COLOR IMAGE RETRIEVAL BASED ON WAVELET SALIENT FEATURES DETECTION

Christophe Laurent, Nathalie Laurent and Muriel Visani

France Telecom R&D - DIH/HDM
4, Rue du Clos Courtel
35512 Cesson Sévigné Cedex - FRANCE
{christophe2.laurent|nathalie.laurent|muriel.visani}@rd.francetelecom.com

ABSTRACT

Nowadays, most of the research works in the area of image retrieval try to build image signature by considering the image as a whole. In this paper, we proposed an alternative based on the detection of some salient points in the image. For this purpose, we propose a new efficient salient point detector based on a wavelet transform. The efficiency of our detector lies in the representation of the wavelet coefficients by a zerotree data structure and by a saliency formulation that does not favor any direction. Thus, the detected salient points are located on region contours whatever their direction. From the detected salient points, we propose to extend to well-known color correlogram to salient features in order to build a saliency-based image retrieval system. This modified correlogram is built in the recently proposed c1c2c3 color model in order to get a photometric invariant color descriptor. Experimental results have shown that our descriptor outperforms the MPEG-7 SCD, based on the construction of a color histogram in the HSV color space.

1. INTRODUCTION

Today, most of the research works in the area of content-based image retrieval aim at giving a global description of an image or a region by designing an image signature considering all pixels of the image or the region of interest. Consequently, all pixels of the image or the region have the same importance during the signature computation. However, it seems natural to consider that some pixels are more perceptually important than others and computing an image signature from features extracted around these pixels may lead to better retrieval results. Moreover, when this approach is applied to object-based queries, we gain robustness against occlusions since we only use a local description of the object of interest.

This approach to content-based image retrieval is said to be salient features-based [18] because the information of the image is condensed into a limited number of feature values. As a result, the salient features must be extracted with precision for greatest saliency and proven robustness. Ideally, one should be able to repeat the extracted salient features whatever the geometric transform applied to the image (rotation, scale change, translation, etc.), the point of view and the imaging conditions. Moreover, most of the image information content should be extracted from the neighborhood of salient features.

Image retrieval based on salient features extraction follows a similar computation flow compared to other image retrieval approaches: first, a robust salient feature extractor must be designed. Then, a salient signature is computed by analyzing image data located in the neighborhood of the extracted features. Finally, a similarity measure must be designed to compare two salient signatures. Obviously, the signature design and the similarity measure greatly depend on the salient feature extractor used.

The salient features can be of different types (edge, junctions, corners, etc.) and a good overview is given in [18].

In this paper, we focus on salient features represented by single points located in image area where the information is considered as perceptually important. A wide variety of salient point[1] detectors have been proposed in literature [14] due to the lack of definition about the concept of salient point. One of the oldest and probably the most used detector is the corner detector proposed by Harris and Stephens in [5]. This detector (and a precise version of it) has been first used in the image retrieval topic by Schmid and Mohr in [13]. Then, it has been extended to cope with color images in [10].

In [1], Bres and Jolion consider that relevant information is located in image area where local contrast is high. For this purpose, they adopt a multi-resolution framework in which they build a contrast pyramid.

In [7], the authors also adopt a multi-resolution scheme in which salient points are those presenting the highest wavelet coefficient values. Consequently, salient points are located on sharp region boundaries. This last approach seems to us the most interesting one for two reasons:

[1] This kind of point is often called *key point* or *point of interest* in literature.

- image contours are more perceptually important than corners that are used in [3, 13];
- salient points detected by a corner detector may be gathered in small image regions in the case of textured images. As a result, the detected points only provide a very local image description.

These drawbacks are naturally avoided by the use of wavelet analysis since texture areas are gradually smoothed by the multi-resolution framework avoiding thus the gathering of salient points in these zones. Regarding the contrast-based detector proposed in [1], we are convinced that the detected points are approximatively the same than those detected by the wavelet approach since a high contrast value generally lead to a contour and thus a high corresponding wavelet coefficient.
In this paper, we propose a new wavelet-based salient point detector. Our approach is more computationnal efficient than the one presented in [7] thanks to a zerotree representation of wavelet coefficients. As a result, the computation time is independent of the wavelet filter size which is not the case for the detector presented in [7]. Moreover, our detector does not favor any contour direction by merging the wavelet coefficients from all detail subbands. Our detector will be detailed in section 2. Due to the limited text size, we will not present in this paper the performance of our detector in terms of repeatability rate but experimental results have shown its good behavior. In section 3, we will propose an example of color image indexing framework based on our salient point detector. In this section, we will extend the well-know color correlogram [6] to salient features and we will work in a photometric invariant color model recently proposed by Gevers and Smeulders in [2]. Finally, our conclusions and perspectives will be discussed in section 4.

