Evaluation of salient point techniques

Abstract

In image retrieval, global features related to color or texture are commonly used to describe the image content. The problem with this approach is that these global features cannot capture all parts of the image having different characteristics. Therefore, local computation of image information is necessary. By using salient points to represent local information, more discriminant features can be computed. In this paper we compare a wavelet-based salient point extraction algorithm with two corner detectors using the criteria: repeatability rate and information content. We also show that extracting color and texture information in the locations given by our salient points provides significantly improved results in terms of retrieval accuracy, computational complexity, and storage space of feature vectors as compared to global feature approaches.

1 Introduction

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 general, the features are often computed from the entire image. The problem with this approach is that these global features cannot handle all parts of the image having different characteristics. Therefore, local computation of image information is necessary. Local features can be computed at different image scales to obtain an image index based on local properties of the image and they need to be discriminant enough to "summarize" the local image information. These features are too time-consuming to be computed for each pixel in the image and therefore, the feature extraction should be limited to a subset of the image pixels, the interest points [9, 11], 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.

Haralick and Shapiro [4] consider a point in an image *interesting* if has two main properties: distinctiveness and invariance. This means that a point should be distinguishable from its immediate neighbors and the position as well as the selection of the interesting point should be invariant with respect to the expected geometric and radiometric distortions. Schmid and Mohr [9] proposed the use of corners as interest points in image retrieval. Different corner detectors are evaluated and compared in [10] and the authors show that the best results are provided by the Harris corner detector [5].

Corner detectors, however, were designated for robotics and shape recognition and they have drawbacks when are applied to natural images. Visual focus points do not need 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 only in corner points, as mathematically defined in most corner detectors. For instance, a smoothed edge may have 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. Corners also gather in textured regions. The problem is that due to efficiency reasons only a preset number of points per image can be used in the indexing process. Since in this case most of the detected points will be in a small region, the other parts of the image may not be described in the index at all.

Therefore, we aim for a set of interesting points called *salient points* that are related to any visual interesting part of the image whether it is smoothed of corner-like. Moreover, to describe different parts of the image the set of salient points should not be clustered in few regions. We believe multi-resolution representation is interesting to detect salient points. We present a salient point extraction algorithm that uses the wavelet transform, which expresses image variations at different resolutions. Our wavelet-based salient points are detected for smoothed edges and are not gathered in texture regions. Hence, they lead to a more complete image representation than corner detectors.

We also compare our wavelet-based salient point detector with the Harris corner detectors used by Schmid and Mohr [10]. In order to evaluate the "interestingness" of the points obtained with these detectors (as was introduced by Haralick and Shapiro [4]) we compute the repeatability rate and the information content. The repeatability rate evaluates the geometric stability of points under different image transformation. Informa-

tion content measures the distinctiveness of a greylevel pattern at an interest point. A local pattern here is described using rotationally invariant combinations of derivatives and the entropy of these invariants is measured for each set of interest points.

We are also interested in using the salient points in a retrieval scenario. Therefore, in a small neighborhood around the location of each point we extract local color and texture features and use only this information in retrieval. It is quite easy to understand that using a small amount of such points instead of all image pixels reduces the amount of data to be processed. Moreover, local information extracted in the neighborhood of these particular points is assumed to be more robust to classic transformations (additive noise, affine transformation including translation, rotation, and scale effects, partial visibility, etc.).

2 Wavelet-based Salient Points

The wavelet representation gives information about the variations in the image at different scales. We would like to extract salient points from any part of the image where something happens 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.

A wavelet is an oscillating and attenuated function with zero integral. We study the image f at the scales (or resolutions) $1/2$, $1/4$, ..., 2^j, $j \in Z$ and $j \leq -1$. 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 non-redundant representation of the image. Second, 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:

$$C(W_{2^j} f(n)) = \{W_{2^{j+1}} f(k), 2n \leq k \leq 2n + 2p - 1\} \quad (1)$$

where p is the wavelet regularity and $0 \leq n < 2^j N$ with N the length of the signal.

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.

