Color Indexing Using Wavelet-based Salient Points

N.Sebe[1] Q.Tian[2] E.Loupias[3] M.S.Lew[1] T.S.Huang[2]

[1]Leiden Institute of Advanced Computer Science, Leiden, The Netherlands {nicu, mlew}@liacs.nl

[2]Beckman Institute, University of Illinois at Urbana-Champaign, USA {qitian, huang}@ifp.uiuc.edu

[3]Laboratoire Reconnaissance de Formes et Vision, INSA-Lyon, France loupias@rfv.insa-lyon.fr

Abstract

Color is an important attribute for image matching and retrieval. Most of the attention from the research literature has been focused on color indexing techniques based on global color distributions. However, these global distributions have limited discriminating power because they are unable to capture local color information. In this paper, we present a wavelet-based salient point extraction algorithm. 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 approaches.

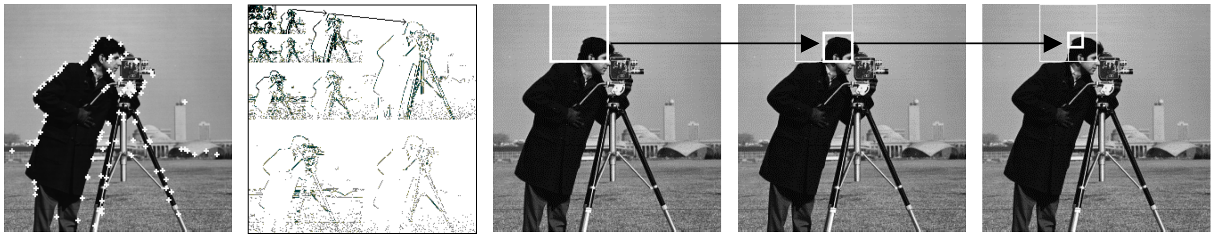
1. Introduction

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 image 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.

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. 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 [6]. Therefore, texture or geometric properties are needed to identify objects [2]. 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. In practice it is necessary to combine color indexing with texture and/or shape indexing methods [5], [4]. Even if texture and shape indexing methods improve, color indexing retains its importance as a computationally simple and fast filter whose output can be processed by computationally more intensive methods. In this context, the major challenge of new color indexing methods is to improve the robustness of finding images with similar color compositions and to increase the retrieval speed.

In color indexing, color histograms are often used because they are feasible in terms of memory usage and provide sufficient accuracy [8]. 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, et al. [7] showed that characterizing one dimensional color distributions with the first three moments is more robust and more efficient than working with color histograms.

Our idea is first to extract salient points in the image and then in their location to extract color features. It is quite easy to understand that the usage of a small set of such points instead of all image 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). From another point of



(a) Salient points (b)Tracked coefficients (c)Spatial support of tracked coefficients

Figure 1. Salient points extraction

view, the underlying idea of salient points is related to the fact that when you look at an image the features you remember will be around some important points [1]. These visual focus points need not to be corners. Salient points should be related to any visually interesting part of the image, regardless whether it is smoothed or corner-like. To describe different parts of the image, the set of salient points should not be clustered in few regions.

In this paper, we present a salient point extraction algorithm using the wavelet transform, which expresses image variations at different resolutions. Wavelet-based salient points are detected for smoothed edges and are not gathered in textured regions. Hence, they lead to a more complete image representation than corners detectors [3].

2. Wavelet-based Salient Points

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 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) $1/2$, $1/4$, $\ldots$, 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, since the wavelets have a compact 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:

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

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. Applying recursively this process, we select a coefficient $W_{2^{-1}} f(n)$ at the finer resolution $1/2$ (Figure 1 (b) and (c)). 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:

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

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. 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. However, this kind of drawback is not likely in natural images and therefore, we used Haar transform in our experiments. In Figure 1 (a) we present the salient points extracted using the Haar transform. Note that our method extracts salient points not only in foreground but also in the background where some smooth details are present.

3. Color Moments

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 process is based on similarity function which uses only these features to determine the similarity 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, as well [7]. 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 p_{ij}, then the index entries related to this color channel are:

$$E_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij}, \qquad \sigma_i = \left(\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^2\right)^{\frac{1}{2}} \tag{3}$$

$$s_i = \left(\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^3\right)^{\frac{1}{3}}$$

where N is the number of pixels in the image.

We were working with the HSV color space so, for each image in the database a 9-dimensional feature vector was considered. Let I and Q be the feature vectors corresponding to two images in the database. The similarity between these two feature vectors is given by:

$$d(I, Q) = (I - Q)^T (I - Q) \tag{4}$$

Note that each vector I and Q is a 9-dimensional feature vector with three color moments for each color channel.

4. Results

The setup of our experiments was the following. First we extracted 50 salient points for each image in the database using Haar wavelet transform and the algorithm described in Section 2. We considered only the pixels in a neighborhood of 3×3 pixels around each salient point which form the image signature. Note that only these 9×50 pixels characterized the image and we used only this information for extracting the feature vectors. For each image signature in the database we computed the color moments. When the user selects a query, the system computes the corresponding feature vector from the query image signature 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.

