# Wavelet-based Salient Points: Applications to Image Retrieval Using Color and Texture Features

Etienne Loupias[1*], Nicu Sebe[2]

[1] Laboratoire Reconnaissance de Formes et Vision,
INSA Lyon, France
loupias@rfv.insa-lyon.fr
[2] Leiden Institute of Advanced
Computer Science,
Leiden University, The Netherlands.
nicu@wi.leidenuniv.nl

**Abstract.** In image retrieval, global features related to color or texture are commonly used to describe the image. The use of interest points in content-based image retrieval allows image index to represent local properties of images. 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. We present a salient point detector that extracts points where variations occur in the image, regardless whether they are corner-like or not. It is based on wavelet transform to detect global variations as well as local ones. We show that extracting the color information in the locations given by these points provides significantly improved retrieval results as compared to the global color feature approach. We also show an image retrieval experiment based on texture features where our detector provides better retrieval performance comparing with other point detectors.

## 1 Introduction

We are interested in content-based image retrieval in general image databases. The query is an image (*iconic search*), and the retrieved images should be *similar* to the query. We assume that high-level concepts (objects, feelings, etc.) cannot be extracted automatically from the image without specific knowledge, and so we use an *image similarity* based on low-level features (such as color, texture and shapes).

An image is "summarized" by a set of features, the *image index*, to allow fast querying. *Local features* are of interest, since they lead to an index based on local properties of the image. This approach is also attractive for sub-image search.

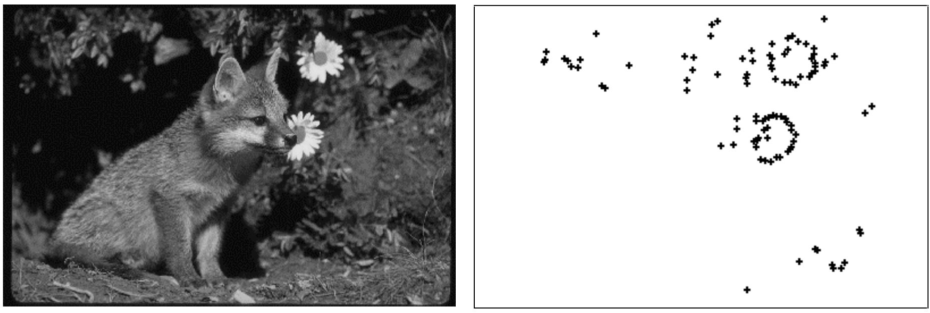
The feature extraction is limited to a subset of the image pixels, the *interest points*, where the image information is supposed to be the most important [9,2,11,1]. This

* Guest period in Leiden University was supported by Région Rhône-Alpes (EURODOC grant)

paper focuses on the selection of points that are significant to compute features for indexing.

Corner detectors are commonly used for indexing [9,11]. They are usually defined as points where the gradient is high in multiple orientations. This definition leads to detectors based on local derivatives [6]. Corner detectors are in general designed for robotics and shape recognition and therefore, they have drawbacks when are applied to natural image retrieval.

**Visual focus points need not be corners:** visual meaningful feature is not necessarily located in a corner point. For instance in Fig. 1, the fur is too smoothed to be detected by a corner detector such as Harris' [6].



(a) Fox image (b) 100 corners (Harris)

**Fig. 1.** Image with smoothed edges. No corners are detected in the fur.

**Corners may gather in small regions:** in various natural images, regions may well contain textures (trees, shirt patterns, etc.), where a lot of corners are detected (*cf.* Fig. 2). As the number of points is preset to limit the indexing computation time, most of the corners are in the same textured region.

(a) Fox image (b) 100 corners (Harris)

**Fig. 2.** Image with texture in the Dutch dress. Corners are gathered in the textured region.

With corner detectors, both examples lead to an incomplete representation, where some parts of the image are not described in the index.

For these reasons, corner points 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 smoothed or corner-like. To describe different parts of the image, the set of interesting point 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*. Wavelet representations, which express image variations at different resolutions, are attractive to extract salient points.

Previous point detectors make use of multiresolution representation. Chen *et al.* consider two different resolutions to extract corners [3]. In image retrieval context, contrast-based points are extracted in [2]. However, a lot of points are also extracted in textured regions because these regions are contrasted. Points are extracted with a specific wavelet in [1]. But since only a given scale is used, different resolutions features cannot be detected.

## 2 From wavelet transform to salient points

The wavelet transform is a multiresolution representation that expresses image variations at different scales. For wavelet theory, see [8].