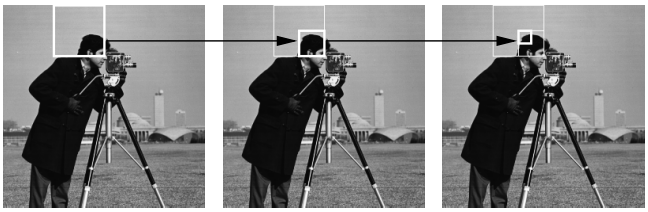


Figure 1: Salient points extraction: spatial support of tracked coefficients

Applying recursively this process, we select a coefficient $W_{2^{-1}} f(n)$ at the finer resolution $1/2$. 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 (Figure 1). We set its saliency value as the sum of the absolute value of the wavelet coefficients in the track:

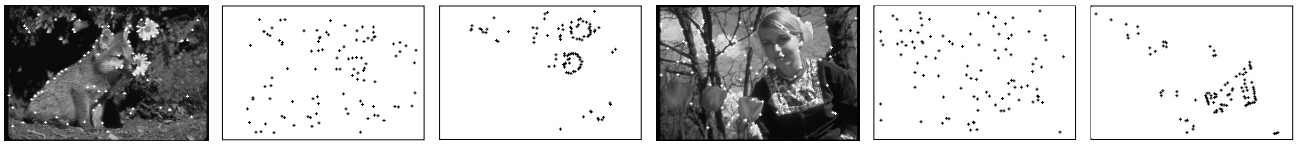
$$saliency = \sum_{k=1}^{-j} |C^{(k)}(W_{2^j} f(n))|, -\log_2 N \leq j \leq -1 \quad (2)$$

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 desired 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 wavelet function, 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 non-overlapping wavelets at a given scale. This can be avoided with the simplest overlapping wavelet, Daubechies4 [2]. Examples of salient points extracted using Daubechies4, Haar, and Harris detectors are shown in Figure 2. Note that while for Harris the salient points lead to an incomplete image representation, for the other two detectors the salient points are detected for smooth edges (as can be seen in the fox image) and are not gathered in texture regions (as can be seen in the girl image). Hence, they lead to a more complete image representation.

3 Repeatability and Information Content

Repeatability is defined by the image geometry. Given a 3D point P and two projection matrices M_1 and M_2, the projections of P into two images I_1 and I_2 are $p_1 = M_1 P$ and $p_2 = M_2 P$. The point p_1 detected in image I_1 is repeated in image I_2 if the corresponding point p_2 is detected in image I_2. To measure the repeatability, a unique relation between p_1 and p_2 has to be established. In the case of a planar scene this relation is defined by an homography: $p_2 = H_{21} p_1$.



Daubechies4 Haar Harris Daubechies4 Haar Harris

Figure 2: Salient points examples. For Daubechies4 and Haar salient points are detected for smooth edges (fox image) and are not gathered in textured regions (girl image).

The percentage of detected points which are repeated is the *repeatability rate*. A repeated point is not in general detected exactly at position p_2, but rather in some neighborhood of p_2. The size of this neighborhood is denoted by ε and repeatability within this neighborhood is called *ε-repeatability*. The set of point pairs (d_2, d_1) which correspond within an ε-neighborhood is $D(\varepsilon) = \{(d_2, d_1) | dist(d_2, H_{21} d_1 < \varepsilon\}$. In this conditions, the repeatability rate is given by:

$$r(\varepsilon) = \frac{|D(\varepsilon)|}{N} \tag{3}$$

where N is the total number of points detected. One can easily verify that $0 \leq r(\varepsilon) \leq 1$.

We would also like to know how much average information content a salient point "has" as measured by its greylevel pattern. The more distinctive the greylevel patterns are, the larger the entropy is. In order to have rotation invariant descriptors for the patterns, we chose to characterize salient points by local greyvalue rotation invariants which are combinations of derivatives. We computed the "local jet" [6] which is consisted of the set of derivatives up to N^{th} order. These derivatives describe the intensity function locally and are computed stably by convolution with Gaussian derivatives. The local jet of order N at a point $\mathbf{x} = (x, y)$ for an image I and a scale σ is defined by: $J^N[I](\mathbf{x}, \sigma) = \{L_{i_1 \ldots i_n}(\mathbf{x}, \sigma) | (\mathbf{x}, \sigma) \in I \times R^+\}$, where $L_{i_1 \ldots i_n}(\mathbf{x}, \sigma)$ is the convolution of image I with the Gaussian derivatives $G_{i_1 \ldots i_n}(\mathbf{x}, \sigma)$, $i_k \in \{x, y\}$ and $n = 0, \ldots, N$.