2. A NEW WAVELET-BASED SALIENT POINT DETECTOR

2.1. Introduction

To detect salient points in an image, our detector proceeds as follows:

- a wavelet transform is firstly performed on the image of interest, resulting in a sub-sampled scale image and a pyramid of detail images;
- the obtained wavelet coefficients are zerotree represented [17] resulting in a hierarchical data structure (tree) of wavelet coefficients;
- this tree is traversed a first time from leaves to the root node by computing at each level the saliency value of each wavelet coefficient;
- from the saliency maps previously computed, the tree is traversed a second time from root to leaves by choosing at each tree level the most salient wavelet coefficient.

All these steps are detailed in the next sections.

2.2. Wavelet Transform

The wavelet transform is a powerful tool approximating a function at different resolution levels [8]. Thus, this theory can describe any function f with a coarse approximation of f and a set of detail functions allowing to perfectly reconstruct the original function f. For a good overview of the wavelet theory, the reader is referred to [8].
In our case, any image I can be considered as a discrete bi-dimensional function sampled over a discrete compact support D_I with $n = 2^k$ $(k \in \mathcal{Z})$ rows and $m = 2^l$ $(l \in \mathcal{Z})$ columns. Moreover, we consider only luminance information that we suppose quantized over 256 values i.e. $I(p) \in [0, 255]$ for each $p \in D_I$.
As mentioned previously, the wavelet transform of I allows a multi-resolution representation of I. At each resolution level 2^j $(j \leq -1)$, I is represented by a coarse approximation of I denoted by $A_{2^j} I$ and by three detail images denoted by $D^s_{2^j} I$ with $s = 1, 2, 3$ and representing respectively vertical, horizontal and diagonal details. These four images are of size $2^{k+j} \times 2^{l+j}$. All these images are obtained from a low pass scaling filter H and a high pass wavelet filter G obtained respectively by dilating and translating a scaling function $\Phi(x)$ and a wavelet function $\Psi(x)$. In our case, we focus on orthogonal wavelets with compact support for which the functions Φ and Ψ are separable, resulting in separable filters H and G. In this case, the wavelet transform can be implemented in a pyramidal framework as illustrated in Figure 1.

2.3. Zerotree Representation of Wavelet Coefficients

Once the wavelet transform is performed up to a fixed resolution 2^r $(r \leq -1)$, we get a coarse approximation image $A_{2^r} I$ and three details images $D^s_{2^j} I$ $(s = 1, 2, 3)$ per resolution level 2^j $(r \leq j \leq -1)$.
We can then construct a hierarchical data structure of wavelet coefficients based on the zerotree approach that has been firstly proposed for image compression in [17]. This allows to build hierarchical relationships between wavelet coefficients as illustrated in Figure 2:

- each pixel $p(x, y) \in A_{2^r} I$ is the root of a tree;
- each root $p(x, y)$ has three children nodes designated by the wavelet coefficients $w^s_{2^r}(x, y)$ $(s = 1, 2, 3)$ located at (x, y) in the corresponding detail subbands $D^s_{2^r} I$;

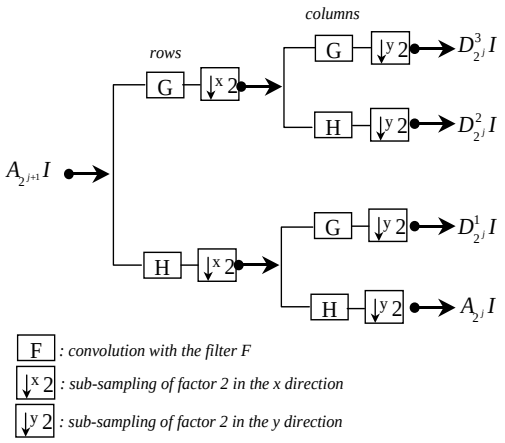


Fig. 1. Pyramidal wavelet transform

- due to the sub-sampling step performed during the wavelet transform (see Figure 1), each wavelet coefficient $w^s_{2^r}(x,y)$ $(s = 1,2,3)$ of the detail subband $D^s_{2^r}I$ corresponds to an area of size 2×2 pixels in the same detail subband at the higher resolution level $D^s_{2^{r+1}}I$. This area is located at $(2x, 2y)$ and all wavelet coefficients belonging to this zone become the children nodes of $w^s_{2^r}(x,y)$.

