

Fast Wavelet-Based Image Characterization for Highly Adaptive Image Retrieval

Gwénolé Quellec, Mathieu Lamard, Guy Cazuguel, *Member, IEEE*, Béatrice Cochener, and Christian Roux, *Fellow, IEEE*

Abstract—Adaptive wavelet-based image characterizations have been proposed in previous works for content-based image retrieval (CBIR) applications. In these applications, the same wavelet basis was used to characterize each query image: This wavelet basis was tuned to maximize the retrieval performance in a training data set. We take it one step further in this paper: A different wavelet basis is used to characterize each query image. A regression function, which is tuned to maximize the retrieval performance in the training data set, is used to estimate the best wavelet filter, i.e., in terms of expected retrieval performance, for each query image. A simple image characterization, which is based on the standardized moments of the wavelet coefficient distributions, is presented. An algorithm is proposed to compute this image characterization almost instantly for every possible separable or nonseparable wavelet filter. Therefore, using a different wavelet basis for each query image does not considerably increase computation times. On the other hand, significant retrieval performance increases were obtained in a medical image data set, a texture data set, a face recognition data set, and an object picture data set. This additional flexibility in wavelet adaptation paves the way to relevance feedback on image characterization itself and not simply on the way image characterizations are combined.

Index Terms—Content-based image retrieval (CBIR), relevance feedback, wavelet adaptation, wavelet transform.

I. INTRODUCTION

OVER THE last decades, the wavelet transform has become a major image characterization tool. One advantage of the wavelet transform over alternative methods is the ability to tune the underlying wavelet basis to users' needs, e.g., to optimize compression, classification, or retrieval performances. Originally, wavelet adaptation was mostly used to approximate a reference signal up to a desired scale [1], [2]. With the advent of the lifting scheme, wavelet adaptation has become more widespread, both for separable [3]–[5] and nonseparable [6]–[8] wavelet transforms. Recently, we have introduced two adaptive

wavelet-based image characterizations [5], [8], which have been successfully applied to different problems [9], [10].

One possible application of wavelet adaptation is content-based image retrieval (CBIR), which is an increasingly popular discipline in computer science [11]–[14]. The goal of CBIR is to automatically select, i.e., in a reference data set, images that resemble a query image. Image characterizations are used to catch similarities between images. Modern retrieval systems usually rely on machine learning to bridge the semantic gap between low-level image characterizations and users' perception of relevance (either through relevance feedback or through offline training) [11]. A popular example of CBIR framework is the bag-of-features (BoF) model [15]. Two wavelet-based CBIR frameworks were proposed in [5] and [8]. The main advantage of wavelet-based CBIR frameworks lies in the ability to tune the wavelet basis and, therefore, image characterizations in order to optimize the retrieval performance (e.g., the precision and the recall). In these frameworks, the same wavelet basis was used to characterize each query image and each reference image; the basis was tuned to maximize the retrieval performance in a manually annotated training data set. We take it one step further in this paper: A different wavelet basis is used to characterize each query image. A regression function, which is tuned to maximize the retrieval performance in the training data set, is used to estimate the best wavelet filter, i.e., in terms of expected retrieval performance, for each query image. Note that comparing image characterizations obtained with different wavelet bases would be inconsistent. As a consequence, using a new wavelet basis for each query image implies characterizing again each reference image, using this new wavelet basis.

With previously proposed approaches [5], [8], characterizing again each reference image would be prohibitive for time reasons. An algorithmic breakthrough was needed. A simple image characterization is presented. Similar to previous approaches, images are characterized by the distribution of the wavelet coefficients at different scales and along different directions [5], [8]. However, the proposed image characterization is lighter: It is simply based on standardized moments of the wavelet coefficient distribution. In order to easily characterize again each reference image, we propose to compute a characterization map as follows: 1) exact image characterizations are computed for a limited number of wavelet filters of given support and 2) approximate characterizations can be computed almost instantaneously for every possible wavelet filter of equal support. Such characterization maps can be computed indifferently for separable or nonseparable wavelet filters. Due to these characterization maps, characterizing an image again using a different wavelet filter does not involve reprocessing

Manuscript received June 25, 2010; revised July 11, 2011 and November 21, 2011; accepted December 11, 2011. Date of publication December 21, 2011; date of current version March 21, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hsueh-Ming Hang.

G. Quellec and M. Lamard are with the LaTIM Inserm Research Unit 1101, 29200 Brest, France (e-mail: gwenole.quellec@inserm.fr).

G. Cazuguel and C. Roux are with LaTIM Inserm Research Unit 1101, 29200 Brest, France, and also with the Department of Image et Traitement de l'Information, Institut Telecom, Telecom Bretagne, Université Européenne de Bretagne, 29200 Brest, France.

B. Cochener is with Centre Hospitalier Universitaire Brest, Service d'Ophthalmologie, 29200 Brest, France.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2011.2180915

the image. In conclusion, using a different wavelet filter for each query image in a CBIR application is now possible.

The remainder of this paper is organized as follows. Section II describes the proposed image characterization given a wavelet filter. Section III addresses the design of characterization maps. Applications to image retrieval are presented in Section IV. The proposed framework is applied to four image data sets in Section V. We end up with a discussion in Section VI.

II. CHARACTERIZING ONE IMAGE WITH A GIVEN WAVELET FILTER

Let $I = (I_{i,j})_{i=0,\dots,M-1,j=0,\dots,N-1}$ be an image of size $M \times N$ pixels. Let $w = (w_{k,l})_{k=-K,\dots,K,l=-L,\dots,L}$ be a wavelet filter of support $(2K+1) \times (2L+1)$. By definition of a wavelet filter, the following relation holds for w [16]:

$$\sum_{u=-K,\dots,K,v=-L,\dots,L} w_{u,v} = 0. \quad (1)$$

By definition, the detail coefficients of the wavelet transform of I , at any location (i, j) and any analysis scale $s \in \mathbb{N}^*$, are given by

$$x_{i,j,s} = \sum_{u=-K,\dots,K,v=-L,\dots,L} w_{u,v} I_{i-su,j-sv}. \quad (2)$$

According to (1), (2) can be rewritten as follows:

$$x_{i,j,s} = \sum_{\substack{u=-K,\dots,K,v=-L,\dots,L, \\ (u,v) \neq (0,0)}} w_{u,v} (I_{i-su,j-sv} - I_{i,j}). \quad (3)$$

We propose to characterize the distribution of detail coefficients $(x_{i,j,s})_{i=0,\dots,M-1,j=0,\dots,N-1}$ in I with standardized moments. In the particular case of texture images, Wouwer *et al.* [17] have shown that the distribution of detail coefficients at any analysis scale can be modeled meaningfully by an unskewed zero-mean generalized Gaussian function; we extended this observation to a more general class of images in previous works [5]. Generalized Gaussian functions have two parameters: α and β , i.e., the scale and shape parameters, respectively. The first four standardized moments (μ , σ , ν , and κ) are related to α and β through these following relations [18]:

$$\mu = \nu = 0 \quad (4)$$

$$\sigma = \alpha \sqrt{\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}} \quad (5)$$

$$\kappa = \frac{\Gamma(5/\beta)\Gamma(1/\beta)}{\Gamma(3/\beta)^2} - 3 \quad (6)$$

where Γ is the Gamma special function. In particular, σ and κ are closely related to α and β , respectively. As a consequence, we propose to characterize the distribution of the detail coefficients at any analysis scale s by its standard deviation $\sigma_{s,K,L}(I)$ and its kurtosis $\kappa_{s,K,L}(I)$, i.e., after subtracting the mean. For short, $\sigma_{s,K,L}(I)$ and $\kappa_{s,K,L}(I)$ are noted σ_s and κ_s , respectively, when I , K , and L are not ambiguous. σ_s and κ_s are given by

$$\sigma_s = \sqrt{m_{s,2}} \quad (7)$$

$$\kappa_s = \frac{m_{s,4}}{\sigma_s^4} - 3 \quad (8)$$

where $m_{s,d}$ is a polynomial function of the wavelet transform of I [and consequently of the wavelet filter coefficients; see (3)] as follows:

$$m_{s,d} = \frac{s^2}{MN} \sum_{i \in \mathbb{I}_s, j \in \mathbb{J}_s} x_{i,j,s}^d \quad (9)$$

where \mathbb{I}_s and \mathbb{J}_s denote the sets $\{0, s, 2s, \dots, M-1\}$ and $\{0, s, 2s, \dots, N-1\}$, respectively. According to (3) and (7)–(9), the complexity of image characterization is in $O(MNKL/s^2)$.

III. BUILDING A CHARACTERIZATION MAP

A. Wavelet Filter Space

Let $\mathbb{W}_{K,L}$ denote the space of all wavelet filters of support $(2K+1) \times (2L+1)$. Let D denote its dimension. Because the central coefficient of each filter $w \in \mathbb{W}_{K,L}$ is constrained by (1), $D = (2K+1)(2L+1) - 1$ (i.e., $\mathbb{W}_{K,L} = \mathbb{R}^D$).

B. Characterization Map and Characterization Derivative Maps

For wavelet adaptation purposes, we propose to compute the characterization (σ_s, κ_s) of I for a given analysis scale $s \in \mathbb{N}^*$ and each filter $w \in \mathbb{W}_{K,L}$. The resulting set of characterizations is referred to as the characterization map of I (given K , L , and s). For wavelet adaptation purposes, it is also useful to compute the first-order derivatives of each characterization. The resulting sets of characterization derivatives are referred to as the characterization derivative maps of I .

