Sparse Overcomplete Gabor Wavelet Representation Based on Local Competitions

Sylvain Fischer, Gabriel Cristóbal, and Rafael Redondo

***Abstract*—Gabor representations present a number of interesting properties despite the fact that the basis functions are nonorthogonal and provide an overcomplete representation or a nonexact reconstruction. Overcompleteness involves an expansion of the number of coefficients in the transform domain and induces a redundancy that can be further reduced through computational costly iterative algorithms like Matching Pursuit. Here, a biologically plausible algorithm based on competitions between neighboring coefficients is employed for adaptively representing any source image by a selected subset of Gabor functions. This scheme involves a sharper edge localization and a significant reduction of the information redundancy, while, at the same time, the reconstruction quality is preserved. The method is characterized by its biological plausibility and promising results, but it still requires a more in depth theoretical analysis for completing its validation.**

***Index Terms*—Biological system modeling, image coding, image edge analysis, overcomplete representations, wavelet transforms.**

I. Introduction

THE GABOR approach involves a representation in terms of elementary functions ("logons") that are simultaneously localized in space and frequency [9]. Gabor functions constitute a family of bandpass filters that have been widely used in image processing in a broad range of applications such as texture analysis, motion estimation, color, etc. They can also be interpreted as a receptive field model for the simple cortical cells [3]. The two-dimensional Gabor elementary functions do not constitute an orthogonal basis, and to obtain exact reconstruction with Gabor filters in both analysis and synthesis filter banks, an *overcomplete basis* of functions is required. The number of coefficients in the transform domain is then larger than the number of pixels in the image domain, introducing a redundancy which in principle might limit the usefulness of the transform for image compression applications.

Orthogonal transforms like the discrete cosine transform (DCT) or wavelet-based schemes are actually the most used transforms for image compression. Nevertheless, they suffer some disadvantages like instability and artificial looking artifacts. Biorthogonal wavelets usually introduce aliasing across the subband analysis. Although such aliasing is canceled at reconstruction, the signal analysis is not optimal (frequency bands are not totally separated) and artifacts can be introduced by the compression.

Manuscript received March 8, 2004; revised January 27, 2005. This work was supported in part under Grants TIC2001-3697-C03-02 and TEC 2004-00 834, in part by the IM3 medical imaging thematic network, and in part by MEC-FPU and CSIC-I3P fellowships. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Fernando M. B. Pereira.

The authors are with the Instituto de Óptica-Consejo Superior de Investigaciones Científicas, 28006, Madrid, Spain (e-mail: sylvain@optica.csic.es).

Digital Object Identifier 10.1109/TIP.2005.860614

Overcomplete Gabor wavelets present the advantage of avoiding any aliasing at the cost of introducing a large number of redundant coefficients. Mallat and Zhang [14] showed that is it possible to represent a signal by means of an overcomplete dictionary of functions limiting at the same time the information redundancy. They proposed an iterative algorithm with the aim to keep only those dictionary functions that best match the signal. Olshausen and Field [16] argued that overcompleteness and such sparse coding algorithms are adequate methods for decorrelating natural images and it could be a strategy employed in the primary visual cortex.

It is important to stress that the problem of sparse coding differs radically from the image compression approaches using orthogonal transforms. In particular, finding the sparsest decomposition from a given overcomplete dictionary is still an unsolved problem. The solution, nonunique and necessarily nonlinear, is generally approximated iteratively through computationally costly algorithms. Thus, sparse coding have been very rarely used for image compression (a review on sparse image coding can be found in [17]).

On the other hand, Grossberg and coworkers simulated the behavior of neurons of the primary visual cortex implementing competitions between neighboring Gabor coefficients. The method sharpens edges and selects the most favored boundary orientations [11]. Although useful for denoising, this algorithm is not invertible and, therefore, cannot be employed directly for image coding.

The motivation of this paper is to investigate the feasibility of a new algorithm based on local competitions for removing the redundancy introduced by an overcomplete Gabor representation. The usefulness of the method is illustrated in the image compression area.

Section II summarizes the Gabor wavelets transform method. Section III describes an iterative algorithm for implementing the competitive strategy. Section IV summarizes the compression stages and Section V presents some image compression results. Concluding remarks are drawn in Section VI.

II. Linear Gabor Wavelets Representations

Log-Gabor filters are used for the wavelets transform (see [5] for a thorough description of log-Gabor properties). They are described in the Fourier domain through the polar coordinates (r, ϕ) as

$$G_k(r,\phi) = \exp\left(-\frac{1}{2}\left(\frac{\log(\frac{r}{r_k})}{\sigma_s}\right)^2 - \frac{1}{2}\left(\frac{\phi-\phi_k}{\sigma_a}\right)^2\right) \quad (1)$$

where (r_k, ϕ_k) is the center of the filter and σ_s, σ_a are, respectively, the scale and angular bandwidths common to all filters.

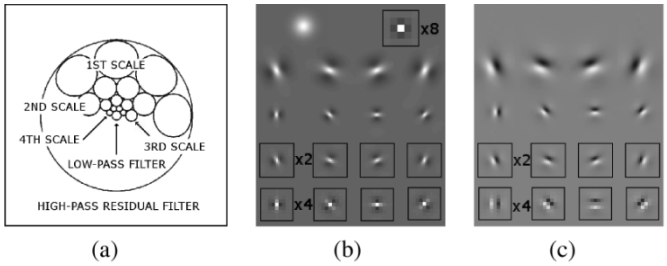

Fig. 1. Gabor's filter set used in this paper. (a) The Fourier domain is covered by log-Gabor filters with four scales, four orientations, a low-pass filter, and a high-pass filter. Only the contours at 60% of the maximum of amplitude are depicted. The high-pass residual filter is represented by the region outside the bigger circle. Note that even scales are shifted by a $\pi/8$ angle so as to obtain a better coverage of the Fourier domain. (b) Real parts of log-Gabor filters on four scales and four orientations are drawn in the space domain. The low-pass and high-pass filters can also be seen, respectively, on the top left and top right of the image. Second-scale, first-scale, and high-pass filters are magnified, respectively, by factors of 2, 4, and 8 for a better visualization. (c) Imaginary parts of log-Gabor filters in the spatial domain. Note low-pass and high-pass filters do not have imaginary parts.