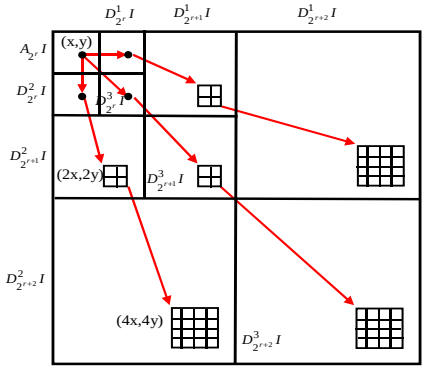


Fig. 2. Zerotree data structure

In a recursive way, we construct the zerotree data structure in which each wavelet coefficient $w^s_{2^u}(x,y)$ $(s = 1,2,3$ and $0 > u > r)$ has four children nodes designated by the wavelet coefficients of $D^s_{2^{u+1}}I$ belonging to the area of size 2×2 pixels and located at $(2x, 2y)$.
Once all the trees are constructed, each wavelet coefficient at the coarsest resolution level $w^s_{2^r}(x,y)$ $(s = 1,2,3)$ corresponds to a region of size $2^{-r} \times 2^{-r}$ pixels in the detail subband $D^s_{2^{-1}}I$.

2.4. Construction of Saliency Maps

From the zerotree data structure previously constructed, we propose to build a set of $-r$ saliency maps (i.e. one saliency map per resolution level). Each saliency map S_{2^j} $(j = -1,\ldots,r)$ should reflect the importance of the wavelet coefficients at the resolution level 2^j. Let us recall that a perceptually important wavelet coefficient should correspond to a pixel located on a contour in the image I. Therefore, a saliency map should satisfy the following properties:

- the more the information content embedded by a wavelet coefficient is perceptually important and the more the associated saliency value must be high;
- a salient wavelet coefficient must have a high saliency value whatever the preferred direction (horizontal, vertical or diagonal) of the corresponding detail subband to which the coefficient belongs. Indeed, we are interested in image contours whatever their direction and for this purpose, we have to merge information coming from each detail subband;
- the saliency value of each wavelet coefficient at the resolution level 2^j must consider the saliency value of its descending nodes in the zerotree data structure.

In order to tsatisfy these properties, the saliency value $S_{2^j}(x,y)$ of a wavelet coefficient located at (x,y) in the resolution level 2^j is given by the following recursion:

$$\begin{cases} S_{2^{-1}}(x,y) = \alpha_{-1}\left(\dfrac{1}{3}\displaystyle\sum_{u=1}^{3}\frac{|w^u_{2^{-1}}(x,y)|}{|Max(D^u_{2^{-1}})|}\right) \\ S_{2^j}(x,y) = \dfrac{1}{2}\left(\alpha_j\left(\dfrac{1}{3}\displaystyle\sum_{u=1}^{3}\frac{|w^u_{2^j}(x,y)|}{|Max(D^u_{2^j})|}\right) + \right. \\ \qquad \left. \dfrac{1}{4}\displaystyle\sum_{u=0}^{1}\sum_{v=0}^{1} S_{2^{j+1}}(2x+u, 2y+v)\right) \end{cases} \quad (1)$$

where $Max(D^s_{2^j})$ $(s = 1,2,3)$ denotes the maximum wavelet coefficient value over the detail subband $D^s_{2^j}$, and α_k (with $k \in [r,-1]$ and $0 \leq \alpha_k \leq 1$) designates a weighting factor allowing to tune the importance of saliency values following the resolution level. In practice, we assign high saliency weights to wavelet coefficients at resolution 2^r since they embed contour information of the biggest image objects that can be considered as the most perceptually important objects (little objects being deleted by the multiresolution framework). Therefore, we generally choose $\alpha_j = 2^{r-j}$.
It can be noted that equation (1) gives normalized saliency values i.e. in the range $[0,1]$.

2.5. Choice of Salient Points

Once all saliency maps are constructed, we propose a method to choose the most salient points in the original image I. For this purpose, we build a hierarchy of saliency values from the $-r$ saliency maps by adopting a similar approach to the construction of the zerotree data structure: 2^{k+l+2r} trees of saliency values can be constructed, each one being rooted at a saliency value of S_{2^r}. As with the zerotree approach, each root node corresponds to an area of size 2×2 saliency values in $S_{2^{r+1}}$. We can thus recursively build the trees in which each node has four children in the saliency map at the next higher resolution. Figure 3 illustrates an example of this construction.