Since characterizing an image with a given wavelet filter can be time-consuming (the complexity is in $O(MNKL/s^2)$; see Section II), we propose to approximate the characterization map and the characterization derivative maps, as described in Sections III-C and III-D. In order to speed up computations further, the invariances of the characterization map and of the characterization derivative maps are studied in Section III-E. Some practical considerations for building and visualizing characterization maps are highlighted in Section III-F.

C. Approximate Characterization Map

We propose to approximate the characterization map and the characterization derivative maps using Taylor expansions. The exact characterization and characterization derivatives are computed for a finite set of wavelet filters, which are referred to as key wavelet filters, and the remainder of each map is approximated using Taylor expansions. The set of key wavelet filters is denoted by $\mathbb{W}_{K,L}^0 \subset \mathbb{W}_{K,L}$. Let $w_0 \in \mathbb{W}_{K,L}^0$ be a key wavelet filter. The Taylor expansion of function f (either σ_s , κ_s , $(\partial\sigma_s/\partial w_{k,l})$, or $(\partial\kappa_s/\partial w_{k,l})$) in the neighborhood of w_0 is given by the following formal relation [19]:

$$f(w) = \sum_{i=0}^{n_T} \frac{f^{(i)}(w_0)}{i!} (w - w_0)^i + o(\|w - w_0\|_2^{n_T}) \quad (10)$$

where n_T denotes the order of the Taylor expansion and $\|\cdot\|_2$ denotes the L2-norm

$$\|w\|_2 = \sqrt{\frac{1}{(2K+1)(2L+1)} \sum_{k=-K, \dots, K, l=-L, \dots, L} w_{k,l}^2}. \quad (11)$$

Note that f is multidimensional: It is a function of the wavelet filter coefficients. As a consequence, the second term of (10) (corresponding to $i = 1$) is given by

$$\nabla f(w_0) \cdot (w - w_0) \quad (12)$$

where ∇f is the gradient of f . Its third term (corresponding to $i = 2$) is given by

$$\frac{1}{2}(w - w_0)^t \cdot \mathbb{H}_f(w_0) \cdot (w - w_0) \quad (13)$$

where \mathbb{H}_f is the Hessian of f .

According to Taylor's theorem [19], (10) holds if f is n_T times differentiable at w_0 . We can check that σ_s and κ_s (and therefore $(\partial\sigma_s/\partial w_{k,l})$ and $(\partial\kappa_s/\partial w_{k,l})$) are infinitely differentiable when $\sigma_s > 0$. In particular, it should be noted that σ_s and κ_s are functions of $w_{u,v}(I_{i-su,j-sv} - I_{i,j})$ terms composed of polynomials, fractions, and square roots [see (3) and (7)–(9)]. The standard deviation σ_s equals 0 when image I is constant everywhere or when $w_0 = 0$; σ_s cannot be strictly negative [see (7) and (9)]. In the trivial case where I is constant everywhere, characterization maps are also constant everywhere; therefore we do not need to compute them. Images are assumed nonconstant in the following sections.

D. Derivatives of the Proposed Characterization With Respect to Filter Coefficients

In order to compute the Taylor expansions above, the first few order derivatives of the proposed image characterization, i.e., the first few order derivatives of (7) and (8), with respect to wavelet filter coefficients $w_{k,l}$, need to be computed for each key wavelet filter $w_0 \in \mathbb{W}_{K,L}^0$. Let $\delta_{s,d}$ and $\delta'_{s,d}$ denote the following polynomial functions of the wavelet transform of I :

$$\delta_{s,d}(k, l) = \frac{s^2}{MN} \sum_{i \in \mathbb{I}_s, j \in \mathbb{J}_s} [(I_{i-sk,j-sl} - I_{i,j}) x_{i,j,s}^d] \quad (14)$$

$$\begin{aligned} \delta'_{s,d}(k, l, u, v) = \frac{s^2}{MN} \sum_{i \in \mathbb{I}_s, j \in \mathbb{J}_s} & [(I_{i-sk,j-sl} - I_{i,j}) \\ & \times (I_{i-su,j-sv} - I_{i,j}) x_{i,j,s}^d]. \end{aligned} \quad (15)$$

The first-order derivatives of the proposed image characterization are given by

$$\frac{\partial\sigma_s}{\partial w_{k,l}} = \frac{\delta_{s,1}(k, l)}{\sigma_s} \quad (16)$$

$$\frac{\partial\kappa_s}{\partial w_{k,l}} = \frac{4}{\sigma_s^6} (\sigma_s^2 \delta_{s,3}(k, l) - \delta_{s,1}(k, l) m_{s,4}). \quad (17)$$

Its second-order derivatives are given by

$$\frac{\partial^2\sigma_s}{\partial w_{k,l}\partial w_{u,v}} = \frac{\delta'_{s,0}(k, l, u, v)}{\sigma_s} - \frac{\delta_{s,1}(k, l) \frac{\partial\sigma_s}{\partial w_{u,v}}}{\sigma_s^2} \quad (18)$$

$$\begin{aligned} \frac{\partial^2\kappa_s}{\partial w_{k,l}\partial w_{u,v}} = \frac{4}{\sigma_s^6} & \left(3\sigma_s^2 \delta'_{s,2}(k, l, u, v) \right. \\ & + 2\delta_{s,1}(u, v) \delta_{s,3}(k, l) \\ & \left. - 4\delta_{s,1}(k, l) \delta_{s,3}(u, v) \right. \\ & \left. - \delta'_{s,0}(k, l, u, v) m_{s,4} \right) \\ & - \frac{6}{\sigma_s^2} \delta_{s,1}(u, v) \frac{\partial\kappa_s}{\partial w_{k,l}}. \end{aligned} \quad (19)$$

Its third-order derivatives, and some computation details, are provided in Appendix A. Higher order derivatives were not used in this paper.

E. Invariances of the Characterization Map and of its Derivatives

In order to reduce the cardinal of $\mathbb{W}_{K,L}^0$, we propose to find invariances in the characterization map and its derivatives.

First, if a wavelet filter $w \in \mathbb{W}_{K,L}$ is multiplied by a positive real number $\lambda \in \mathbb{R}^+$, then according to (3), (9), and (14), $\delta_{s,d}(k, l)$ and $m_{s,d}$ are multiplied by λ^d , $\forall s, k$, and l . Consequently, if w is multiplied by λ , then σ_s is multiplied by λ , and κ_s is unchanged [see (7) and (8)]. As for the derivatives, if w is multiplied by λ , then $(\partial\sigma_s/\partial w_{k,l})$ is unchanged, and $(\partial\kappa_s/\partial w_{k,l})$ is divided by λ [see (16) and (17)]. This invariance analysis implies that we only need to compute the characterization map and its derivatives for wavelet filters w on the unit sphere $\|w\|_2 = 1$.

Second, if w is multiplied by -1 , then $\delta_{s,d}(k, l)$ and $m_{s,d}$ are multiplied by $(-1)^d$, $\forall s, k$, and l [see (3), (9), and (14)]. Consequently, if w is multiplied by -1 , then σ_s and κ_s are unchanged [see (7) and (8)], and $(\partial\sigma_s/\partial w_{k,l})$ and $(\partial\kappa_s/\partial w_{k,l})$ are multiplied by -1 [see (16) and (17)]. This invariance analysis implies that we only need to compute the characterization map and its derivatives on one half of the unit sphere $\|w\|_2 = 1$.

F. Practical Considerations for Building the Characterization Map

The proposed procedure for building the characterization map is illustrated in Fig. 1 for $K = 0$ and $L = 1$ ($D = 2$).

The first step in building the characterization map is to extract a set $\mathbb{W}_{K,L}^0 \subset \mathbb{W}_{K,L}$ of key wavelet filters. The exact characterizations and characterization derivatives associated with these key filters are computed, as described in Sections II and III-D, respectively. The elements of $\mathbb{W}_{K,L}^0$ should be selected approximately uniformly in one half of the unit sphere $\|w\|_2 = 1$ (see Section III-E). A good $\mathbb{W}_{K,L}^0$ set can be obtained as described hereafter [see also Fig. 1(a)]. For each dimension $d = 1, \dots, D$ of the wavelet space, a set of n^{D-1} key wavelet filters, i.e., $n \in \mathbb{N}^*$, is generated as follows:

- the d th unconstrained coefficient of each filter w is set to 1;
- the i th unconstrained coefficient of each filter w , with $i \in \{1, 2, \dots, d-1, d+1, \dots, D\}$, is in $\{\tan(-(\pi/4)) + j(\pi/2n), j = 0, \dots, n-1\}$ [see Fig. 1(a)];
- each filter w is divided by $\|w\|_2$.

