

# A PERFECTLY INVERTIBLE RANK ORDER CODER

Khaled Masmoudi<sup>1</sup>, Marc Antonini<sup>1</sup> and Pierre Kornprobst<sup>2</sup>

<sup>1</sup>*I3S laboratory, UNS-CNRS, Sophia-Antipolis, France*

<sup>2</sup>*NeuroMathComp Team Project, INRIA, Sophia-Antipolis, France*

**Keywords:** Rank Order Code, Bio-inspired Image Coding, Frame Theory, Scalability.

**Abstract:** Our goal is to revisit rank order coding by proposing an original exact decoding procedure for it. Rank order coding was proposed by Simon Thorpe et al. to explain the impressive performance of the human visual system in recognizing objects. It is based on the hypothesis that the retina represents the visual stimulus by the order in which its cells are activated. A classical rank order coder/decoder was then designed (Van Rullen and Thorpe, 2001) involving three stages: (i) A model of the stimulus transform in the retina consisting in a redundant filter bank analysis; (ii) A sorting stage of the filters according to their activation degree; (iii) A straightforward decoding procedure that consists in a weighted sum of the most activated filters. Focusing on this last stage, it appears that the decoding procedure employed yields reconstruction errors that limit the model Rate/Quality performances when used as an image codec. Attempts made in the literature to overcome this issue are time consuming and alter the coding procedure or are infeasible for standard size images and lacking mathematical support. Here we solve this problem in an original fashion by using the frames theory, where a frame of a vector space designates an extension for the notion of basis. Our contribution is threefold. First we add an adequate scaling function to the filter bank under study that has both a mathematical and a biological justification. Second, we prove that the analyzing filter bank considered is a frame, and then we define the corresponding dual frame that is necessary for the exact image reconstruction. Finally, to deal with the problem of memory overhead, we design an original recursive out-of-core blockwise algorithm for the computation of this dual frame. Our work provides a mathematical formalism for the retinal model under study and specifies a simple and exact reverse transform for it. Furthermore, the framework presented here can be extended to several models of the visual cortical areas using redundant representations.

## 1 INTRODUCTION

Neurophysiologists made substantial progress in better understanding the early processing of visual stimuli. Especially, several efforts proved the ability of the retina to code and transmit a huge amount of data under strong time and bandwidth constraints (Thorpe, 1990; Meister and Berry, 1999; Gollisch and Meister, 2008). The retina does this by means of a neural code consisting in a set of electrical impulses: the spikes (Rieke et al., 1997). Based on these results, our aim is to use the computational neuroscience models that mimic the retina behavior to design novel lossy coders for static images.

In the literature, numerous retina models are proposed. Although there is no clear evidence on how the spikes encode a given visual stimulus, we assume in this paper that the relevant encoding feature is the order in which the retina neurons emit their first spike. This assumption was motivated by Thorpe et al. neu-

rophysiologic results on ultra-rapid stimulus categorization (Thorpe, 1990; Thorpe et al., 1996). Authors showed that still image classification can be achieved by the visual cortex within very short latencies of about 150 ms or even faster. As an explanation, it was stated that: *There is information in the order in which the cells fire*, and thus the temporal ordering can be used as a code. This code termed as rank order coding (ROC) was further studied in (Van Rullen and Thorpe, 2001; Thorpe, 2010) and yielded the conception of a classical retina model (Van Rullen and Thorpe, 2001).

However, one major limit of the ROC coder defined in (Van Rullen and Thorpe, 2001) prevents us from its use for the design of image codecs. The proposed decoding procedure is inaccurate. Indeed, the bio-inspired retina model that generates the spikes is based on a redundant filter bank image analysis, where the considered filters are not strictly orthogonal. Thus, the filter overlap yields reconstruction errors that limit the Rate/Quality performance (Perrinet

et al., 2004; Bhattacharya and Furber, 2010; Masmoudi et al., 2010). In this work we tackle explicitly this issue. Attempts made in the literature to correct this overlap are time consuming and alter the coding procedure (Perrinet et al., 2004; Bhattacharya and Furber, 2010) or are infeasible for standard size images and lacking mathematical support (Bhattacharya and Furber, 2007; Bhattacharya and Furber, 2010). In this paper, we give an original solution relying on the so-called frames theory (Duffin and Schaeffer, 1952; Kovacevic and Chebira, 2008). The mathematical concept of frames extends the notion of basis to sets of filters which are linearly dependent. Their use for redundant signal analysis and reconstruction has already been experienced with success (Burt and Adelson, 1983; Rakshit and Anderson, 1995; Do and Vetterli, 2003). Our method requires the computation of a reverse operator once for all and keeps the bio-inspired coding scheme unchanged. We moreover verify that the reconstructed image that we get is equal to the original stimulus.