In order to obtain invariance under the group $SO(2)$ (2D image rotation), Koenderink and van Doorn [6] compute differential invariants from the local jet:

$$\vec{\nu}[0 \ldots 3] = \begin{bmatrix} L_x L_x + L_y L_y \\ L_{xx} L_x L_x + 2 L_{xy} L_x L_y + L_{yy} L_y L_y \\ L_{xx} + L_{yy} \\ L_{xx} L_{xx} + 2 L_{xy} L_{xy} + 2 L_{yy} L_{yy} \end{bmatrix} \tag{4}$$

The computation of entropy requires a partitioning of the space of $\vec{\nu}$. Partitioning is dependent on the distance measure between descriptors and we consider the approach described by Schmid, et al. [10]. The distance we used is the Mahalanobis distance given by: $d_M(\vec{\nu}_1, \vec{\nu}_2) = \sqrt{(\vec{\nu}_1 - \vec{\nu}_2)^T \Lambda^{-1} (\vec{\nu}_1 - \vec{\nu}_2)}$, where $\vec{\nu}_1$ and $\vec{\nu}_2$ are two descriptors and Λ is the covariance of $\vec{\nu}$. The covariance matrix Λ is symmetric and positive definite. Its inverse can be decomposed into $\Lambda^{-1} = P^T D P$ where D is diagonal and P an orthogonal matrix. Furthermore, we can define the square root of Λ^{-1} as $\Lambda^{-1/2} = D^{1/2} P$ where $D^{1/2}$ is a diagonal matrix whose coefficients are the square roots of the coefficients of D. The Mahalanobis distance can then be rewritten as: $d_M(\vec{\nu}_1, \vec{\nu}_2) = ||D^{1/2} P (\vec{\nu}_1 - \vec{\nu}_2)||$. The distance d_M is the norm of difference of the normalized vectors: $\vec{\nu}_{norm} = D^{1/2} P \vec{\nu}$. This normalization allows us to use equally sized cells in all dimensions. This is important since the entropy is directly dependent on the partition used. The probability of each cell of this partition is used to compute the entropy of a set of vectors $\vec{\nu}$.

In the experiments we used a set of 1000 images taken from the Corel database and we compared 4 salient point detectors. In Section 2 we introduced two salient point detectors using wavelets: Haar and Daubechies4. For benchmarking purposes we also considered the Harris corner detector [5] and a variant of it called PreciseHarris, introduced by Schmid, et al. [10]. The difference between the last two detectors is given by the way the derivatives are computed. Harris computes derivatives by convolving the image with the mask [-2 -1 0 1 2] whereas PreciseHarris uses derivatives of the Gaussian function instead.

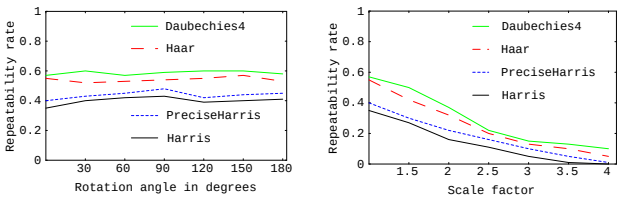
3.1 Results for repeatability

Before we can measure the repeatability of a particular detector we first had to consider typical image alterations such as image rotation and image scaling. In both cases, for each image we extracted the salient points and then we computed the average repeatability rate over the database for each detector.

In the case of image rotation the rotation angle varied between 0° and 180°. The repeatability rate in a ε=1 neighborhood for the rotation sequence is displayed in Figure 3(a).

The detectors using wavelet transform (Haar and Daubechies4) give better results compared with the other ones. Note that the results for all detectors are not very dependent on image rotation. The best results are provided by Daubechies4 detector.

In the case of scale changes, for each image we considered a sequence of images obtained from the original image by reducing the image size so that the image was aspect ration preserved. The largest scale factor used



(a) (b)

Figure 3: Repeatability rate for image rotation (a) and scale change (b) (ε=1)

was 4. The repeatability rate for scale change is presented in Figure 3(b).

All detectors are very sensitive to scale changes. The repeatability is low for a scale factor above 2 especially for Harris and PreciseHarris detectors. The detectors based on wavelet transform provide better results compared with the other ones.

3.2 Results for information content

In these experiments we also considered random points in our comparison [7]. For each image in the database we computed the mean number m of salient points extracted by different detectors and then we selected m random points using a uniform distribution.