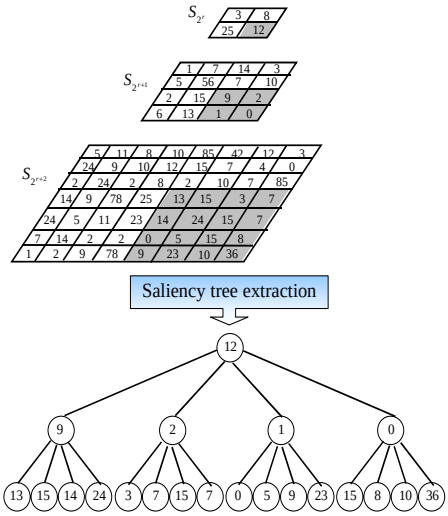


Fig. 3. Saliency tree construction

To localize the most salient points in the original image I, we proceed as follows:

1. the 2^{k+l+2r} trees are sorted in the decreasing order of the saliency value associated to the root node;
2. the most salient branch is selected in each of the 2^{k+l+2r} trees.

The first step is justified by the recursive definition of the saliency (see equation (1)) showing that the saliency values contained in the map S_{2^r} embed the saliency values of all descending nodes in the saliency hierarchy.
To select the most salient branch in the second step, we traverse each tree from the root to the leaves by selecting at each level the node with the highest saliency value (see Figure 4).

In this way, we obtain 2^{k+l+2r} lists of $-r$ saliency values, each list corresponding to the most salient branch of a saliency tree:

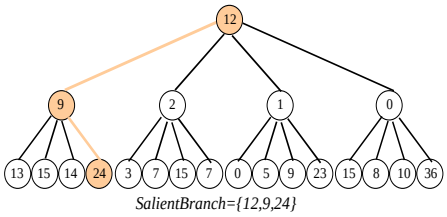
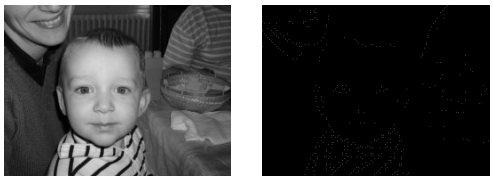


Fig. 4. Extraction of a salient branch from the example given in Figure 3

$$SalientBranch(i) = \{S_{2^r}(x_1^i, y_1^i), \ldots, S_{2^{-1}}(x_{-r}^i, y_{-r}^i)\}$$

with $i \in [0, 2^{k+l+2r} - 1]$ and
$(x_k, y_k)_{k>1} = ArgMax\{S_{2^{r+k-1}}(2x_{k-1}+u, 2y_{k-1}+v), 0 \leq u, v \leq 1\}$.
We can thus select up to 2^{k+l+2r} salient pixels in the image I. These pixels are those with the highest gradient value among all pixels located in the area of size 2×2 pixels and located at $(2x_{-r}, 2y_{-r})$.
In practice, we often need $p < 2^{k+l+2r}$ salient pixels. In this case, the p chosen pixels are those generated by the first p salient branches due to the ordering of the salient tree in decreasing order of their root saliency value.

Figure 5 illustrates the salient point detection results. Figure 5(a) shows the original image and Figure 5(b) shows the 400 salient points detected with $r = -2$ and the Haar wavelet basis.



(a) (b)

Fig. 5. Salient point detection results. (a)original image (b)the detected 400 salient points

3. TOWARDS COLOR IMAGE RETRIEVAL BY USING WAVELET SALIENT POINTS

3.1. Problem Statement and Previous Works

The use of salient points for content-based image retrieval need the design of (1) a signature computed from information extracted in the neighborhood of salient points and (2)

a similarity measure allowing to compare two salient signatures. Let us recall that the main task of designing an image signature is to bridge the gap between image semantics and pixel representation, that is, to create a better correlation with image semantics. To reach this goal, the detected salient points must be located in image area where information content is perceptually relevant and the features extracted around the salient points should constitute a compact representation of the perceptually relevant information contained in the entire image.
In [3, 13], each salient point is described by a combination of differential invariants to obtain invariance under the group SO(2) of similitudes. While only grayvalue information is used in [13], Gouet and Boujemaa use in [3] the Harris color detector [10] allowing them to use only the first order invariants whereas invariants up to the third order are needed in [13]. As a result, a feature vector of size 9 in [13] and 8 in [3] is extracted from each salient point.
In [7, 15, 16, 20], a different kind of approach is proposed in which the same number of salient points must be extracted from all images in the database. Wolf et al. [20] extract a region of size 32×32 pixels around each salient point and submit it to a Gabor filter bank with 3×8 filters (i.e. 3 scales and 8 orientations). Thus, each salient point is described by a texture feature of size 24 denoting the Gabor filter responses. In [7], Loupias et al. adopt the same method but by using a wavelet-based salient point detector. In [16], the signature is composed of the third first color moments (in the HSV color space) computed in a small neighborhood of the detected salient points. In [15], the authors complete this approach by adding a texture feature computed by Gabor filters in a 9×9 neighborhood of salient points.