A wavelet is an *oscillating* and *attenuated* function (its integral is equal to zero). We study the image $f$ at the scales (or resolutions) ½, ¼, … $2^j$, $j \in \mathbf{Z}$ and $j \leq -1$. The wavelet *detail image* $W_{2^j} f$ is the convolution of the image with the wavelet function dilated at different scales.

Here we consider *orthogonal wavelets*, which lead to a complete and non-redundant representation of the image. A wavelet can also have a *compact support*: its value is zero outside a bounded interval. The simplest orthogonal compactly supported wavelet is the Haar wavelet, which is the discontinuous step function. Daubechies proposed wavelets, with any regularity $p$ ($p > 1$), that are also orthogonal and compactly supported [4].

The wavelet representation gives information about the variations in the signal 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 signal 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 (as shown in Fig. 3).

Since we use wavelets with a compact support, we know from which signal points each wavelet coefficient at the scale $2^j$ was computed. We can study the wavelet coefficients for the same points at the finer scale $2^{j+1}$. Indeed 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 $2^j$

(see [7] for details). We call this set of coefficients the *children* $C\left(W_{2^j} f(n)\right)$ of the coefficient $W_{2^j} f(n)$. The children set in one dimension[1] is:

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

$0 \le n < 2^j N$ ($N$ is the length of the signal, $p$ the wavelet regularity).

**Fig. 3.** Salient point extraction: spatial support of tracked wavelet coefficients

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 $C\left(W_{2^j} f(n)\right)$. In our salient point extraction algorithm, we consider this maximum, and look at its highest child. Applying recursively this process, we select a coefficient $W_{2^{-1}} f(n)$ at the finer resolution ½ (*cf.* Fig. 4). Hence, this coefficient only 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:

$$saliency = \sum_{k=1}^{-j} \left| C^{(k)}\left(W_{2^j} f(n)\right) \right|,\ 0 \le n < 2^j N,\ -\log_2 N \le j \le -1$$

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 the fastest for execution. Some localization drawbacks can appear with Haar due to its non-overlapping wavelets at a given scale. This

---

[1] For clarity we use one-dimensional signals. Extension to two dimensions and signals with length not restricted to a power of 2, in addition to algorithm complexity, are discussed in [6].

can be avoided with the simplest overlapping wavelet, Daubechies 4. However, this kind of drawback is not likely in natural images.

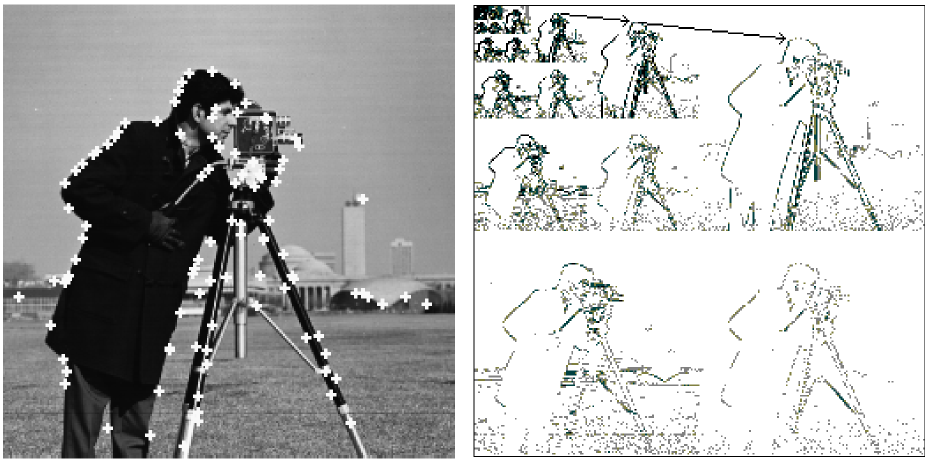



**Fig. 4.** Haar transform. (a) 100 Haar salient points in the Cameraman image. (b) Tracked coefficients in the Haar transform of the Cameraman image

