Abstract

We propose in this article a Content-Based Image Retrieval method for diagnosis aid in medical fields. In the proposed system, images are indexed in a generic fashion, without extracting domain-specific features: a signature is built for each image from its wavelet transform. These image signatures characterize the distribution of wavelet coefficients in each subband of the decomposition. A distance measure is then defined to compare two image signatures and thus retrieve the most similar images in a database when a query image is submitted by a physician. To retrieve relevant images from a medical database, the signatures and the distance measure must be related to the medical interpretation of images. As a consequence, we introduce several degrees of freedom in the system so that it can be tuned to any pathology and image modality. In particular, we propose to adapt the wavelet basis, within the lifting scheme framework, and to use a custom decomposition scheme. Weights are also introduced between subbands. All these parameters are tuned by an optimization procedure, using the medical grading of each image in the database to define a performance measure. The system is assessed on two medical image databases: one for diabetic retinopathy follow up and one for screening mammography, as well as a general purpose database. Results are promising: a mean precision of 56.50%, 70.91% and 96.10% is achieved for these three databases, when five images are returned by the system.

Index terms—content-based image retrieval, custom wavelet, lifting scheme, diabetic retinopathy, mammography

1 INTRODUCTION

Images have always been used in medicine for teaching, diagnosis, and management purposes. Now medical imaging systems produce more and more digitized images in all medical fields: visible,
ultrasound, X-ray tomography, MRI, nuclear imaging, etc. These images are very interesting for diagnostic purposes: they are directly related to the patient pathology and medical history, and may be used for diagnosis aid. However the amount of images we can access nowadays is so huge that database systems require efficient indexing to enable fast access to images in databases. Automatic image indexing using image digital content (Content-Based Image Retrieval - CBIR) is one of the possible and promising solutions to effectively manage image databases (Nastar, 1997). In recent years, CBIR has been applied to several applications such as industrial (Tao et al., 2008), forensic (AlGarni and Hamiane, 2008), remote sensing (Hafiane et al., 2006; Xie, 2004) and medical applications (Müller et al., 2004). In a medical context, the potential applications of CBIR are categorization (i.e. retrieving images of the same anatomical regions) (Rahman et al., 2007; Horsthemke et al., 2007; Greenspan and Pinhas, 2007), data mining (Siadat et al., 2005), retrieving visually similar lesions (Kim et al., 2006; El-Naqa et al., 2004; Doyle et al., 2007; Balmashnova et al., 2007), typically for educational purposes, and finally diagnosis aid (Dy et al., 2003; Pourghassem and Ghassemian, 2008; Antani et al., 2004), which is the application we will focus on in this article. The purpose is to let an expert carry out a search in the database, without formulating a semantic description of the image he is examining: the expert simply places the image as a query to a system, which retrieves the most similar images in the database. Finally, the expert chooses the most similar images, from his point of view, in the set of retrieved images, and accesses the associated information.

To compare the query image to another image in the database, feature vectors characterizing their numerical content, called signatures, are extracted. Then, comparing these two images simply comes down to comparing their signatures. In recent CBIR systems, images are mostly characterized visually by their shape (Pourghassem and Ghassemian, 2008; Dai and Zhang, 2005; Zhang and Lu, 2005; Antani et al., 2004), their edges (Rahmani et al., 2008; Rahman et al., 2007; Banerjee and Kundu, 2003; Han and Guo, 2003), textural features (Rahman et al., 2007; Kokare et al., 2006; Chun et al., 2003; Rallabandi and Rallabandi, 2008; Pourghassem and Ghassemian, 2008; Oliveira et al., 2007; Muneeswaran et al., 2006; Yu et al., 2005) or color histograms (Rahman et al., 2007; Yap and Parameswaran, 2006; Khanh et al., 2003; Konstantinidis et al., 2005). Several features proposed in the literature are invariant to translation (Bishnu et al., 2005), rotation (Kokare et al., 2006; Bishnu et al., 2005; Rallabandi and Rallabandi, 2008; AlGarni and Hamiane, 2008; Sastry et al., 2004; Xie, 2004), size (Bishnu et al., 2005; AlGarni and Hamiane, 2008; Lo et al., 2007) or illumination (Greenspan and Pinhas, 2007). In medical applications, rotation and size invariance are mostly necessary if we want to retrieve images similar in shape: typically, if we want to retrieve anatomical regions (Rahman et al., 2007; Horsthemke et al., 2007) or if we want to characterize lesions or patterns within a user-defined ROI (Kim et al., 2006) or within an automatically segmented region (Yu et al., 2005; Siadat et al., 2005; Antani et al., 2004). In this article, we propose a diagnosis aid system intended to retrieve images with the same pathology and at the same disease severity level, not simply representing the same anatomical region. In that purpose, we have to characterize images by the lesions they contain. And we do not want the user to define ROIs: medical images sometimes contain hundreds of lesions, it would be to much time consuming and physicians would not use such a system. Also, we do not want to rely on segmentation to characterize the lesions, since it is hard to design a robust segmentation algorithm for each type of lesions. So we propose to characterize images by their textural content, at different scales: it allows us to characterize lesions of different sizes, in a generic fashion. Our approach is invariant to translation. For many medical imaging modalities, such as eye fundus imaging or mammography, images are correctly oriented and scaled, due to the acquisition protocol; so, rotation and size invariance are not real issues in
The outstanding challenges of CBIR in medical applications are the following. First, how to bridge the gap between the extracted low-level features and the high-level perception of similarity of the physicians (Rahman et al., 2007)? We addressed this issue by automatically adapting the system parameters to a high-level criterion, using a supervised learning procedure. Then, how to integrate the proposed retrieval systems with the compression and archiving standards (Greenspan and Pinhas, 2007)? We addressed this issue by defining a wavelet-based image signature that may be extracted as images are compressed by a wavelet-based compression standard, such as JPEG 2000 (Taubman and Marcellin, 2001). Then, how to combine visual features with medical metadata (Greenspan and Pinhas, 2007; Antani et al., 2002; le Bozec et al., 2000; Shao et al., 2004)? And finally, how to efficiently retrieve images in large databases (Oliveira et al., 2007; Dy et al., 2003; Iakovidis et al., 2008)? We addressed the last two issues in a previously published article (Quellec et al., 2008): we used knowledge discovery trees to retrieve complete patient files consisting of possibly several images with metadata, thanks to wavelet-based image signatures, which enables a fast retrieval and an increased retrieval performance.

Wavelets have been widely used in image processing since the publication of the JPEG 2000 standard, including in the CBIR field (Xiong and Huang, 2002). We explore in this article two approaches to define image signatures and their associated distance measures from the wavelet transform of images. Both approaches are based on a specific model for wavelet coefficient distributions: the first one is based on histograms and the second one on generalized Gaussian functions. They are generic and automatically tuned to adapt to the studied images and pathologies. To adapt them, we propose to use a custom wavelet basis and a custom decomposition scheme. The two approaches are evaluated on two medical image databases: one for diabetic retinopathy follow up and one for screening mammography. To show the method is generic, it was also evaluated on a general purpose database.

