he has worked at the M.I.T. Lincoln Laboratory, the IBM Thomas J. Watson Research Center, and the Rheinishes Landes Museum, Bonn, Germany, and held visiting professor positions at the Swiss Institutes of Technology, Zurich and Lausanne, and the University of Hannover, West Germany, INRS-Telecommunications of the University of Quebec, Montreal, Canada, and University of Tokyo, Japan. He has served as a consultant to numerous industrial firms and government agencies both in the U.S. and abroad. His professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 12 books and over 300 papers in network theory, digital filtering, image processing, and computer vision. He is a founding editor of the *International Journal of Computer Vision, Graphics, and Imaging Processing*, and editor of the Springer Series in Information Sciences, published by Springer-Verlag. Dr. Huang is a Fellow of the International Association of Pattern Recognition and the Optical Society of America, and has received a Guggenheim fellowship, an A.V. Humboldt Foundation Senior U.S. Scientist Award, and a fellowship from the Japan Association for the Promotion of Science. He received the IEEE Acoustics, Speech, and Signal Processing Society's Technical Achievement Award in 1987, the Society Award in 1991, recognized as a pioneer in signal processing at 1998 IEEE ICASSP, IEEE Third Millennium Medal in 2000, and Honda Lifetime Achievement Award in 2000.