For each detector we computed the salient points for the set of images and characterized each point by a vector of local greyvalue invariants (cf. Eq. (4)). The invariants were normalized and the entropy of the distribution was computed. The cell size in the partitioning was the same in all dimensions and it was set to 20. The σ used for computing the greylevel invariants was 3.

The results are given in Table 1. The detector using the Daubechies4 wavelet transform has the highest entropy and thus the salient points obtained are the most distinctive. The results obtained for Haar wavelet transform are almost as good. The results obtained with PreciseHarris detector are better than the ones obtained with Harris but worse than the ones obtained using the wavelet transform. Unsurprisingly, the results obtained for all of the salient points detectors are significantly better than those obtained for random points. The difference between the results of Daubechies4 and random points is about a factor of two.

Detector	Entropy
Haar	6.0653
Daubechies4	6.1956
Harris	5.4337
PreciseHarris	5.6975
Random	3.124

Table 1: The information content for different detectors

In summary, the most "interesting" salient points were detected using the Daubechies4 detector. These points have the highest information content and proved to be the most robust to rotation and scale changes. Therefore, in our next experiments we will consider this detector and as benchmark the PreciseHarris corner detector.

4 Content-based retrieval

Our next goal is to use the salient points in a content-based retrieval scenario. We consider a modular approach: the salient points are first detected for each image in the database and then feature vectors are extracted from a small neighborhood around each salient point. This approach assures the independence of the salient point extraction techniques and the feature extraction procedure and gives the user the liberty to use any features he wants for a specific application. In our experiments we use color moments and Gabor texture features to construct the feature vectors for each image. Of course the wavelet coefficients used during the salient point detection can be also used in constructing the feature vectors.

The number of salient points extracted will clearly influence the retrieval results. We performed experiments (not presented here due to space limitation) in which the number of salient points varied from 10 to several hundreds and found out that when using more than 50 points, the improvement in accuracy we obtained did not justify the computational effort involved. Therefore, in the experiments, 50 salient points were extracted for each image.

4.1 Features

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, as well [13]. Moreover, Stricker, et al. [13] showed that characterizing one dimensional color distributions with the first three moments is more robust and more efficient than working with color histograms. Because most of the information is concentrated on the low-order moments, we consider only the first moment (mean), the second moment (variance), and the third central moment (skewness) for each color channel (the HSV color space was used).

Color by itself is used to encode functionality (sky is blue, forests are green) and in general will not allow us to determine the identity of an object. Therefore, in practice, it is necessary to combine color features with texture and/or shape features. Recently there was a strong push to develop multi-scale approaches to the texture problem. Smith and Chang [12] used the statistics (mean and variance) extracted from the wavelet sub-

bands as texture features. Ma and Manjunath [8] evaluated the texture image annotations by various wavelet transform representations and found out that Gabor wavelet transform was the best among the tested candidates, which matched the human vision study results [1]. In addition to good performances in texture discrimination and segmentation, the justification for Gabor filter is also supported through psychophysical experiments.

4.2 Setup

For feature extraction, we considered the set of pixels in a small neighborhood around each salient point. In this neighborhood we computed the color moments (in a 3×3 neighborhood) and the Gabor moments (in a 9×9 neighborhood). For convenience, this approach is denoted as the Salient W (wavelet) approach when Daubechies4 detector is used and as the Salient C (corner) approach when the PreciseHarris corner detector is used. For benchmarking purposes we also considered the results obtained using the color moments and the wavelet moments [12] extracted over the entire image (denoted as Global CW approach) and the results obtained using the color moments and the Gabor moments [8] extracted over the entire image (denoted as Global CG approach).

The overall similarity distance D_j for the j^{th} image in the database is obtained by linearly combining the similarity distance of each individual feature, $S_j(f_i)$:

$$D_j = \sum_i W_i S_j(f_i) \qquad j = 1, \ldots, N \tag{5}$$

where N is the total number of images in the database and $S_j(f_i)$ is defined as:

$$S_j(f_i) = (\mathbf{x}_i - \mathbf{q}_i)^T (\mathbf{x}_i - \mathbf{q}_i) \tag{6}$$

where $\mathbf{x}_i$ and $\mathbf{q}_i$ are the i^{th} feature (e.g. $i = 1$ for color and $i = 2$ for texture) vector of the j^{th} image in the database and the query, respectively. The low-level feature weights W_i for color and texture in Eq. (5) are set to be equal.