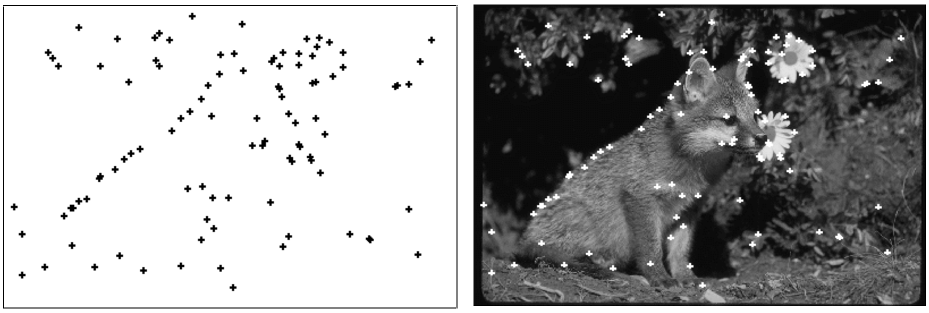
## 3 Examples

The salient points detected with the Haar transform are presented for the images used in Fig. 1 and Fig. 2 (*cf.* Fig. 5). Salient points are detected for smoothed edges (*cf.* Fig. 5.a) and are not gathered in textured regions (*cf.* Fig. 5.b). Hence they lead to a more complete image representation than corner detectors. Similar behavior can be observed with Daubechies 4 wavelets.

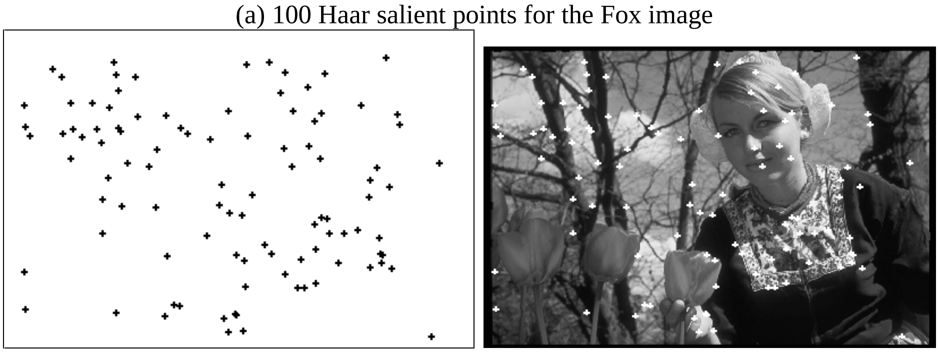
Repeatability of the detection under typical alterations is a common evaluation criterion for corner detectors. Repeatability of our detector is comparable to other detectors. However this criterion may not be relevant in our context, because features stability is more important than geometric stability for image retrieval.

## 4 Evaluation for image retrieval

The best way to evaluate points detectors for image retrieval is to compare the retrieval results. In color indexing, global color distributions are mainly used to index images. We show in the next section that extracting the color information in the locations given by salient points improve the retrieval results as compared to the global color feature approach. Then we use salient points for image retrieval based on texture features and we show an experiment where our detector provides better retrieval performance comparing with other point detectors.



(a) 100 Haar salient points for the Fox image



(b) 100 Haar salient points for the Dutch image

**Fig. 5.** Haar salient points examples. We can that notice salient points are detected in smooth features like the fur and not gathered in textured region. For each image the detected points are superimposed on the original image to evaluate salient points location.

### 4.1 Color

The setup of our experiments was the following. First we extracted 100 salient points for each image in the database using *Haar* wavelet transform and the algorithm described in Section 2.

For feature extraction, we considered only the pixels in a 3×3 neighborhood around each salient point in forming an image signature. For each image signature we computed the color moments and stored them is a feature vector [10]. Since 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. We were working with the HSV color space and for each image in a database a 9-dimensional feature vector was considered. When the user selects a query, the system computes the corresponding feature vector and compares it with the feature vectors in the database. For benchmarking purposes we also considered the results obtained using color moments over the entire image.

In the first experiment we considered a database of 479 images of color objects such as domestic objects, tools, toys, food cans, etc [5]. As ground truth we used 48 images of 8 objects taken from different camera viewpoints (6 images for a single object). 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. 6 we present an example of a query image and the similar images from the database.

**Query Image**

| | | | | | |
|---|---|---|---|---|---|
| **Salient** | **1** | **2** | **6** | **12** | **18** |
| **Global moments** | **1** | **4** | **12** | **27** | **41** |

**Fig. 6.** Example of images of one object taken from different camera viewpoints. The ranks of individual image were obtained using salient point information (Salient) and the global color moments method (Global moments)

The salient point approach outperforms the global color method. 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. When the global color moments were used the influence of the background became important so the retrieval results were worse.

In the next experiment, we used a database of 1505 various natural images. They cover a wide range of natural scenes, animals, buildings, construction sites, textures and paintings. As test set we considered 50 images which were grouped in 5 classes (10 images in a class): Airplane, Car, Flower, Lion, Bird. Fig. 7 shows an example of the retrieved images from a query using the salient point approach.