The setup of the article is as follows. The signatures and distance measures are described in section 2. Section 3 introduces our approach to search for an optimal wavelet basis. We discuss the use of an optimal wavelet transform for both image indexing and compression in section 4. Sections 5 and 6 explain how our method is evaluated and calibrated. And the method is applied to the three databases in section 7. We end with a discussion and conclusion in section 8.

2 Proposed image signatures and distance measures

We propose to build image signatures from their wavelet transform. The wavelet transform is indeed convenient since it is highly tunable: in particular we can use a custom basis of functions, called wavelets, to decompose images. This wavelet basis will be tuned in order to improve the retrieval performance of the system. To implement the wavelet transform, we use the lifting scheme Sweldens (1996), which is particularly suited to design a custom wavelet basis (see section 2.2). It is also computationally efficient, hence its use in compression, for instance. Two different image signatures are then extracted from these decompositions in section 2.4. The associated distance measures are described in section 2.5. But first, let us set out the basics of multiresolution analysis.

2.1 Multiresolution analysis

A multiresolution approximation is a nested sequence of linear spaces \((V_j)_{j \in \mathbb{Z}}\) for approximating functions \(f \in L^2(\mathbb{R})\) such that:
\[ V_j \subset V_{j+1}, \forall j \in \mathbb{Z} \]
\[ \bigcup_{j=-\infty}^{+\infty} V_j \text{ is dense in } L_2(\mathbb{R}) \] and \[ \bigcap_{j=-\infty}^{+\infty} V_j = \{0\} \]
\[ f(x) \in V_j \iff f(2x) \in V_{j+1}, \forall j \in \mathbb{Z} \]
\[ f(x) \in V_j \Rightarrow f(x - 2^{-k}) \in V_{j}, \forall j, k \in \mathbb{Z} \]

The approximation of a function \( f \in L_2(\mathbb{R}) \) at a resolution \( 2^j \) is defined as the projection of \( f \) on \( V_j \). There is a function \( \Phi \in L_2(\mathbb{R}) \), called scaling function, such that for any \( j \in \mathbb{Z} \), \( (x \mapsto \Phi_{j,k}(x) = \sqrt{2^j} \Phi(2^j x - k))_{k \in \mathbb{Z}} \) is a basis of \( V_j \). The scaling function \( \Phi \) satisfies the refinement relation of equation 1.

\[ \Phi(x) = 2 \sum_{k=-\infty}^{\infty} h_k \Phi(2x - k) \] (1)

The additional information available in the approximation of \( f \) at the resolution \( 2^{j+1} \) as compared with the resolution \( 2^j \), is given by the projection of \( f \) on the orthogonal complement of \( V_j \) in \( V_{j+1} \), noted \( W_j \). There is a function \( \Psi \in L_2(\mathbb{R}) \), called wavelet function, such that for any \( j \in \mathbb{Z} \), \( (x \mapsto \Psi_{j,k}(x) = \sqrt{2^j} \Psi(2^j x - k))_{k \in \mathbb{Z}} \) is a basis of \( W_j \). The wavelet function \( \Psi \) satisfies the refinement relation of equation 2.

\[ \Psi(x) = 2 \sum_{k=-\infty}^{\infty} g_k \Phi(2x - k) \] (2)

The wavelet and the scaling filters are noted \( G = (g_k)_{k \in \mathbb{Z}} \) and \( H = (h_k)_{k \in \mathbb{Z}} \), respectively. \( \{V_j\}_{j \in \mathbb{Z}} \) and \( \{W_j\}_{j \in \mathbb{Z}} \) constitute what we call a multiresolution analysis framework.

### 2.2 Biorthogonal wavelets and the lifting scheme

To reconstruct a signal decomposed by the procedure above, we define a dual multiresolution analysis framework \( (\{\tilde{V}_j\}_{j \in \mathbb{Z}}, \{\tilde{W}_j\}_{j \in \mathbb{Z}}) \) associated with a dual scaling function \( \tilde{\Phi} \) and a dual wavelet function \( \tilde{\Psi} \), that are biorthogonal to \( \Phi \) and \( \Psi \) (see equation 3).

\[
\begin{align*}
&< \tilde{\Phi}_{j,k}, \Phi_{j',k'} > = \delta_{kk'} \\
&< \tilde{\Psi}_{j,k}, \Phi_{j',k'} > = \delta_{kk'} \\
&< \tilde{\Psi}_{j,k}, \Phi_{j,k'} > = 0 \\
&< \tilde{\Phi}_{j,k}, \Phi_{j,k'} > = 0
\end{align*}
\] (3)

\( \tilde{\Phi} \) and \( \tilde{\Psi} \) satisfy refinement relations such as equations 1 and 2, with coefficients \( \tilde{H} = (\tilde{h}_k)_{k \in \mathbb{Z}} \) and \( \tilde{G} = (\tilde{g}_k)_{k \in \mathbb{Z}} \) respectively. A function \( f \in L_2(\mathbb{R}) \) can be written as follows:

\[ x \mapsto f(x) = \sum_{j,k \in \mathbb{Z}} < f, \tilde{\Phi}_{j,k} > \tilde{\Phi}_{j,k}(x) \]
\[ = \sum_{k \in \mathbb{Z}} < f, \tilde{\Phi}_{j,k} > \tilde{\Phi}_{j,k}(x) \] (4)
\[ + \sum_{j \leq J, k \in \mathbb{Z}} < f, \tilde{\Psi}_{j,k} > \tilde{\Psi}_{j,k}(x), J \in \mathbb{Z} \]
The lifting scheme allows the construction of compactly supported biorthogonal wavelets with compactly supported duals, i.e. such that the refinement filters \( H, G, \tilde{H} \) and \( \tilde{G} \) are finite filters (a finite number of coefficients are non-zero). It relies on a simple relationship between all multiresolution analysis frameworks that share the same scaling function: let \( \{ \Phi^0, \tilde{\Phi}^0, \Psi^0, \tilde{\Psi}^0 \} \) be an initial set of biorthogonal scaling and wavelet functions, then a new set \( \{ \Phi, \tilde{\Phi}, \Psi, \tilde{\Psi} \} \) can be found according to equation 5 (Sweldens, 1996).

\[
\begin{align*}
\Phi(x) &= \Phi^0(x) \\
\tilde{\Phi}(x) &= 2 \sum_k \hat{h}_k \tilde{\Phi}(2x - k) + \sum_k s_{-k} \tilde{\Psi}(x - k) \\
\Psi(x) &= 2 \sum_k \hat{g}_k \Phi(2x - k) \\
\tilde{\Psi}(x) &= \Psi^0(x) - \sum_k s_k \Phi^0(x - k)
\end{align*}
\]  

(5)

Coefficients \( s_k \) of equation 5 can be freely chosen. Hence we can start with an initial set of biorthogonal filters and generate new biorthogonal filters by the previous relation. As a consequence, the lifting scheme is particularly suited to design a custom wavelet basis: we can generate several biorthogonal filters, using an optimization procedure, in order to find the wavelet basis maximizing a criterion (the retrieval performance of the system, in our application).