4.3 Results

In the first experiment we used the same 1000 images from the Corel database as in Section 3. As ground truth we used 146 images divided in 7 classes: airplane (25 images), bird (27 images), car (18 images), tiger (18 images), flower (19 images), mountain (19 images), and church paintings (20 images). Figure 4 shows the results for airplane class using the color moments extracted from the 3×3 neighborhood of the salient points.

In order to test the retrieval results for each individual class, we randomly picked 15 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. Only the color feature was used. Thus the comparison was between the Salient W and C approaches and the Global approach. The results are given in Table 2.

Figure 4: An experimental result using the color moments extracted from the 3×3 neighborhood of the Daubechies4 salient points (rank from the left to right and from top to bottom, the top left is the query image)

Class	Salient W	Salient C	Global
Airplane	100	93.33	93.33
Bird	93.33	93.33	86.66
Tiger	86.66	86.66	80
Car	73.33	66.66	60
Flower	73.33	66.66	60
Church painting	93.33	93.33	93.33
Mountain	93.33	93.33	93.33

Table 2: Retrieval accuracy (%) for each individual class using 15 randomly chosen images from each class as queries

From this experiment we can see that for some classes that are mainly composed of a single object on a simple background (e.g., Bird or Airplane where the background represented the blue sky), the Salient and the Global approaches have similar performances. For the Global approach, the color moments were extracted from the entire image and therefore, the color information was mainly determined by the dominant background, e.g., blue sky. In this sense, the birds were found to be similar because of the background, not the objects themselves. In the Salient W approach, the salient points were mostly found on the boundaries of the objects. The local color moments around the neigh-

borhood of the salient points were extracted and they represented the object information instead of the background. Here, the images were found to be similar in terms of the object, not the background. Similar things were happening with the Salient C approach. However, the results are slightly worse than in the case of Salient W approach. In summary, the Salient approaches capture more accurately the user's concept than the Global approach in terms of object finding. When the classes have complex background (e.g., Tiger, Car) the retrieval conditions are more difficult and the Salient approaches perform much better than the Global approach. When the images show more global variations, (e.g., Church Painting, Mountain), all approaches perform very well showing that the Salient approaches can still capture global image information (background) as well.

In our second experiment we considered a database of 479 images (256×256 pixels in size) of color objects such as domestic objects, tools, toys, food cans, etc [3]. As ground truth we used 48 images of 8 objects taken from different camera viewpoints (6 images for a single object). Both color and texture information were used. The Salient approaches, the Global CW approach, and the Global CG approach were compared. Color moments were extracted either globally (the Global CW and Global CG) or locally (the Salient approaches). For wavelet texture representation of the Global CW approach, each input image was first fed into a wavelet filter bank and was decomposed into three wavelet levels, thus 10 de-correlated subbands. For each subband, the mean and standard deviation of the wavelet coefficients were extracted. The total number of wavelet texture features was 20. For the Salient approaches, we extracted Gabor texture features from the 9×9 neighborhood of each salient point. The dimension of the Gabor filter was 7×7 and we used 2 scales and 6 orientations/scale. The first 12 features represented the averages over the filter outputs and the last 12 features were the corresponding variances. Note that these features were independent so that they had different ranges. Therefore, each feature was then Gaussian normalized over the entire image database. For the Global CG approach, the global Gabor texture features were extracted. The dimension of the global Gabor filter was 61×61. We extracted 36 Gabor features using 3 scales and 6 orientations/scale. The first 18 features were the averages over the filters outputs and the last 18 features were the corresponding variances.

We expect the salient point methods 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 Figure 5 we show an example of a query image and the similar images from the database retrieved with various ranks.

Salient W	1	2	4	15	22
Salient C	1	3	6	17	25
Global CW	2	7	12	25	33
Global CG	2	5	9	20	27

Figure 5: Example of images of one object taken from different camera viewpoints and the corresponding ranks of each individual image using different approaches

The Salient point approaches outperform both the Global CW approach and the Global CG approach. Even when the image was taken from a very different viewpoint, the salient points captured the object details enough so the similar image was retrieved with a good rank. The Salient W approach again shows better retrieval performance than the Salient C approach. The Global CG approach provides better performance than the Global CW approach. This fact demonstrates that Gabor feature is a very good feature for texture characterization. Moreover, it should also be noted that: (1) the Salient point approaches only use the information from a very small part of the image, but still achieve a good representation of the image. For example, in our object database $9 \times 9 \times 50$ pixels were used to represent the image. Compared to the Global approaches (all 256×256 pixels were used), they only use less than 1/16 of the whole image pixels. (2) Compared to the Global CG approach, the Salient approaches have much less computational complexity (see Table 4).