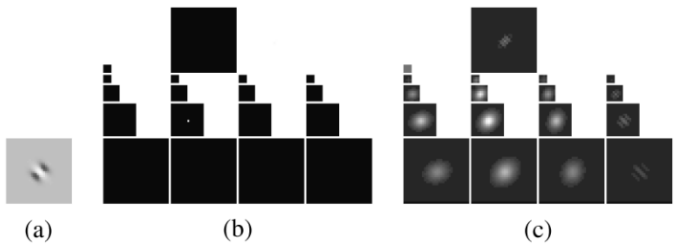

Fig. 2. (a) Comparison between two exact reconstruction pyramids. The 32×32 original image $\mathbf{x}$ is composed of an unique Gabor function that can be coded by the pyramid $\mathbf{h}$ shown in (b). (b) $\mathbf{h}$ pyramid contains a single nonzero coefficient. The low-pass and high-pass channels are located on the first row. The bottom row represents the first scale, that is, the highest frequency channels. The different orientations are arranged in columns. Each bandpass coefficient contains a real and an imaginary part. The modulus is shown here. (c) $\mathbf{g}$ pyramid is the linear transform of $\mathbf{x}$ (explanations are given in the text).

All filters have one octave-scale bandwidth and $\pi/4$ angular bandwidth.

The Gabor wavelet transform consists of filtering in the Fourier domain the source image by a set of log-Gabor filters located on four scales and four orientations. Note that every two scales are shifted in orientation for obtaining an hexagonal-like arrangement which covers more uniformly the Fourier domain [10]. An additional low-pass filter is used for transmitting the luminance and a high-pass filter allows encoding the highest frequencies [15] (all of those 18 filters are represented in space and Fourier domain in Fig. 1). For shortness, each filter output will be called a *channel*.

Filter coefficients below a fixed threshold are zeroed out for allowing an aliasing-free downsampling. Each filter is slightly corrected for obtaining exact reconstruction (the correction consists in dividing by the square sum of all the filters). A *sparse downsampling* strategy significantly reduces the transform domain size. The procedure is based on shifting the frequencies of each channel to the center of the Fourier domain, diminishing the Nyquist frequency [24]. The transform domain is arranged as a *pyramid* composed by channels of different sizes due to the downsampling operation (see Fig. 2 as an illustrative example). The inverse transform consists in summing up each channel filtered by the corresponding log-Gabor function (more details can be found in [6] and [7]).

This scheme aims at avoiding an excessive expansion of the transform domain for two main reasons: first, because sparse coding algorithms do not permit to recover all the redundancy, and, second, because, even if most of the coefficients will be finally zeroed out, such a large population of zeros still represents information for coding. An excessive expansion would then introduce irremovable redundancy. In practice, the expansion factor will be chosen around 10 what simultaneously permits limiting the expansion, avoiding any aliasing and preserving the smoothness of the filters.

III. Local Competition Algorithm

Because the transformation is overcomplete, many pyramids provide the same reconstructed image by the inverse transform. Therefore, overcompleteness is tantamount to transformation nonuniqueness. Being aware of the nonuniqueness, different iterative algorithms have already been proposed for searching for sparse representations [1], [2], [13], [14], [18] (reviews can be found in [2], [17], [22]). Drawbacks of these algorithms are high computational cost and some difficulties for finding the sparsest representations [2], [17]. Thus, they are not frequently used for image processing applications.

In the particular case of the overcomplete Gabor wavelets, it seems that an important part of the redundancy can be found between neighboring transform coefficients. The reason is that close Gabor functions are nonorthogonal and, therefore, they respond redundantly to the same image features.

For removing such redundancy and for obtaining a sparse representation, we propose here an iterative algorithm based on competitions between neighboring coefficients.

A. Nonorthogonality of Neighboring Gabor Functions

Let us consider an image $\mathbf{x}$ composed of a single log-Gabor function [Fig. 2(a)], which can be coded by a pyramid $\mathbf{h}$ containing a single nonzero coefficient [Fig. 2(b)]. Applying the linear log-Gabor wavelets transform on $\mathbf{x}$ does not provide the pyramid $\mathbf{h}$ but a pyramid $\mathbf{g}$ [drawn in Fig. 2(c)]. This pyramid $\mathbf{g}$ contains a large number of coefficients spread out in the spatial and frequency neighborhood of the original Gabor function. Both pyramids $\mathbf{h}$ and $\mathbf{g}$ can be used for exact reconstruction of $\mathbf{x}$ by the inverse transform. Nevertheless, $\mathbf{h}$ is a much sparser representation: After a quantization stage, some entropy calculations give 19 bits for $\mathbf{h}$ and 1697 bits for $\mathbf{g}$. The goal of the local competition algorithm is then to obtain the $\mathbf{h}$ pyramid.

Coefficients represented Fig. 2(c) can be considered as the projection of the coefficient Fig. 2(b) over the whole set of log-Gabor functions. Thus, from Fig. 2(c), it can be seen how the transform coefficients are not orthogonal to their spatial and frequency neighbors.

Note also that the linear transform does not optimally localize the image features: The single Gabor function is detected by a large set of coefficients of different scales, orientations, and positions. From the $\mathbf{g}$ pyramid, it is even difficult to say how many Gabor functions are present in the image (a). From the $\mathbf{h}$ pyramid, the answer is clearly one.

B. Algorithm Description

In order to obtain a sparse representation, the proposed algorithm implements a competition mechanism between neighboring coefficients by zeroing out the lowest coefficients and by concentrating the amplitude in the highest ones (previous versions of the algorithm appear in [6]–[8]). As in other sparse coding algorithms or in orthogonal transform-based image compression, the objective is to obtain as much zeros as possible among the transform coefficients [17], [13].

The algorithm is based on building at each iteration an exact reconstruction pyramid $\mathbf{h}_n$ with fewer nonzero coefficients. Each iteration is composed of the following steps.