The present work brings three main contributions: (i) We add to the original retinal filter bank an adequate scaling function, and (ii) we provide an original mathematical demonstration that the filter bank thus defined is a frame. Then we propose an algorithm for errorless reconstruction through the construction of a so-called “dual frame”; (iii) We solve the technical issue of memory overhead that prevented the use of frames for high dimension spaces, with a novel out-of-core algorithm that computes the dual frame.

This paper is organized as follows: In Section 2, we present the bio-inspired retina model under study. Then in Section 3 we introduce a matrix-based formalism for the retinal transform and use it to define an exact decoding scheme through the construction of a dual frame. Finally in Section 4 and Section 5, we summarize and discuss the results and show the gain that we obtain in terms of Rate/Quality tradeoff.

## 2 THE RANK ORDER CODER

This Section summarizes the three stages of the ROC coder/decoder defined in (Van Rullen and Thorpe, 2001). First we present in Section 2.1 the image transform as performed by a bio-inspired retina model. Then we give the specification of the subsequent neurons rank ordering in section 2.2. Finally we describe the decoding procedure and show its inaccuracy in Section 2.3.

### 2.1 The Image Transform in the Bio-inspired Retina Model

Neurophysiologic experiments have shown that, as for classical image coders, the retina encodes the stimulus representation in a transform domain. The retinal stimulus transform is performed in the cells of the outer layers. Quantitative studies such as (Field, 1994; Rodieck, 1965) have proven that the outer cells processing can be approximated by a linear filtering. In particular, neurophysiologic experiments conducted in (Field, 1994) yielded the largely adopted DoG filter which is a weighted difference of spatial Gaussians that is defined as follows:

$$DoG(x, y) = w^c G_{\sigma^c}(x, y) - w^s G_{\sigma^s}(x, y), \quad (1)$$

where  $w^c$  and  $w^s$  are the respective positive weights of the center and surround components of the receptive fields,  $\sigma^c$  and  $\sigma^s$  are the standard deviations of the Gaussian kernels  $G_{\sigma^c}$  and  $G_{\sigma^s}$ , such that  $\sigma^c < \sigma^s$ . The DoG cells can be arranged in a dyadic grid  $\Gamma$  of  $K$  layers to sweep all the stimulus spectrum as shown in Figure 1(b) (Van Rullen and Thorpe, 2001; Perrinet et al., 2004; Masmoudi et al., 2010). As in the retina topology, the cells density and scale are inversely proportional. This keeps the model strongly inspired from the mammalian retina, though the authors in (Van Rullen and Thorpe, 2001) do not claim biological plausibility. Each layer  $0 \leq k < K$  in the grid  $\Gamma$ , is tiled with filtering cells, denoted by  $DoG_k$ , having a scale  $k$  and generating a transform subband  $B_k$  such that:

$$DoG_k(x, y) = w^c G_{\sigma_k^c}(x, y) - w^s G_{\sigma_k^s}(x, y), \quad (2)$$

where  $\sigma_{k+1}^c = \frac{1}{2}\sigma_k^c$  and  $\sigma_{k+1}^s = \frac{1}{2}\sigma_k^s$ . Each  $DoG_k$  filter has a size of  $(2M_k + 1)^2$ , with  $M_k = 3\sigma_k^s$ . Authors in (Van Rullen and Thorpe, 2001) chose the biologically plausible parameters established in (Field, 1994)  $w^c = w^s = 1$ ,  $\sigma_k^c = \frac{1}{3}\sigma_k^s \forall k$ , and  $\sigma_{K-1}^c = 0.5$  pixel.

In order to measure the degree of activation  $c_{kij}$  of a given retina cell, such that  $(k, i, j) \in \Gamma$ , we compute the convolution of the original image  $f$  by the  $DoG_k$  filter. Yet each layer  $k$  in the dyadic grid  $\Gamma$  is undersampled with a pace of  $2^{K-k-1}$  pixels with an original offset of  $\lfloor 2^{K-k-2} \rfloor$  pixels, where  $\lfloor \cdot \rfloor$  is the floor operator. Having this, we define the function  $u_k$ , such that the  $c_{kij}$  coefficients are computed at the locations  $(u_k(i), u_k(j))$  as follows:

$$u_k(i) = \lfloor 2^{K-k-2} \rfloor + 2^{K-k-1}i, \forall k \in \llbracket 0, K-1 \rrbracket. \quad (3)$$