Table 3 shows the retrieval accuracy for the object database. Each of the 6 images from the 8 classes was considered as query image and the average retrieval accuracy was calculated.

Top	6	10	20
Salient W	61.2	74.2	84.7
Salient C	59.7	73.8	83.2
Global CW	47.3	62.4	71.7
Global CG	58.3	73.4	82.8

Table 3: Retrieval accuracy (%) using 48 images from 8 classes for object database

Results in Table 3 show that using the salient point information the retrieval results are significantly improved (>10%) compared to the Global CW approach. When compared to the Global CG approach, the retrieval accuracy of salient approach is up to 2.9%, 0.8%,

and 1.9% higher in the top 6, 10, and 20 images, respectively. The Salient approaches have therefore, similar performance comparing with the Global CG approach but much lower computational complexity (see Table 4 for texture and Table 5 for color) and 33.3% less storage space of feature vectors than the Global CG approach. Although the global wavelet texture features are fast to compute, their retrieval performance is much worse than the other methods. Therefore, in terms of overall retrieval accuracy, computational complexity, and storage space of feature vectors, the Salient approaches are the best among all the approaches.

In our third experiment we considered a database consisted of 4013 various images covering a wide range of natural scenes such as animals, buildings, paintings, mountains, lakes, and roads. In order to perform quantitative analysis, we randomly chose 15 images from a few categories, e.g., building, flower, tiger, lion, road, forest, mountain, sunset and use each of them as queries. For each category, we measured how many hits, i.e. how many similar images to the query were returned in the top 20 retrieved images.

Figure 6 shows the average number of hits for each category using the Global CW approach, the Global CG approach, and the Salient W approach. Clearly the Salient approach has similar performance comparing with the Global CG approach and outperforms the Global CW approach when the first five categories are considered. 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 because now the images exhibit more global characteristics and therefore, the global approaches can capture better the image content.

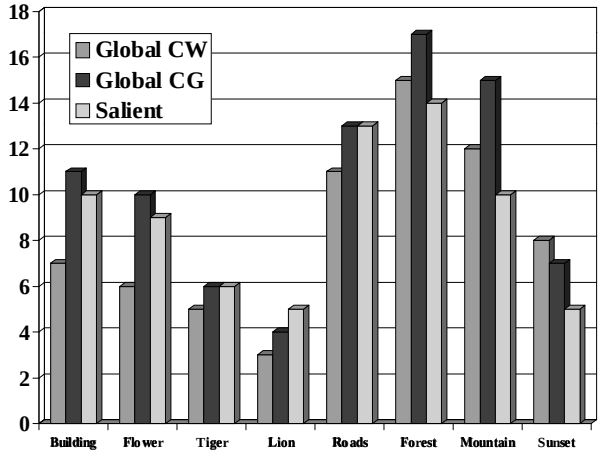


Figure 6: The average number of hits for each category using the global color and wavelet moments (Global CW), the global color and Gabor moments (Global CG) and the Salient W approach (Salient)

As noted before, the Salient approach uses only a very small part of the image to extract the features. Therefore, comparing with the global approaches the Salient approach has much less computational complexity. Table 4 shows the computation complexity for the two image databases using the Salient point approach and 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 time spent on the salient point extraction and the Gabor texture feature extraction from the neighborhood of the salient points. The time spent in the salient point extraction is varying within 2% when Salient W and Salient C approaches are considered. Therefore, in Table 4 only the Salient W approach is presented. From Table 4, the average computational complexity ratio of the Global approach to the Salient approach is about 5.84 (average of 4.67 and 7.02) for the listed two image databases. It can be inferred that the computational complexity difference will be huge when a very large database, e.g., millions of images, is used.

Database	1	2
Description	Object	Scenery images
Number	479	4013
Image size	256×256	360×360
Salient points extraction	23.9 min	225 min
Salient Gabor feature extraction	7.98 min	108 min
Salient total time	31.88 min	333 min
Global Gabor feature extraction	149 min	2340 min
Ratio (Global/Salient)	4.67	7.02

Table 4: The computational complexity comparison between the salient approach and the global approach for extracting texture feature using Gabor filters

When the computational complexity of the color feature extraction is compared, the Salient approach is much faster than the Global approach. Color moments were extracted from neighborhood of the 50 salient points. If we consider that the computational cost for the salient points extraction has already been counted in the Gabor texture feature extraction step, then it will not be counted for color feature extraction. Table 5 summarizes the results for the same databases used in Table 4. As one can see, the computational complexity difference is very large.