1) **Linear transform.** The linear transform is computed as $\mathbf{g} = \mathbf{A}\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^N$ is the source image consisting of N pixels; $\mathbf{g} \in \mathbb{R}^M$ is a pyramid represented as a vector of M real values; the matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the direct transform; and $\mathbf{A}^\dagger \in \mathbb{R}^{N\times M}$ is the generalized inverse transform. Because of the transform is overcomplete $(M > N), \mathbf{A}$ is not strictly invertible and it is important to remark that $\mathbf{A}\mathbf{A}^\dagger$ is not the identity. Remember, nevertheless, that the linear transform provides exact reconstruction (see Section II), then $\mathbf{A}^\dagger\mathbf{A}$ is the identity and $\mathbf{A}^\dagger\mathbf{g} = \mathbf{x}$.

   The first iteration is initialized with $\mathbf{h}_1 = \mathbf{g}$.

2) **Coefficient selection.** Coefficients must satisfy two criteria in order to be selected. First, they must be local maxima along the normal to the filter direction. The second criterion consists in making the coefficients growing up to a threshold. Coefficients are selected when they pass the threshold θ, which is chosen as the maximum absolute value of $\mathbf{g}(\theta = \max|\mathbf{g}|)$. The growing is implemented through a sequence of pyramids called $(\mathbf{\Sigma}_n)_{n\in\mathbb{N}}$. $\mathbf{\Sigma}_1 = 0$ and $\mathbf{\Sigma}_n$ is increased at each iteration by adding $\mathbf{h}_{n-1}$ as follows (the scalar η_n will be further described)

$$\mathbf{\Sigma}_n = \mathbf{\Sigma}_{n-1} + \eta_{n-1} \cdot \mathbf{h}_{n-1}. \tag{2}$$

   $\mathbf{h}_n$ coefficients are selected when their corresponding $\mathbf{\Sigma}_n$ coefficients pass above a threshold θ in absolute value. Through that process, the largest $\mathbf{h}_n$ coefficients will be first selected, and at each step, some smaller coefficients will be added up to the selection.

   An *approximation pyramid* $\mathbf{a}_n$ is obtained by zeroing out all the nonselected coefficients of $\mathbf{h}_{n-1}$

$$\mathbf{a}_n(k) = \begin{cases} \mathbf{h}_{n-1}(k), & \text{if } |\mathbf{\Sigma}_n(k)| > \theta \\ 0, & \text{otherwise.} \end{cases}. \tag{3}$$

   Nonselected coefficients will be called as *residual pyramid* and denoted $\mathbf{r}_n$. We have

$$\mathbf{h}_{n-1} = \mathbf{a}_n + \mathbf{r}_n. \tag{4}$$

   $\eta_1 \in]0, 1[$ is proportional to the speed of convergence. A large η_1 implies a quick convergence and a small η_1 provides more refined results. The parameter η_1 is chosen as a fixed value (e.g., $\eta_1 = 0.02$) and η_n can be progressively increased through iterations for speeding up the convergence as, e.g., $\eta_n = \eta_1(\theta/\max|\mathbf{r}_n|)$, where $\max|\mathbf{r}_n|$ is the maximum absolute value of $\mathbf{r}_n$.

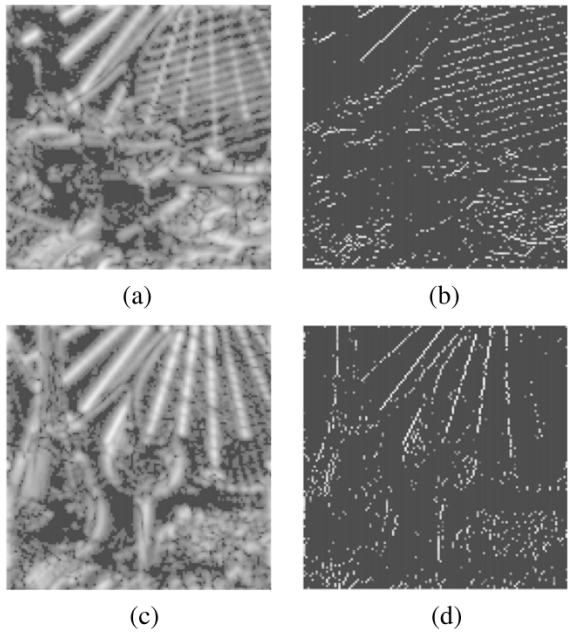

Fig. 3. Comparison between linear and local-competition pyramids of the ISO standard image "Bike" [the original image is represented Fig. 5(a)]. (a) Third -orientation, second-scale channel from the linear transform. This channel is sensitive to $(\pi/8)$ (east-northeast) edges (see the filter Fig. 1). (b) Same channel after 220 iterations of the local-competition algorithm. (c) Fourth-orientation, second-scale channel from the linear transform, this channel is sensitive to $(3\pi/8)$ (north-northeast) edges. (d) Same channel after the local competition algorithm. More details are given in the text.

3) **Coefficient adjustment.** A new pyramid $\mathbf{h}_n$ is defined as follows (for $\mathrm{n} > 1$)

$$\mathbf{h}_n = \mathbf{a}_n + \mathbf{A}\mathbf{A}^\dagger\mathbf{r}_n \tag{5}$$

   which can also be written as

$$\mathbf{h}_n = \mathbf{h}_{n-1} - \mathbf{r}_n + \mathbf{A}\mathbf{A}^\dagger\mathbf{r}_n. \tag{6}$$

   Because $\mathbf{A}^\dagger\mathbf{A}$ is the identity, it is straightforward to see that $\mathbf{h}_n$ has the exact reconstruction

$$\mathbf{A}^\dagger\mathbf{h}_n = \mathbf{A}^\dagger\mathbf{h}_{n-1} - \mathbf{A}^\dagger\mathbf{r}_n + \mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger\mathbf{r}_n \tag{7}$$

$$\mathbf{A}^\dagger\mathbf{h}_n = \mathbf{A}^\dagger\mathbf{h}_{n-1}. \tag{8}$$

   Because $\mathbf{h}_1 = \mathbf{g}$ and $\mathbf{A}^\dagger\mathbf{g} = \mathbf{x}$, we have $\mathbf{A}^\dagger\mathbf{h}_n = \mathbf{x}$ for any n. In summary, at each iteration n, the $\mathbf{h}_n$ pyramid has exact reconstruction.