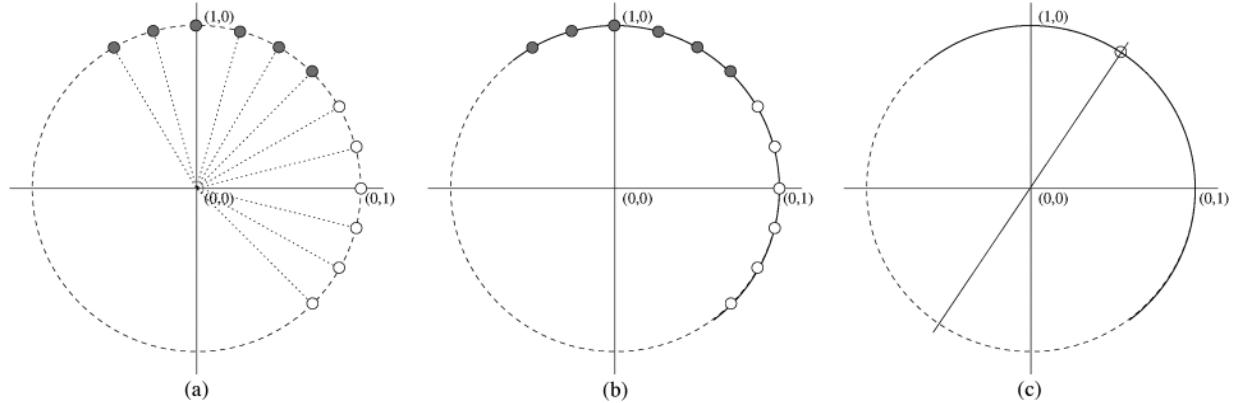


Fig. 1. Characterization map building. In this example ($K = 0$, $L = 1$, and therefore, $D = 2$), wavelet filters w have three coefficients: $w_{0,-1}$, $w_{0,0}$ and $w_{0,1}$ [$w_{0,0}$ is constrained by (1)]. In each figure, the x axis represents the value of the first unconstrained wavelet filter coefficient, i.e., $w_{0,-1}$. The y axis represents the value of the second unconstrained coefficient: $w_{0,1}$. The value of the constrained wavelet coefficient $w_{0,0}$ is not represented. In this example, the number of key wavelet filters is controlled by $n = 6$, and the portion of the map obtained by Taylor expansions is the unit circle restricted to the $[-(\pi/4) - (\pi/4n); (\pi/4) - (\pi/4n)]$ angle interval. In (a) and (b), the two sets of n^{D-1} key wavelet filters described in Section III-F are indicated by circles of two different colors. In (b) and (c), the solid arc indicates the portion of the map obtained by Taylor expansions. In (c), the solid straight line indicates the portion of the map that can be obtained, through invariance analysis, from the characterization estimated at the location of the small circle. (a) Exact computation (cf. Section II). (b) Approximate computation (cf. Section III-C). (c) Invariances (cf. Section III-E).

$W_{K,L}^0$ is the union of these D sets: Its cardinal is therefore Dn^{D-1} . If $D = 2$, then the key wavelet filters are selected exactly uniformly in one half of the unit sphere.

The second step is to approximate the characterizations and characterization derivatives in the remainder of the half unit sphere $\|w\|_2 = 1$ [see Fig. 1(b)]. For each filter w in this set, image characterizations are approximated by Taylor expansions; the closest key wavelet filter plays the role of w_0 (see Section III-C). Taylor expansions can be used in this step because $\|w_0\|_2 = 1$; therefore, $w_0 \neq 0$ (see Section III-C).

The last step is to compute the characterizations and characterization derivatives in the remainder of the $W_{K,L}$ set [see Fig. 1(c)]. Let λ be the L2-norm of filter w in this set. Let $\delta \in \{-1, 1\}$ be a variable indicating whether (w/λ) belongs to the half unit sphere above ($\delta = 1$) or not ($\delta = -1$). The characterizations and characterization derivatives are obtained from $\delta(w/\lambda)$ according to the invariance analysis of Section III-E.

The proposed procedure is applied, in the particular case $D = 2$, to a real-world image in Fig. 2. Characterization map building is assessed in Section V-B.

The proposed wavelet-based image characterization is applied to image retrieval in the following section.

IV. IMAGE RETRIEVAL

Let I_q be a query image and \mathcal{R} be a data set of reference images. In this section, characterization maps are used to rank images $I \in \mathcal{R}$ in increasing order of distance to I_q . Then, the first $\mathcal{K} \in \mathbb{N}^*$ images, which are noted $I_1(I_q), I_2(I_q), \dots, I_{\mathcal{K}}(I_q)$, are retrieved. In this paper, the goal is to retrieve a small set of \mathcal{K} highly relevant images. Therefore, retrieval systems are tuned in a training set in order to maximize the precision at \mathcal{K} , i.e., the fraction of images, among $\{I_1(I_q), I_2(I_q), \dots, I_{\mathcal{K}}(I_q)\}$, that belong to the same category as I_q . If, on the contrary, one would like to retrieve all potentially relevant images in the reference data set, then the recall at a large \mathcal{K} should be maximized instead [11]. Achieving high precision is challenging when the number of categories is high or when the semantic

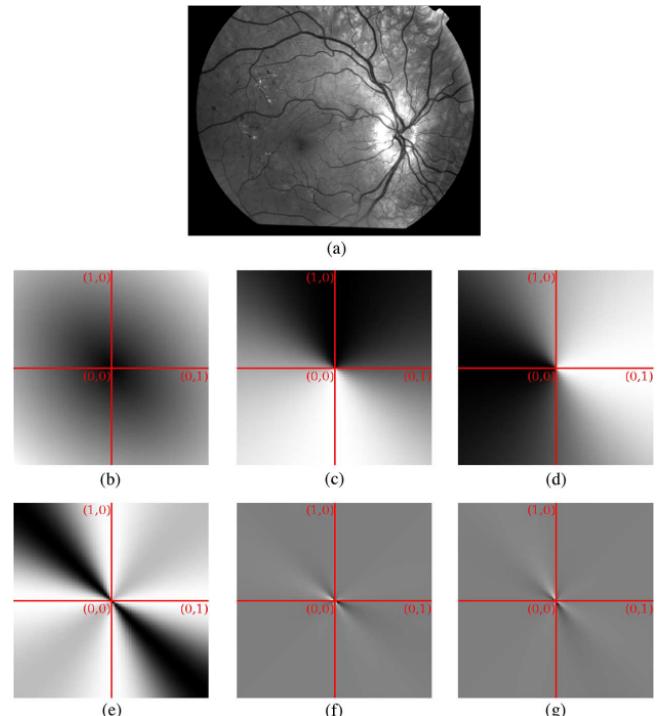


Fig. 2. Example of characterization map, in (b) and (e), and of characterization derivative maps, in (c), (d), (f) and (g), obtained for image (a) at the analysis scale $s = 2$. The scenario described in Fig. 1 was reused ($K = 0$ and $L = 1$); therefore wavelet filters w have three coefficients: $w_{0,-1}$, $w_{0,0}$, and $w_{0,1}$. In each map, the intensity of pixel (x, y) , where $x = w_{0,-1}$ and $y = w_{0,1}$, is proportional to the value of $f(w)$, where f is either σ_2 [in (b)], κ_2 [in (e)], or one of their derivatives. In (b) and (e), black means 0. In (c), (d), (f), and (g), medium gray means 0. (a) Retinal Image. (b) σ_2 . (c) $(\partial\sigma_2)/\partial w_{0,-1}$. (d) $(\partial\sigma_2)/\partial w_{0,0}$. (e) κ_2 . (f) $(\partial\kappa_2)/\partial w_{0,-1}$. (g) $(\partial\kappa_2)/\partial w_{0,1}$.

gap between category assignments and image characteristics (textures, colors, etc.) is wide.

We propose to characterize each image I by the following signature $\mathcal{S}(I) = (\sigma_{s,K,L}(I), \kappa_{s,K,L}(I))_{s \in \mathbb{N}^*, K \in \mathbb{N}, L \in \mathbb{N}}$. In signature $\mathcal{S}(I)$, features $\sigma_{s,K,L}(I)$ and $\kappa_{s,K,L}(I)$ are extracted

at analysis scale s , using a wavelet filter $w^{s,K,L}$ of support $(2K+1) \times (2L+1)$. Based on these signatures, the distance between I_q and another image $I \in \mathcal{R}$ is defined as follows:

$$\begin{aligned} d(I_q, I) = & d_H(I_q, I) \\ & + \left[\sum_{s,K,L} \left(\zeta_{s,K,L} [\sigma_{s,K,L}(I_q) - \sigma_{s,K,L}(I)]^2 \right. \right. \\ & \left. \left. + \eta_{s,K,L} [\kappa_{s,K,L}(I_q) - \kappa_{s,K,L}(I)]^2 \right) \right]^{\frac{1}{2}} \end{aligned} \quad (20)$$

where $d_H(I_q, I)$ denotes the L2-distance between the intensity histogram of I_q and that of I [5]: $d_H(I_q, I)$ is used to compare the low-frequency component of I_q and that of I . Each component in the distance measure is weighted by a real number: $\zeta_{s,K,L} \in \mathbb{R}^+$ or $\eta_{s,K,L} \in \mathbb{R}^+$, where $s \in \mathbb{N}^*$, $K \in \mathbb{N}$, and $L \in \mathbb{N}$.

