

Research Article

Sparse Approximation of Images Inspired from the Functional Architecture of the Primary Visual Areas

Sylvain Fischer,^{1,2} Rafael Redondo,¹ Laurent Perrinet,² and Gabriel Cristóbal¹

¹Instituto de Óptica - CSIC, Serrano 121, 28006 Madrid, Spain

²INCM, UMR 6193, CNRS and Aix-Marseille University, 31 chemin Joseph Aiguier, 13402 Marseille Cedex 20, France

Received 1 December 2005; Revised 7 September 2006; Accepted 18 September 2006

Recommended by Javier Portilla

Several drawbacks of critically sampled wavelets can be solved by overcomplete multiresolution transforms and sparse approximation algorithms. Facing the difficulty to optimize such nonorthogonal and nonlinear transforms, we implement a sparse approximation scheme inspired from the functional architecture of the primary visual cortex. The scheme models simple and complex cell receptive fields through log-Gabor wavelets. The model also incorporates inhibition and facilitation interactions between neighboring cells. Functionally these interactions allow to extract edges and ridges, providing an edge-based approximation of the visual information. The edge coefficients are shown sufficient for closely reconstructing the images, while contour representations by means of chains of edges reduce the information redundancy for approaching image compression. Additionally, the ability to segregate the edges from the noise is employed for image restoration.

Copyright © 2007 Sylvain Fischer et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Recent works on multiresolution transforms showed the necessity of using overcomplete transformations to solve drawbacks of (bi-)orthogonal wavelets, namely their lack of shift invariance, the aliasing between subbands, their poor resolution in orientation and their insufficient match with image features [1–4]. Nevertheless the representations from linear overcomplete transforms are highly redundant and consequently inefficient for such tasks needing sparseness as, for example, for image compression. Several *sparse approximation* algorithms have been proposed to address this problem by approximating the images through a reduced number of decomposition functions chosen in an overcomplete set called *dictionary* [5–8] (see reviews in [6, 9]). In some very particular cases there exist algorithms achieving the optimal solutions. In the general case, two main classes of algorithms are available: matching pursuit (MP) [5, 10] which recursively chooses the most relevant coefficients in all the dictionary and basis pursuit (BP) [6] which minimizes a penalizing function corresponding to the sum of the amplitude of all coefficients. Both these algorithms perform iteratively and globally through all the dictionary. They are computationally costly algorithms which generally only achieve approximations of the optimal solutions.

We propose here to build a new method for sparse approximation of natural images based both on classical image processing criteria and on the known physiology of the primary visual cortex (V1) of primates. The rationale behind the biological modeling is the plausibility that V1 could accomplish an efficient coding of the visual information and a certain number of similarities between V1 architecture and recent image processing algorithms: first, the receptive field (RF) of V1 simple cells can be modeled through oriented Gabor-like functions [11], arranged in a multiscale structure [12], similarly to the Gabor-like multiresolutions. Second, V1 supposedly carries out a sparse approximation procedure [13]. And finally, interactions between V1 cells such as inhibitions between neighboring cells and facilitation between coaligned and collinear cells have been described by physiological and psychophysical studies [14–16]. These interactions have been shown efficient for image processing in applications such as contour extraction and image restoration [17–21]. We propose here the hypothesis that lateral interactions deal not only with contour extraction or noise segregation but also allow to achieve sparse approximations of natural images.

The present model is also based on previous image processing works on denoising, edge extraction, and compression. Denoising by wavelet thresholding is nowadays a

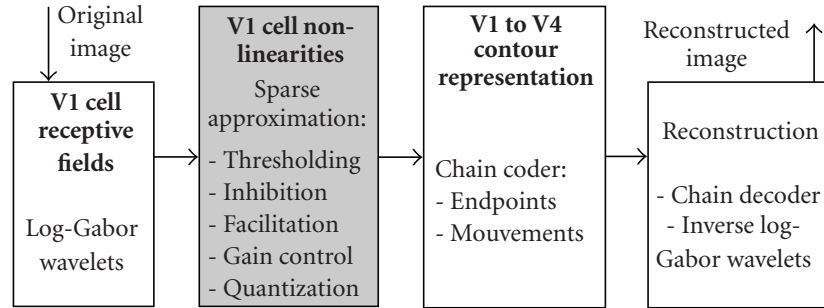


FIGURE 1: Scheme of the algorithm. The lossy parts, that is, the operations inducing information losses, are depicted with gray color.

popular method, and it was shown that overcomplete transforms which preserve the translation invariant property are more efficient than (bi-)orthogonal wavelets [1, 22]. An augmented resolution in orientation was also shown to be important [4], as well as a better match between edges of natural images and the wavelet shape [4]. According to such studies we previously proposed log-Gabor wavelets as a candidate for an efficient noise segregation [23, 24]. Denoising was also shown to be improved by taking into account the adjacent neighborhood of transform coefficients [25] or thanks to inhibition/facilitation interactions [17]. Denoising is also known to be linked with compression, where (bi-)orthogonal wavelets are the golden standard with JPEG-2000. A compression based on edge extraction was proposed by Mallat and Zhang [26], while the possibility to reconstruct images from their edges was studied in [27]. Several authors proposed a separated coding of edges and residual textures generally by means of sparse approximation algorithms [28–30]. Various usual and popular edge extraction methods proceed through a first step of filtering through oriented kernels before applying an oriented inhibition or nonlocal maxima suppression and some hysteresis or facilitation processes to reinforce coaligned edge segments [17, 19, 20, 31].

We propose here a unified algorithm for denoising, edge extraction, and image compression based on a new sparse approximation strategy for natural images. The second objective of this study is to approach visual cortex understanding and image processing. From the image processing point of view, one important novelty consists in achieving denoising and sparse approximation based on multiscale edge extraction. From the mathematical point of view, the selection of the sparse subdictionary through local operations and in a noniterative manner is an important novelty. Compared with our previous work implementing oriented inhibition on log-Gabor wavelets [8], the improvements consist here in the implementation of facilitative interactions and in proposing a further redundancy reduction through a contour encoding. From the neuroscience point of view, the model aims at reproducing some of the behaviors observed in the visual cortex and to fix the unknown parameters thanks to image processing criteria (this last optimization takes sense since we consider the visual cortex as an efficient visual processing system optimized under evolutionary pressure). It proposes

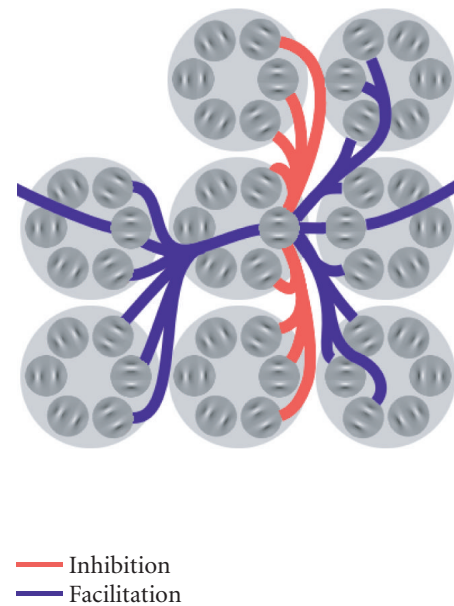


FIGURE 2: Schematic structure of the primary visual cortex implemented in the present study. Simple cortical cells are modeled through log-Gabor functions. They are organized in pairs in quadrature of phase (dark-gray circles). For each position the set of different orientations compose a pinwheel (large light-gray circles). The retinotopic organization induces that adjacent spatial positions are arranged in adjacent pinwheels. Inhibition interactions occur towards the closest adjacent positions which are in the directions perpendicular to the cell preferred orientation and toward adjacent orientations (light-red connections). Facilitation occurs towards coaligned cells up to a larger distance (dark-blue connections).

a computational hypothesis about how the primary visual areas could achieve a noise robust sparse approximation of the visual information under the form of edges and contours.

The paper is structured as follows: Section 2 describes the model implementation. Section 3 presents the results on edge extraction, image compression, and denoising in comparison with state-of-the-art image processing algorithms. Conclusions are drawn in Section 4.

TABLE 1: Correspondences between visual cortex physiology and image processing operations defined in the different sections.

Visual cortex structures	Image processing	Section
Simple and complex cells	log-Gabor fcts.	Section 2.1
Even-sym. simple cell	$\Re(\mathbf{h}(x, y, s, r))$	
Odd-sym. simple cell	$\Im(\mathbf{h}(x, y, s, r))$	
Pair of simple cells	$\mathbf{h}(x, y, s, r)$	
Complex cell	$ \mathbf{h} (x, y, s, r)$	
Pinwheel	$\mathbf{h}(x, y, s, \cdot)$	
Retinotopic organization	x, y arrangement	
Spike threshold	CSF (\mathbf{h}_2)	Section 2.2
Oriented inhibition	Edges (\mathbf{h}_3)	Section 2.3
Facilitation across scales	Parents (\mathbf{f}_1)	Section 2.4
Facilitation across space	Chain length (\mathbf{f}_2)	Section 2.5
Set of spiking cells	Subdictionary \mathbf{h}_4	Section 2.5
Gain control	Amplitude (a_k)	Section 2.6
Hypercomplex cells	Endpoints	Section 2.7
Contour shape	Movements	
Contour representation	Chain coding	

2. MODEL IMPLEMENTATION

The present study proposes a novel sparse approximation strategy which can at the same time be interpreted as a model of the primary visual areas. The model summarized in Figures 1, 2, and Table 1 also incorporates a contour representation and a reconstruction module. It is composed by successive steps which analyze and integrate the visual information from local features to increasing larger ones. First, simple cell and complex cell receptive fields are modeled by log-Gabor functions as described in Section 2.1. Then nonlinear behaviors of V1 cells such as spike thresholding (Section 2.2), inhibition (Section 2.3), facilitation (Sections 2.4 and 2.5), gain control (Section 2.6) are implemented. Finally a contour representation is proposed in Section 2.7.