4) **Iteration loop.** In the current implementation the number of iterations is fixed as, e.g., $n_{\max} = 5/\eta_1$. While $n \leq n_{\max}$, the algorithm is iterated by going back to step 2.

Fig. 3 compares two channels of the linear pyramid with the ones obtained after 220 iterations of the algorithm. In both linear channels high-amplitude coefficients spread around edges so as the channels look blurred. Moreover many of the edges are duplicately represented in both channels, e.g., vertical lines can be seen in the diagonal channel [Fig. 3(a)] and diagonal lines in the vertical channel [Fig. 3(c)]. After applying the local-competition algorithm [Figs. 3(b) and (d)], lines are thinner and enhanced what indicates better localization of the edges. Moreover, channels are more selective in orientation: fewer edges appear in both local-competition channels.

C. Computational Cost and Biological Plausibility

The algorithm computation only requires Fourier transforms and scalar operations using a limited number of iterations which is not dependent on the image size. Algorithm complexity is then $O(N\log(N))$ where N is the number of pixels of the input image. Other redundancy removal algorithms, like Matching Pursuit [14], are based on selecting only one coefficient per iteration. The present algorithm performs simultaneous competitions throughout all the image by selecting many coefficients at the same iteration. This causes a significant reduction on the computational complexity. That is, for selecting 22, 120 coefficients only 220 iterations are required (see Table I). Moreover, Gabor wavelets can be implemented as spatial convolutions with filters of limited support. In such case, the computational complexity would be $O(N)$. Such schemes would be particularly suitable to be implemented using a neural network approach.

We implemented the method as a biologically plausible model based on the following classical studies. Gabor functions are a model for the *simple cortical cells* [3]. Sparse coding has been described as a strategy employed by the primary visual cortex, in particular, at the level of simple cells [16], [18]. Local competitions could model the *masking effect*, which is responsible of visual stimuli diminishing when they are placed close to higher ones [12], [21]. Spatial and orientational competitions have been implemented by Grossberg *et al.* [11] in a neural network modeling *simple, complex, and hypercomplex cells*.

D. Some Additional Comments About the Algorithm

The present algorithm does not search for an orthogonal basis of decomposition; therefore, selected coefficients may be nonorthogonal. The orthogonal condition is relaxed since Gabor functions are particularly inappropriate to build orthogonal basis.

Each iteration modifies $\mathbf{h}_n$ following (6). That is, the residual $\mathbf{r}_n$ is removed from $\mathbf{h}_{n-1}$ while $\mathbf{A}\mathbf{A}^{\dagger}\mathbf{r}_n$ is added for compensating the induced error. It is straightforward to see that $\mathbf{r}_n$ and $\mathbf{A}\mathbf{A}^{\dagger}\mathbf{r}_n$ have the same reconstruction by $\mathbf{A}^{\dagger}$; therefore, the modification (6) does not impair the reconstruction.

$\mathbf{A}\mathbf{A}^{\dagger}\mathbf{r}_n$ is the projection of $\mathbf{r}_n$ over all the transform functions. Thus, (6) replaces part of the residual by its projection on the selected coefficients. This is why, at each iteration, a part of the amplitude migrates from nonselected to neighboring selected coefficients, making selected coefficient growing and nonselected decreasing. Such migration is achieved without forcing selected coefficients to receive the total amount of the amplitude, nor forcing nonselected ones to be zeros, as it would be done when searching for an orthogonal basis of decomposition.

It is outside the scope of the present paper to establish a proof of convergence. Note, nevertheless, that, experimentally, the algorithm always converged.

IV. Image Compression Scheme

A compression application based on the current scheme is described in the following. It contains three stages: the Gabor wavelets transform, the quantization, and the entropy calculation. The transform is of exact reconstruction; thus, the quantization is the only lossy part of the compression. The quantization stage is based on a *contrast sensitivity function* (CSF), which gives the sensitivity of the overall visual system at each frequency. The CSF proposed by Rust [20] has been implemented with a low-pass correction for avoiding the low-pass frequency decay of the CSF [4]. The CSF is indeed directly applied to the Fourier transform of the input image for giving all frequencies the same perceptual weight [23]. All coefficients are then quantized by the same nonlinear steps following the method described in [25].

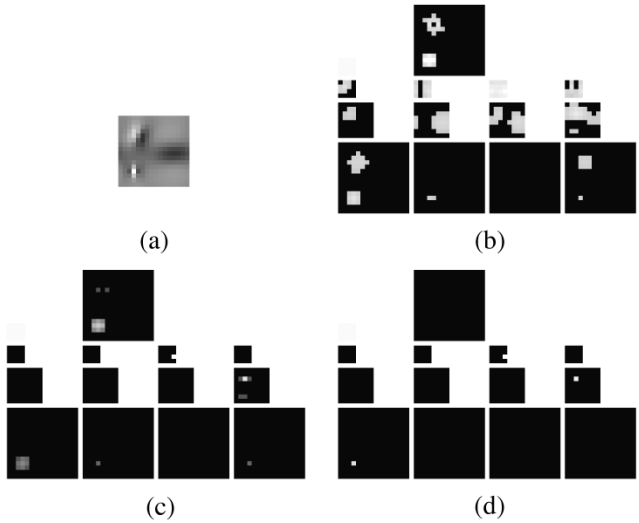

Fig. 4. Results of the algorithm for a basic example. (a). Original image (16×16 pixels) is composed of three Gabor functions: one in the first-scale (phase 0), one in the second-scale (phase $-(\pi/2)$), and one in the third-scale (phase π). The following insets shows the quantized pyramids. (b) Linear transform (before applying the first iteration) contains 197 nonzeros. Entropy is evaluated as 4.57 b.p.p., the root-mean square error (RMSE) of the reconstruction is $\mathrm{RMSE}=0.021$. (c) After 80 iterations, only 45 coefficients are kept, most of them close to zero. One second-scale and one third-scale coefficient appear enhanced. Entropy is 3.35 b.p.p., $\mathrm{RMSE}=0.026$. (d) After 140 iterations, the algorithm has been able to code the image by only 19 coefficients (including the 16 coefficients of the low-pass channel): It has localized the three original Gabor functions. Entropy is 0.39 b.p.p., $\mathrm{RMSE}=0.016$.