All these different approaches show that salient points can be used in very different ways for content-based image retrieval. Obviously, it seems natural to think that all indexing methods proposed in literature can be adapted to the saliency-based indexing model. However, it is important to favorably take advantage of the information located in the neighborhood of the salient points since it represents perceptually relevant information.
In this paper, we present a new approach motivated by the following observations:

1. the use of differential invariants only provide a very local description of the salient points and differential invariants are known to be sensitive to noise (although the authors of [3] have limited this drawback by working only with the first order invariants);

2. we are not interested in methods using a fixed number of salient points. Obviously, this number depends on many factors such as the image size, image content, imaging conditions, etc. and fixing the same number for all images inevitably spoils retrieval results;

3. none of the existing methods consider the structure of image content in the neighborhood of salient points. The only structural constraint used in previous approaches is the geometric constraint between salient points proposed in [3, 13]. It seems important to us to strongly consider color spatial structure as an important property of salient point neighborhood.

To satisfy these observations, our retrieval scheme uses an adaptive approach to determine the number of salient points to be extracted and compute a color descriptor based on the correlogram introduced in [6] and adapted to salient features. This approach is detailed in the next sections.

3.2. Adaptive Computation of the Number of Salient Points to be Extracted

In this section, we propose a method determining automatically the number p of salient points sufficient to provide a good representation of the image content.
For this purpose, we consider the list λ of the the 2^{k+l+2r} saliency values of S_{2^r} that have been ordered in the decreasing order of saliency during the choice of the salient points (cf. section 2.5):

$$\lambda = \{S_{2^r}(x_1, y_1), \ldots, S_{2^r}(x_{k+l+2r}, y_{k+l+2r})\}$$

with $S_{2^r}(x_t, y_t) \geq S_{2^r}(x_{t+1}, y_{t+1})$, $1 \leq t \leq k+l+2r$.
We then denote by $\xi(\lambda)$ the total energy embedded in the saliency values of λ:

$$\xi(\lambda) = \sum_{i=1}^{k+l+2r} S_{2^r}(x_i, y_i).$$

Finally, we compute a cumulative histogram $\mathcal{H}_\lambda$ of saliency values that gives, for each ordered saliency index s ($1 \leq s \leq k+l+2r$), the percentage of $\xi(\lambda)$, reached up to s:

$$\mathcal{H}_\lambda(s) = \frac{100}{\xi(\lambda)} \sum_{i=1}^{s} S_{2^r}(x_i, y_i)$$

with $1 \leq s \leq k+l+2r$.
By using an energy treshold τ, we can determine the number p of salient points to be extracted according to:

$$p = Arg\{\mathcal{H}_\lambda(s) \geq \tau\}.$$

In practice, $30\% \leq \tau \leq 50\%$ provides a good image representation.

3.3. Photometric Invariant Salient Color Signature

From the initial paper of Swain and Ballard [19], object recognition by using photometric information has attracted

a big amount of research. Indeed, color histograms are known to be relatively robust against various image transforms such as translation, rotation, scale changes, occlusions and view position. However, color histograms are not robust against lighting changes [19] and using color histograms without a color constancy algorithm can lead to drastically poor results when the images of the database have been taken under unconstrained imaging conditions. In this paper, we have chosen to use the c1c2c3 color model recently proposed by Gevers and Smeulders in [2] which is known to be photometric invariant under the assumption of a dichromatic reflection model with white light source for matte, dull objects. This color model is defined by:

$$\begin{array}{l} c1 = arctan(\frac{R}{max\{G,B\}}) \\ c2 = arctan(\frac{G}{max\{R,B\}}) \\ c3 = arctan(\frac{B}{max\{R,G\}}) \end{array} \quad (2)$$

where R, G and B denotes respectively red, green and blue pixel values.