Claypoole et al. (1998) proposed a filter bank adapted from equation 5. It uses the lazy wavelet (Sweldens, 1996) as initial biorthogonal filter and two filters \( P = (p_i)_{i \in \mathbb{N}} \) and \( U = (u_i)_{i \in \mathbb{N}} \) are defined from the \( s_k \) coefficients. One step in the lazy wavelet transform consists in splitting the signal into its odd and even indexed samples. One step of the wavelet decomposition is given as follows:

- **split**: the signal \( x[n] \) is split into its odd \( x_o[n] \) and even \( x_e[n] \) coefficients.
- **predict**: we generate the wavelet coefficients \( d[n] \) as the error in predicting \( x_o[n] \) from \( x_e[n] \) using predictor operator \( P \): \( d[n] = x_o[n] - P(x_e[n]) \).
- **update**: combine \( x_e[n] \) and \( d[n] \) to obtain scaling coefficients representing a coarse approximation to the original signal \( x[n] \). This is done by using an update operator \( U \): \( c[n] = x_e[n] + U(d[n]) \).

The filter bank is depicted in figure 1. Let \( N_P \) and \( N_u \) be the length of the linear filters \( P \) and \( U \). (Claypoole et al., 1998) demonstrated that, to form a biorthogonal wavelet, filters \( P \) and \( U \) only
have to satisfy the following conditions:

\[
\begin{align*}
    \sum_{i=1}^{N_p} p_i &= 1 \\
    \sum_{i=1}^{N_u} u_i &= \frac{1}{2}
\end{align*}
\]  

(6)

There are then \( N_p + N_u - 2 \) undetermined coefficients. Wavelets designed by the lifting scheme have a support length equal to \( s/t \) given in equation 7, where \( s \) is the support of the low-pass filter and \( t \) is the support of the high-pass filter.

\[
\begin{align*}
    s &= 2(N_p + N_u) - 3 \\
    t &= 2N_p - 1
\end{align*}
\]  

(7)

For instance, the following wavelets have been evaluated in the proposed system:

- the Le Gall 5/3 biorthogonal wavelet, which is used in the JPEG 2000 standard part I (support=5/3) (Le Gall and Tabatabai, 1988)
- the Daubechies 9/7 biorthogonal wavelet, also in the JPEG 2000 standard (support=9/7)
- the orthogonal Haar wavelet (support=2/2)
- Cubic B-spline wavelet (support=4/2) (Cohen et al., 1992)
- the Daubechies 4-tap orthogonal wavelet (support=4/4) (Daubechies, 1988)
- the Daubechies 6-tap orthogonal wavelet (support=6/6) (Daubechies, 1988)
- Adapted wavelets (see section 3)

2.3 Decomposition of images

In the case of images, a separable wavelet decomposition is commonly used; it means that the unidimensional filtering operations described above are performed on both rows and columns separately (the same filters are used along both directions). Three subimages of coefficients are obtained at each level \( l \): \( lHL \), \( lLH \) and \( lHH \), and a unique approximation subimage at the last level \( N \): \( NLL \). These coefficients are obtained from the decomposition of \((l-1)LL\) (0LL is the raw image):

- \( lHL \) contains high frequency coefficients along columns and low frequency coefficients along lines,
- \( lLH \) contains low frequency coefficients along columns and high frequency coefficients along lines,
- \( lHH \) contains high frequency coefficients along columns and along lines,
- \( lLL \) contains low frequency coefficients along columns and along lines, it is decomposed if \( l < N \).
Subimages of coefficients are called subbands. An image decomposed on $N_l$ levels thus consists of $3 \times N_l + 1$ subbands. An example of an image and its wavelet transform is shown in figure 2.

Such a decomposition scheme is said to be pyramidal. Other decomposition schemes are conceivable, indeed it is possible to decompose other subbands of level $(l-1)$ in addition to $(l-1)LL$, and the choice of the subbands to decompose can be different at each scale. JPEG 2000 standard part II (JPEG, 2001), for instance, lets us specify any decomposition scheme. In particular, the following four schemes are specified in the standard, they are illustrated in figure 3:

- pyramidal decomposition
- SPACL decomposition (Signal Processing And Coding Lab, University of Arizona)
- packet decomposition
- FBI decomposition (Federal Bureau of Investigation, for fingerprint image compression)

Typically, the interest of using a different decomposition scheme is to find a basis relative to which the given image or collection of images has the lowest information cost, for compression purposes. Also, some subbands may emerge in which it is easier to discriminate between images of different classes, for classification or image retrieval purposes.
2.4 Image signatures

In order to characterize the image texture at different scales, we characterize the distribution of the wavelet coefficients in each subband of such a decomposition.

2.4.1 Histogram Representation

The first approach we consider to define an image signature is thus to build a histogram of the transformed image coefficients in each subband. If images are decomposed with a pyramidal scheme on $N_l$ levels, they consist of $3 \times N_l + 1$ subbands: there are 3 subbands of details at each scale $l \leq N_l$ (IH, IHL and LH) plus an approximation (NLL), $3 \times N_l + 1$ histograms are thus built. The signature is a vector formed by the concatenation of these histograms. The size of the resulting signature is relatively large. Indeed, let $NB$ be the number of bins in each histogram, if $N_l=4$ and $NB=32$, the signature is consists of $NB \times (3 \times N_l + 1) = 416$ coefficients. For this reason, we propose an alternative method, in the next paragraph, that alleviates the problem.

2.4.2 Generalized Gaussian Representation

To characterize textures, Wouwer et al. (1999) proposed to model the distribution of transformed coefficients in each subband with generalized Gaussian functions. A generalized Gaussian function may be expressed as in equation 8:
\[
\begin{aligned}
  p(x; \alpha, \beta) &= \frac{\beta}{2\alpha \Gamma\left(\frac{1}{\alpha}\right)} e^{-\left(\frac{|x|}{\beta}\right)^\alpha} \\
  \Gamma(z) &= \int_0^\infty e^{-t^{\alpha-1}} dt, \ z > 0
\end{aligned}
\] (8)

We propose to adapt this representation to a larger range of images. In the three studied databases, we notice that the representation holds for each subband except the last (the low frequency subband), as figure 4 shows. It stands to reason: the histogram of an approximation of the image is a smoothed version of the image histogram.

As a consequence, we propose to build signatures consisting in the maximum likelihood estimator \((\hat{\alpha}, \hat{\beta})\) of a generalized Gaussian distribution in each subband except the last and an histogram of the transformed image coefficients in the last subband. For the same scenario as in the last paragraph, the signature consists of \(2 \times (3 \times N_l) + NB = 56\) coefficients. To compute the maximum likelihood estimator, we applied the method proposed by Do and Vetterli (2002), extending the work of Varanasi and Aazhang (1989). This method relies on the Newton-Raphson root finding algorithm (Kay, 1993), that sometimes diverges on our images. This phenomenon is more or less important depending on the wavelet basis used to decompose images. As a consequence, we applied a robust root finding method proposed in (Press et al., 1992c): a hybrid method combining Newton-Raphson and the bisection algorithm.