2.1. Simple and complex cell receptive fields

The first step of the implementation consists in modeling the receptive fields of the simple cell population through the log-Gabor wavelet transform W which has been proposed in our previous studies [8, 23, 24]. The transform consists in filtering the given input image \mathbf{x} by a set of log-Gabor kernels ($\mathbf{G}_{(s,r)}(s,r)$ where s is the scale which ranges from 1 to 5 for edge extraction and denoising (and from 1 to 6 for compression) and r indexes the orientations ranging from 1 to 6. The scheme also includes a residual low-pass filter. All those kernels are shown in Figure 3 for the 5 scales, 6 orientation case. Each filter output is called a *channel*. It represents the response of a set of cells having a particular orientation and scale and covering the full range of positions (eventually decimated for the coarsest scales). The transform coefficients are organized in 4-dimensional arrays, called *pyra-*

mids, $\mathbf{h}(x, y, s, r)$ where x, y, s, r denote the position in x , in y , the scale, and the orientation, respectively. \mathbf{h} coefficients are complex-valued, the real parts $\Re(\mathbf{h})$ correspond to the receptive fields (RF) of even-symmetric simple cells (i.e., with cosine shape) as shown in Figure 3(b). The imaginary parts $\Im(\mathbf{h})$ correspond to odd-symmetric (i.e., sine shape) RF shown in Figure 3(c). Hence, each coefficient represents the amplitude of a pair of simple cells in quadrature of phase localized in the same position, orientation, and scale (illustrated as dark-gray discs in Figure 2). The activities of simple cells are then calculated as (where \otimes is the 2D convolution in x, y)

$$\mathbf{h}(x, y, s, r) = \mathbf{G}_{(s,r)}(x, y) \otimes \mathbf{x}(x, y). \quad (1)$$

The activities $|\mathbf{h}|$ of the *complex cells* are defined as the square quadratic sum of the pairs of simple cells $\Re(\mathbf{h})$ and $\Im(\mathbf{h})$, that is, the modulus of the log-Gabor wavelet coefficients \mathbf{h} . Such definition is consistent with previous models [19, 32].

The log-Gabor wavelets are not described in details here, for a thorough study including justifications of their biological plausibility please refer to [8, 23, 24]. Nevertheless it is worth stressing here some important characteristics of the log-Gabor wavelets. (1) The transform is linear and is translation invariant. It allows exact reconstruction and is self-invertible (it is a tight frame): the pseudoinverse is also the transposed operator noted W^T and $W W^T \mathbf{x} = \mathbf{x}$ for any image \mathbf{x} . (2) It is overcomplete by a factor R around $(14/3)n_t$ where n_t is the number of orientations (i.e., $R \simeq 28$ for 6 orientations). Such an overcompleteness factor R is consistent with the redundant number of simple cells in comparison with the number of photoreceptors in the retina. It is also acceptable for sparse approximation algorithms which currently deal with much more redundant transforms (see, e.g., [28]). (3) The elongated shape and the phase, scale, and orientation arrangement of the filters properly model the receptive fields present in the V1 simple cell population.

2.2. Spike threshold

Those complex cells whose activities do not reach a certain spike rectification threshold are considered as inactive. The contrast sensitivity function (CSF) proposed in [33] is implemented here to model this thresholding. $\text{CSF}(s, r)$ establishes the threshold of detection for each channel (s, r) , that is, the minimum amplitude for a coefficient to be visible for a human observer. All the nonperceptible coefficients are then zeroed out.

In presence of noise, the CSF is known to modify its response to filter down the highest frequencies (see [34] for a model of such behavior). This change in the CSF is modeled here by lowering the spike threshold depending on the noise level. The new threshold level is determined according to classical image processing methodologies for removing noise: the noise variance $\sigma_{(s,r)}^2$ induced in each channel (s, r) is evaluated following the method proposed in [25] (if the noise variance in the source image is not known, it is evaluated as in [35]). The spike threshold is set up experimentally to $1.85\sigma_{(s,r)}$. This threshold allows to eliminate most of the

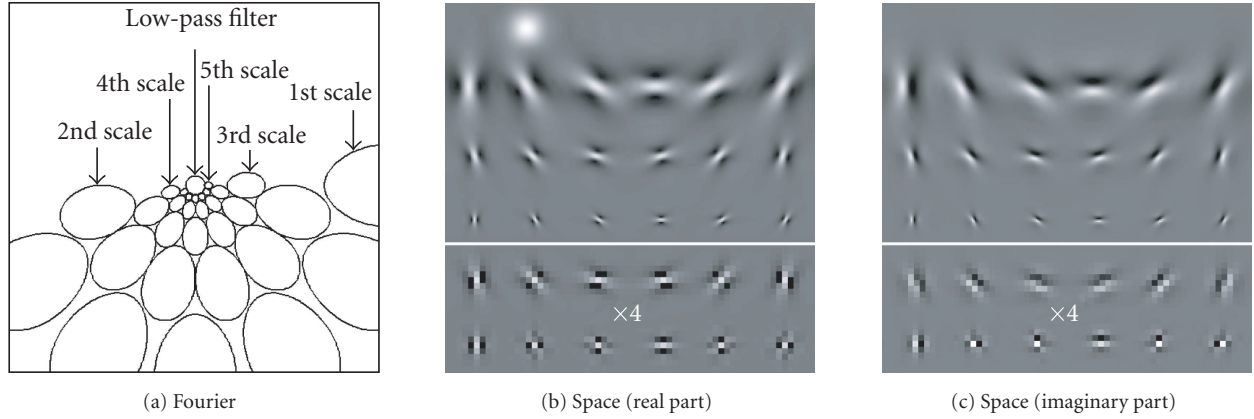


FIGURE 3: Multiresolution scheme with 6 orientations and 5 scales. (a) Schematic contours of the filters in the Fourier domain. The Fourier domain origin (DC component) is located at the center of the inset and the highest frequencies lie on the border. (b) Real part of the filters in the space domain. Scales are arranged in rows and orientations in columns. The two first scales are drawn at the bottom magnified by a factor of 4 for a better visualization. The low-pass filter is drawn in the upper-left part. (c) The imaginary part of the filters is shown in the same arrangement. The low-pass filter does not have an imaginary part.

apparent noise apart from a few residual noise features. This threshold is set to a low value so as to preserve a larger part of the signal while the processes of facilitation (Sections 2.4 and 2.5) will refine the denoising by removing the residual artifacts. The activities of simple cells after spike thresholding are calculated as \mathbf{h}_2 :

$$\mathbf{h}_2(x, y, s, r) = \begin{cases} \mathbf{h}(x, y, s, r) & \text{if } |\mathbf{h}|(x, y, s, r) \geq \max(\text{CSF}(s, r), 1.85\sigma_{(s,r)}), \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

2.3. Oriented inhibition

The inhibition step is designed according to energy models [19, 32] which implement nonlocal maxima suppression between complex cells for extracting edges and ridges. A very similar strategy is also deployed in classical image processing edge extraction methods like in the Canny operator [31] which marks edges at local maxima after the filtering through oriented kernels. As indicated by the light-gray connections in Figure 2 the inhibition occurs toward the direction perpendicular to the edge, that is to the filter orientation. It zeroes out the closest adjacent orientations and positions which have lower activity (no inhibition across scales is implemented here). The implementation of the oriented inhibition is not detailed more here since it does not differ substantially from the classical implementations proposed in [19, 31]. The inhibition operation can be summarized by the following equation (where (v_x, v_y) points to an adjacent pixel in the direction perpendicular to the channel preferred

orientation):

$$\mathbf{h}_3(x, y, s, r) = \begin{cases} \mathbf{h}_2(x, y, s, r) & \text{if } |\mathbf{h}_2|(x, y, s, r) \\ & \geq \max_{(\delta_v, \delta_r) \in \{-1, 0, 1\}^2} \{ |\mathbf{h}_2|(x + \delta_v v_x, y + \delta_v v_y, s, r + \delta_r) \}, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

It is worth to note that the shape of the filter is critical here for an accurately localized, nonredundant and noise-robust detection [31]. Figure 4 illustrates that log-Gabor filters are adequate for extracting both edges and ridges by non-local maxima suppression: (1) both edges and ridges induce local-maxima in the modulus of the log-Gabor coefficients and (2) that the modulus monotonously decreases on both sides of edges and ridges without creating extra local-maxima (the modulus response is monomodal).

After inhibition is performed, most coefficients are set to zero and the remaining coefficients already show a strong similitude with the multiscale edges and ridges perceived by visual inspection (see Figure 5(c)). It is remarkable moreover that coefficients appear in *chains*, that is in clusters of coefficients lying within a single scale which are adjacent in position and eventually in orientation. Those chains closely follow the contours perceived by visual inspection of the image. Moreover they appear mainly continuous, while only a few

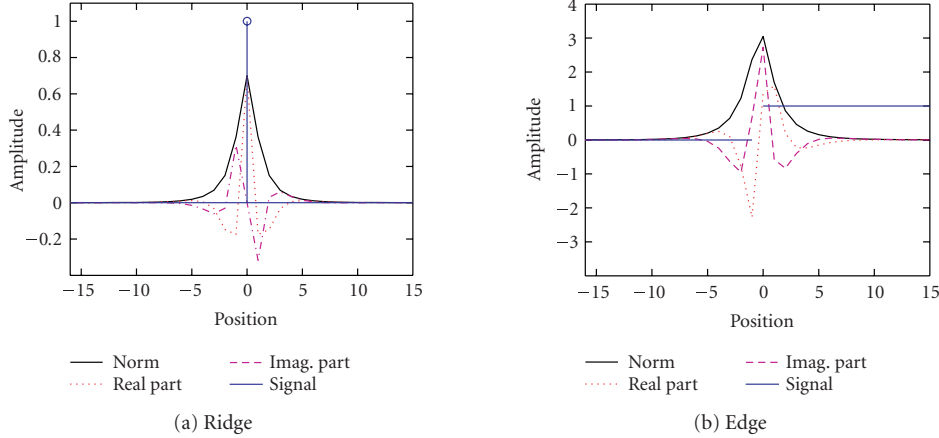


FIGURE 4: Log-Gabor wavelet response to edges and ridges. (a) Response of a 1D complex log-Gabor filter to an impulse (ridge): the modulus (black continuous curve) of the response monotonously decreases away from the impulse. It implies that the ridge is situated just on the local maximum of the response. On the contrary the real (dot) and imaginary (dash-dot) parts present various local-maxima and minima which makes them less suitable for ridge localization. (b) Same curves for a step edge.

gaps are cutting off the contours. Some isolated nonzero coefficients also remain due to noise as well as irrelevant or less salient edges. Facilitation interactions will now allow to evaluate the saliency and reliability of such coefficients.