Another drawback of color histograms is that they do not consider the spatial structure of color values which is of prime importance for content-based image retrieval. To overcome this drawback, several alternatives to classical color histograms have been proposed in literature, including the correlogram [6] and more generally the geometric histograms [12], or the color coherence vector [11]. In this paper, we have extended the notion of correlogram to salient features. Let us recall that the color correlogram of an image I is defined by:

$$\gamma^{(l)}_{\sigma_i,\sigma_j}(I) = \underset{\substack{p_1 \in I_{\sigma_i} \\ p_2 \in I}}{Pr} \{p_2 \in I_{\sigma_j},\ |p_1 - p_2| = l\}$$

where:

- σ_i, σ_j denote color values quantized into m possible values $[\sigma_1, \sigma_2, \ldots, \sigma_m]$;
- $|p_1 - p_2|$ denote the distance between pixels p_1 and p_2 using the L^∞ norm;
- $p \in I_{\sigma_i}$ is similar to $I(p) = \sigma_i$ (i.e. p is of color σ_i);
- $l \in [1, \ldots, d]$ is a set of distances of interest. A large d leads to expensive computation and storage requirement but a small d might compromise the robustness of the color signature.

Literally speaking, $\gamma^{(l)}_{\sigma_i,\sigma_j}(I)$ gives the probability that a pixel of color σ_j is at distance l from a pixel of color σ_i.

In practice, we use the autocorrelogram $\alpha^{(l)}_{\sigma}(I)$ for image retrieval due to its low storage requirement ($O(md)$ instead of $O(m^2 d)$ for the correlogram):

$$\alpha^{(l)}_{\sigma}(I) = \gamma^{(l)}_{\sigma,\sigma}(I)$$

To obtain a photometric invariant salient correlogram, we proceed as follows:

- we first apply a c1c2c3 color transform (cf. equation 2) to the original image I to obtain a photometric invariant color image I^{c1c2c3};
- the image I^{c1c2c3} is then uniformly quantized into q bins per color axe resulting into a quantized image $\overline{I}^{c1c2c3}$ with q^3 possible color values;
- a RGB color mask J is created from I by inserting the p detected salient points (p being fixed by the method presented in section 3.2) along with their k-neighborhood (k being small in practice);
- all pixels of J that have a local saturation and intensity smaller than 5% of the total range are discarded from the autocorrelogram computation since they are known to have instable color values [2];
- a salient color correlogram is defined by:

$$\overline{\gamma}^{(l)}_{\sigma_i,\sigma_j}(I) = \underset{\substack{p_1 \in J \\ p_1 \in \overline{I}^{c1c2c3}_{\sigma_i} \\ p_2 \in \overline{I}^{c1c2c3}}}{Pr} \{p_2 \in \overline{I}^{c1c2c3}_{\sigma_j},\ |p_1 - p_2| = k\}$$

  The corresponding salient autocorrelogram can be obviously deduced:

$$\overline{\alpha}^{(l)}_{\sigma}(I) = \overline{\gamma}^{(l)}_{\sigma,\sigma}(I).$$

The salient autocorrelogram thus constitutes the feature vector of the image I and contains $q^3 d$ components. Let us point out that this size can be however quite important. Gevers and Smeulders have indeed shown in [2] that good retrieval results can be expected by using the c1c2c3 color model with at least 16 bins per color axe. Moreover, Huang et al. have shown that using the autocorrelogram with $d = 4$ possible distances provide good retrieval results. Combining these two parameters produces a salient feature vector of size 16384 per image which may drastically increase the retrieval execution time. However, experimental results have shown that most components of this vector have null values and we can therefore easily run-length encode it so that only non null values remain. Using this encoding allowed us to greatly improve retrieval time.

Once the salient signatures are computed, a similarity measure must be proposed to compare the corresponding images. In [6], it is shown that the distance d_1, defined by:

$$|I - I'|_{\overline{\alpha},d_1} = \sum_{\substack{\sigma \in [\sigma_1,\sigma_m] \\ l \in [1,d]}} \frac{|\overline{\alpha}^{(l)}_{\sigma}(I) - \overline{\alpha}^{(l)}_{\sigma}(I')|}{1 + \overline{\alpha}^{(l)}_{\sigma}(I) + \overline{\alpha}^{(l)}_{\sigma}(I')}$$

outperforms classical Minkowski metrics. This distance is in fact a weighted version of the L^1 metric and has a theoretical justification as shown in [6]. We have thus decided to use this distance to measure the retrieval performances of our salient descriptor.