2.5 Distance Measures

Each of the signatures presented above are associated with a particular distance measure, as described in the following paragraphs. The proposed distance measures are designed in order to profit from the characteristic of their associated signature.

2.5.1 Distance Between Histogram Based Signatures

The distance used to compare two images \(Im_1\) and \(Im_2\) by the first approach is base on the \(l_1\)–distance between histograms. Varela (2004) compared other distance measures between histograms, on several medical databases, and this distance measure turned out to give the highest precision. The distance measure is given in equations 9 and 10:

\[
d(Im_1, Im_2) = \sum_{i=1}^{3N_l+1} \lambda_i(H_i^1 - H_i^2)
\] (9)

\[
H_i^1 - H_i^2 = \sum_{j=1}^{NB} H_i^1(j) - H_i^2(j)
\] (10)

where \(H_i^n(j)\) is the value of the \(j^{th}\) bin of the \(i^{th}\) normalized histogram of image \(n\) and \((\lambda_i)_{i=1..3N_l+1}\) is a set of tunable weights.

2.5.2 Distance Between Generalized Gaussian Based Signatures

Do and Vetterli (2002) used the Kullback-Leibler divergence to compare two generalized Gaussian distributions. It measures the difference from a query probability distribution \(q\) to an arbitrary
probability distribution $i$, defined by parameters $\theta_q$ and $\theta_i$, respectively. Its general form is given in equation 11:

$$D(p(X;\theta_q)||p(X;\theta_i)) = \int p(X;\theta_q) \log \frac{p(X;\theta_q)}{p(X;\theta_i)} dx$$  \hspace{1cm} (11)$$

Injecting equation 8 in 11, we obtain equation 12:

$$D(p(X;\alpha_1,\beta_1)||p(X;\alpha_2,\beta_2))$$

$$= \log \left( \frac{\beta_1}{\beta_2} \frac{\alpha_1}{\alpha_2} \right) + \left( \frac{\alpha_1}{\alpha_2} \frac{\beta_2}{\beta_1} \right) - \frac{1}{\beta_1} \hspace{1cm} (12)$$

The proposed distance measure between two images $Im_1$ and $Im_2$ is a combination of equations 12 and 10 (it is not actually a distance measure since it is not symmetric):

$$d(Im_1, Im_2) = \sum_{i=1}^{3N_l} \lambda_i D(p(X;\alpha_i,1,\beta_i,1)||p(X;\alpha_i,2,\beta_i,2)))$$

$$+ (H_{3N+1} - H_{3N+1}^2) \hspace{1cm} (13)$$

where $(\alpha_{i,n}, \beta_{i,n})$ is the maximum likelihood estimator for the subband $i$ of image $n$.

The two couples (signature,distance measure) presented in this section will be compared experimentally in section 7.4. But first, we will see in the following section how the wavelet transform can be tuned in order to adapt these generic image representations to a given database.

### 3 Wavelet Adaptation

A major interest of wavelet decompositions is that we can choose the wavelet basis used to decompose images. Thus it can be adapted to the images and pathologies studied. However, the design of matched wavelet bases is usually a difficult task, as several constraints have to be satisfied simultaneously, in particular perfect reconstruction and vanishing moments. Moreover, these properties might not be sufficient to process specific signals or images. Several studies have thus been carried out to adapt a wavelet to a signal of reference, with different constraints on the wavelet properties and on the adaptation criteria. Usually, the goal is to minimize the $L^2$ norm between the signal and an approximation derived from its wavelet decomposition (Tewfik et al., 1992; Gupta et al., 2005), for compression purposes, for instance. These methods have been applied to medical classification problems, in order to match waveform signals (Maitrot et al., 2005; de Sobral Cintra et al., 2004). However, in our case, we do not intend to match a signal of reference: we want to find the wavelet basis that maximizes the retrieval performance of our system.

#### 3.1 Adapting the wavelet basis with a high level criterion

We propose in this article a method to find, within the lifting scheme framework, the wavelet basis that maximizes a given high level criterion, such as the retrieval performance of our system. The lifting scheme is convenient since it generates biorthogonal wavelet bases, and it is known that biorthogonality implies perfect reconstruction of decomposed images (Mallat, 1999). Indeed, any couple of filters $P/U$ (of size $N_p$ and $N_u$, respectively) satisfying the simple relations of equation 6 defines a biorthogonal wavelet filter bank.
3.2 First solution: adapting biorthogonal wavelets

A first solution to adapt the wavelet filter, within the lifting scheme framework, is thus to browse the space of biorthogonal filter banks, by an optimization procedure, and to select the filter bank that maximizes the adaptation criterion. The first \( N_p - 1 \) (resp. \( N_u - 1 \)) coefficients of \( P \) (resp. \( U \)) can be freely chosen, and the last coefficient of both vectors are computed in order to satisfy equation 6; as a consequence, we have \( N_p + N_u - 2 \) degrees of freedom. The optimization procedure we used is described in appendix A: it is based on a genetic algorithm and Powell’s direction set method.

3.3 Second solution: adapting biorthogonal wavelets with vanishing moments

In an image compression application, the \( N_p + N_u - 2 \) degrees of freedom may be tuned in order to minimize the prediction error. To achieve this goal, Claypoole et al. (1998) proposed a solution to eliminate the first \( N_p \) moments of the primal wavelet \( \Psi \) and the first \( N_u \) moments of the dual wavelet \( \tilde{\Psi} \). In our application, we wish to find a filter bank that would imply a high retrieval performance. However, it can be useful to design wavelets with several vanishing moments: it makes it possible to reduce the dimension of our optimization problem. Another interest is discussed in section 4. So, we propose a solution to find a trade-off between these two goals: maximizing the retrieval performance and eliminating the lowest order moments of the wavelet. First, we eliminate the first \( M_p \) (resp. \( M_u \)) moments of the primal (resp. dual) wavelet, where \( M_p \leq N_p \) and \( M_u \leq N_u \); it sets \( M_p \) and \( M_u \) degrees of freedom to design filters \( P \) and \( U \), respectively. And the remaining \( N_p - M_p \) and \( N_u - M_u \) degrees of freedom, respectively, are set in order to maximize our high level-criterion. Note that the first solution described above is a particular case of this second solution, where \( M_p = M_u = 1 \); the same optimization procedure (described in appendix A) is used to find the \((N_p - M_p) + (N_u - M_u)\) free coefficients left after eliminating the first wavelet moments. The way we adapt the predict filter \( (P) \) and the update filter \( (U) \) is detailed in the next sections.

3.3.1 Adapting the predict filter \((P)\)

In order to control the regularity of the primal wavelet, we will eliminate its first \( M_p \leq N_p \) moments:

\[
\int t^n \Psi(t) dt = 0, \quad n = 0..M_p - 1
\]  

(14)

In that purpose, we define the following matrix \( V \), of size \( M_p \times N_g \), where \( N_g = 2N_p + 1 \) is the size of the wavelet filter \( G \) (see equation 7):

\[
V(k, l) = (l + 1 - N_p)^k, \quad k = 0..M_p - 1, \quad l = 0..N_g - 1
\]  

(15)

\[
V.G = 0
\]  

(16)

If the wavelet filter \( G \) satisfies equation 16, then the first \( M_p \) moments of the primal wavelet vanish (each line in \( V \) eliminates one moment) (Claypoole et al., 1998).