2.4. Facilitation across scales

Facilitation interactions have been described in V1 as excitative connections between co-oriented, coaxial, aligned neighboring cells [14, 36]. Psychophysical studies and the Gestalt psychology determined that coaligned or cocircular stimuli are more easily detected and more perceptually salient [15, 16]. Studies of natural image statistics also show that statistically edges tend to be coaligned and cocircular [37, 38]. Experimentally we observe that log-Gabor coefficients arranged in chains of coaligned coefficients or present across different scales correspond to reliable and salient edges. Moreover, the probability that remaining noise features could be responsible for chains of coefficients is decreasing with the chain length. Thus a facilitation reinforcing cocircular cells conforms a noise segregation process. For all those reasons a facilitation across scale is set up to reinforce co-oriented cells across scales (under the conditions described in the next paragraph) and a facilitation in space and orientation reinforce chains of coaligned coefficients (Section 2.5).

The facilitation across scales consists in favoring those coefficients located where there exist also noninhibited coefficients at coarser scales. In practice, the *parent* coefficient \mathbf{h}_p (i.e., the one in the coarser scale) must be located in the same spatial location (tolerating a spatial deviation of one coefficient), in an adjacent orientation channel and be compatible in phase (i.e., it must have a difference lower than $2\pi/3$ in phase). $\mathbf{f}_1(x, y, s, r) = 1$ indicates that the coefficient (x, y, s, r) has a parent (otherwise $\mathbf{f}_1(x, y, s, r) = 0$). The

calculation of \mathbf{f}_1 can be summarized as follows:

$$\mathbf{h}_p(x, y, s, r) = \max_{(\delta_x, \delta_y, \delta_r) \in \{-1, 0, 1\}^3} \{\mathbf{h}_3(x + \delta_x, y + \delta_y, s + 1, r + \delta_r)\},$$

$$\mathbf{f}_1 = \begin{cases} 1 & \text{where } (h_3 \neq 0) \text{ and } (h_p \neq 0) \\ & \text{and } \left(\text{angle}(h_3, h_p) < \frac{2\pi}{3} \right), \\ 0 & \text{elsewhere.} \end{cases} \quad (4)$$

It is then straightforward to calculate the presence of *grand-parents* (noted $\mathbf{f}_1(x, y, s, r) = 2$), where the parent coefficient has itself a parent.

Kovesi showed that phase congruency of log-Gabor coefficients across scales is efficient for extracting edges [39]. It is remarkable to note (see Figure 5(c)) that many edges and ridges extracted are closely repeated across scales with coefficients linked by parent relationships. This regularity is due in part to the good behavior the log-Gabor wavelets is promising for the decorrelation and efficient coding of contours.

2.5. Facilitation across space and orientation

As proposed in Yen and Finkel's V1 model [20], we implement a saliency measurement linked with the *chain length* defined as the number of coefficients composing the chain. It is calculated for each coefficient and consists in counting the number of coefficients forward n_f and backward n_b along the chain. The successive coefficients must be coaligned along the preferred orientation of the channel tolerating a maximal variation of 53° . The compatibility in phase is also checked, that is, two successive coefficients are not considered to belong to the same chain only if they have a difference of phase superior to $2\pi/3$. The number of coefficients is counted in each direction to a maximum of l_{\max} coefficients

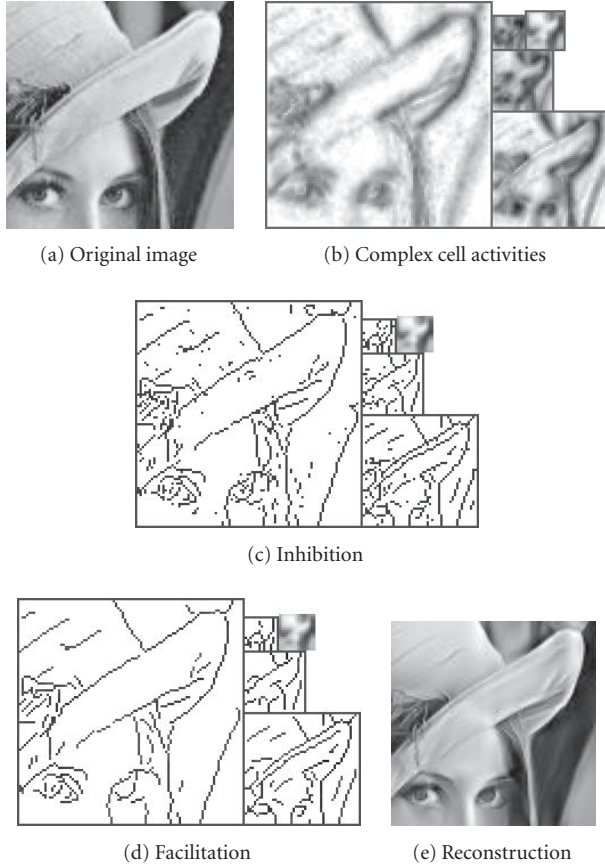


FIGURE 5: Successive steps modeling V1 architecture as a sparse approximation strategy. (a) 96×96 detail of the “Lena” image. (b) Complex cell activities are modeled as the log-Gabor coefficient modulus (Section 2.1). All the orientations are overlaid so that one inset is shown for each scale. The different scales have different sizes due to the downsampling applied. From the largest to the smallest the insets correspond respectively to the 2nd, 3rd, 4th, low-pass and 5th scale. The first scale is not represented. (c) Remaining coefficients after the inhibition step (Section 2.3). (d) The facilitation step (Sections 2.4-2.5) preserves the coefficients arranged in sufficiently long chains and having parent coefficients within coarser scales. The remaining cells conform the sparse approximation of the image. It is composed by a subdictionary including the most salient multiscale edges and the low-pass version of the image. (e) The gain control step (Section 2.6) assigns an amplitude to the subdictionary edges. Then the inverse log-Gabor wavelet transform reconstructs an approximation of the image.

(with $l_{\max} = 16$. The different parameters are chosen experimentally). The saliency is finally calculated in the following form which permits to obtain a constant response along each chain:

$$f_2(x, y, s, r) = \min(l_{\max}, n_f + n_b). \quad (5)$$

Finally the facilitation consists in retaining those coefficients which fulfill the following two criteria (while the other coefficients are zeroed out to be considered as noise or less

salient edges). First they must pass a certain length threshold depending of the scale and the presence of parent coefficients. Typically the chain length threshold is chosen as 16, 16, 8, 4, 2, respectively, for the scales 1, 2, 3, 4, 5, half of these lengths if coefficients have a parent, and a fourth of these lengths if they have a grandparent. Second, the amplitude must overpass a spike threshold corresponding to twice the CSF threshold defined in Section 2.2. Each coefficient is selected with its chain neighbors which implies that chains are selected or rejected entirely (see the final selection Figure 5(d)). This second condition is equivalent to the Canny hysteresis [31]. As a summary, the facilitation process can be approximated by the equation

$$\mathbf{h}_4(x, y, s, r) = \begin{cases} \mathbf{h}_3(x, y, s, r) & \text{if } (f_2(x, y, s, r) \geq 2^{6-s-f_1(x,y,r,s)}) \\ & \text{and } (|\mathbf{h}_3|(x, y, s, r) \geq 2\text{CSF}(s, r)), \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

The facilitation implementation is not described here in more detail since it does not incorporate strong improvements over the algorithms existing in the literature. Moreover small changes in the implementation do not strongly impair the final results.

Both the chain length and CSF thresholds are chosen depending on the application since for high compression rates the thresholdings must be severe while for image denoising most edges should be preserved which requires more permissive thresholds. The first scale edges are less reliable because of the intrinsic lower orientation selectivity of the filters close to the Nyquist frequency. In the present implementation edges selected in the second scale will also be those used for the first scale.

Additionally, for further increasing the sparsity, some coefficients can be periodically ruled out along chains. If the induced hollows are sufficiently narrow they will not be perceptible in the reconstructed image thanks to the important overlapping between log-Gabor functions. This is the case, for instance, when one every two or two every three coefficients are zeroed (as it will be shown in Section 3.2 and Figures 8, 9). This strategy will be exclusively adopted for image compression tasks where highly sparse approximations are required.

2.6. Gain control

In this section both the image \mathbf{x} , the log-Gabor wavelet transform $\mathbf{h} = W\mathbf{x}$, and the \mathbf{h}_4 pyramid are treated as 1D vectors (for such a purpose the 2D or 4D vectors are concatenated into 1D vectors). We have $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{h} \in \mathbb{R}^M$, $\mathbf{h}_4 \in \mathbb{R}^M$, $W \in \mathbb{R}^{M \times N}$, and $W^T \in \mathbb{R}^{N \times M}$, N being the number of pixels in the image and M the size of the dictionary (with $M > N$).

The previous steps of thresholding, inhibition, and facilitation allowed to extract a set of active cells corresponding to multiscale edges. They define a set of selected coefficients

called *subdictionary* from which an approximation of the image will be reconstructed. Let us assume $D \in \mathbb{R}^{M \times M}$ the diagonal matrix defined on the dictionary space and which eigenvalues are 1 on the selected subdictionary and 0 elsewhere. We call $a_0 = \mathbf{h}_4$ the *approximation* and $r_0 = \mathbf{h} - \mathbf{h}_4$ the *residual*:

$$\begin{aligned} a_0 &= DW\mathbf{x} \quad (= \mathbf{h}_4), \\ r_0 &= (1 - D)W\mathbf{x} \quad (= \mathbf{h} - \mathbf{h}_4). \end{aligned} \quad (7)$$

The gain control aims at adapting the amplitude of the a_0 coefficients for obtaining the closest possible reconstruction through the W^T operation. We know that $\mathbf{h} = a_0 + r_0$ reconstructs exactly the image with $W^T\mathbf{h} = \mathbf{x}$. Nevertheless it can be verified experimentally that a_0 (the sparsified version of \mathbf{h}) only reconstructs a very smoothed version of \mathbf{x} : the a_0 coefficients need to be enhanced for a closer reconstruction.