Several parameters need to be tuned in (20): $\zeta_{s,K,L}$, $\eta_{s,K,L}$, and $w_{k,l}^{s,K,L}$ ($k = -K, \dots, K, l = -L, \dots, L$), $\forall s, K$, and L . These parameters are referred to as the distance parameters. They are tuned in a training image data set \mathcal{T} to maximize the average precision at \mathcal{K} . Note that the same data set can be used as the reference data set \mathcal{R} and as the training data set \mathcal{T} . During the training phase, $\mathcal{R} = \mathcal{T} \setminus \{I_q\}$, $I_q \in \mathcal{T}$. After the training phase, $\mathcal{R} = \mathcal{T}$, $I_q \notin \mathcal{T}$.

Two image retrieval scenarios are considered: 1) a single set of distance parameters is used regardless of the query image (i.e., adaptive image retrieval; see Section IV-A), or 2) a different set of distance parameters is used for each query image (i.e., highly adaptive image retrieval; see Section IV-B). Note that adaptive image retrieval simply is a faster version of previously proposed approaches [5], [8]. From a user perspective, adaptive and highly adaptive image retrieval operate similarly to these previous approaches; the only differences are that adaptive image retrieval is much faster and that highly adaptive image retrieval is both much faster and potentially more precise. The training phase is described hereafter for each scenario. Section IV-C explains how the system processes a query.

A. Adaptive Image Retrieval: Training Procedure

In adaptive image retrieval, distance d is tuned, by a gradient descent, in order to increase the average precision at \mathcal{K} among all query images $I_q \in \mathcal{T}$. Each parameter is modified, in turn, until the average precision at \mathcal{K} in \mathcal{T} converges.

Let p denote the parameter currently being modified (either $\zeta_{s,K,L}$, $\eta_{s,K,L}$, or $w_{k,l}^{s,K,L}$). Optimizing the average precision at \mathcal{K} is not straightforward: Its derivative, with respect to p , is either zero or undefined everywhere on \mathbb{R}^+ . We got around this difficulty as explained hereafter. Let $I_1^-(I_q) \in \mathcal{R}$ be the image minimizing $d(I_q, I_1^-(I_q))$ among images from a different category than I_q . Let $I_1^+(I_q) \in \mathcal{R}$ be the image minimizing $d(I_q, I_1^+(I_q))$ among images from the same category as I_q and such that

$$d(I_q, I_1^+(I_q)) > d(I_q, I_1^-(I_q)). \quad (21)$$

If (21) can be inverted (i.e., $d(I_q, I_1^+(I_q)) \leq d(I_q, I_1^-(I_q))$) for one or several images $I_q \in \mathcal{T}$, then the average precision at \mathcal{K} is expected to increase. In that purpose, parameter p is modified as follows:

$$p = p - \delta_p \sum_{I_q \in \mathcal{T}} \frac{\partial}{\partial p} \frac{d(I_q, I_1^+(I_q))}{d(I_q, I_1^-(I_q))} \quad (22)$$

with $\delta_p \in \mathbb{R}^+$. Equation (22) is iterated until p converges (note that $I_1^-(I_q)$ and $I_1^+(I_q)$ are updated at each iteration).

Whenever p is a wavelet filter coefficient, the characterization derivative maps (see Fig. 2) are used in the gradient descent. In that case, (22) follows from:

$$\begin{aligned} \frac{\partial d(I, J)}{\partial w_{k,l}^{s,K,L}} = & \frac{1}{d(I, J)} \\ & \times \left[\zeta_{s,K,L} \left(\frac{\partial \sigma_{s,K,L}(I)}{\partial w_{k,l}^{s,K,L}} - \frac{\partial \sigma_{s,K,L}(J)}{\partial w_{k,l}^{s,K,L}} \right) \right. \\ & \times (\sigma_{s,K,L}(I) - \sigma_{s,K,L}(J)) \\ & + \eta_{s,K,L} \left(\frac{\partial \kappa_{s,K,L}(I)}{\partial w_{k,l}^{s,K,L}} - \frac{\partial \kappa_{s,K,L}(J)}{\partial w_{k,l}^{s,K,L}} \right) \\ & \left. \times (\kappa_{s,K,L}(I) - \kappa_{s,K,L}(J)) \right]. \end{aligned} \quad (23)$$

B. Highly Adaptive Image Retrieval: Training Procedure

In highly adaptive image retrieval, a different set of distance parameters is used for each query image I_q . Precisely, distance parameters (wavelet filters and weights) are allowed to vary continuously in signature space. A continuous regression function is used to map $\bar{\mathcal{S}}(I_q)$, i.e., the initial signature of I_q (obtained by an initial set of filters $\bar{w}^{s,K,L}$), to new distance parameters for I_q : a new set of weights $(\zeta_{s,K,L}(I_q), \eta_{s,K,L}(I_q))_{s \in \mathbb{N}^*, K \in \mathbb{N}, L \in \mathbb{N}}$ and a new set of filters $(w^{s,K,L}(I_q))_{s \in \mathbb{N}^*, K \in \mathbb{N}, L \in \mathbb{N}}$. The initial set of filters $\bar{w}^{s,K,L}$ is obtained by the training procedure of the adaptive image retrieval (see Section IV-A). Once new distance parameters are obtained for I_q , $\mathcal{S}(I_q)$, which is the final signature of I_q , is computed using $(w^{s,K,L}(I_q))_{s \in \mathbb{N}^*, K \in \mathbb{N}, L \in \mathbb{N}}$. For consistency, the signature of each reference image also needs to be computed using $(w^{s,K,L}(I_q))_{s \in \mathbb{N}^*, K \in \mathbb{N}, L \in \mathbb{N}}$.

The continuity property implies that two images with similar initial signatures are mapped to similar distance parameters. Let p denote one distance parameter (either $\zeta_{s,K,L}$, $\eta_{s,K,L}$, or $w_{k,l}^{s,K,L}$). The following the continuity equation holds for p :

$$\begin{aligned} \forall \varepsilon > 0 \quad \exists \mu > 0 \quad \forall I, J \\ \|\bar{\mathcal{S}}(I) - \bar{\mathcal{S}}(J)\|_2 \leq \mu \Rightarrow \|p(I) - p(J)\|_2 \leq \varepsilon. \end{aligned} \quad (24)$$

The regression functions are trained on \mathcal{T} . For each image $I_q \in \mathcal{T}$, the distance parameters for I_q are tuned to maximize the precision at \mathcal{K} for I_q while respecting the continuity constraint. Let p denote one distance parameter (either $\zeta_{s,K,L}$, $\eta_{s,K,L}$, or $w_{k,l}^{s,K,L}$).

$w_{k,l}^{s,K,L}$). In order to respect the continuity constraint, (22) is replaced by the following equation:

$$p(I_q) = p(I_q) - \delta_p \sum_{I \in \mathcal{R}} e^{-\lambda_c \|\bar{\mathcal{S}}(I_q) - \bar{\mathcal{S}}(I)\|_2} \frac{\partial}{\partial p} \frac{d(I, I_1^+(I))}{d(I, I_1^-(I))} \quad (25)$$

where $\lambda_c \in \mathbb{R}^+$ is a positive real number. One first advantage of the continuity constraint is that, with an appropriate value for λ_c , we can prevent the system from overfitting the training data. A second advantage is that, after training, we can easily define the regression function r_p for each distance parameter p through multivariate interpolation [20]. For query images $I_q \notin \mathcal{R}$, $p(I_q)$ is given by the following equation:

$$p(I_q) = r_p(\bar{\mathcal{S}}(I_q)) \quad (26)$$

where

$$r_p : x \mapsto \frac{\sum_{I \in \mathcal{R}} e^{-\lambda_c \|x - \bar{\mathcal{S}}(I)\|_2} p(I)}{\sum_{I \in \mathcal{R}} e^{-\lambda_c \|x - \bar{\mathcal{S}}(I)\|_2}}. \quad (27)$$

C. Processing a Query: Scenario of Operation

Adaptive Image Retrieval: First, the query image $I_q \notin \mathcal{R}$ is decomposed using the optimal wavelet filters $(w^{s,K,L})_{s \in \mathbb{N}^*, K \in \mathbb{N}, L \in \mathbb{N}}$ obtained during the training phase, and the signature $\mathcal{S}(I_q)$ is computed. Second, references images $I \in \mathcal{R}$ are ranked in increasing order of distance to I_q , using the optimal distance weights $(\zeta_{s,K,L}, \eta_{s,K,L})_{s \in \mathbb{N}^*, K \in \mathbb{N}, L \in \mathbb{N}}$ obtained during the training phase. Third, the first K images are retrieved.

Highly Adaptive Image Retrieval: First, the query image $I_q \notin \mathcal{R}$ is decomposed using the initial wavelet filters $(\bar{w}^{s,K,L})_{s \in \mathbb{N}^*, K \in \mathbb{N}, L \in \mathbb{N}}$ obtained during the training phase, and the initial signature $\bar{\mathcal{S}}(I_q)$ is computed. Second, the final distance parameters (wavelet filters and weights) are obtained through multivariate interpolation [see (26)]. Third, the final signature of I_q and of each reference image $I \in \mathcal{R}$ are computed using the final set of wavelet filters $(w^{s,K,L}(I_q))_{s \in \mathbb{N}^*, K \in \mathbb{N}, L \in \mathbb{N}}$. Fourth, references images are ranked in increasing order of distance to I_q using the final distance weights. Fifth, the first K images are retrieved.