We would like to transform this system in order to determine filter \( P \). Comparing the two equivalent filter banks depicted in figure 1, we see that filter \( G \) can be expressed as a function of \( P \). Indeed, \( d[n] \), the high-pass filtered version of the input signal \( x[n] \), satisfies the relations \( d[n] = G(x[n]) = x_o[n] - \)
$P(x_{[n]})$: as a consequence, if for instance $P = (p_0, p_1, p_2, p_3)$, then $G = (-p_0, 0, -p_1, 1, -p_2, 0, p_2)$ (see figure 5 (a)).

Let the central coefficient in $G$ be indexed by 0; we notice that the even coefficients in $G$ are equal to 0, except the central coefficient which is equal to 1. As a consequence, we can reduce the dimension of the system above. Let $C$ be the matrix containing the even coefficients in $V$; equation 16 becomes:

$$C: \begin{pmatrix} p_0 & \ldots & p_{N_p-1} \end{pmatrix}^t = \begin{pmatrix} 1 & 0 & \ldots & 0 \end{pmatrix}^t$$

The right-hand side of this equation comes from the product of the central column of $V$, i.e. $(0^0 = 1, 0^1 = 0, 0^2 = 0, \ldots)^t$, by the 1 in the central coefficient of $G$ when solving equation 16 (the convention $0^0 = 1$ was proposed by Claypoole et al. (1998)). Notice that the constraint associated with the first line in $C$ is equivalent to the first biorthogonality relation (see equation 6).

So we have $M_p \leq N_p$ equations and $N_p$ unknowns. We propose to generate coefficients $(p_i)_{i=0\ldots N_p-M_p-1}$ by the optimization module, in order to maximize the retrieval performance of the system (see appendix A) and to determine coefficients $(p_i)_{i=N_p-M_p\ldots N_p-1}$ thanks to equation 17. As a consequence, we have to solve a linear system $A.x = b$ of dimension $M_p$, where:

- $A$ is a square matrix of size $M_p$, which consists of the last $M_p$ columns in $C$.
- Let $B$ be the matrix containing the first $N_p - M_p$ columns in $C$. The right-hand side $b$ is worth $ \begin{pmatrix} 1 & 0 & \ldots & 0 \end{pmatrix}^t - B: \begin{pmatrix} p_0 & \ldots & p_{N_p-M_p-1} \end{pmatrix}^t$.

We propose to solve this system by the LU method (Press et al., 1992a).

### 3.3.2 Adapting the update filter (U)

In order to control the regularity of the dual wavelet, we will eliminate its first $M_u \leq N_u$ moments:

$$\int t^n \hat{\Psi}(t) dt = 0, n = 0..M_u - 1$$

As in the previous section, we design a matrix $V'$ of size $M_u \times N_q$ to eliminate the $M_u$ moments of $\hat{G}$. Since, we work with biorthogonal wavelets, filters $\hat{G}$ and $H$ are related by equation 19 (Mallat, 1999) and $N_q = N_h = 2(N_p + N_u) - 3$ (see equation 7).
\[
\hat{g}_k = (-1)^k h_k, \forall k
\]  
\(V'\) is thus defined by equation 20 and \(\tilde{G}\) must satisfy equation 21.

\[
V'(k, l) = (l + 2 - N_p - Nu)^k, k = 0..M_u - 1, l = 0..N_h - 1
\]

\[
V'.\tilde{G} = 0
\]

We can express this system as a function of filters \(P\) and \(U\). Indeed, \(c[n]\), the low-pass filtered version of the input signal \(x[n]\), satisfies the relations \(c[n] = H(x[n]) = x_e[n] + U(d[n]) = x_e[n] + U(x_o[n] - P(x_e[n]))\): as a consequence, if for instance \(P = (p_0, p_1)\) and \(U = (u_0, u_1)\), then \(H = (-p_0 u_0, u_0, 1 - p_1 u_0 - p_0 u_1, u_1, -p_1 u_1)\) (see figure 5 (b)) and according to equation 19, \(\tilde{G} = (-p_0 u_0, -u_0, 1 - p_1 u_0 - p_0 u_1, -u_1, -p_1 u_1)\). Suppose that filter \(P\) has already been determined by the method proposed in section 3.3.1, then we can express equation 21 as a function of \(U\). We propose a convenient solution below.

Let \(C'\) be the matrix of size \(M_u \times N_u\) defined in equation 22, in which \(g_{i}^{0}\) is the factor of \(u_j\) in the expression of \(\hat{g}_i\) (for instance, if \(\hat{g}_i = 1 - p_1 u_0 - p_0 u_1\), then \(g_{i}^{0} = -p_1, \hat{g}_{i}^{1} = -p_0, g_{i}^{2} = 0\), etc.).

\[
C'(k, l) = \sum_u V'(k, u)\hat{g}_{u}^{l}
\]

\[
C'.u = \begin{pmatrix} -1 & 0 & \ldots & 0 \end{pmatrix}^t
\]

\[
(0^0 = 1, 0^1 = 0, 0^2 = 0, \ldots)^t, \text{by the 1 in the central coefficient of } \tilde{G}.
\]

So we have \(M_u \leq N_u\) equations and \(N_u\) unknowns. We propose to generate coefficients \((u_j)_{j=0..N_u-M_u-1}\) by the optimization module, in order to maximize the retrieval performance of the system (see appendix A) and to determine coefficients \((u_j)_{j=N_u-M_u..N_u-1}\) thanks to equation 23. As a consequence, we have to solve a linear system \(A'.x = b'\) of dimension \(M_u\), where:

- \(A'\) is a square matrix of size \(M_u\), which consists of the last \(M_u\) columns in \(C'\).

- Let \(B'\) be the matrix containing the first \(N_u - M_u\) columns in \(C'\). The right-hand side \(b'\) is worth \(\begin{pmatrix} -1 & 0 & \ldots & 0 \end{pmatrix}^t - B'.\begin{pmatrix} u_0 & \ldots & u_{N_u-M_u-1} \end{pmatrix}^t\).

We also propose to solve this system by the LU method.

This concludes the description of the method we propose to adapt the wavelet basis with some high level criterion. We will now see how to adapt it in order to maximize the performance of a retrieval system.

### 4 Image indexing and compression

We envisage to extract the image signatures described above as images are compressed to be stored in a database: the same decomposition would be used for both indexing and compressing images. With this in mind, the use of the lifting scheme is particularly convenient, since it is used in compression standards such as JPEG 2000. In our application, we wish to find a filter bank that
would imply a high retrieval performance and that may be used to compress images efficiently before they are stored in the database. That is why the second solution we proposed to adapt the wavelet (see section 3.3) is particularly suited: we build a wavelet basis with a desired number of vanishing moments, which ensures a good compression rate (Claypoole et al., 1998). In other words, the custom wavelet we build can be used to store images in a database efficiently.