This enhancement could be realized through a fixed gain factor. But for a better reconstruction, we adopt a strategy close to matching pursuit [5] which plausibility as biological model has been explored in [7]. MP selects at each iteration the largest coefficient which is added to the approximation while its projection on the other dictionary functions is subtracted from the *residual*. This projection, which depends on the correlation between dictionary functions, can be interpreted as a lateral interaction [7]. Here as a difference with MP, the residual r_0 is projected on the subspace \mathcal{V} spanned by the subdictionary. We do not know the projection operator P^* that realizes this operation. Thus the projector $P = WW^T$ that projects the residual on the whole transform space is iteratively used instead¹:

$$\begin{aligned} a_k &= a_{k-1} + DP r_{k-1}, \\ r_k &= (1 - D)P r_{k-1}. \end{aligned} \quad (8)$$

By the self invertible property we have $W^T P = W^T W W^T = W^T$ and it comes that

$$\begin{aligned} W^T(a_k + r_k) &= W^T(a_{k-1} + P r_{k-1}) \\ &= W^T(a_{k-1} + r_{k-1}). \end{aligned} \quad (9)$$

Iteratively and using again the self-invertible property and (8) we have finally

$$W^T(a_k + r_k) = W^T(a_0 + r_0) = W^T W \mathbf{x} = \mathbf{x}. \quad (10)$$

Hence, $W^T(a_k + r_k)$ reconstructs exactly the source image \mathbf{x} for any k .

It is also straightforward to show that a_k and r_k converge: let Q be defined as $Q = (1 - D)P$. We have now

$$\begin{aligned} a_k &= a_0 + DP \sum_{q=1}^k r_{q-1} = a_0 + DP \left(\sum_{q=1}^k Q^q \right) r_0, \\ r_k &= Q^k r_0, \end{aligned} \quad (11)$$

P and D being projections, $\|Qe\| \leq \|e\|$ for any vector e (where $\|\cdot\|$ is the quadratic norm). Moreover any vector e' which verifies $\|Qe'\| = \|e'\|$ is an eigenvector of P (with eigenvalue 1) and of D (with eigenvalue 0), then of Q (with eigenvalue 1). We deduce that (a) $DPQ^q e' = 0$; and (b) the eigenvalues of Q different than 1 are strictly smaller than 1. Hence for any r_0 , $DP(\sum_{q=1}^k Q^q) r_0$ and a_k converge, and from (b) we have the r_k convergence. The convergence is moreover exponential with a factor corresponding to the highest eigenvalue of Q which is strictly smaller than 1.

In practice we observe that the algorithm converges with regularity, a_k and r_k becoming stable in around 40 iterations. If the dictionary has been adequately selected, most of the residual coefficients dramatically decrease their amplitude and the selected coefficients encode almost all the image information (e.g., the reconstruction of Lena is shown in Figure 5(e)). But because some edges and ridges can lack in the dictionary, in particular around corners, crossing and textures, a second pass of thresholding, inhibition and facilitation can also be advantageously deployed on the residual for selecting new edge coefficients.

Concerning the overall computational complexity, all the thresholding, inhibition, and facilitation steps are computed by local operations consisting in convolutions by small kernels (mainly 3×3). The linear and inverse log-Gabor wavelet transforms W and W^T are computed in the Fourier domain but could also be implemented as convolutions in space domain, which is a biologically plausible implementation. In such a case the algorithm would consist in a fixed number of local operations. The computational complexity would then be as low as $O(N)$, where N is the number of pixels in the image.

2.7. Contour representation

The former processes allowed to approximate the visual information through continuous chains of active cells representing contour segments (see Figure 5(d)). The next step in the integration of the visual information would be to build an efficient representation of such chains. For such purpose V1 hypercomplex or end-stopped cells [19, 40, 41] which respond preferentially to ridge endings, abrupt corners and other types of junctions and crossings could play an important role since such features are known to be determinant in perception of contours. Descriptions of integrated contours could also take place in higher visual areas like V2 and V4 which are supposed to provide increasingly complex descriptions of visual shapes. For instance, recent advances have shown that cells in V4 area may respond to curvature degree (concavity) and to angles between aggregated curved segments [42].

In this first implementation we choose to represent contours by their endpoints, called *chain heads*, simulating hypercomplex cells and the contour shape through elementary displacements called *movements*. This shape representation through successive movements is not biologically inspired but it corresponds to a relatively simple and classical image processing method called *chain coding*. In future

¹ It is direct that P is linear and $P^2 = P$, hence P is a projector.

implementations a full biological model representing contours through shape parameters such as curvatures and angles could advantageously be set up.

The contour representation aims at further integrating the visual information simultaneously for providing a description more easily exploitable by the highest visual areas in tasks such as object recognition and for reducing the redundancy by removing higher-order correlations [34]. The chain coder will be evaluated here for redundancy reduction, that is for image compression.

The present chain coder has been specially adapted from [43] to log-Gabor channels features. Chain coding has been many times revisited for efficient representation of contours, whose main precursor was Freeman [44]. He proposed to link the nonzero adjacent pixels by elementary movements. The chains are represented by three data sets: *head locations* which are the starting point of chains, *movements* which are the displacement directions to trace chains, and *amplitudes* which are the values of log-Gabor coefficients.

(i) Head locations

The vertical and horizontal coordinates of the heads are coded considering the distance between the current head and the previous coded head. The compressing benefit comes from the idea of avoiding to code always the absolute location within channels. *Prefix codes* compress efficiently such relative distances according to their probabilities. Since channels are scanned by rows, short vertical differences are more probable than long ones, whereas horizontal differences are almost equiprobable.

(ii) Movements

Only movements not implicated in the inhibition are possible. Thus, only two or three movements (pointing to the channel orientation) are possible. These movements together with an additional movement to mark the end of chain are coded by prefix codes.

(iii) Amplitudes

The Gabor modulus is quantified using steps depending on the contrast sensitivity function (CSF) [33], while the phase is quantized in 8 values ($-3\pi/4$, $-\pi/2$, $-\pi/4$, 0 , $\pi/4$, $\pi/2$, $3\pi/4$, π). Data to code is the difference between the value of a link and the previous one (prediction error). Moreover, head amplitudes, which are used as offsets, can also be predicted, although their correlation is not so high. Two predictive codings (module/phase) for head's amplitudes and two for link's amplitudes are then encoded by *arithmetic coding*.

Furthermore, natural contours usually present complex shapes which are unable to be covered by a single channel: they spread across different orientation channels and even across scales. For this reason we concatenate adjoining chains by their end(starting)-points jumping from one to another oriented channel (not necessarily contiguous). Note this concatenation procedure implies the use of special labels

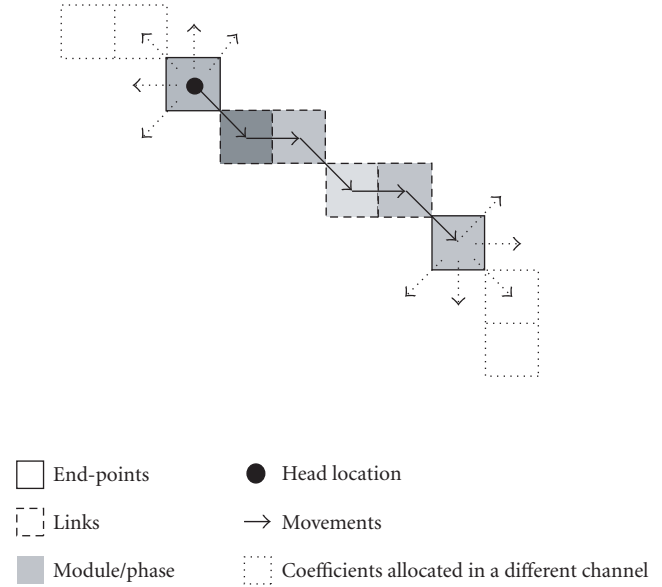


FIGURE 6: Scheme proposed for contour representation.

to indicate to which channel belongs the chain to concatenate. Figure 6 depicts a scheme of the proposed contour representation. Future implementations will envisage to concatenate chains across scales taking into account the strong predictability of contours across scales.

Additionally the residual low-pass channel is coded by a simple neighboring and causal predictor followed by an arithmetic coding stage. An outstanding report about the here mentioned codings can be found in [45].

3. RESULTS

3.1. Edge and ridge extraction

Examples of contours extracted by the spike threshold, inhibition and facilitation processes are shown in Figures 5 and 7. The different orientations are summed up so that edges belonging to a same scale are drawn together. Results can be compared with Figures 7(d) and 7(e) which show the edges extracted by the Canny operator. The proposed model presents the following advantages. (1) It extracts both edges and ridges while Canny only extracts edges drawing generally two edges where there is one ridge. It consequently often yields unrealistic solutions. (2) It is able to reconstruct the image from the multiscale edges which is a warranty of the nearly completeness of the edge information (see Figures 5(e) and 7(c), 7(h)). Indeed since reconstruction is now possible, the quality of reconstruction from the edges could be considered as a measure of the accurateness of the edge extraction. Such measurement would be a great use since it is generally complicated to evaluate edge extraction methods due to the lack of a "ground truth." Reconstruction quality will be discussed in the next sections

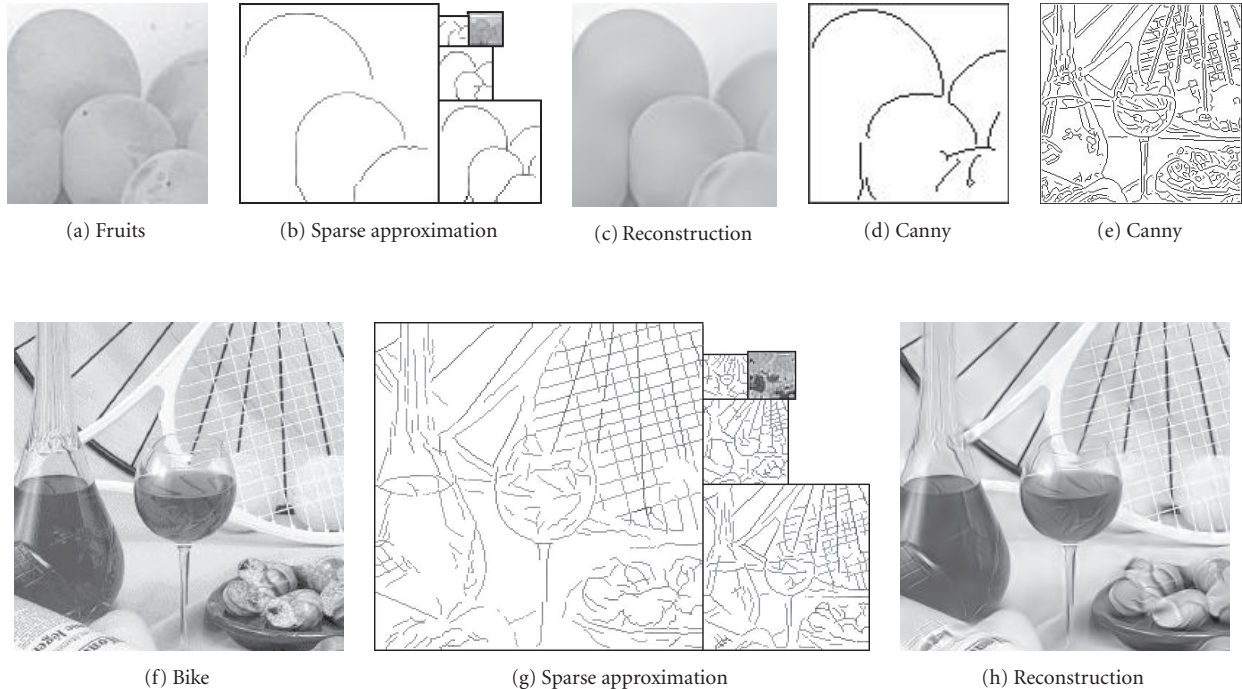


FIGURE 7: Extraction of multiscale edges and reconstruction. (a) 96×96 pixels tile of the image “Fruits.” (f) 224×224 pixels tile of the image “Bike.” (b), (g) Edges extracted by the proposed model. The gray level indicates the amplitude of the edges given by the gain control mechanism. (c), (h) Reconstruction from edges. (d), (e) Edges extracted by Canny method.