D. Characterization Maps and Time Complexity

During Training: In both adaptive and highly adaptive image retrieval, the signatures of each image $I_q \in \mathcal{T}$ and of each image $I \in \mathcal{R} (= \mathcal{T} \setminus \{I_q\})$ are evaluated for an arbitrary large number of wavelet filters. Computing characterization maps for each image $I \in \mathcal{T}$ is therefore very useful. However, there is no need to compute the entire characterization maps all at once: Whenever an image characterization needs to be evaluated for a new wavelet filter w , the closest key wavelet filter w_0 is searched for, the image characterization is evaluated for w_0 , and it is evaluated for w through Taylor expansions and invariance analysis (see Sections III-D and III-E).

After Training (Reference Images): In highly adaptive image retrieval, characterization maps computed for reference images $I \in \mathcal{R}$ are still very useful to process new query images $I_q \notin \mathcal{T}$.

Indeed, the signature $\mathcal{S}(I)$ of each image $I \in \mathcal{R}$ needs to be evaluated again using the optimal wavelet filters $w^{s,K,L}(I_q)$ computed for I_q (see Section IV-C), which is very fast if characterization maps are already available for I .

After Training (Query Images): There is no need to compute characterization maps for query images $I_q \notin \mathcal{T}$. In adaptive image retrieval, we only need to compute $\mathcal{S}(I_q)$, i.e., the signature associated with the optimal set of filters $(w^{s,K,L}(I_q))_{s \in \mathbb{N}^*, K \in \mathbb{N}, L \in \mathbb{N}}$. In highly adaptive image retrieval, we only need to compute $\bar{\mathcal{S}}(I_q)$, which is the signature associated with the initial set of filters $(\bar{w}^{s,K,L}(I_q))_{s \in \mathbb{N}^*, K \in \mathbb{N}, L \in \mathbb{N}}$, and $\mathcal{S}(I_q)$, which is the signature associated with the final set of filters $(w^{s,K,L}(I_q))_{s \in \mathbb{N}^*, K \in \mathbb{N}, L \in \mathbb{N}}$ (see Section IV-C).

To summarize, characterization maps should be computed for reference/training images but not for test images. As a consequence, the additional flexibility introduced in the proposed framework does not imply a dramatic retrieval time increase.

In addition, should the optimal wavelet basis be further fine tuned (e.g., to address relevance feedbacks from the user), there is no need to compute the updated signatures from scratch: new image characterizations can be estimated through Taylor expansions (where $w_0 = w^{s,K,L}(I_q)$) and invariance analysis (see Sections III-D and III-E).

V. APPLICATIONS

After an introduction to the four data sets under study in Section V-A, characterization map building is assessed in Section V-B. The retrieval performance of the proposed retrieval systems is then assessed in Section V-C.

A. Data Sets

Caltech101: This data set was collected in September 2003 at the California Institute of Technology.¹ It consists of 9144 pictures of objects belonging to 101 categories. This is about 40 to 800 images per category; most categories have about 50 images. The size of each image is roughly 300×200 pixels. Examples of images from two categories (“water_lilly” and “hedgehog”) are given in Fig. 3.

MESSIDOR: This data set was collected in three ophthalmology departments in France for research on automated diabetic retinopathy screening.² It consists of 1200 eye fundus color photographs. Images were acquired using a color video 3CCD camera on a Topcon TRC NW6 nonmydriatic retinograph with a 45° field-of-view. The size of each image is either 1440×960 , 2240×1488 , or 2304×1536 pixels. In a disease screening context, clinicians classify images into two categories: normal and pathological images. In MESSIDOR, 546 images were marked as normal and 654 as pathological.

Face Database: This data set was collected by AT&T Laboratories Cambridge for research on face recognition.³ It consists of 400 images from 40 categories. Each category consists of ten images from the same person at different times, with different lighting conditions, different facial expressions (open/

¹http://www.vision.caltech.edu/Image_Data_sets/Caltech101/

²<http://messidor.crihan.fr/index-en.php>

³<http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>

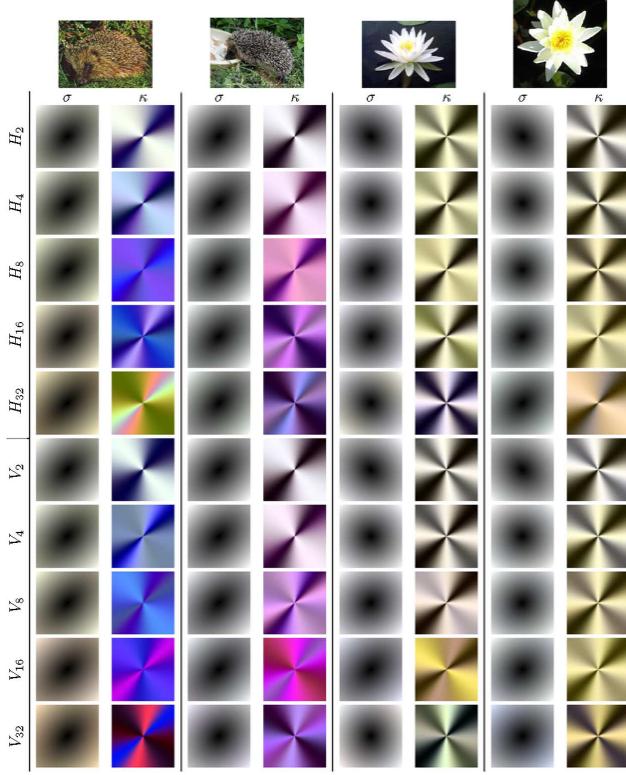


Fig. 3. Examples of characterization maps computed, at different analysis scales and orientations, for images in the Caltech101 data set. Color characterization maps should be interpreted as follows: The red channel of a color characterization map was computed from the red channel of the input image (the same goes for the green channel and the blue channel). In row labels, V stands for vertical wavelet filter ($K = 1, L = 0$), H stands for horizontal wavelet filter ($K = 0, L = 1$), and the subscript indicates the analysis scale (s). Characterization maps all look roughly alike. However, they differ in terms of angular frequency and phase, for instance. In addition, evolution across analysis scales differs from one image to another. Therefore, it is usually possible to select one pixel location (i.e., one filter) in each map to derive a discriminative feature vector. In addition, because a weight vector is used, the most discriminative maps/filters can be emphasized.

closed eyes, smiling/not smiling), and facial details (glasses/no glasses). The size of each image is 92×112 pixels.

VisTex: This data set was collected by the Media Laboratory at the Massachusetts Institute of Technology for research on texture recognition.⁴ It consists of categorized texture images representative of real-world conditions. Categories consisting of less than five elements (“Clouds,” “Grass,” “Misc,” “WheresWaldo,” and “Wood”) were discarded in this paper. As a consequence, 14 categories, consisting of 152 images altogether, were selected. The size of each image is 512×512 pixels.

For training purposes, each data set \mathcal{D} above was divided into five subsets: $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_5$. Each category in \mathcal{D} was equally divided into each of these subsets at random. The retrieval performance was assessed by a so-called fivefold cross-validation strategy: for each fold $i = 1, \dots, 5$, a training subset $\mathcal{T} = \mathcal{D} \setminus \mathcal{S}_i$ and a test subset $\overline{\mathcal{T}} = \mathcal{S}_i$ were defined. During retrieval performance assessment, an image was considered relevant if it belonged to the same category as the query image. Final results are reported on the union of the five test subsets (see Table II).

⁴<http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>

TABLE I
RELATIVE APPROXIMATION ERROR OF THE
CHARACTERIZATION (DERIVATIVE) MAPS

| | | $\{\sigma_s^{(i)}, \kappa_s^{(i)}, i = 1..Δ\}$ | | |
|---|---|--|--|--|
| s | n | Taylor expansion order n_T | | |
| | | 1 | 2 | 3 |
| 2 | 3 | 7.50×10^{-3} $\pm 1.63 \times 10^{-3}$ | 1.03×10^{-3} $\pm 3.53 \times 10^{-4}$ | 5.75×10^{-4} $\pm 3.30 \times 10^{-4}$ |
| | 6 | 3.29×10^{-3} $\pm 6.77 \times 10^{-4}$ | 3.57×10^{-4} $\pm 1.39 \times 10^{-4}$ | 1.12×10^{-4} $\pm 6.48 \times 10^{-5}$ |
| 4 | 3 | 7.15×10^{-3} $\pm 1.39 \times 10^{-3}$ | 9.80×10^{-4} $\pm 3.86 \times 10^{-4}$ | 4.07×10^{-4} $\pm 1.86 \times 10^{-4}$ |
| | 6 | 3.17×10^{-3} $\pm 6.05 \times 10^{-4}$ | 3.19×10^{-4} $\pm 1.31 \times 10^{-4}$ | 7.65×10^{-5} $\pm 3.44 \times 10^{-5}$ |
| | | $\{\frac{\partial \sigma_s}{\partial w_{k,l}}^{(i)}, \frac{\partial \kappa_s}{\partial w_{k,l}}^{(i)}, k = -K..K, l = -L..L, i = 1..Δ\}$ | | |
| s | n | Taylor expansion order n_T | | |
| | | ∅ | 1 | 2 |
| 2 | 3 | ∅ | 2.80×10^{-2} $\pm 1.06 \times 10^{-2}$ | 1.59×10^{-2} $\pm 9.30 \times 10^{-3}$ |
| | 6 | ∅ | 1.37×10^{-2} $\pm 5.56 \times 10^{-3}$ | 4.70×10^{-3} $\pm 2.79 \times 10^{-3}$ |
| 4 | 3 | ∅ | 2.76×10^{-2} $\pm 1.09 \times 10^{-2}$ | 1.22×10^{-2} $\pm 5.80 \times 10^{-3}$ |
| | 6 | ∅ | 1.31×10^{-2} $\pm 5.30 \times 10^{-3}$ | 3.49×10^{-3} $\pm 1.63 \times 10^{-3}$ |