The salient points were able to capture the image details, even if the background was different and the position of the animal changed. In order to test the retrieval results for each individual class, we randomly picked 5 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 which were retrieved in top 15 images. The results are given in Table 1.

Note that for classes where the background was complex (Car, Flower) the results were worse than for the other classes. However, the salient points captured the details of the foreground objects and therefore the results were significantly better than in the case of using global color moments.

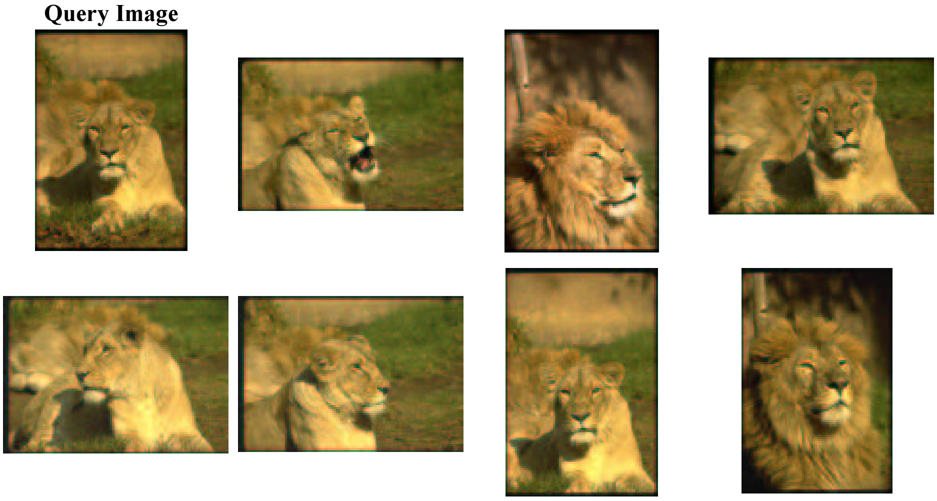



**Fig. 7.** Retrieved images from a query using the salient point approach. Match quality decreases from the top left to the bottom right

| Class | Salient | Global |
|---|---|---|
| Airplane | 94 | 88 |
| Bird | 88 | 82 |
| Car | 74 | 62 |
| Flower | 72 | 58 |
| Lion | 90 | 82 |

**Table 1.** Retrieval accuracy (%) for each individual class using 5 randomly chosen images from each class as queries

### 4.2 Texture

In the texture experiments, Gabor features are computed for regions around the extracted points (32×32) for 3 scales and 8 orientations. Features for each extracted point are used to build a set of 24 histograms (one histogram for each scale and orientation). We use two-dimensional histograms to take into account the spatial coherence in our representation, as described in [12].

We use the same database of 1505 various natural images as in the second color experiment. In this case we were not only interested to retrieve images with similar color (see Fig. 7), so we built another test set where each image belongs to an instinctive (and subjective) category (animals, flowers, landscapes, buildings, cities, etc). Very heterogeneous categories and images too different from the rest of the category were removed from the test set. Finally, we have a test set of 577 images in 9 classes.

For benchmarking purposes we compare the results obtained with different detectors. We considered two wavelet-based detectors (Haar and Daubechies4), the Harris corner detector [6], the contrast-based detector proposed in [2] and a detector based on random points. We present the recall-precision graph, computed from different numbers of return images *n*. The system retrieves *r* images that belong to the same class *C* as the query ($r \leq n$). There are $N_C$ images in the class *C* of the query. Then $P = r / n$ is the precision and $R = r / N_C$ the recall for this query. We use each test set image as a query, and use the average recall and precision for the graph (*cf.* Fig. 8).