For evaluating the compression rate, an entropy calculation is performed as

$$H=-\frac{M}{N}\sum_{q}p(q)\log_2 p(q) \tag{9}$$

where q is the quantization level, $p(q)$ is its probability, N is the number of pixels of the image, and M is the number of coefficients of the transform domain. The factor $(M/N)>1$ is due to the transform domain expansion. (Note that, because every scale has different statistics, the entropy of each scale is separately calculated before to be summed up; for the low-pass filter, a basic DPCM is applied). Entropy is a measurement of the amount of information contained in the source to be coded. The difference of entropy between the linear and the local competition pyramids can then be considered as a direct evaluation of the redundancy reduction (with the condition that reconstruction qualities are similar). Moreover, the entropy gives the theoretical limit (in bit per pixels, b.p.p.) of the compression rate and it is a good evaluation measure, although somewhat optimistic, of the file size that can be produced by an efficient coder.

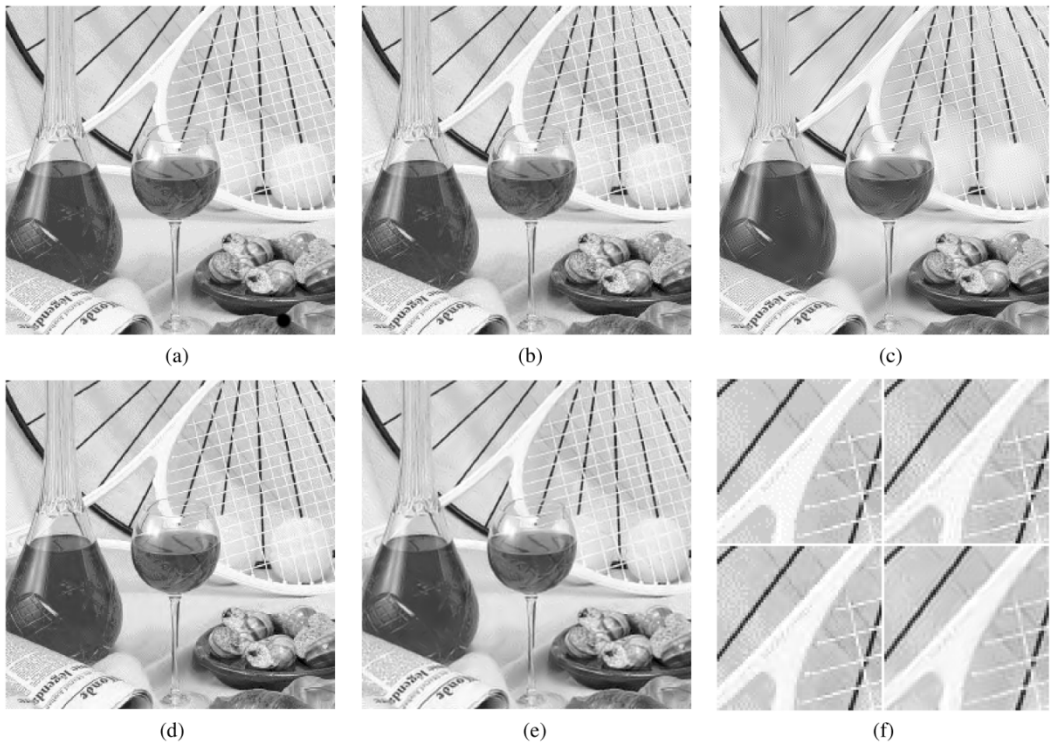

Fig. 5. Compression results through the different methods. Every image is compressed at 2.08 b.p.p. excepted the image (c) at 7.39 b.p.p. (results are summarized Table I). (a). The original image is a 256×256 pixel detail of the ISO standard image "Bike." (b) The compression by JPEG yields to a $\mathrm{RMSE} = .028$. (c) Compression through log-Gabor linear transform at 7.39 b.p.p. $(\mathrm{RMSE} = .031)$. (d) Compression by JPEG-2000 $(\mathrm{RMSE} = .017)$. (e) Compression by the local-competition algorithm $(\mathrm{RMSE} = .033)$. Although the mathematical error is higher than with the linear transform, the perceptual quality improves so as get close to the quality offered by JPEG, in particular, edges are better preserved than with the linear transform. (f) A detail of the images (a), (b), (d), and (e) is zoomed by a factor of 2 (zoomed images are arranged in the same manner as images (a), (b), (d), and (e): upper-left is the original; upper-right is JPEG; lower-left is JPEG-2000; lower-right is local competition).

The quality of the reconstructed image is mathematically evaluated using RMSE. However, it is well known that such mathematical error calculation does not correlate well with the perceptual quality evaluated by human observers. Therefore, direct observation will provide a better insight about the performance of the scheme.

V. Results

Some preliminarily compression results are here presented as they offer a simple and direct evaluation of the method. Nevertheless, sparse Gabor wavelets would be of interest not only for image compression, but also for image analysis since they could provide an optimal localization and a nonredundant representation of the features.

Fig. 4 shows an example for a basic image containing only three Gabor functions. The algorithm is able to localize these Gabor functions and to reduce the entropy from 4.57 b.p.p. to 0.39 b.p.p. In the meantime, the RMSE stays stable and even decreases from 0.021 to 0.016. It is to stress that in the linear transform image details can never be represented as single isolated Gabor functions, but always as mixture of them. The optimal localization can only be achieved by a sparse algorithm (see Fig. 4).

In the following, the algorithm is applied to 256×256 pixels images ($N = 65\,536$ pixels) with 8-bit resolution. The transform domain contains $M = 566\,272$ real coefficients, the expansion factor being then $(M/N) = 8.64$. It is worth noting that the reconstruction is exact if the quantization stage is not included.