5 Retrieval Performance

To tune the proposed retrieval system for a given database, we search for the system settings that maximize the retrieval performance. The retrieval performance depends on the classification of each image in the database, defined a priori by the users, in order to catch their perception of similarity between images.

Let $N$ be the number of images in the database and $V_n$ the relevance of an image $n$ for a query: $V_n = 1$ if $n$ and the query image belong to the same class, $V_n = 0$ otherwise. The usual way to evaluate a retrieval system performance is the following (Müller et al., 2001): 1) each image in the database is presented to the system as a query, 2) the system ranks the $N - 1$ remaining images by increasing order of the distance measure, 3) for each integer $k = 1..N-1$ the following values are computed:

- true positive detections: the number of relevant images returned $A_k = \sum_{n=0}^{k-1} V_n$
- false positive detections: the number of irrelevant images returned $B_k = \sum_{n=0}^{k-1} (1 - V_n)$
- false negative detections: the number of relevant images not returned by the system $C_k = \sum_{n=0}^{N-1} V_n - A_k$

From these values, the following standard performance values are derived:

- recall: the fraction of relevant images that are returned $R_k = \frac{A_k}{A_k + C_k}$
- precision: the fraction of images returned that are relevant $P_k = \frac{A_k}{A_k + B_k}$

To evaluate the system, precision-recall curves are usually plotted: they consist in linking consecutive points $(P_k, R_k)$ for $k = 1..N-1$. However, to optimize the system settings, a numerical value is needed: we measured the retrieval performance of the system by the mean precision evaluated at a cut-off rank $k$ of interest, which is called mean precision at $k$.

Note that, to retrieve images in a large database, the query image shall not be compared with each image in the database: to speed up the retrieval process, some efficient search algorithms may be applied (Quellec et al., 2008; Oliveira et al., 2007; Dy et al., 2003; Iakovidis et al., 2008).

6 Retrieval system Calibration

To tune the retrieval system, the following parameters have to be defined:

- $P_1$: the model used to represent the distribution of coefficients on each subband (either histograms or generalized Gaussians),
• $P_2$: the number of decomposition levels $N_l$,
• $P_3$: the wavelet used,
• $P_4$: the decomposition scheme (either pyramidal, packet or SPACL decomposition).

Each parameter is evaluated independently, the other three are set to a default value. The following default values are used:

• for $P_1$: generalized Gaussian model,
• for $P_2$: $N_l = 3$,
• for $P_3$: Daubechies 9/7 wavelet,
• for $P_4$: pyramidal decomposition.

Each time a parameter tuple $\{P_1, P_2, P_3, P_4\}$ is evaluated, the optimal weights between the subband distance measures are searched for. To find the weights, we used the optimization procedure that was also used to find the wavelet coefficients (described in appendix A). In that purpose, the database is divided into a training and an evaluating set. The weights are set in order to maximize the mean precision at $k$ on the training set. Then, we compute the mean precision at $k$ on the evaluation dataset, using the weights found on the training set.

Finally, to design an optimal wavelet basis within the lifting scheme framework, we use the optimal values found for parameters $P_1$, $P_2$ and $P_4$. Notice that, to find the optimal wavelet basis, we have to conduct a two-stage optimization: each time a wavelet basis is evaluated, we search for the optimal weight vector, as illustrated in figure 6. The wavelet basis is also designed in order to maximize the mean precision at $k$ on the training dataset.
7 Application

The retrieval method proposed below is evaluated on two medical databases, and on a face database to show its genericity. The databases are first described and the results we obtained are given in section 7.4.

7.1 Databases

7.1.1 Diabetic Retinopathy Database (DRD)

This database has been built at the LaTIM laboratory\(^1\) (Inserm U650) for research on diabetic retinopathy follow up. Diabetes is a metabolic disorder characterized by sustained inappropriate high blood sugar levels. This progressively affects blood vessels in many organs, which may lead to serious renal, cardiovascular, cerebral and also retinal complications. In the latter case, the pathology, namely diabetic retinopathy, can cause blindness. The database consists of 78 patient files: one for each diabetic patient seen at Brest University Hospital from June 2003 to April 2007 for an eye test. It contains 1261 photographs altogether, with associated anonymous information on the pathology. Images have a definition of 1280 pixels/line for 1008 lines/image. They were acquired by experts using a Topcon Retinal Digital Camera (TRC-501A) connected to a computer and are lossless compressed. Four types of images are obtained: color, red free, blue-light images and the angiographic sequences. An image series is given in figure 7 as an illustration. The disease severity level, ranging from 0 to 5, according to ICDRS classification (Wilkinson et al., 2003), was determined by experts for each patient.

7.1.2 Digital Database for Screening Mammography (DDSM)

The DDSM project (Heath et al., 1998), involving the Massachusetts General Hospital, the University of South Florida and the Sandia National laboratories, has built a mammographic image database for research on breast cancer screening. It consists of 2500 patient files. Each one includes two images of each breast, along with some associated patient information. The database consists of 10000 images. These images have a varying definition, of about 2000 pixels/line for 5000 lines/image. An example of image sequence is given in figure 8. Each patient file has been graded by a physician. Patients are then assigned one of these labels: normal, benign and cancer.

7.1.3 Face Database (FD)

The database consists of 400 images: ten photographs of 40 distinct subjects. For some of them, the images were taken at different times, with different lightings, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position. Images have a definition of 92 pixels/line for 112 lines/image. The database can be viewed on the website\(^2\). Figure 9 shows the 10 face images of a subject. Even though it is not our main purpose, this database can be classified easily. Images belong to the same class if and only if they represent the same subject.

---

\(^1\)http://latim.univ-brest.fr
\(^2\)www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
Figure 7: Photograph series of a patient eye. Images (a), (b) and (c) are photographs obtained by applying different color filters on the camera lens. Images (d) to (j) form a temporal angiographic series: a contrast agent is injected and photographs are taken at different stages (early (c), intermediate (d), (f), (g), (h), (j) and late (i)). For the intermediate stage, photographs from the periphery of the retina are available.

Figure 8: Mammographic image sequence of the same patient. (a) and (b) are images of the left breast, (c) and (d) are images of the right one.
7.2 Class Definition

Concerning the two medical databases, we are mainly interested in the medical grade of each image. As a consequence, we associate each possible grade with a class: six classes are then defined in DRD and three in DDSM. In the face database, we associate each person with a class, so we have 40 classes of ten elements. The goal of the system is thus to retrieve images belonging to the same class.

7.3 Objectives of the retrieval system

In a medical application, the retrieval system is intended to be used to assist the diagnosis of physicians. Such a system would be used as follows: if a physician has a doubt on his diagnosis, he could send the acquired images to the retrieval system, that will display the $k$ most similar images along with their medical interpretations. The physician would then analyze the retrieved images and, reasoning by analogy, these images may help him confirm or invalidate his diagnosis for the new case. In order to limit the time required by physician to analyze the retrieved cases, ophthalmologists in Brest University Hospital want the system to return only the five most similar images for a query. For comparison purposes, the same number of images is displayed for the other two databases. As a consequence, we want the five retrieved cases to be as relevant to the query as possible: we will tune the method parameters in order to maximize the mean precision when $k=5$ images are returned by the system, i.e. the mean precision at $k=5$.