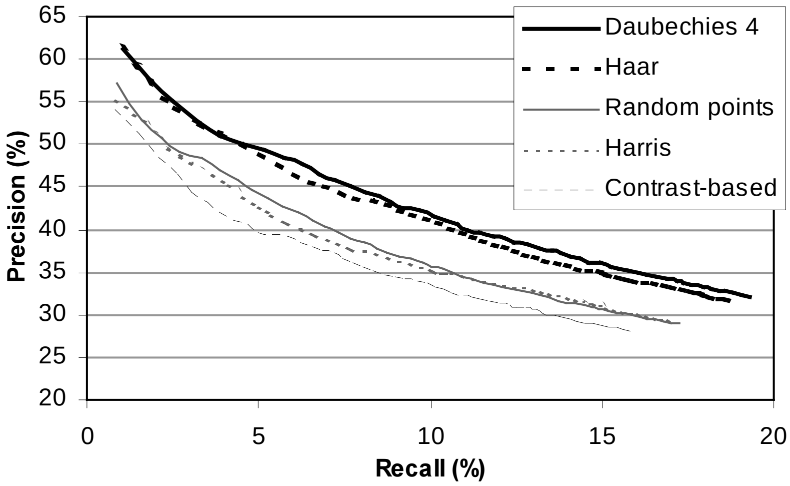



**Fig. 8.** Retrieval results with Gabor texture features for the database with 1505 natural images

We observe that the wavelet-based salient points perform better than other detectors for these features and this database. Daubechies 4 has better performances than Haar but is computationally more expensive. Random points are also used in the experiment: we randomly select points, and compute the Gabor features around these points. Their good result can be explained by their spreading in the image. For that reason they lead to a more complete representation of the image than some detectors. Obviously, the random points are very unlikely to be located in corners or edges point, but they are spread enough to represent these variations in the index. Good result of random points for indexing was observed with other databases and other local features [7]. These experiments show that the points spreading can be as important as the points location for image indexing (depending on the features). However, wavelet-based salient points, which are simultaneously spread and located, perform better than random points.

## 5 Discussion

We presented a salient point detector based on wavelets. The wavelet-based salient points are interesting for image retrieval, because they are located in many visual features (whether they are corner-like or not), without gathering in textured regions. We presented an experiment of color retrieval where using salient points is an interesting alternative to global approaches, and another retrieval experiment with Gabor features where our method performs better than other point detectors from the literature.

We used the Haar transform for point extraction, which is simple but may lead to bad localization. Daubechies wavelets avoid this drawback, but are not symmetric. Since orthogonality is not required in our approach, we could extend it to other wavelets that are compactly supported and symmetric.

Since points performance for indexing depends on the image database, detector choice for a specific database should be investigated, as well as random points relevance for local features extraction. Wavelets are also attractive to extract image features for indexing. These local features would be more related to our salient points.

## References


1. S. Bhattacharjee and T. Ebrahimi, "Image Retrieval Based on Structural Content", *Workshop on Image Analysis for Multimedia Interactive Services*, Heinrich-Hertz-Institut (HHI) Berlin, Germany, May 31 - June 1 1999.
2. S. Bres and J.-M. Jolion, "Detection of Interest Points for Image Indexation", *3rd Int. Conf. on Visual Information Systems, Visual99*, Amsterdam, The Netherlands, June 2-4 1999, pp. 427-434.
3. C.-H. Chen, J.-S. Lee and Y.-N. Sun, "Wavelet Transformation for Gray-level Corner Detection", *Pattern Recognition*, 1995, Vol. 28, No. 6, pp. 853-861.
4. I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets", *Communications on Pure and Applied Mathematics*, 1988, Vol. 41, pp. 909-996.
5. T. Gevers and A. Smeulders, "Color-based Object Recognition", *Pattern Recognition*, 1999, Vol. 32, No. 3: pp. 453-464.
6. C. Harris and M. Stephens, "A Combined Corner and Edge Detector", *Proc. of 4th Alvey Vision Conference*, 1988, pp. 147-151.
7. E. Loupias and N. Sebe, "Wavelet-based Salient Points for Image Retrieval", *RR 99.11*, Laboratoire Reconnaissance de Formes et Vision, INSA Lyon, November 1999.
8. S. Mallat, "A Theory for Multiresolution Signal Decomposition : The Wavelet Representation", *IEEE Trans. on PAMI*, July 1989, Vol. 11, No. 7, pp. 674-693.
9. C. Schmid and R. Mohr, "Local Grayvalue Invariants for Image Retrieval", *IEEE Trans. on PAMI*, May 1997, Vol. 19, No. 5, pp. 530-535.
10. M. Stricker and M. Orengo, "Similarity of Color Images", *SPIE - Storage and Retrieval for Image and Video Databases*, 1995.
11. T. Tuytelaars and L. Van Gool, "Content-based Image Retrieval Based on Local Affinely Invariant Regions", *3rd Int. Conf. on Visual Information Systems, Visual99*, Amsterdam, The Netherlands, 2-4 June 1999, pp. 493-500.
12. C. Wolf, J.-M. Jolion, W. Kropatsch and H. Bischof, "Content Based Image Retrieval Using Interest Points and Texture Features", *to appear in Proceedings of 15th ICPR*, 2000.