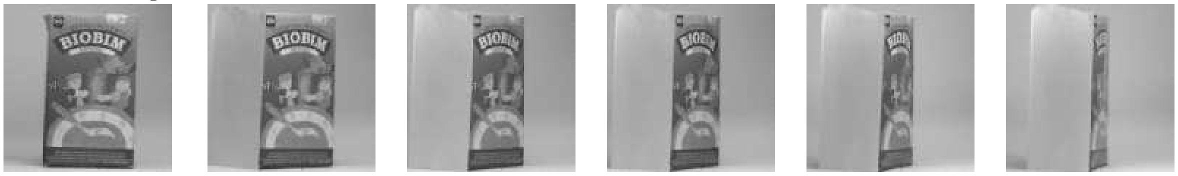
The problem is formulated as follows: Let $\mathcal{Q}_1, \cdots, \mathcal{Q}_n$ be the query images and for the i-th query $\mathcal{Q}_i$, $\mathcal{I}_1^{(i)}, \cdots, \mathcal{I}_m^{(i)}$ be the images similar with $\mathcal{Q}_i$ according to the ground truth. The retrieval method will return this set of answers with various ranks. As an evaluation measure of the performance of the retrieval method we used precision vs. recall at different scopes: For a query $\mathcal{Q}_i$ and a scope $s > 0$, the recall r is defined as $|\{\mathcal{I}_j^{(i)} | rank(\mathcal{I}_j^{(i)}) \leq s\}|/m$, and the precision p is defined as $|\{\mathcal{I}_j^{(i)} | rank(\mathcal{I}_j^{(i)}) \leq s\}|/s$.

In the first experiment we considered a database consisting of 500 images of color objects such as domestic objects, tools, toys, food cans, etc. 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 changes because the salient points are located around the object boundary and capture the details inside the object, neglecting the noisy background. In Figure 2 we present an example of a query image and the similar images from the database. The salient point method outperformed the global color moments 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. When the global color moments were used the influence of the background became important so the retrieval results were worse.

Figure 3 shows the precision-recall graphs. The curve corresponding to salient points method is above the one corresponding to the global color moments method, indicating that using salient points information is more effective.

In the second experiment, we used a database of 1505 various natural images. They cover a wide range of nature scenes, animals, buildings, construc-

Query Image



Salient	1	2	4	15	22
Global moments	2	7	12	25	33

Figure 2. Example of images of one object taken from different camera viewpoints. The ranks of individual images were obtained using salient points information (Salient) and the global color moments method (Global moments)

Query Image

Figure 4. Retrieved images from a query using salient points. Match quality decreases from the top left to the bottom right

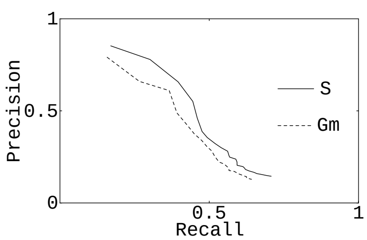


Figure 3. Precision/Recall for color objects database using salient points information (S) and the global color moments method (Gm)

tion 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. In Figure 4 an example is given of a query from the Lion class and its retrieval results. Note that 7 out of 8 retrieved images were from the same class as the query image. 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.

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

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

Note that for the 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.

In the next experiment, each of the 50 images from the test set were considered as queries and the retrieval accuracy was calculated. The idea of this experiment was to retrieve images from the same class as the query image. The retrieval accuracy was given by the average percentage of correct images from the same class as the query which were retrieved in top n images. These results are presented in Table 2.

Top	10	15	20
Salient	70.5	83.6	92.3
Global moments	60.7	72.7	79.5

Table 2. Retrieval accuracy (%) using 50 images from 5 classes

Note that again using the salient point information the retrieval results were improved. In average 9 out of 10 images from the same class were retrieved in top 20 using the salient points.

5. Discussion and Conclusions

A wavelet-based salient points extraction algorithm was presented. The salient points are interesting for image retrieval because they are located in visual focus points and thus they can capture the local image information. Extracting the color information in the locations given by the salient points provided significantly improved retrieval results as compared to the global color feature approach.

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 and second, they should not be clustered in few regions. To accomplish these demands we used a Haar-based salient points extraction algorithm which is fast and captures the image details at different resolutions.

Two experiments were conducted. In the first experiment, we investigated the retrieval of color objects images taken from different viewpoints. The salient points method proved to be more robust to the viewpoint changes because the salient points were located around the objects boundaries and captured better the details inside the objects, neglecting the background influence. In the second experiment, a database of natural images was used. The test images were grouped in different classes based on the subject depicted in each of them. Some classes were mainly composed of a single object on a simple background (e.g. Airplane, Bird where the background represented the sky), while the others (Lion, Car, Flower) had a more complex background making the retrieval difficult. When the background was roughly the same for all the images in the class, then the global color moments performed reasonable well but still worse than using the salient points method. When the background was more complex, the global color moments had worse retrieval results. However, the salient point method captured the foreground object details more accurately and therefore, the retrieval results were improved.

In conclusion, the image retrieval results 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 features information and therefore, they provide a better characterization of the image content. In future work we plan on extracting texture and shape information in the locations of the salient points making the retrieval more accurate.

References

[1] S. Bres and J. Jolion. Detection of interest points for image indexation. *Visual99*, pages 427–434, 1999.

[2] M. Flicker, 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. *IEEE Computer*, 28(9):23–32, 1995.

[3] E. Loupias and N. Sebe. Wavelet-based salient points for image retrieval. *TR. 99.11, Laboratoire Reconnaissance de Formes et Vision, INSA Lyon*, 1999.

[4] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Tools for content-based manipulation of image databases. *SPIE - Storage and Retrieval for Image and Video Databases*, 2185, 1994.

[5] J. Smith. *Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis.* PhD thesis, Columbia University, February 1997.

[6] M. Stricker and A. Dimai. Spectral covariance and fuzzy regions for image indexing. *SPIE - Storage and Retrieval for Image and Video Databases*, 1997.

[7] M. Stricker and M. Orengo. Similarity of color images. *SPIE - Storage and Retrieval for Image and Video Databases*, 1995.

[8] M. Swain and D. Ballard. Color indexing. *IJCV*, 7(1):11–32, 1991.