3.4. Retrieval Results

To measure the performances of our photometric invariant salient autocorrelogram, we have used an image database with $N_1 = 2000$ images that have been extracted from several TV programs by a temporal segmentation software. It is important to note that the so built database is a general-purpose database in which images have been shot with various imaging conditions. Due to the lack of space, we do not include in this paper an example view of our database content. We have thus a high degree of changes between images of the database including lighting changes and geometric transforms (zoom, pan, etc.). We can thus consider that this database constitutes a very difficult working basis. From these 2000 key frames, we have extracted $N_2 = 18$ images Q_i $(1 \leq i \leq 18)$ that will serve as queries to measure retrieval results. These query images have been chosen so that there exist similar images in the database but with some transformations such as changes in viewing position, imaging conditions, geometric transforms, etc. We are thus not interested in queries for which approximately the same images can be found in the database. For each query Q_i, we manually searched all similar images to Q_i in the database, resulting in an image list S_i of size $|S_i|$ that represents the ground truth for Q_i.
We have also decided to compare our salient descriptor to the SCD (*Scalable Color Descriptor*) proposed by the MPEG-7 standard [9]. Let us recall that SCD builds a color histogram by uniformly quantizing the HSV color space into 256 bins. This descriptor has been chosen for the following reasons: firstly, SCD is based on the HSV color space which is known to be robust to lighting changes [2]; secondly, SCD is based on a color histogram construction that is known to be relatively robust to most of the classical image transforms [19].
We have not implemented the compression step required by SCD allowing us to use the powerful quadratic distance proposed in [4] to compare the obtained color histograms. This distance can be considered close to a perceptual distance since it considers the similarity between histogram bins.
To obtain a measure that considers the rank of each retrieved image, we do not use the classical precision/recall measures. Instead, we define a ranking percentile metric by:

$$\overline{r}_i = \frac{1}{|S_i|} \sum_{k=1}^{|S(i)|} \frac{N_3 - Rank(S_i^k)}{N_3 - k}$$

where N_3 denotes the number of results displayed to the user, $Rank(I)$ denotes the position of the image I in the result list displayed to the user ($Rank(I) \leq N_3$) and S_i^k denotes the k^{th} element of S_i.
This ranking measure ranges from $\overline{r}_i = 0$ for the worst possible match (i.e. none relevant image is found) to $\overline{r}_i = 1$ for a perfect match (i.e. all relevant images are found in the $|S_i|$ first positions).

The parameters chosen for our salient autocorrelogram are the following:

- the wavelet transform is performed up to the resolution 2^{-2} and the Haar wavelet basis is used;
- the energy threshold τ used to determine the number of detected salient points is set to $\tau = 30\%$;
- the c1c2c3 color model is quantized into 32 bins per color axe;
- all pixels belonging to the $k = 2$-neighborhood of salient points are considered during the color mask computation (cf. section 3.3);
- the distances of interest for the autocorrelogram computation are set to $\{1, 3, 5, 7\}$.

Finally, we present the first $N_3 = 25$ retrieval results to the user (i.e. 1,25% of the database).
The figure 6 illustrates the retrieval results obtained by our salient descriptor compared to those obtained by SCD.

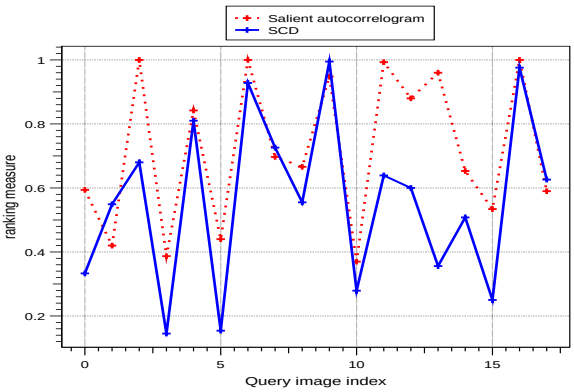


Fig. 6. Retrieval results and comparison with SCD

For 14 of the 18 queries, our salient detector outperforms SCD with a ranking gain ranging from 2,45% to 186%, the average gain being 68,57%. For the four remaining cases for which SCD performs better than our descriptor, the gain ranges from 4% to 23% with an average gain of 9,5%. After a deeper analysis of these last four cases, we are convinced that they are caused by:

- the fine quantization we have imposed in the c1c2c3 color model (32 bins per color axe) whereas SCD performs a coarse quantization into a total number of 256 bins. In few cases, a too much fine quantization can spoil retrieval results;
- the use of the d_1 distance does not consider the similarity between bins of the salient descriptor instead of the quadratic distance. In fact, one should expect better results with a more perceptual distance.