7.4 Results

To assess the performance of the method, each database is divided into a training and an evaluation dataset. The size of these datasets are chosen according to the number of images available for each database. For DRD and FD, the training dataset consists of 50% of the whole database; on DDSM, it consists of 5% of the database. For both medical databases, the training and the evaluation datasets have the same proportion of images from each class. For FD, all images in a given sequence are assigned to the same set. The images are assigned to the two datasets at random, with respect
Table 1: Mean precision at five (in %) and standard deviation (std)

<table>
<thead>
<tr>
<th>database</th>
<th>DRD</th>
<th>DDM</th>
<th>FD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>std</td>
<td>mean</td>
</tr>
<tr>
<td>N_j</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>42.88</td>
<td>19.73</td>
<td>63.97</td>
</tr>
<tr>
<td>1</td>
<td>51.58</td>
<td>18.96</td>
<td>67.43</td>
</tr>
<tr>
<td>2</td>
<td>52.42</td>
<td>18.93</td>
<td>67.72</td>
</tr>
<tr>
<td>3</td>
<td>54.17</td>
<td>18.78</td>
<td>68.23</td>
</tr>
<tr>
<td>4</td>
<td>54.12</td>
<td>18.89</td>
<td>68.37</td>
</tr>
<tr>
<td>5</td>
<td>53.22</td>
<td>19.74</td>
<td>67.99</td>
</tr>
<tr>
<td>wavelet basis</td>
<td>Le Gall 5/3</td>
<td>52.90</td>
<td>19.37</td>
</tr>
<tr>
<td></td>
<td>Daubechies 9/7</td>
<td>54.17</td>
<td>18.78</td>
</tr>
<tr>
<td></td>
<td>Haar</td>
<td>49.95</td>
<td>19.20</td>
</tr>
<tr>
<td></td>
<td>Cubic B-spline</td>
<td>53.30</td>
<td>18.77</td>
</tr>
<tr>
<td></td>
<td>Daubechies 4</td>
<td>56.10</td>
<td>19.54</td>
</tr>
<tr>
<td></td>
<td>Daubechies 6</td>
<td>56.14</td>
<td>19.95</td>
</tr>
<tr>
<td>model</td>
<td>histograms</td>
<td>51.77</td>
<td>18.73</td>
</tr>
<tr>
<td></td>
<td>gen. Gaussians</td>
<td>54.17</td>
<td>18.78</td>
</tr>
<tr>
<td>decomposition</td>
<td>pyramidal</td>
<td>54.17</td>
<td>18.78</td>
</tr>
<tr>
<td></td>
<td>wavelet packet</td>
<td>51.36</td>
<td>19.97</td>
</tr>
<tr>
<td>scheme</td>
<td>SPACL</td>
<td>53.81</td>
<td>19.22</td>
</tr>
<tr>
<td>adapted wavelet</td>
<td></td>
<td>56.50</td>
<td>19.26</td>
</tr>
</tbody>
</table>

to the constraints above. The same training and evaluation datasets are used for each experiment. For wavelet adaptation, the size of filters $P$ and $U$, respectively $N_p$ and $N_u$, are both set to 4. At first, these filters only verify the biorthogonality relations (see section 3.2). The mean precision at $k=5$ is given in table 1. To study the influence of the training dataset cardinality, we repeat the experiment with a training dataset consisting of 25% of the whole database for DRD and FD, and of 2.5% for DDSM. The scores obtained with an adapted wavelet are reported in table 2. It emerges that the performance of the method increases with the size of the training dataset, we can thus expect an improvement as the databases will grow. We show in table 3 that we can also successfully adapt the wavelet if we eliminate some moments (see section 3.3). The experiment has been carried out on FD.

The optimal weight vectors and $P/U$ filters are reported in tables 4 and 5. The wavelet and scaling filters, respectively $G$ and $H$, corresponding to filters $P$ and $U$, are plotted in figure 10. The mean computation time to retrieve the 5 closest images is given in table 6 for the two types of signatures: histogram based (H) or generalized Gaussian based (GG). All experiments were conducted using an AMD Athlon 64-bit based computer running at 2 GHz, with the default values for $P_2$, $P_3$ and $P_4$, for comparison purposes.
Table 2: Precision at 5 with a smaller training dataset

<table>
<thead>
<tr>
<th></th>
<th>DRD</th>
<th>DDSM</th>
<th>FD</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>52.42</td>
<td>70.75</td>
<td>94.40</td>
</tr>
<tr>
<td>std</td>
<td>21.89</td>
<td>26.24</td>
<td>10.47</td>
</tr>
</tbody>
</table>

Table 3: Mean precision at 5 with vanishing moments

<table>
<thead>
<tr>
<th></th>
<th>$M_u=1$</th>
<th>$M_u=2$</th>
<th>$M_u=3$</th>
<th>$M_u=4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_p=1$</td>
<td>95.50</td>
<td>95.60</td>
<td>95.20</td>
<td>95.00</td>
</tr>
<tr>
<td>$M_p=2$</td>
<td>96.00</td>
<td>96.10</td>
<td>95.20</td>
<td>95.00</td>
</tr>
<tr>
<td>$M_p=3$</td>
<td>95.40</td>
<td>95.10</td>
<td>95.10</td>
<td>95.10</td>
</tr>
<tr>
<td>$M_p=4$</td>
<td>95.20</td>
<td>95.20</td>
<td>95.10</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4: Optimal weight vectors. The weights are gathered level by level: \{HH,HL,LH\}.

<table>
<thead>
<tr>
<th>subbands</th>
<th>DRD</th>
<th>DDSM</th>
<th>FD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st level</td>
<td>{2.030,0.3072,0.6189}</td>
<td>{17.73,43.32,13.78}</td>
<td>{0.4416,1.672,2.861}</td>
</tr>
<tr>
<td>2nd level</td>
<td>{22.62,3.553,2.705}</td>
<td>{26.31,8.748,3.524}</td>
<td>{0,0,0}</td>
</tr>
<tr>
<td>3rd level</td>
<td>{3.091,6.177,1.201}</td>
<td>{4.150,3.124,4.574}</td>
<td>{0,0,0}</td>
</tr>
<tr>
<td>4th level</td>
<td>{0,0,0}</td>
<td>{17.62,28.15,3.892}</td>
<td>{0,0,0}</td>
</tr>
</tbody>
</table>

LL

<table>
<thead>
<tr>
<th></th>
<th>DRD</th>
<th>DDSM</th>
<th>FD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st level</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5: Optimal filters $P$ and $U$.