B. Assessing Characterization Map Building

Characterization map building was assessed on a data set of 40 images: Ten images were randomly selected from each data set above. For each image, the exact and approximate value for σ_s , κ_s , $(\partial \sigma_s / \partial w_{k,l})$, and $(\partial \kappa_s / \partial w_{k,l})$ were computed for $\Delta = 100$ equally spaced wavelet filters on one half of the unit sphere. The approximate values were obtained by the procedure described in Section III-F. The experiment was performed for $D = 2$ at two analysis scales: $s = 2$ and $s = 4$. Only the red channel of color images was used in this experiment. Table I reports the per-image relative approximation error of $\{\sigma_s^{(i)}, \kappa_s^{(i)}, i = 1, \dots, \Delta\}$ and $\{(\partial \sigma_s / \partial w_{k,l})^{(i)}, (\partial \kappa_s / \partial w_{k,l})^{(i)}, k = -K, \dots, K, l = -L, \dots, L, i = 1, \dots, \Delta\}$. In this paper, the relative approximation error of a vector was defined as the L2-norm of the approximation error divided by the L2-norm of the exact vector.

In the remainder of this paper, two setups are studied:

S1 $D = 2$ and $n = 6$ ($|W_{K,L}^0| = 12$);

S2 $D = 4$ and $n = 3$ ($|W_{K,L}^0| = 108$);

a Taylor expansion of order $n_T = 3$ is used to approximate σ_s and κ_s , and a Taylor expansion of order $n_T = 2$ is used to approximate $(\partial \sigma_s / \partial w_{k,l})$ and $(\partial \kappa_s / \partial w_{k,l})$.

The average computation times for building one characterization map and the associated characterization derivative maps is 0.0476 s for Caltech101, 1.53 s for MESSIDOR, 0.00667 s for the Face Database, and 0.168 s for VisTex (using $D = 2$ and $s = 2$). One core of an Intel Xeon E5520 processor, running at 2.27 GHz, was used in all experiments. Note that computation times, for building one characterization map or one characterization derivative map, are in $O((MNKL/s^2)Dn^{D-1})$ (see Sections II and III-F).

TABLE II
PRECISION AT FIVE ON THE TEST SUBSET

| Dataset (§V-A) | | Caltech101 | MESSIDOR | Face Database | VisTex |
|---------------------------|--|-------------------|-------------------|-------------------|-------------------|
| S1 setup | <i>adaptive image retrieval</i> (§IV-A) | 0.404 ± 0.007 | 0.812 ± 0.012 | 0.952 ± 0.006 | 0.539 ± 0.029 |
| | <i>highly adaptive image retrieval</i> (§IV-B) | 0.512 ± 0.006 | 0.875 ± 0.009 | 0.969 ± 0.006 | 0.643 ± 0.038 |
| S2 setup | <i>adaptive image retrieval</i> (§IV-A) | 0.415 ± 0.006 | 0.816 ± 0.012 | 0.956 ± 0.005 | 0.543 ± 0.029 |
| | <i>highly adaptive image retrieval</i> (§IV-B) | 0.523 ± 0.006 | 0.879 ± 0.009 | 0.974 ± 0.005 | 0.649 ± 0.038 |
| wavelet-based methods | separable lifting-scheme [5] | 0.420 ± 0.006 | 0.817 ± 0.012 | 0.955 ± 0.005 | 0.539 ± 0.029 |
| | non-separable lifting-scheme [8] | 0.435 ± 0.006 | 0.828 ± 0.011 | 0.958 ± 0.005 | 0.580 ± 0.028 |
| | dual-tree CWT [21] | 0.421 ± 0.006 | 0.785 ± 0.012 | 0.950 ± 0.005 | 0.549 ± 0.028 |
| non wavelet-based methods | BoF (SIFT) [15], [22] | 0.287 ± 0.006 | 0.601 ± 0.014 | 0.864 ± 0.016 | 0.271 ± 0.040 |
| | BoF (SURF) [15], [23] | 0.299 ± 0.005 | 0.543 ± 0.013 | 0.695 ± 0.024 | 0.346 ± 0.035 |

C. Retrieval Performance

The following filter supports and analysis scales were used in all experiments (see Fig. 3 for the S1 setup):

$$\begin{aligned} H_2 s &= 2/S1: K = 1, L = 0/S2: K = 2, L = 0. \\ H_4 s &= 4/S1: K = 1, L = 0/S2: K = 2, L = 0. \\ H_8 s &= 8/S1: K = 1, L = 0/S2: K = 2, L = 0. \\ H_{16} s &= 16/S1: K = 1, L = 0/S2: K = 2, L = 0. \\ H_{32} s &= 32/S1: K = 1, L = 0/S2: K = 2, L = 0. \\ V_2 s &= 2/S1: K = 0, L = 1/S2: K = 0, L = 2. \\ V_4 s &= 4/S1: K = 0, L = 1/S2: K = 0, L = 2. \\ V_8 s &= 8/S1: K = 0, L = 1/S2: K = 0, L = 2. \\ V_{16} s &= 16/S1: K = 0, L = 1/S2: K = 0, L = 2. \\ V_{32} s &= 32/S1: K = 0, L = 1/S2: K = 0, L = 2. \end{aligned}$$

The following values were used for δ_p [see (22)]: $\delta_p = 10^{-5}$ if p is a weight, and $\delta_p = 5 \times 10^{-2}$ if p is a wavelet filter coefficient. λ_c was found by cross-validation on each training subset [see (25) and (27)].

The precision at $K = 5$ [5], obtained on the test subset \bar{T} of each data set by both experiments, is reported in Table II. It is compared with three other wavelet-based CBIR methods on the same data sets (with the same subsets S_1, S_2, \dots, S_5): separable lifting-scheme-based optimization [5], nonseparable lifting-scheme-based optimization, [8] and dual-tree complex wavelet transform (CWT) [8], [21]. A comparison with the original BoF model [15], which has been one of the most popular CBIR frameworks, was also included. In our BoF implementation, based on the OpenCV library,⁵ features were detected and described using the scale-invariant feature transform (SIFT) [22] or speeded-up robust features [23]; term-frequency-inverse-document-frequency (TF-IDF) weighting was used to rank reference images. The reader is referred to [5] for additional comparisons with nonwavelet-based CBIR methods. Precision at five versus query time are reported in Fig. 4; every algorithm was implemented in C++, except for dual-tree CWT.⁶

VI. DISCUSSION

A novel image characterization framework, which is based on adaptive separable or nonseparable wavelet transforms, has

⁵<http://opencv.willowgarage.com/wiki/> (BOWKMeansTrainer, BOWImgDescriptorExtractor, SiftFeatureDetector/SurfFeatureDetector and SiftDescriptorExtractor/SurfDescriptorExtractor classes)

⁶<http://eeweb.poly.edu/iselesni/WaveletSoftware/dt2D.html> (the software was run with GNU Octave)

been proposed in this paper. It was presented for 2-D images, but it could be generalized to any type of n -D digital signals. This framework allows fast image characterization, particularly when images need to be characterized several times, using different wavelet filters. This feature is particularly useful for wavelet adaptation.

The proposed framework was applied to wavelet adaptation for CBIR. Two CBIR methods were presented. The first method, i.e., adaptive image retrieval, sped up previously proposed CBIR methods based on wavelet adaptation [5], without decreasing the retrieval performance. The second method, i.e., highly adaptive image retrieval, takes full advantage of the proposed image characterization framework: During each query, a different wavelet filter is used, which means each reference image has to be characterized again with a new wavelet filer. Highly adaptive image retrieval increased the retrieval performance significantly, while maintaining low computation times (see Fig. 4). These improvements were observed in four different image data sets: Caltech101, MESSIDOR, Face Database, and VisTex. A simple gradient descent was used in this paper for wavelet adaptation. However, the proposed image characterizations may be used jointly with more advanced wavelet adaptation techniques.

In all data sets, highly adaptive image retrieval performed significantly better than dual-tree CWT [21], in which wavelet filters of large support are used to analyze images. In highly adaptive image retrieval, several adapted filters of small support are used. The number and the high adaptivity of these filters compensates for the difference in support size.