TABLE 2: Compression results in terms of PSNR for Lena, Boats, and Barbara.

Image	bpp	JPEG	JPEG2K	Our model
Lena	0.93	22.94	26.09	22.38
Boats	0.55	24.09	27.21	24.06
Barb	0.64	24.62	28.68	24.50

both in cases where few edges are selected (image compression, Section 3.2) or when most of the edges are preserved (image denoising, Section 3.3).

3.2. Redundancy reduction

The sparse approximation and the chain coding are applied to several test images as summarized in Figures 8, 9, 10, and 11 and Table 2. Such experiments aim at evaluating the abilities of the model to reduce the redundancy of the visual information. Redundancy reduction can be measured as the abilities of the model for image compression measured in terms compression rate (in bpp, bit per pixel), mathematical error, and perceptual quality (i.e., visual inspection). JPEG and JPEG-2000 are, respectively, the former and the actual golden standards in terms of image compression. They are then the principal methods to compare the model with. Additionally, a comparison with MP is included in Figures 9 and 10.

The sparse approximation applied to a tile of “Lena” shown in Figure 8(a) induces the selection of a subdictionary shown in Figure 8(e). The chain coding compresses the image at 0.93 bpp and the reconstruction is shown in Figure 8(d). The comparison at the same bit rate with both JPEG and JPEG-2000 compressed images are shown in Figures 8(b)-8(c). Other results at 1.03 and 0.56 bpp for the image “Bike” are shown in Figures 9 and 10, where an additional comparison with MP is included.

As shown in Figure 10(a) the compression standards provide better results in terms of the peak-signal-to-noise ratio (PSNR)² at bit rates higher than 1 bpp for the image “Bike.” In contrast at bit rates lower than 1 bpp, the current model provides better PSNR than JPEG, and at bit rates lower than 0.3 bpp better than JPEG-2000.

Nevertheless it is well known that mathematical errors are not a reliable estimation of the perceptual quality. Since images are almost exclusively used by humans, it is important to evaluate the perceptual quality by visual inspection. Moreover as the proposed scheme models the primary visual areas, it is hoped that the distortions introduced present similarities with those produced by the visual system. Then one important expectation is that the distortions introduced

² The PSNR is measured in dB as $\text{PSNR} = -20 \log_{10}(\text{RMSE})$ where RMSE is the root mean square error between the original and the reconstructed image.

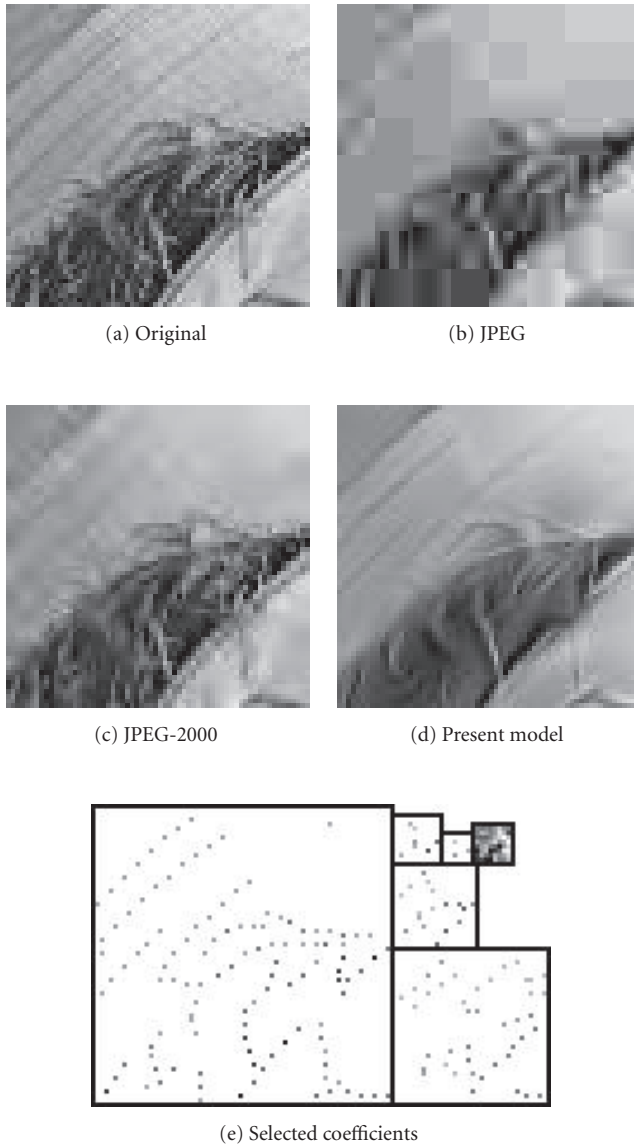


FIGURE 8: Compression of “Lena” at 0.93 bpp. (a) 64×64 original image. (b) In the JPEG-compressed image most of the contours and textures disappeared while block artifacts are salient. (c) Many details of the JPEG-2000 image are smoothed, in particular the strips and hairs of the hat. Moreover artifacts appear specially on diagonal edges. (d) In the image compressed through sparse approximation, the disappearance of visual details does not yield high frequency artifacts. (e) Selected subdictionary (here 2 every 3 coefficients have been zeroed along chains as proposed in Section 2.5).

by the model would appear less perceptible. This objective is important since a requirement of the lossy compression algorithms is the ability to introduce errors in a low perceptible manner.

A first remarkable property of the model is the lack of high-frequency artifacts. In contrast to JPEG or JPEG-2000, no ringing, aliasing, nor blocking effects appear. As a second good property, the continuity of contours appear particularly preserved. Finally, the gradients of luminance are

preserved smooth thanks to the elimination of isolated coefficients. For those reasons, the reconstructed images tend to look natural even when the mathematical error is significantly higher. Compared with MP, the model provides a more structured arrangement of the selected coefficients (compare Figure 9(b) with Figure 9(c)), which induces more continuity of the contours in the reconstruction and reduces the appearance of isolated artifacts.

Reconstruction quality appears worst in junctions, crossings, and corners of the different scales (see also Figure 11(a) for an image containing many of such features). This can be explained by the good adequacy of log-Gabor functions for matching edges and ridges and their worst match with junction and crossing features. One can argue that the present sparse approximation method should be completed by the implementation of junctions/crossing detectors as other models do [19]. Nevertheless this lies out of the scope of the present paper.

The second problem concerns textures which are generally not well treated by edge extraction methods. One of the worst cases is the pure sinusoidal pattern which in some conditions does not even induce local-maxima in the modulus of complex log-Gabor functions. Nevertheless in the majority of cases, textures can be considered as sums of edges. For example in Figure 8 the bristles of Lena’s hat form a texture and at least the most salient bristles are reproduced. In the same manner the texture constituted by the hat striation is not reproduced integrally but the most salient striations are preserved (note moreover that the striations also tend to disappear in the JPEG and JPEG-2000 compressed images). For further improving the reconstruction quality, and to extract more edges, a few additional passes of sparse approximation can be deployed. For example, a second pass allows the extraction of a significant part of the textures in Barbara’s scarf and in its chair as shown in Figure 11(h). Nevertheless the method does not allow to capture so much sparse approximations for textures than it does with contours. The compression quality at the same rate is then significantly lower. As future improvements, it could then be advantageous to deal with textures through a separate dedicated mechanism exploiting the texture statistical regularities as those proposed, for example, in [29, 46], or more simply using a standard wavelet coder as proposed in [28, 30]. Such improvements stay nevertheless out of the scope of the present study.

The reduction of information quantity between the sparse approximation and the chain coding can be evaluated as around 34% through classical entropy calculations (data available in [47]). As the chain coder does not introduce information losses (the reconstruction is the same), the information quantity reduction is uniquely due to a redundancy reduction. Thus chain coding offers a significant redundancy reduction which shows the importance of applying an additional transform for grouping selected coefficients in further decorrelated clusters like chains. This is an important advantage on MP which induces a sparse approximation less structured then harder to further decorrelate.