TABLE I
Compression Results for the Image "Bike" (See Also Figs. 5 and 6)

b.p.p.	Compression scheme	non-zero coef.	RMSE	Comments
-	both linear and local comp.	566272	0	exact reconstruction (without quantization)
7.39	linear Gabor w.	111839	.031	some blurring on edges
2.08	JPEG	-	.028	good perceptual quality
	JPEG-2000	-	.017	very good perceptual q.
	linear Gabor w.	26253	.089	high level of blurring
	local comp.	22120	.033	good perceptual quality
0.57	JPEG	-	.060	ringing and block artifacts
	JPEG-2000	-	.046	ringing
	linear Gabor w.	6619	.091	very high level of blurring
	local comp.	6709	.056	blurring

For the 256×256 detail of the ISO standard image "Bike" (see Fig. 5 and Table I), the linear transform needs 111 839 nonzero coefficients if we want a good quality of reconstruction. The competition algorithm with 220 iterations provides a

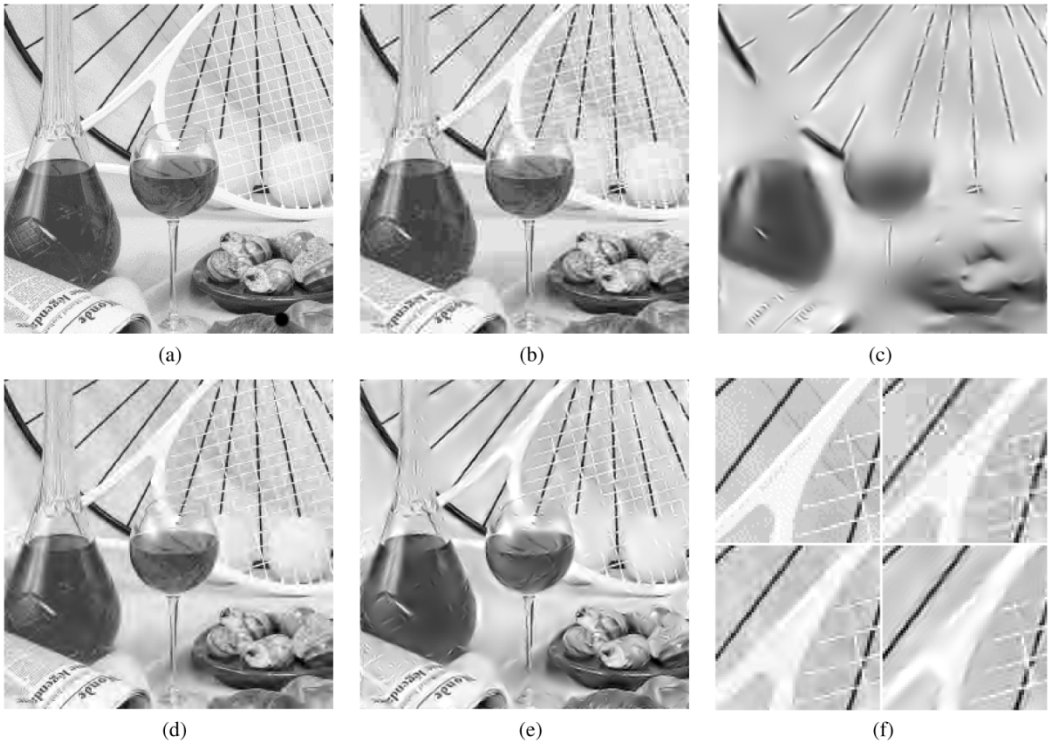

Fig. 6. Results at high compression rates: all images are compressed at 0.57 bpp. (a). Original image "Bike." (b) In the image compressed by JPEG, high-frequency artifacts are observed, in particular, ringing and block artifacts $(\mathrm{RMSE} = .060)$. (c) Compression through linear transform $(\mathrm{RMSE} = .105)$. (d) In the image compressed by JPEG-2000, ringing artifacts can be observed $(\mathrm{RMSE} = .046)$. (e) Compression through local competition. Artifacts consist principally of blurring and the quality is close or even better than the JPEG one $(\mathrm{RMSE} = .056)$. Note the mathematical error is lower than with JPEG (see also results Table I). (f) Some details are zoomed in the same manner as in Fig. 5.

very similar reconstruction quality with only 22 927 nonzero coefficients (a 80% reduction factor). Entropy decreases then from 7.39 b.p.p. to 2.08 b.p.p. (72% of reduction).

It is remarkable that in the transform domain edges are strongly enhanced and at the same time their localizations are refined. This is reflected on the fact that the magnitude of coefficients located along the edges are weighted by a factor up to 10 after finishing the competition algorithm (see the coefficients represented in the transform domain Fig. 3). Such edge enhancement also implies that, for the same quantization steps, the local competition algorithm provides much higher quality than the linear transform.

At higher compression rates (Fig. 6), the smooth shape of the filters together with the lack of aliasing produces a lower level of artifacts, appearing mainly blurring as the most significant artifact. There is low ringing and minor presence of high-frequency artifacts. One can argue that the proximity to biological models also helps to make the distortions less visible.

In comparison with the classical linear transform, the local competition method allows an important reduction of the redundancy without quality losses. At the current stage of implementation, the method does not outperform the JPEG-2000 compression standard, but results seem to be close to the ones offered by the former JPEG DCT-based having the additional bonus of strongly limiting the level of high-frequency artifacts.

Similar results have been obtained with classical images as "Goldhill" (Table II and Fig. 7). The algorithm should be evaluated through a larger set of images to validate the results. Nevertheless, as the main purpose of this study is to establish the feasibility of the method, such statistical evaluation of the algorithm remains out of the current scope.