The number of Gabor texture features used in the Salient approach and the Global approach were 24 and 36, respectively. This does not have a big effect for small database. However, for very large image databases, the storage space used for these texture features will surely

Database	1	2
Computation Cost (Global/Salient)	145	288

Table 5: The computation complexity of color features for the Salient approach and the Global approach

make big difference. As to the color features, both approaches have the same number of features.

5 Discussion

In this paper we compared a wavelet-based salient point extraction algorithm with two corner detectors using the criteria: repeatability rate and information content. Our points have more information content and better repeatability rate than the Harris corner detector. Moreover, the detectors have significantly more information content than randomly selected points.

We also show that extracting color and texture information in the locations given by our salient points provides significantly improved results in terms of retrieval accuracy, computational complexity, and storage space of feature vectors as compared to global feature approaches. Our salient points are interesting for image retrieval because they are located in visual focus points and therefore, they capture the local image information.

For content-based retrieval, a fixed number of salient points (50 points in this paper) were extracted for each image. Color moments and Gabor moments were extracted from the 3×3 and the 9×9 neighborhood of the salient points, respectively. For benchmark purpose, the Salient point approaches were compared to the global color and wavelet moment (Global CW) approach and the global color and Gabor moments (Global CG) approach.

Three experiments were conducted and the results show that: (1) the Salient approaches have better performance than the Global CW approach. The Salient approaches 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 approaches have similar performance compared to the Global CG approach in terms of the retrieval accuracy. However, the Salient approaches achieve the best performance in the overall considerations of retrieval accuracy, computational complexity, and storage space of feature vectors. The last two factors will have very important influence for very large image databases; (3) Better retrieval results are obtained when Daubechies4 salient points are used compared with Harris corners. This shows that our wavelet-based points can capture better the image content.

Our experimental results also show that the global Gabor features perform much better than the global wavelet features. This fact is consistent with the results of the other researchers in the field proving that Gabor features are very good candidates for texture characterization [1].

In conclusion, the content-based image retrieval can be 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.

In our future work, we plan to extract shape information in the location of the salient points making the retrieval more accurate. We also intend to automatically determine the optimal number of the salient points needed to be extracted for each image.

References

[1] J. Beck, A. Sutter, and A. Ivry. Spatial frequency channels and perceptual grouping in texture segregation. *Comp Vis Graph Imag Process*, 37:299–325, 1987.

[2] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–996, 1988.

[3] T. Gevers and A. Smeulders. PicToSeek: Combining color and shape invariant features for image retrieval. *IEEE Trans Imag Process*, 20(1):102–119, 2000.

[4] R. Haralick and L. Shapiro. *Computer and Robot Vision II*. Addison-Wesley, 1993.

[5] C. Harris and M. Stephens. A combined corner and edge detector. *Alvey Vis Conf*, pages 147–151, 1999.

[6] J.J. Koenderink and A.J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.

[7] E. Loupias and N. Sebe. Wavelet-based salient points: Applications to image retrieval using color and texture features. In *Visual'00*, pages 223–232, 2000.

[8] W.Y. Ma and B.S. Manjunath. A comparison of wavelet transform features for texture image annotation. *Int Conf on Imag Process*, 2:256–259, 1995.

[9] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans on Patt Anal and Mach Intell*, 19(5):530–535, 1997.

[10] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interst point detectors. *I J Comp Vis*, 37(2):151–172, 2000.

[11] N. Sebe, Q. Tian, E. Loupias, M.S. Lew, and T.S. Huang. Color indexing using wavelet-based salient points. In *IEEE Workshop on Content-based Access of Image and Video Libraries*, pages 15–19, 2000.

[12] J.R. Smith and S.-F. Chang. Transform features for texture classification and discrimination in large image databases. *Int Conf on Imag Process*, 3:407–411, 1994.

[13] A. Stricker and M. Orengo. Similarity of color images. *SPIE - Storage and Retrieval for Image and Video Databases III*, 2420:381–392, 1995.