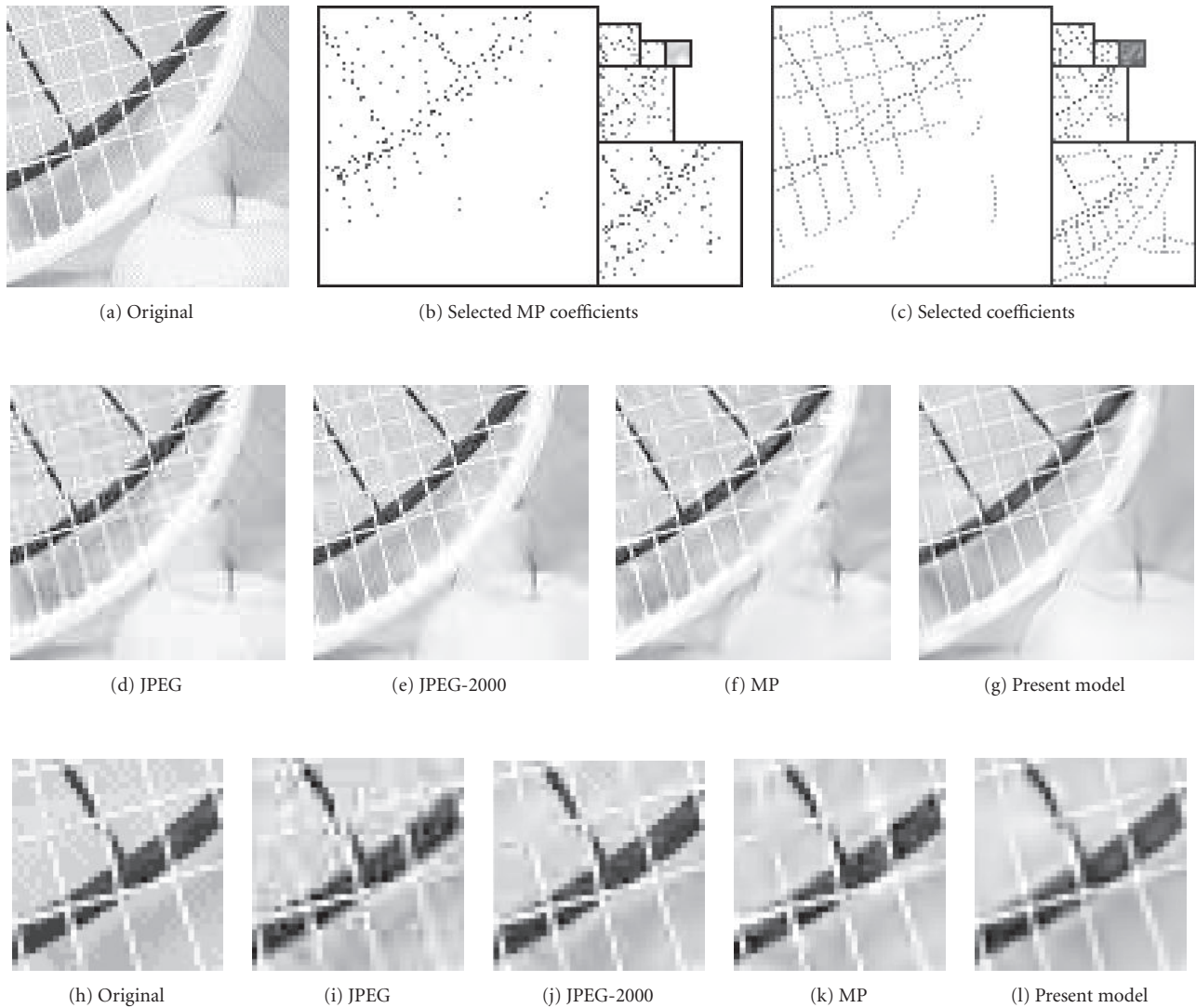


FIGURE 9: Compression results at 1.03 bpp. (a) 96×96 tile of the “Bike” image. (b) Coefficients selected by the MP algorithm. (c) Coefficients selected through the sparse approximation steps. (d) Compression with JPEG, PSNR = 25.73 dB. (e) Compression with JPEG-2000, PSNR = 29.61 dB. (f) Reconstruction by the MP algorithm, PSNR = 25.03 dB. (g) Compression by the proposed model, PSNR = 26.05 dB. (h), (i), (j), (k), (l) 36×36 zoom tile for original, JPEG, JPEG-2000, MP, and the model, respectively.

3.3. Noise elimination

Denosing results are presented in Figures 12, 13, and 14 in comparison with the standard method by wavelet shrinkage [22] (orthogonal and undecimated wavelets “Db4” are used) and the GSM model using steerable pyramids [25]. For all methods the noise level is supposed to be known and the implementation proposed in [25] is used both for the GSM and the wavelet shrinkage methods. In denoising the quality of reconstruction is important, then no edges should be missed in the sparse approximation. Consequently the sparse approximation steps are deployed two additional times on the reconstruction error, so as to extract the residual edges not detected in the first passes.

It is worth to note first that the method is able to extract and reconstruct almost all the image features. For example, the reconstruction of the image boats (Figure 12(e)) incorporates almost all the original image features. Nevertheless some few edges are lost, for example, close to intricate junctions (see also Lena’s right eye and the upper part of the hat border in Figure 13(f)). Thus, at very low noise level the method cannot compete with other denoising methods due to that approximated reconstruction.

In contrast, long edges and lines are particularly preserved. For example, the boat wires in Figure 12 are particularly well preserved while they tend to disappear and to be smoothed out by the other methods. This contour preservation remains at high noise levels, where it allows for the image

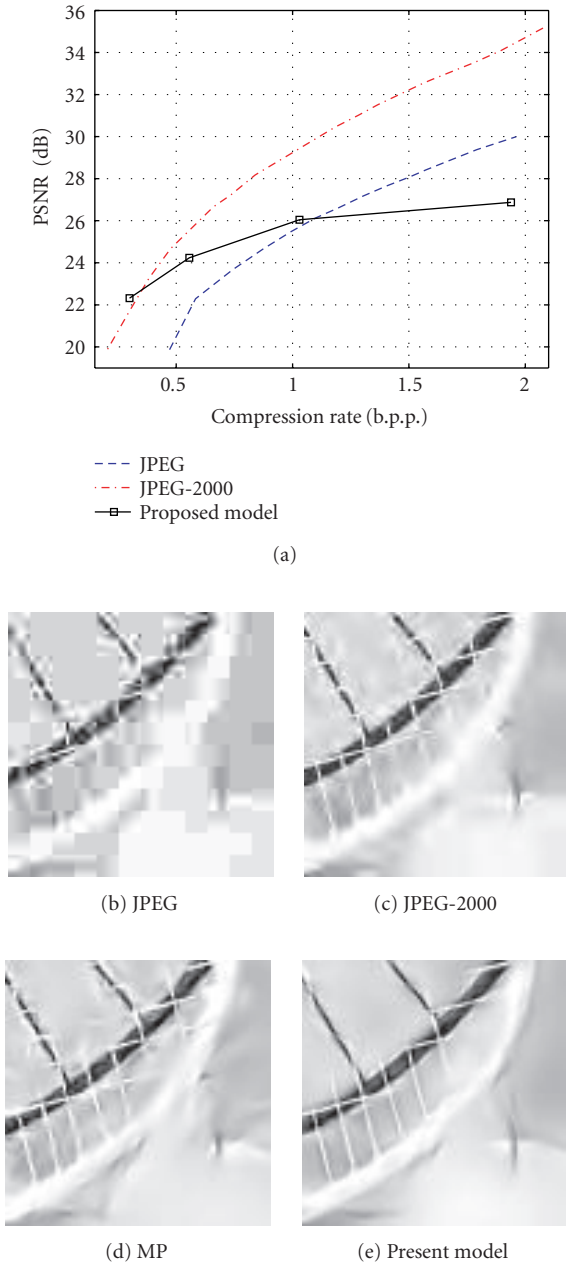


FIGURE 10: Compression results for the image Bike (original in Figure 9). (a) Evolution of the PSNR for different compression rates. The proposed method can offer a reconstruction competitive with the compression standards at very high compression rates. (b), (c), (d), (e) Compression results at 0.56 bpp, respectively, for JPEG, JPEG-2000, MP, and the proposed model.

Lena a significant gain over the other methods (the difference is around 0.6 dB with GSM, see Figure 14(a)). Figures 14(b)–14(f) show that contours are preserved sharper than in the other methods also at very high noise level.

Moreover as in the compression application, an important quality of the model is to yield reconstructions without high frequency artifacts. This allows in particular the

preservation of smooth gradients of luminance (see, e.g., Lena’s skin in Figure 13).

For explaining the results, it is worth noting that an important difference between methods resides in the thresholding mechanism. Wavelet shrinkage only considers the amplitude of the coefficients, retaining the highest ones as signal and eliminating the smallest coefficients as noise. The GSM model considers the 3×3 neighborhood and the parent coefficient in the thresholding decision. In contrast the proposed model takes into account larger neighborhoods by considering that contours are arranged in long chains of coaligned edges while noise is spatially incoherent.

4. CONCLUSIONS

We proposed a sparse approximation inspired from biological knowledge on V1 cortical cells and constructed following image processing criteria. It consists in a log-Gabor wavelet transform modeling V1 receptive fields followed by steps of thresholding, inhibition, facilitation and gain control modeling V1 nonlinearities and lateral interactions between cells. Those steps are able to extract continuous chains of coefficients located on edges and ridges of the image, achieving an efficient contour extraction. Such procedure is incorporated in a sparse approximation scheme which selects uniquely those contour coefficients for building an approximation of the image. As an additional advantage of the method, the redundancy of sparse approximation can be further reduced by predictively encoding the chains of coefficients.

The redundancy reduction abilities allows the compression of images preserving particularly the perceptual quality and approaching the results obtained by the standard image compression algorithms at high or very high compression rates. In parallel the ability for extracting contours shows promising results for image denoising since it preserves particularly long lines and contours and at the same time it reduces the appearance of artifacts. Best results are obtained at high noise levels.

Those encouraging results confirm the potential of overcomplete transforms and sparse approximation algorithms for image processing and in particular for compression applications. The present study shows that overcomplete transforms can offer important advantages in terms of perceptual quality in particular for avoiding the appearance of artifacts and preserving smooth gradients and continuous sharp contours. Another significant advantage is a high interpretability of transform coefficients in terms of edges and contours. It is remarkable also that the computational cost is reduced through the use of pure local operations and the noniterative selection of the subdictionary. Moreover, the efficiency of the scheme for visual processing argues for the plausibility that similar processes could take place in the primary visual cortex.