In all data sets, highly adaptive image retrieval performed significantly better than the original BoF model [15], whatever feature detection and description method was used (SIFT or SURF, see Fig. 4). As expected, this improvement was particularly noticeable in VisTex and MESSIDOR, where texture is the most discriminative feature. One reason why lower precision rates have been achieved on the Caltech101 data set is that the semantic gap is wider: the “Butterfly” category, for instance, contains pictures of real butterflies among flowers, pictures of butterfly-shaped chocolates, stylized butterfly drawings on uniform backgrounds, etc. Another reason is that there are much more categories (101). In line with a major trend in image characterization for CBIR (BoF, multiple-instance learning [24], etc.) [21], we plan to use the

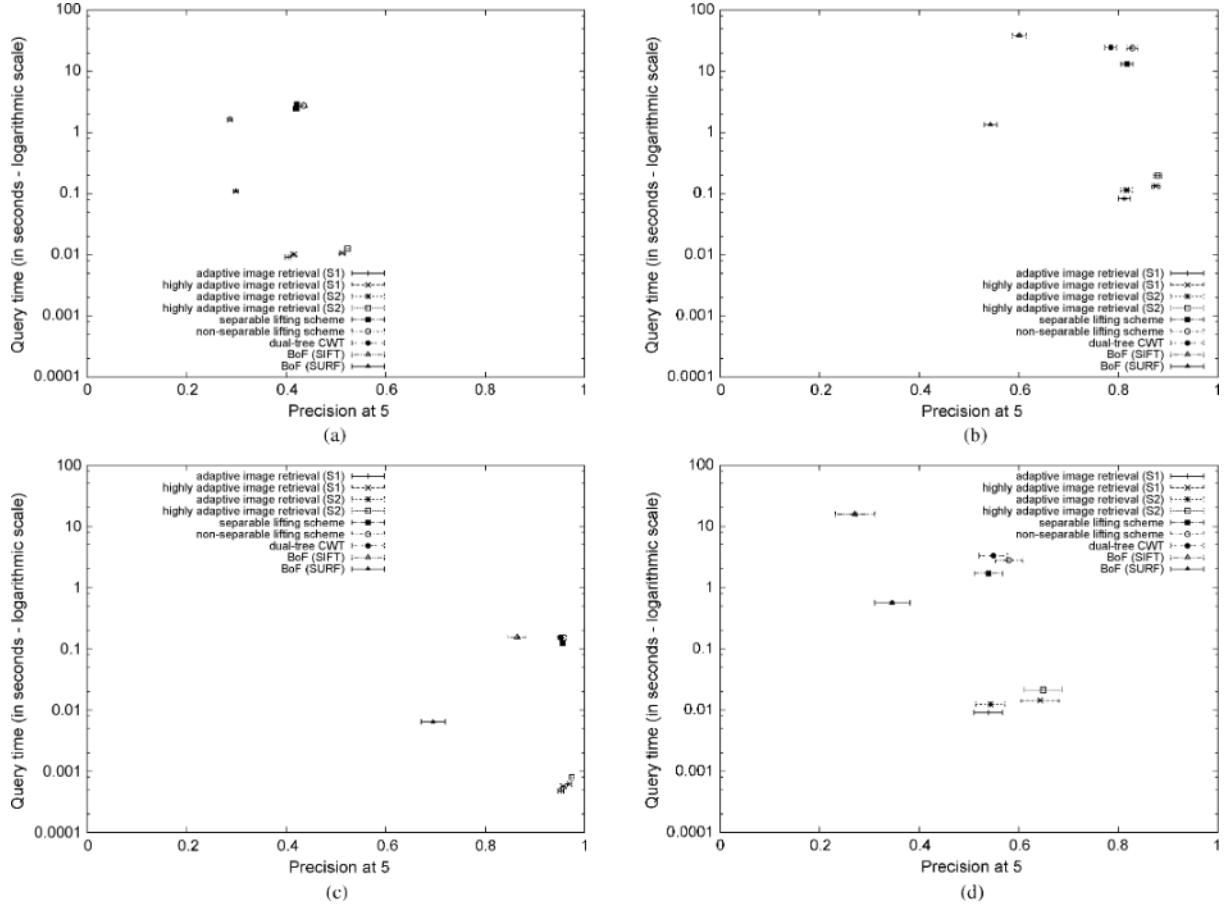


Fig. 4. Precision at five versus query time (in logarithmic scale). Query time includes image processing time and search time; training time is not included. Error bars represent confidence intervals: They are seldom visible along the y axis. (a) Caltech101. (b) MESSIDOR. (c) Face Database. (d) VisTex.

proposed framework in future works to characterize regions of interest, detected by SIFT or SURF for instance, instead of characterizing images as a whole. The proposed framework may also be used for relevance feedback, which is one of the most active research fields for CBIR systems [25]–[28]. In relevance feedback, the user iteratively selects relevant and irrelevant images among retrieved reference images, and the distance measure is updated: Basically, the weight of each image feature is modified. In this context, the wavelet filter could simply be updated after each user feedback: The proposed framework would therefore allow relevance feedback on image features themselves and not only on the way these features are combined, which would be novel. Through experiments on highly adaptive image retrieval, we have shown that adapting the wavelet filter to each query image increases the retrieval performance, without considerably increasing computation times. This observation gives us some insights on potential performance increases through relevance feedback.

We believe that the proposed framework also has great potential for classification and compression. Adapting a CBIR approach for classification purposes is straightforward. As for compression, the characterization maps could be used to find the wavelet filter that should be used to optimally compress the image at each scale s : The highest intensity in κ_s 's map

indicates the wavelet filter maximizing kurtosis at this scale, indicating the highest SNR (or the highest contrast) [29].

In this paper, two features have been selected to build characterization maps: the standard deviation and the kurtosis. Should additional features be used, differentiability checks would have to be performed to ensure that characterization maps can still be approximated by Taylor expansions (see Section III-C). However, most features are smooth and continuous functions of the data; otherwise, they could not characterize noisy data robustly.

This study has two limitations. First, for simplicity, no low-pass filter was used before subsampling images between two analysis scales; however, it does not seem to affect performances (see Table II: adaptive image retrieval (setup S2) versus separable lifting scheme). Second, although the framework was presented for nonseparable wavelets of arbitrary support, it was only applied to filters of small support (three or five taps) in this paper. A solution is presented in Section IV-D to compute characterization maps progressively; this should facilitate the use of larger filters, which might improve retrieval performances further.

As a conclusion, a novel image characterization framework was presented, and its suitability for CBIR was shown. This study paves the way to efficient relevance feedback on image features themselves, which should enable interactive image search with higher performance.

APPENDIX A
DERIVATIVES OF THE PROPOSED CHARACTERIZATION:
COMPUTATION DETAILS

A. First- and Second-Order Derivatives

Equations (16)–(19) are obtained from lower order derivatives by usual differentiation rules and the following relations:

$$\frac{\partial m_{s,d}}{\partial w_{k,l}} = d\delta_{s,d-1}(k, l) \quad (28)$$

$$\frac{\partial \delta_{s,d}}{\partial w_{u,v}}(k, l) = d\delta'_{s,d-1}(k, l, u, v). \quad (29)$$

B. Third-Order Derivatives

Let A denote the following expression of two filter coefficients $w_{k,l}$ and $w_{u,v}$ appearing in (19):

$$A(k, l, u, v) = 3\sigma_s^2 \delta'_{s,2}(k, l, u, v) + 2\delta_{s,1}(u, v)\delta_{s,3}(k, l) - 4\delta_{s,1}(k, l)\delta_{s,3}(u, v) - \delta'_{s,0}(k, l, u, v)m_{s,4}. \quad (30)$$

Its derivative with respect to a third filter coefficient $w_{p,q}$ is given by

$$\begin{aligned} \frac{\partial A}{\partial w_{p,q}}(k, l, u, v) = & 6\sigma_s \frac{\partial \sigma_s}{\partial w_{p,q}} \delta'_{s,2}(k, l, u, v) \\ & + 6\sigma_s^2 \delta''_{s,1}(k, l, u, v, p, q) \\ & + 2\delta'_{s,0}(u, v, p, q)\delta_{s,3}(k, l) \\ & + 6\delta_{s,1}(u, v)\delta'_{s,2}(k, l, p, q) \\ & - 4\delta'_{s,0}(k, l, p, q)\delta_{s,3}(u, v) \\ & - 12\delta_{s,1}(k, l)\delta'_{s,2}(u, v, p, q) \\ & - 4\delta'_{s,0}(k, l, u, v)\delta_{s,3}(p, q) \end{aligned} \quad (31)$$

where $\delta''_{s,d}$ is a polynomial function of the wavelet transform of I defined as

$$\begin{aligned} \delta''_{s,d}(k, l, u, v, p, q) = & \frac{s^2}{MN} \sum_{i \in I_n, j \in J_n} \\ & \times \left[(I_{i-sk,j-sl} - I_{i,j}) \right. \\ & \times (I_{i-su,j-su} - I_{i,j}) \\ & \times (I_{i-sp,j-sq} - I_{i,j}) \times x_{i,j,s}^d \left. \right]. \end{aligned} \quad (32)$$

The following relation holds:

$$\frac{\partial \delta'_{s,d}}{\partial w_{p,q}}(k, l, u, v) = d\delta''_{s,d-1}(k, l, u, v, p, q). \quad (33)$$