TABLE II
COMPRESSION RESULTS FOR THE IMAGE "GOLDHILL" (SEE ALSO FIG. 7)

b.p.p.	Scheme	Coef.	RMSE	Comments
1.55	JPEG	-	.020	good perceptual quality
	JPEG-2000	-	.014	very good perceptual q.
	linear Gabor w.	20157	.071	high level of blurring
	local c. 194 iter.	18843	.022	good perceptual quality

Gabor wavelets combined with the local competition algorithm gather a number of interesting properties which can be summarized as follows. First, the transformation is *nonlinear*, invertible and allows an *exact reconstruction*. Second, the inverse transform is just performed by the Gabor wavelet linear inverse transform, what makes it very fast. Third, the *information contents* and the *redundancy* are highly reduced. Fourth, *visual events (edges) are enhanced, sharper and thinner localized* in the transform domain. It can be argued that such local competition response presents a number of similarities with the *masking effect*. Finally, the computational complexity $O(N \cdot \log(N))$ is reasonably low and can be even reduced to $O(N)$ if Gabor wavelets are implemented in the space domain through filters of limited support.

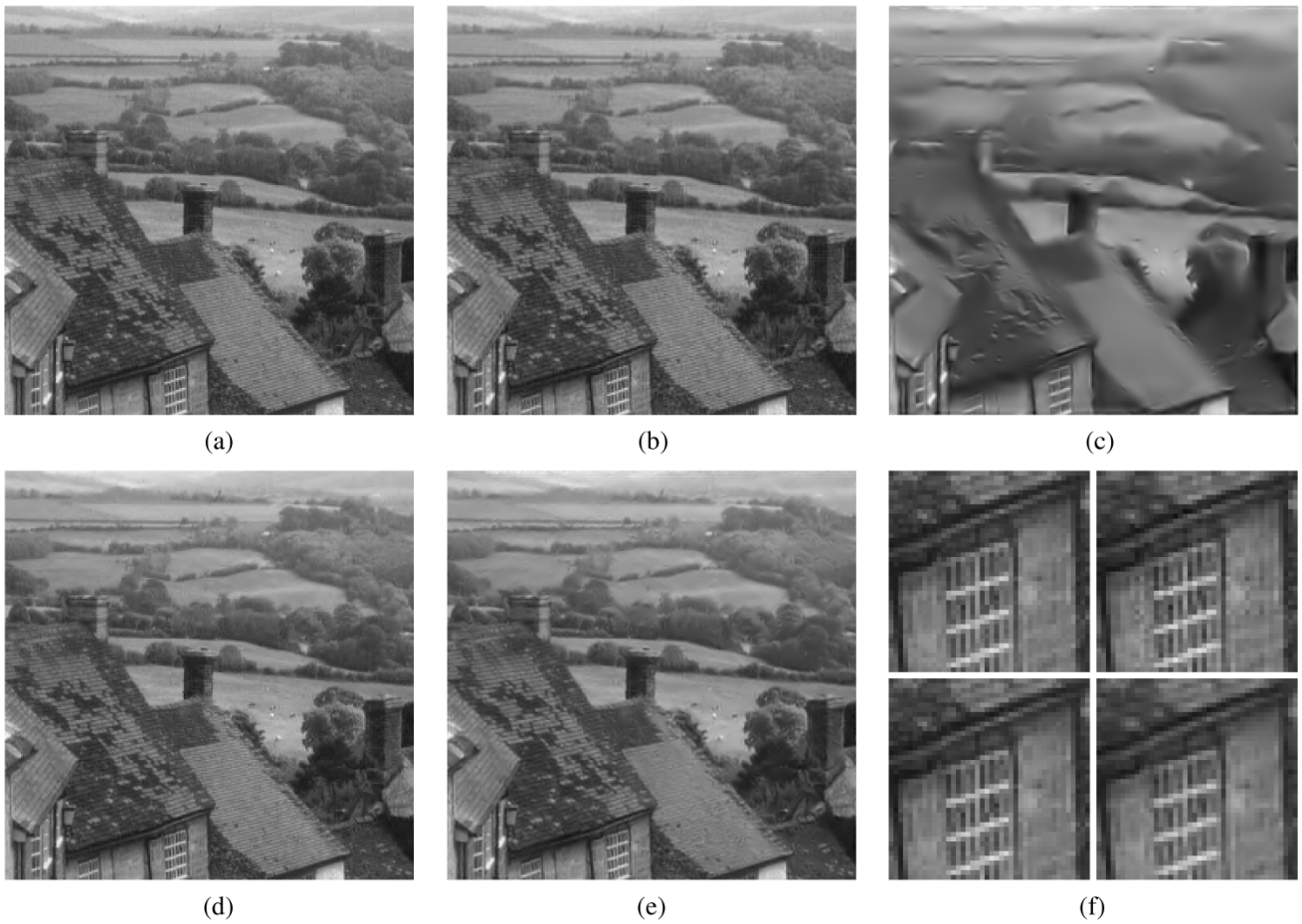

Fig. 7. Comparison between compression methods. Every image is here compressed at 1.55 b.p.p. (see also Table II). (a). Original image "Goldhill" (detail). (b) Compression by JPEG ($\mathrm{RMSE} = .020$). (c) Compression by log-Gabor linear transform ($\mathrm{RMSE} = .071$). (d) Compression by JPEG-2000 ($\mathrm{RMSE} = .014$). (e) Image compressed by log-Gabor local competition transform ($\mathrm{RMSE} = 0.022$). (f) Zoom of (a), (b), (d), and (e) image details. Arrangement is the same as in Fig. 5.

VI. Conclusions and Further Work

The local-competition Gabor wavelets transformation described in this paper is composed by a linear Gabor wavelets followed by a nonlinear iterative algorithm consisting of local-competitions between coefficients which aim is reducing the amount of information to be encoded. The resulting transform is aliasing-free and provides exact reconstruction by the linear inverse transform. Although the transformation begins with an expansion of the number of coefficients by almost a factor of 10, the proposed algorithm allows to reduce the redundancy for achieving compression rates similar to those given by the JPEG-DCT compression standard. Due to the low level of artifacts, the good properties of Gabor functions and the proximity of the method to Human Visual System models, the perceptual quality is particularly preserved. Because the amplitude is concentrated in a reduced number of coefficients, edges appear thinner localized and enhanced in the transform domain. It has also been seen that, in opposition to the linear transform, the local competition algorithm allows for achieving the theoretical optimal limit of localization (being able to represent features by single isolated Gabor coefficients). It is interesting to remark that the competition between coefficients can be interpreted as a visual masking phenomena. The enhancement of selected coefficients is counteracted by decreasing the amplitude of their neighbors, what would reveal a masking effect.