Among further improvements, dedicated end-stopping operators dealing with the extraction of junctions, corners, and crossings could be implemented, which should improve the performance in the proximity of such features. Another important improvement would consist in incorporating a

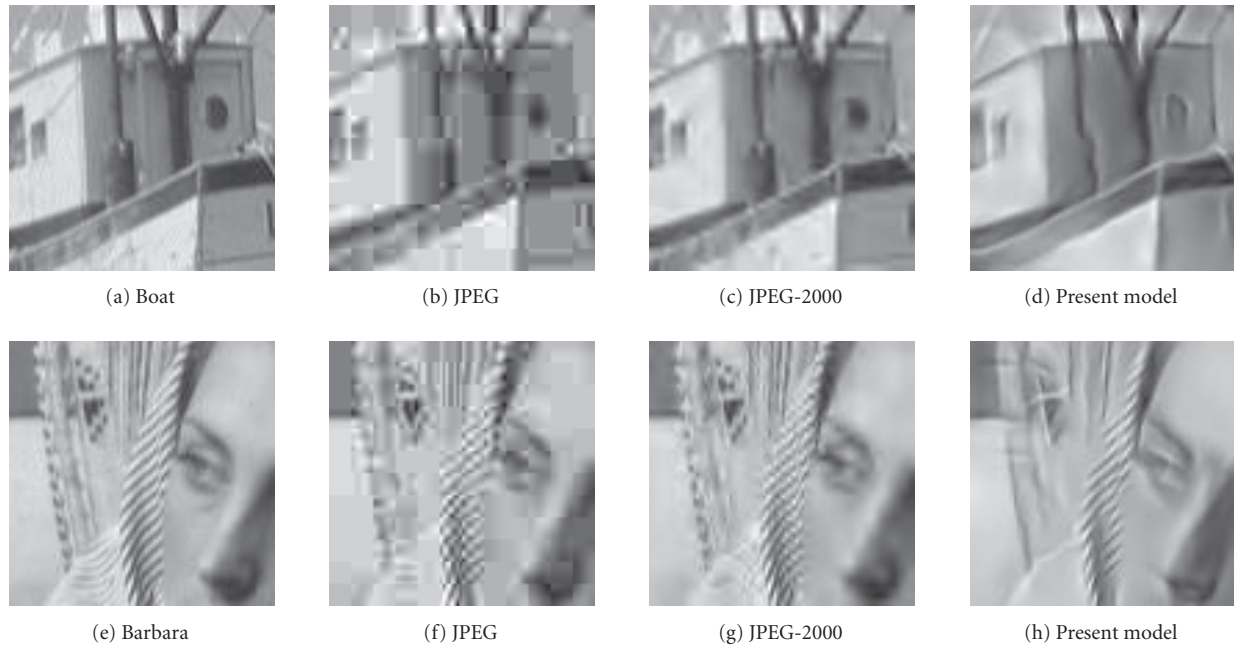


FIGURE 11: Compression results of “Boats” at 0.55 bpp and of “Barbara” at 0.64 bpp. (a) This 96×96 tile of “Boats” image contains many junctions and corners, which are difficult features to be captured by the model. (b) Compression with JPEG. (c) Compression with JPEG-2000. (d) Compression using sparse approximation and chain coding. (e) 96×96 tile of “Barbara” image. This image contains textures which are also difficult features to be encoded by the model. (f) Compression with JPEG. (g) Compression with JPEG-2000. (h) Compression using the proposed model.

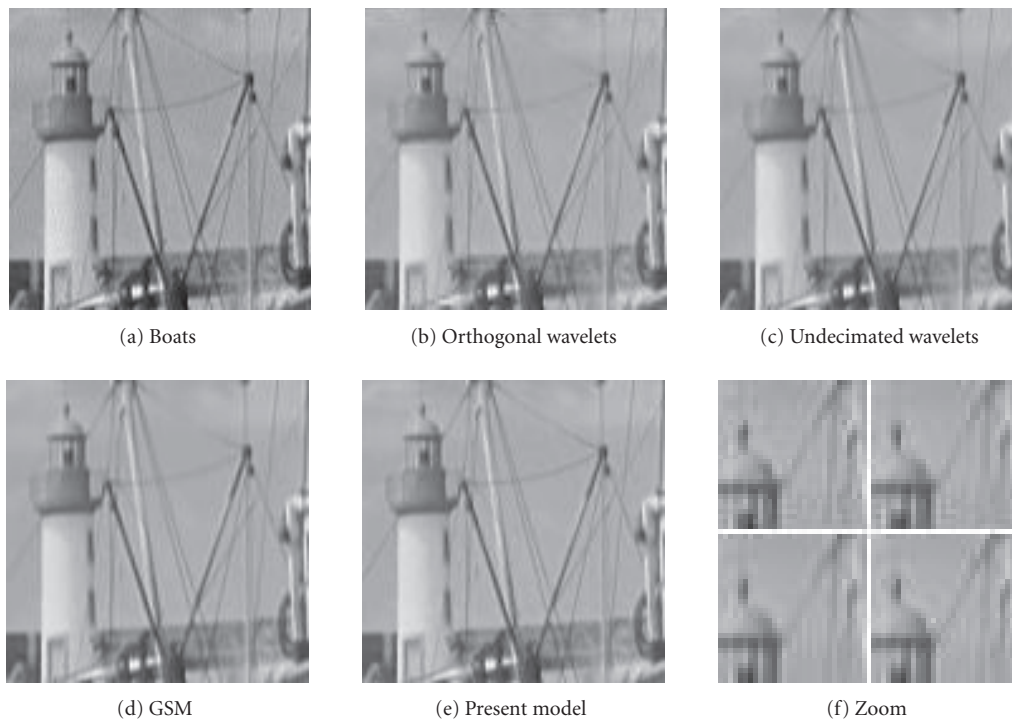


FIGURE 12: Denoising results at low noise level. (a) The original image “Boats” (120×120 pixels) already contains a low level of noise. The image is denoised using (b) orthogonal wavelet, (c) undecimated wavelet, (d) the GSM model and (e) the present model. The proposed method preserves particularly long lines and edges, as, for example, the wires at the top and at the left of the image. Different details appear smoothed by the other methods while they remain sharp for the proposed method. (f) 30×30 zoom for original, undecimated, GSM, and the present model (left to right and top to bottom).

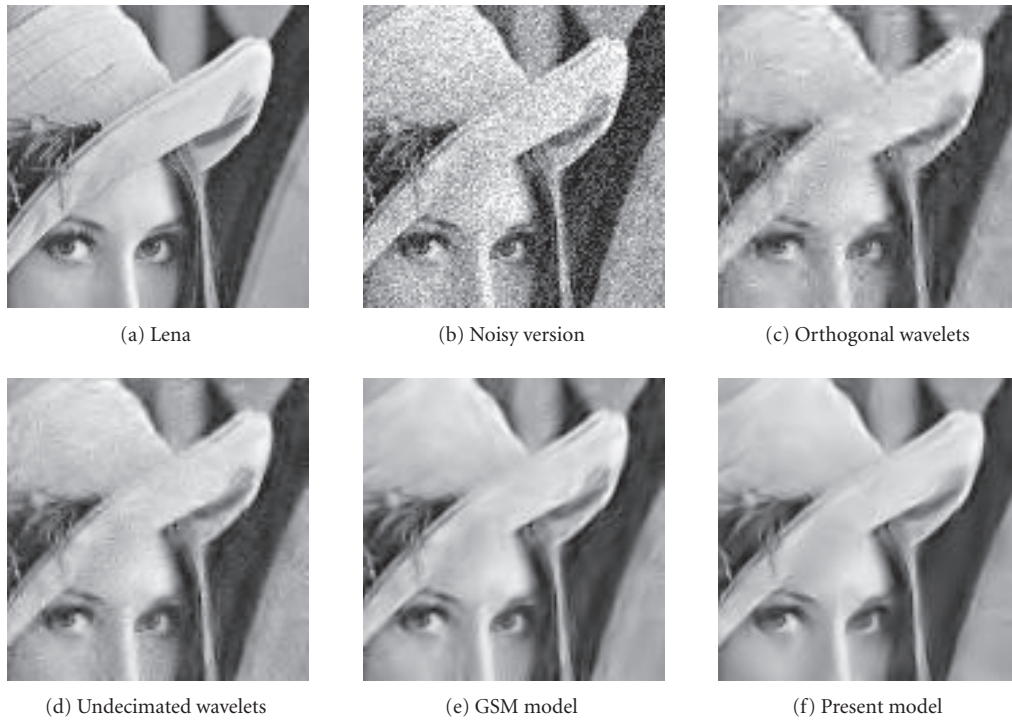


FIGURE 13: Denoising results with Lena image at medium noise level. (a) 112×112 detail of the image “Lena.” (b) Same image corrupted by Gaussian noise for a PSNR of 20.22 dB. (c) After denoising using orthogonal wavelets a high level of artifacts appears. (d) The quantity and strength of artifacts is reduced thanks to the use of undecimated wavelets. (e) The GSM model allows an additional reduction of the number of artifacts. (f) The proposed model also allows to reduce the appearance of artifacts, preserving particularly smooth gradients of luminance. Nevertheless the model shows difficulties in capturing some intricate features in particular close to junctions (see, e.g., the right eye) and to adjacent parallel lines (e.g., upper end of the hat border).

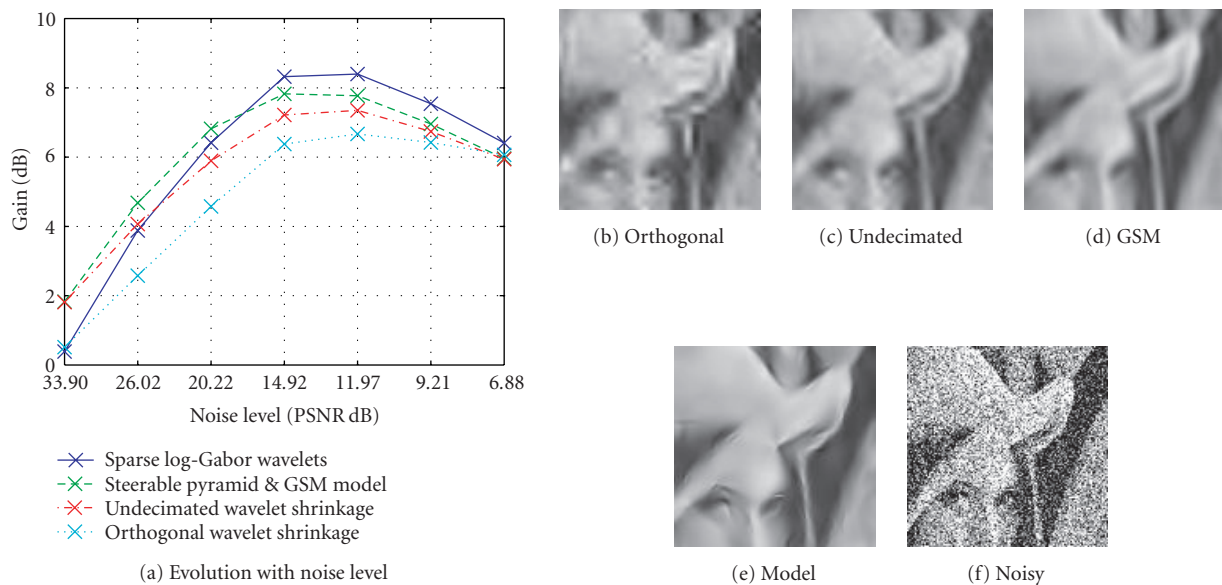


FIGURE 14: Denoising results at different noise levels. (a) The evolution of the denoising results is plotted for Lena (original in Figure 13) as the gain (i.e., the improvement in terms of PSNR), as a function of the noise level (also measured as PSNR). For low noise levels (33.90 and 26.02 dB), the model yields poorer gain than the other methods. This could be due to the imperfect reconstruction offered by the sparse approximation method. For high noise level (PSNR lower than 15 dB), the model offers a higher gain than the other methods. Bottom insets show denoising results using (b) orthogonal wavelets, (c) undecimated wavelets, (d) the GSM model, and (e) the present model on a noisy version of “Lena” (f) which PSNR is 11.97 dB.

separated texture representation as already proposed by several authors. Many improvements are also possible in all the different steps of the algorithm, in particular to improve the selection of coefficients by incorporating a statistical framework linking the different saliency measurements (chain length, presence of parent coefficients, and coefficient amplitude), or for further exploiting the predictability of the coefficients across scales for image compression.