Finally, the third-order derivatives of the characterization are given as follows:

$$\begin{aligned} \frac{\partial^3 \sigma_s}{\partial w_{k,l} \partial w_{u,v} \partial w_{p,q}} = & \frac{1}{\sigma_s^2} \left(\begin{array}{c} \delta'_{s,0}(k, l, u, v) \frac{\partial \sigma_s}{\partial w_{p,q}} \\ -\delta'_{s,0}(k, l, p, q) \frac{\partial \sigma_s}{\partial w_{u,v}} \\ -\delta_{s,1}(k, l) \frac{\partial^2 \sigma_s}{\partial w_{u,v} \partial w_{p,q}} \end{array} \right) \\ & - \frac{2\delta_{s,1}(p, q)}{\sigma_s^4} \left(\begin{array}{c} \delta'_{s,0}(k, l, u, v) \sigma_s \\ -\delta_{s,1}(k, l) \frac{\partial \sigma_s}{\partial w_{u,v}} \end{array} \right) \end{aligned} \quad (34)$$

$$\begin{aligned} \frac{\partial^3 \kappa_s}{\partial w_{k,l} \partial w_{u,v} \partial w_{p,q}} = & -\frac{24}{\sigma_s^7} \frac{\partial \sigma_s}{\partial w_{p,q}} A(k, l, u, v) \\ & + \frac{4}{\sigma_s^6} \frac{\partial A}{\partial w_{p,q}}(k, l, u, v) \\ & + \frac{\partial \kappa_s}{\partial w_{k,l}} \left(\begin{array}{c} \frac{12}{\sigma_s^3} \frac{\partial \sigma_s}{\partial w_{p,q}} \delta_{s,1}(u, v) \\ -\frac{6}{\sigma_s^2} \delta'_{s,0}(u, v, p, q) \end{array} \right) \\ & - \frac{6}{\sigma_s^2} \frac{\partial^2 \kappa_s}{\partial w_{k,l} \partial w_{p,q}} \delta_{s,1}(u, v). \end{aligned} \quad (35)$$

REFERENCES

- [1] A. H. Tewfik, D. Sinha, and P. Jorgensen, "On the optimal choice of a wavelet for signal representation," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 747–765, Mar. 1992.
- [2] A. Gupta, S. D. Joshi, and S. Prasad, "A new method of estimating wavelet with desired features from a given signal," *Signal Process.*, vol. 85, no. 1, pp. 147–161, Jan. 2005.
- [3] R. L. Claypoole, R. G. Baraniuk, and R. D. Nowak, "Adaptive wavelet transforms via lifting," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 1998, vol. 3, pp. 1513–1516.
- [4] G. Piella, B. Pesquet-Popescu, H. J. A. M. Heijmans, and G. Pau, "Combining seminorms in adaptive lifting schemes and applications to image analysis and compression," *J. Math. Imag. Vis.*, vol. 25, no. 2, pp. 203–226, Sep. 2006.
- [5] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Wavelet optimization for content-based image retrieval in medical databases," *Med. Image Anal.*, vol. 14, no. 2, pp. 227–241, Apr. 2010.
- [6] D. Serši and M. Vranki, "Adaptation of a 2-D nonseparable wavelet filter bank with variable number of zero moments," in *Proc. Visualization, Imag., Image Process.*, Sep. 2002, pp. 257–260.
- [7] D. Serši and M. Vranki, "Adaptation in the quincunx wavelet filter bank with applications in image denoising," in *Proc. Int. Workshop Spectral Methods Multirate Signal Process.*, 2004, pp. 245–252.
- [8] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Adaptive nonseparable wavelet transform via lifting and its application to content-based image retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 25–35, Jan. 2010.
- [9] M. D. Abràmoff, J. M. Reinhardt, S. R. Russell, J. C. Folk, V. B. Mahajan, M. Niemeijer, and G. Quellec, "Automated early detection of diabetic retinopathy," *Ophthalmology*, vol. 117, no. 6, pp. 1147–1154, Jun. 2010.
- [10] G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abràmoff, and M. Sonka, "Three-dimensional analysis of retinal layer texture: Identification of fluid-filled regions in SD-OCT of the macula," *IEEE Trans. Med. Imag.*, vol. 26, no. 6, pp. 1321–1330, Jun. 2010.
- [11] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.
- [12] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol. 40, no. 2, pp. 1–60, Apr. 2008.
- [13] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications—Clinical benefits and future directions," *Int. J. Med. Inform.*, vol. 73, no. 1, pp. 1–23, Feb. 2004.
- [14] C. B. Akgül, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: Current status and future directions," *J. Digit Imaging*, vol. 24, no. 2, pp. 208–222, Apr. 2010.
- [15] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. Int. Conf. Comput. Vis.*, 2003, pp. 1470–1477.
- [16] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.
- [17] G. van de Wouwer, P. Scheunders, and D. van Dyck, "Statistical texture characterization from discrete wavelet representations," *IEEE Trans. Image Process.*, vol. 8, no. 4, pp. 592–598, Apr. 1999.
- [18] S. Nadarajah, "A generalized normal distribution," *J. Appl. Stat.*, vol. 32, no. 7, pp. 685–694, Sep. 2005.
- [19] M. Hazewinkel, "Taylor's theorem," in *Encyclopaedia of Mathematics*. College Station, TX: Springer-Verlag, 2001.
- [20] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proc. ACM Nat. Conf.*, 1968, pp. 517–524.

- [21] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, Nov. 2005.
- [22] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 1150–1157.
- [23] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool, "Surf: Speeded up robust features," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.
- [24] R. Rahmani, S. A. Goldman, H. Zhang, S. R. Cholleti, and J. E. Fritts, "Localized content-based image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 1902–1912, Nov. 2008.
- [25] M. Ferecatu and D. Geman, "A statistical framework for image category search from a mental picture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 1087–1101, Jun. 2009.
- [26] E. Cheng, F. Jing, and L. Zhang, "A unified relevance feedback framework for web image retrieval," *IEEE Trans. Image Process.*, vol. 18, no. 6, pp. 1350–1357, Jun. 2009.
- [27] M. R. Azimi-Sadjadi, J. Salazar, and S. Srinivasan, "An adaptable image retrieval system with relevance feedback using kernel machines and selective sampling," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1645–1659, Jul. 2009.
- [28] W. Bian and D. Tao, "Biased discriminant Euclidean embedding for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 545–554, Feb. 2010.
- [29] M. Bethge, "Factorial coding of natural images: How effective are linear models in removing higher-order dependencies?," *J. Opt. Soc. Amer. A, Opt. Image Sci. Vis.*, vol. 23, no. 6, pp. 1253–1268, Jun. 2006.



Gwénolé Quellec was born in Saint-Renan, France, on November 29, 1982. He received the Engineering degree in computer science and applied mathematics from ISIMA, Clermont-Ferrand, France, in 2005, the M.S. degree in image processing from the University of Clermont-Ferrand II, Clermont-Ferrand, in 2005, and the Ph.D. degree in signal processing from Telecom Bretagne, Brest, France, in 2008.

He is currently a Research Associate with the LaTIM Inserm Research Unit 1101, Brest, France. His research interests include retinal image processing, content-based image retrieval, and information fusion for medical applications.



Mathieu Lamard was born in Bordeaux, France, on May 18, 1968. He received the M.S. degree in applied mathematics from the University of Bordeaux, Bordeaux, in 1995, and the Ph.D. degree in signal processing and telecommunication from the University of Rennes, France, in 1999.

In 2000, he joined the LaTIM Inserm Research Unit 1101, Brest, France, where he is currently a Research Associate. His research interests include image processing, 3-D reconstruction, content-based image retrieval, and information fusion for medical applications.



Guy Cazuguel (M'83) received the Engineering degree from the Ecole Nationale de l'Aviation Civile, Toulouse, France, in 1975, the M.S. degree in advanced automatics, and the Ph.D. degree in signal processing and telecommunications from the University of Rennes I, Rennes, France, in 1976 and 1994, respectively.

He is currently a Professor with the Department of Image and Information Processing, Telecom Bretagne, Brest, France. His research interests include image analysis and content-based image retrieval in medical applications, within the LaTIM Inserm Research Unit 1101.



Béatrice Cochener received the M.D. degree in ophthalmology from Nancy University, Nancy, France, in 1992.

She is currently a Professor and the Head with the University Eye Clinic, Brest, France. Together with J. Colin, she developed a very active anterior segment surgery practice. She is a Specialist of refractive techniques in vision correction. She participated in three books on surgical techniques and has published more than 30 peer-reviewed journal articles. Her research interests include imaging research, clinical evaluation, and anterior segment surgery teaching.

Dr. Cochener is the Vice President of the French Implant and Refractive Surgery Society SAFIR, the President of the French Academy of Ophthalmology, and is a member of the Editorial Board of the *Journal Français d'Ophthalmologie*.



Christian Roux (F'05) received the Aggregation degree in physics from the Ecole Normale Supérieure, Cachan, France, in 1978 and the Ph.D. degree in signal processing from the Institut National Polytechnique, Grenoble, France, in 1980.

In 1982, he joined Telecom Bretagne, Brest, France. He is the founding Director of LaTIM Inserm Research Unit 1101, Brest. He is the author of more than 150 papers, holds nine patents, and contributed to the creation of three startup companies developing medical technologies. His current research interests include medical information processing, spatial and functional information modeling, and analysis in medical images.

Dr. Roux was the recipient of the INSERM Basic Research Award in 2006. He has been the founding Co-Chair of the IEEE EMBS International Summer School on Biomedical Imaging since 1994 and served as the President of the IEEE Engineering in Medicine and Biology Society in 2001. He is a member of the Editorial Board of the *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY* and of the *PROCEEDINGS OF THE IEEE*.