Although compression results are still below JPEG-2000 performances, further investigation can be pursued first for image analysis, because the transform could permit to achieve the unique properties of optimal localization and nonredundant Gabor representation, and second for image compression where the low level of artifacts could make the method appropriate for those applications where the highest perceptual fidelity is required as for example in medical imaging. On the other hand, it is expected to obtain significantly higher compression rates through the improvement of the coefficient selection strategy. In particular, additional criteria, such as competitions between orientations or cooperations between spatially aligned Gabor functions, could be implemented (both are described in [11] as "orientational competition" and "long-range cooperation"). The long-range cooperation would have the additional advantage to locate selected coefficients along continuous curves. Such curves can be predictively encoded, e.g., by chain coding [19], which will increase, even more, the compression rate.

ACKNOWLEDGMENT

The authors would like to thank M. Keil, L. Perrinet, and B. Beferull-Lozano for the useful discussions during the development of this work.

REFERENCES

[1] F. Bergeaud and S. Mallat, "Matching pursuit of images," in *Proc. IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, 1994, pp. 330–333.
[2] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," in *SIAM Rev.*, vol. 43, 2001, pp. 129–159.
[3] J. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vis. Res.*, vol. 20, pp. 847–856, 1980.
[4] M. Eckert and A. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Process.*, vol. 70, pp. 177–200, 1998.
[5] D. Field, "Relation between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2394, 1987.
[6] S. Fischer, "Modelos Perceptuales en la Cuantificación y Codificación de Imágenes," M.S. thesis, Escuela Téc. Sup. Ing. Telecom., Univ. Politécnica Madrid, Madrid, Spain, 2000.
[7] S. Fischer and G. Cristóbal, "Minimum entropy transform using Gabor wavelets for image compression," in *Proc. Int. Conf. Image Analysis*, 2001, pp. 428–433.
[8] S. Fischer, G. Cristóbal, and R. Redondo, "Sparse overcomplete Gabor wavelets based on local competitions: A model for simple cells?," presented at the ECOVISION Workshop, May 2004.
[9] D. Gabor, "Theory of communication," *J. Inst. Elect. Eng.*, vol. 93, pp. 429–457, 1946.
[10] M. Gross and R. Koch, "Visualization of multidimensional shape and texture features in laser range data using complex-valued gabor wavelets," *IEEE Trans. Vis. Comput. Graph.*, vol. 1, no. 1, pp. 44–59, 1995.
[11] S. Grossberg, E. Mingolla, and J. Williamson, "Synthetic aperture radar processing by a multiple scale neural system for boundary and surface representation," *Neural Netw.*, vol. 8, no. 7-8, pp. 1005–1028, 1995.
[12] G. Legge and J. Foley, "Constrast masking in human vision," *J. Opt. Soc. Amer. A*, vol. 12, pp. 1458–1471, 1980.
[13] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.
[14] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
[15] O. Nestares, R. Navarro, J. Portilla, and A. Tabernero, "Efficient spatial-domain implementation of a multiscale image representation based on gabor functions," *J. Elect. Imag.*, vol. 7, no. 1, pp. 166–173, 1998.
[16] B. Olshausen and D. Field, "Sparse coding with an overcomplete basis set: A strategy employed by VI?," *Vis. Res.*, vol. 37, pp. 3311–3325, 1997.
[17] A. Pece, "The problem of sparse image coding," *J. Math. Imag. Vis.*, vol. 17, no. 2, pp. 89–108, 2002.
[18] L. Perrinet, M. Samuelides, and S. Thorpe, "Sparse spike coding in an asynchronous feed-forward multi-layer neural network using matching pursuit," *Neurocomput.*, vol. 57, pp. 125–134, 2004.
[19] R. Redondo and G. Cristóbal, "Lossless chain coder for gray edge image," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2, 2003, pp. 201–204.
[20] B. Rust and H. Rushmeier, "A new representation of the contrast sensitivity function for human vision," in *Proc. Int. Conf. Image, Science System, Technology*, 1997, pp. 1–15.
[21] C. Stromeyer and B. Julesz, "Spatial-frequency masking in vision: Critical bands and spread of masking," *J. Opt. Soc. Amer. A*, vol. 62, no. 10, pp. 1221–1232, 1972.
[22] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[23] A. Watson and J. Solomon, "Model of visual contrast gain control and pattern masking," *J. Opt. Soc. Amer. A*, vol. 14, no. 9, pp. 2379–2391, 1997.
[24] R. P. Wurtz, "Multilayer Dynamic Link Networks for Establishing Image Point Correspondences and Visual Object Recognition," Ph.D. dissertation, Phys. Dept., Rurh-Univ., Bochum, Germany, 1994.
[25] W. Zeng, S. Daly, and L. Shawmin, "Point-wise extended visual masking for JPEG-2000 image compression," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2000, pp. 657–660.

Sylvain Fischer received the degrees from the Ecole National Supérieure des Télécommunications, Paris, France, and the Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain, in 2000. He is currently pursuing the Ph.D. degree at the Instituto de Óptica (CSIC), Madrid.

His areas of interest include vision modeling, sparse coding, and perceptual contour extraction.

Gabriel Cristóbal received the M.Sc. and Ph.D. degrees in telecommunication engineering from the Universidad Politecnica of Madrid, Madrid, Spain, in 1979 and 1986, respectively.

He is currently a Research Scientist with the Instituto de Optica, Madrid. His current research interest are in joint representations, vision modeling, and image compression.

Rafael Redondo received the degree from the Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain, in 2002. He is currently pursuing the Ph.D. degree at the Instituto de Óptica (CSIC), Madrid, where he works on vision and image processing.

Among his research fields are vision modeling, image and volumetric coding, and time-frequency representations applied to pattern recognition and image fusion.