ACKNOWLEDGMENTS

Thanks to Laura Rebollo-Neira, Sandrine Anthoine, and Nader Yeganefar for discussions on the mathematical aspects of sparse approximation. This work has been supported in part by the grants TEC2004-00834, PI040765, TEC2005-24046-E, and TEC2005-24739-E. SF, RR and LP are supported by grants from MEC-FPU, CSIC-I3P and EC IP project FP6-015879, "FACETS," respectively.

REFERENCES

- [1] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, 1992.
- [2] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.
- [3] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, 2001.
- [4] D. L. Donoho and A. G. Flesia, "Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics?" *Network: Computation in Neural Systems*, vol. 12, no. 3, pp. 371–393, 2001.
- [5] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [6] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [7] L. Perrinet, M. Samuelides, and S. Thorpe, "Coding static natural images using spiking event times: do neurons cooperate?" *IEEE Transactions on Neural Networks*, vol. 15, no. 5, pp. 1164–1175, 2004.
- [8] S. Fischer, G. Cristóbal, and R. Redondo, "Sparse overcomplete Gabor wavelet representation based on local competitions," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 265–272, 2006.
- [9] A. E. C. Pece, "The problem of sparse image coding," *Journal of Mathematical Imaging and Vision*, vol. 17, no. 2, pp. 89–108, 2002.
- [10] L. Perrinet, "Feature detection using spikes: the greedy approach," *Journal of Physiology Paris*, vol. 98, no. 4–6, pp. 530–539, 2004.
- [11] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America. A, Optics and Image Science*, vol. 2, no. 7, pp. 1160–1169, 1985.
- [12] R. L. De Valois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vision Research*, vol. 22, no. 5, pp. 545–559, 1982.
- [13] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?" *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [14] M. K. Kapadia, G. Westheimer, and C. D. Gilbert, "Spatial distribution of contextual interactions in primary visual cortex and in visual perception," *Journal of Neurophysiology*, vol. 84, no. 4, pp. 2048–2062, 2000.
- [15] S. Mandon and A. K. Kreiter, "Rapid contour integration in macaque monkeys," *Vision Research*, vol. 45, no. 3, pp. 291–300, 2005.
- [16] R. F. Hess, A. Hayes, and D. J. Field, "Contour integration and cortical processing," *Journal of Physiology Paris*, vol. 97, no. 2–3, pp. 105–119, 2003.
- [17] S. Grossberg, E. Mingolla, and J. Williamson, "Synthetic aperture radar processing by a multiple scale neural system for boundary and surface representation," *Neural Networks*, vol. 8, no. 7–8, pp. 1005–1028, 1995.
- [18] T. Hansen, W. Sepp, and H. Neumann, "Recurrent long-range interactions in early vision," in *Emergent Neural Computational Architectures Based on Neuroscience*, S. Wermter, J. Austin, and D. Willshaw, Eds., vol. 2036 of *LNAI*, pp. 127–138, Springer, Heidelberg, Germany, 2001.
- [19] F. Heitger, L. Rosenthaler, R. Von der Heydt, E. Peterhans, and O. Kubler, "Simulation of neutral contour mechanisms: from simple to end-stopped cells," *Vision Research*, vol. 32, no. 5, pp. 963–981, 1992.
- [20] S.-C. Yen and L. H. Finkel, "Extraction of perceptually salient contours by striate cortical networks," *Vision Research*, vol. 38, no. 5, pp. 719–741, 1998.
- [21] R. VanRullen, A. Delorme, and S. J. Thorpe, "Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation," *Neurocomputing*, vol. 38–40, no. 1–4, pp. 1003–1009, 2001.
- [22] R. R. Coifman and D. Donoho, "Translation-invariant denoising," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds., vol. 103 of *Lecture Notes in Statistics*, pp. 125–150, Springer, New York, NY, USA, 1995.
- [23] S. Fischer, F. Sroubek, L. Perrinet, R. Redondo, and G. Cristóbal, "Self-invertible 2D log-Gabor wavelets," *International Journal of Computer Vision*, to appear.
- [24] S. Fischer, R. Redondo, L. Perrinet, and G. Cristóbal, "Sparse Gabor wavelets by local operations," in *Bioengineered and Bioinspired Systems II*, R. A. Carmona, Ed., vol. 5839 of *Proceedings of SPIE*, pp. 75–86, Sevilla, Spain, May 2005.
- [25] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.
- [26] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, 1992.
- [27] J. H. Elder, "Are edges incomplete?" *International Journal of Computer Vision*, vol. 34, no. 2–3, pp. 97–122, 1999.
- [28] L. Peotta, L. Granai, and P. Vandergheynst, "Image compression using an edge adapted redundant dictionary and wavelets," *Signal Processing*, vol. 86, no. 3, pp. 444–456, 2006, special issue on Sparse Approximations in Signal and Image Processing.

- [29] J.-L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1570–1582, 2005.
- [30] M. Wakin, J. Romberg, H. Choi, and R. Baraniuk, "Image compression using an efficient edge cartoon + texture model," in *Proceedings of Data Compression Conference (DCC '02)*, pp. 43–52, Snowbird, Utah, USA, April 2002.
- [31] J. Canny, "Computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [32] M. C. Morrone and D. C. Burr, "Feature detection in human vision: a phase-dependent energy model," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 235, no. 1280, pp. 221–245, 1988.
- [33] B. W. Rust and H. E. Rushmeier, "A new representation of the contrast sensitivity function for human vision," in *Proceedings of the International Conference on Imaging Science, Systems, and Technology (CISST '97)*, H. R. Arabnia, Ed., pp. 1–15, Las Vegas, Nev, USA, June 1997.
- [34] J. J. Atick, "Could information theory provide an ecological theory of sensory processing?" *Network: Computation in Neural Systems*, vol. 3, no. 2, pp. 213–251, 1992.
- [35] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.
- [36] W. H. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick, "Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex," *Journal of Neuroscience*, vol. 17, no. 6, pp. 2112–2127, 1997.
- [37] N. Krüger, "Collinearity and parallelism are statistically significant second-order relations of complex cell responses," *Neural Processing Letters*, vol. 8, no. 2, pp. 117–129, 1998.
- [38] W. S. Geisler, J. S. Perry, B. J. Super, and D. P. Gallogly, "Edge co-occurrence in natural images predicts contour grouping performance," *Vision Research*, vol. 41, no. 6, pp. 711–724, 2001.
- [39] P. Kovsi, "Phase congruency detects corners and edges," in *Proceedings of the 7th International Conference on Digital Image Computing: Techniques and Applications (DICTA '03)*, pp. 309–318, Sydney, NSW, Australia, December 2003.
- [40] D. Hubel, *Eye, Brain, and Vision*, Scientific American Library Series, W. H. Freeman, New York, NY, USA, 1988.
- [41] A. Dobbins, S. W. Zucker, and M. S. Cynader, "Endstopping and curvature," *Vision Research*, vol. 29, no. 10, pp. 1371–1387, 1989.
- [42] A. Pasupathy and C. E. Connor, "Population coding of shape in area V4," *Nature Neuroscience*, vol. 5, no. 12, pp. 1332–1338, 2002.
- [43] R. Redondo and G. Cristóbal, "Lossless chain coder for gray edge images," in *Proceedings of IEEE International Conference on Image Processing (ICIP '03)*, vol. 2, pp. 201–204, Barcelona, Spain, September 2003.
- [44] H. Freeman, "On the encoding of arbitrary geometric configurations," *IRE Transactions on Electronic Computers*, vol. 10, pp. 260–268, 1961.
- [45] G. P. Howard, "The design and analysis of efficient lossless data compression systems," Tech. Rep. CS-93-28, Department of Computer Science, Brown University, Providence, RI, USA, 1993.
- [46] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Inter-*

national Journal of Computer Vision, vol. 40, no. 1, pp. 49–70, 2000.

- [47] S. Fischer, *New contributions in overcomplete image representations inspired from the functional architecture of the primary visual cortex*, Ph.D. thesis, Technical University Madrid High Technical School of Telecommunication Engineering, Department of Electronic Engineering, Spain, 2007.

Sylvain Fischer received the M.S. degree in telecommunication engineering from ENST, Telecom Paris, France and ETSIT-UPM, Madrid, Spain in 2000. He is finishing the Ph.D. in the Instituto de Óptica, CSIC, Madrid. His current research interests include vision modeling and sparse approximation.



Rafael Redondo received in 2002 his Engineering degree from ETSIT (Universidad Politécnica de Madrid, Spain) focused on developing new image compression methods based on vision models. He currently works as an Ph.D. student at Instituto de Óptica (CSIC) since 2001. Among his research fields are vision modeling, image and volumetric coding algorithms and time-frequency representations applied to pattern recognition, image fusion and compression.



Laurent Perrinet received an Engineering degree from SUPAERO, in Toulouse (France) with a focus on signal and image processing using artificial neural networks and a Ph.D. degree in computational neuroscience. He is currently a Researcher at INCM-CNRS in Marseille (France). His research interests focus on bridging lower-level neural computations with functional models of inference and spatio-temporal integration aiming at understanding low- to mid-level visual perception.



Gabriel Cristóbal received the M.S. and Ph.D. degrees in telecommunication engineering from the Universidad Politécnica de Madrid, Madrid, Spain, in 1979 and 1986, respectively. He is currently a Research Scientist with the Instituto de Óptica, Madrid. His current research interests are in joint representations, vision modeling, and image compression.