<table>
<thead>
<tr>
<th></th>
<th>DRD</th>
<th>DDSM</th>
<th>FD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>0.005319</td>
<td>-0.3874</td>
<td>0.09967</td>
</tr>
<tr>
<td>$U$</td>
<td>-0.4908</td>
<td>0.4068</td>
<td>0.4585</td>
</tr>
<tr>
<td>$P$</td>
<td>0.5987</td>
<td>0.4627</td>
<td>0.5924</td>
</tr>
<tr>
<td>$U$</td>
<td>-0.09482</td>
<td>0.01792</td>
<td>-0.1506</td>
</tr>
</tbody>
</table>

Table 6: Computation Time

<table>
<thead>
<tr>
<th></th>
<th>DRD</th>
<th>DDSM</th>
<th>FD</th>
</tr>
</thead>
<tbody>
<tr>
<td>wavelet transform</td>
<td>0.22s</td>
<td>1.99s</td>
<td>0.005s</td>
</tr>
<tr>
<td>computing histograms</td>
<td>0.03s</td>
<td>0.22s</td>
<td>0.0002s</td>
</tr>
<tr>
<td>distance with each image in the database</td>
<td>0.31s</td>
<td>2.26s</td>
<td>0.109s</td>
</tr>
<tr>
<td>overall time</td>
<td>0.56s</td>
<td>4.47s</td>
<td>0.114s</td>
</tr>
<tr>
<td>estimating ($\hat{\alpha}, \hat{\beta}$)</td>
<td>4.35s</td>
<td>33.90s</td>
<td>0.03s</td>
</tr>
<tr>
<td>distance with each image in the database</td>
<td>0.16s</td>
<td>1.16s</td>
<td>0.056s</td>
</tr>
<tr>
<td>overall time</td>
<td>4.73s</td>
<td>37.05s</td>
<td>0.091s</td>
</tr>
</tbody>
</table>
The proposed method is compared to a classic CBIR method: the retrieval of similar images based on gray level histograms. It is a particular case of the proposed method: when we take \( N_l = 0 \) in the wavelet decomposition scheme, it leads to the classic comparison of gray level histograms. The score obtained with this method is thus the first line of table 1. The proposed and the classic methods are compared on figure 11.

Finally, as an illustration, the precision-recall curves obtained for each evaluated value of the system parameters (\( P_1, P_2, P_3 \) and \( P_4 \) - see section 6) are given in figure 12. We observe that, on medical databases, the precision-recall curves are not influenced much by any parameter, except for small recall values (see also table 1). The reason is that the learning procedure focuses on small recall values: it is supervised by the mean precision at five.
Figure 11: Comparison between the proposed method (right bars) and the classic histogram based method (left bars)
Figure 12: Precision-recall Curves
8 DISCUSSION AND CONCLUSION

In this article, two kinds of generic wavelet-based image signatures, with associated distance measures, have been evaluated in a CBIR system. They take advantage of the flexibility inherent in the wavelet transform framework to adapt the system to any specialized database. In particular, a way to adapt the wavelet transform to a high level criterion, within the lifting scheme framework, is proposed in this paper. It makes it possible to generate any wavelet transform, respecting the biorthogonality relations, and with a desired number of vanishing moments. A controlled random search, based on a genetic algorithm, is then performed in the predict and update filter space in order to find the optimal wavelet transform, and a similar search is performed in the distance weight vector space in order to maximize the precision of the system. The adapted wavelet transforms could also be used to store images in a compressed format, such as JPEG 2000, as their signature is computed. Indeed, as table 3 shows, a high mean precision score can still be achieved while eliminating a desired number of moments, which ensures a good compression rate. It is particularly suited to JPEG 2000 standard part II (JPEG, 2001), which allows the user to define a custom wavelet basis and a custom decomposition scheme.

The system efficiency, measured with the mean precision when five images are returned by the system, i.e. the mean precision at five, is satisfactory: it reaches 56.50% on DRD, 70.91% on DDSM and 96.10% on FD. On DRD for instance, it means that on average two or three images out of the five retrieved by the system are relevant for a query. Compared with a classic histogram based method, the improvement is notable, particularly on medical databases (see figure 11): for instance, we observe an increase of more than 30% on DRD.

It emerges from the precision-recall curves and the mean precision at five that the most influential parameters are the wavelet basis and the number of decomposition levels, which justifies our interest in finding an optimal basis within the lifting scheme framework. The optimization procedure always leads to a “better” wavelet basis than the standard bases evaluated. As for the decomposition scheme, it does not significantly influence the precision; as a consequence, we did not explore this trail any further. However, one could use the best basis algorithm (Coifman and Meyer, 1991) to find an optimal decomposition scheme. The best model to represent the distribution of coefficients (histograms or generalized Gaussian functions) vary from one database to another, but it generally does not influence much the system efficiency on the medical databases. However the generalized Gaussian model is preferred since it produces smaller signatures, and according to Occam’s razor principle, the best model is the simplest.

From a computational point of view, both models have their advantage. On the one hand computing the signatures is almost 20 times faster with the histogram based approach on average, but on the other hand it is twice as slow to compute the distance measure. As a consequence, the larger the database size is, the more adequate the generalized Gaussian based approach is; and the larger the query image definition is, the more adequate the histogram based approach is.

The proposed retrieval system is better suited if the criteria used to classify the databases are closely related to the image content. Thus, in the face database, two photographs of the same persons obviously share common features at each scale, so performance are good. Similarly, in DRD and DDSM, the medical grading of the pathology depends on the presence and the number of lesions in images, which affect the distribution of the wavelet transform coefficients at some lesion-specific scales. Moreover, in DDSM, the presence of lesions affects the distribution along specific directions: some directions are more discriminant (see table 4) because the layout of lesions is anisotropic. So, by assigning a larger weight to discriminant subbands, a good retrieval efficiency can be achieved.
Finally, considering the results obtained on the two medical databases, the system is precise enough to be used in a diagnosis aid system.

Appendix A - Optimization procedure

When we search for a weight vector between subbands or the most adapted wavelet, we have no preconceived ideas about either the fitness function or the free coefficients. Then we start the search by a controlled random process. Genetic algorithms (Goldberg, 1989) are amongst the quickest and most popular methods for finding rough estimations of the best local maxima within a given region of space. The general principle of genetic algorithms is to:

- generate a random population of solutions
- combine previous good solutions (crossover) and slightly modify some of them (mutation) to make the population evolve
- keep the fittest solutions at each generation

In particular, we used the steady state algorithm (at each generation, new individuals are created and added to the former population, the worst individuals of the resulting population being removed), with the following parameters:

- population size = 50
- maximum number of generations = 30
- selection methods: tournament selector (2 individuals are selected for crossover with a probability proportional to their score, and only the fittest is actually kept)
- crossover probability (CR) = 70% (CR: probability for two selected individuals to combine their genes and form two new individuals, otherwise they are cloned)
- mutation probability (MU) = 60% (MU: probability for a gene to be swapped with another after a crossover)

This configuration was adopted for its fast convergence.

Once approximations of the fitness function optima are found by a genetic algorithm, we use the best approximations as initial points for descents to reach the actual local minima. We used the Powell’s Direction Set method (Press et al., 1992b). The interest of this method is that it does not require computing the function gradient, which is unknown. The descent leads to a typical improvement of 2 or 3% in precision over genetic algorithms alone.

References


URL http://www.library.cornell.edu/nr/bookcpdf.html

URL http://www.library.cornell.edu/nr/bookcpdf.html

URL http://www.library.cornell.edu/nr/bookcpdf.html