4. CONCLUSIONS AND PERSPECTIVES

In this paper, we have firstly proposed an efficient wavelet-based salient point detector that we have then used to build a saliency-based image retrieval system. The efficiency of the proposed detector lies in a zerotree representation of the wavelet coefficients and in the formulation of the saliency that does not favor any direction. Then, our proposed salient color descriptor has shown a good behaviour compared to SCD proposed by MPEG-7 due to the use of a photometric color model and to the extension of the well-known correlogram allowing to consider the spatial structure of colors.
Now, some improvements can be proposed. Firstly, the salient points are detected by working only on the luminance pixel values. With this approach, high lighting changes may generate high saliency values and thus false salient points. This scheme can thus be improved by detecting salient points directly in a photometric invariant color space. Secondly, the d_1 similarity measure is not a perceptual metric and we are convinced that the extension of a perceptual distance to salient features can lead to better retrieval results. Finally, combining salient correlogram with other features such as texture can greatly improve results.

5. REFERENCES

[1] Bres S. and Jolion J.M. Detection of Interest Points for Image Indexation. In Springer Verlag, editor, 3^{rd} *Int. Conf. on Visual Information Systems*, Lecture Notes in Computer Science, pages 427–434, Amsterdam, June 1999.

[2] Gevers T. and Smeulders A.W.M. Color Based Object Recognition. *Pattern Recognition*, 32:453–464, 1999.

[3] Gouet V. and Boujemaa N. Object-based Queries Using Color Points of Interest. In *IEEE Workshop on Content-Based Access of Image and Video Libraries*, Hawaii, December 2001.

[4] Hafner J., Sawhney H.S., Equitz W., Flickner M., and Niblack W. Efficient Color Histogram Indexing for Quadratic Form Distance Functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):729–736, July 1995.

[5] Harris C. and Stephens M. A Combined Corner and Edge Detector. In 4^{th} *Alvey Vision Conf.*, pages 147–151, 1988.

[6] Huang J., Kumar S.R., Mitra M., and Zhu W.J. Image Indesing using Color Correlograms. In *Proc. of the* $16^t h$ *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 762–768, 1997.

[7] Loupias E., Sebe N., Bres S., and Jolion J.M. Wavelet-Based Salient Points for Image Retrieval. In *IEEE Int. Conf. on Image Processing*, Vancouver, September 2000.

[8] Mallat S.G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, July 1989.

[9] Manjunath B.S., Ohm J.R., Vasudevan V.V., and Yamada A. Color and Texture Descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):703–715, June 2001.

[10] Montesinos P., Gouet V., and Deriche R. Differential Invariants for Color Images. In *Proc. of the IAPR Conf. on Pattern Recognition*, Brisbane, August 1998.

[11] Pass G. and Zabih R. Histogram Refinement for Content-Based Image Retrieval. In *Proc. of the IEEE Workshop on Applications of Computer Vision*, pages 96–102, Sarasota, December 1996.

[12] Rao A., Srihari R.K., and Zhang Z. Geometric Histogram: A Distribution of Geometric Configurations of Color Subsets. *Internet Imaging*, 3(964):91–101, 2000.

[13] Schmid C. and Mohr R. Local Grayvalue Invariants for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, May 1997.

[14] Schmid C., Mohr R., and Bauckhage C. Evaluation of Interest Point Detectors. *Int. Journal of Computer Vision*, 37(2):151–172, June 2000.

[15] Sebe N. and Lew M.S. Salient Points for Content-based Retrieval. In *Proc. of the British Machine Vision Conference*, Manchester-UK, 2001.

[16] Sebe N., Tian Q., Loupias E., Lew M.S., and Huang T.S. Color Indexing Using Wavelet-based Salient Points. In *Proc. of the IEEE Workshop on Content-Based Access of Image and Video Libraries*, pages 15–19, Hilton Head-South California, June 2000.

[17] Shapiro J.M. Embedded Image Coding Using Zerotrees of Wavelet Coefficients. *IEEE Transactions on Signal Processing*, 41(12):3445–3462, December 1993.

[18] Smeulders A.W.M., Worring M., Santini S., Gupta A., and Jain R. Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.

[19] Swain M.J. and Ballard D.H. Color Indexing. *Int. Journal of Computer Vision*, 7(1):11–32, 1991.

[20] Wolf C., Jolion J.M., Kropatsch W., and Bischof H. Content based Image Retrieval using Interest Points and Texture Features. In *Proc. of the Int. Conf. on Pattern Recognition*, volume 4, pages 234–237, Barcelona, September 2000.