$u_k$  can be seen as an undersampling function. We notice that  $u_{K-1}(i) = i$ , and that the other functions  $(u_k)_{k \in \llbracket 0, K-2 \rrbracket}$  are undersampled versions of  $u_{K-1}$ . We

also define  $N_k = \left\lfloor \frac{N - |N - 2^{K-k-2}|}{2^{K-k-1}} \right\rfloor$ , such that  $N_k^2$  is the number of cells in  $B_k$ , the  $k^{\text{th}}$  layer of  $\Gamma$ . We also notice here that  $N_{K-1} = N$ . Having these definitions,  $c_{kij}$  is then computed as follows:

$$c_{kij} = \sum_{x=u_k(i)-M_k, y=u_k(j)-M_k}^{x=u_k(i)+M_k, y=u_k(j)+M_k} \text{DoG}_k(u_k(i) - x, u_k(j) - y) f(x, y). \quad (4)$$

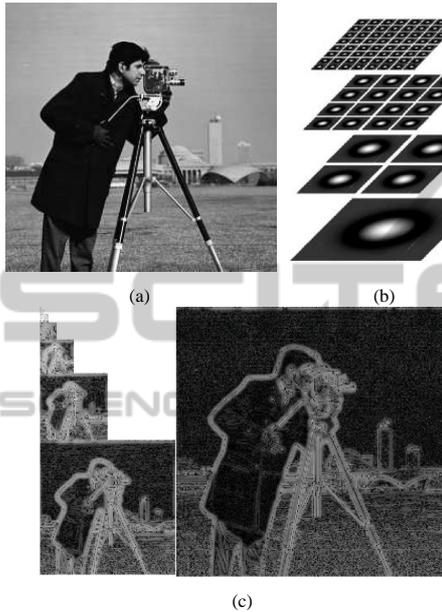


Figure 1: 1(a): *cameraman*. The image  $f$  size is  $257 \times 257$  pixels. 1(b): Example of a dyadic grid of DoG's used for the image analysis (from (Van Rullen and Thorpe, 2001)). 1(c): The transform result  $c$  with the different generated subbands  $B_k$  (here shown in a logarithmic scale).

This transform generates a vector  $c$  of  $(\frac{4}{3}N^2 - 1)$  coefficients  $c_{kij}$  for an  $N^2$ -sized image (if  $N$  is a power of 2). This architecture is similar to a Laplacian pyramid (Burt and Adelson, 1983). An example of such a transform performed on the *cameraman* test image is shown in Figure 1.

## 2.2 The Generation of the ROC

Thorpe et al (Thorpe, 1990; Thorpe et al., 1996) proposed that the order in which spikes are emitted encodes the stimulus. This yielded the ROC which relies on the following simplifying assumptions:

- i) From stimulus onset, only the first spike emitted is considered.
- ii) The time to fire of each cell is proportional to its degree of excitation.
- iii) Only the order of firing of the neurons encodes the stimulus.

The neurons responses  $(c_{kij})_{kij \in \Gamma}$  (cf. Equation (4)) are then sorted in the decreasing order of their amplitude  $|c_{kij}|$ .

The final output of this stage, the ROC, is the sorted list of  $N_s$  couples  $(p, c_p)$  such that  $|c_p| \geq |c_{p+1}|$ , with  $p$  being the index  $p(k, i, j) = k.N_k^2 + i.N_k + j$  and  $N_k^2$  the number of cells in the subband  $B_k$ . Such a code gives a biologically plausible interpretation to the rapidity of the visual stimuli processing in the human visual system. Indeed, it seems that most of the processing is based on feed-forward mechanism before any feedback occurs (Thorpe et al., 1996). The generated series  $(p, c_p)_{0 \leq p < N_s}$  is the only data transmitted to the decoder. Note that in some implementations as (Van Rullen and Thorpe, 2001; Perrinet et al., 2004), the exact values of the coefficients  $c_p$  are omitted and recovered through a look-up-table (but this is out of the scope of this paper).

## 2.3 Decoding Procedure of the ROC

We consider the set of the first  $N_s$  generated spikes forming the ROC of a given image  $f$ . The estimation  $\tilde{f}_{N_s}$  of the original input  $f$  is defined (Van Rullen and Thorpe, 2001) by:

$$\tilde{f}_{N_s}(x, y) = \sum_{p(k,i,j)=0}^{N_s-1} c_p \text{DoG}_k(u_k(i) - x, u_k(j) - y). \quad (5)$$

Equation (5) gives a progressive reconstruction depending on  $N_s$ . Indeed, one can restrict the code to the most valuable coefficients  $c_p$ , i.e the most activated neurons of the retina. This feature makes the coder be scalable (Masmoudi et al., 2010).

An example of such a reconstruction is given in Figure 2, with all the retina cells taken into account. Figure 2 also shows that the retina model decoding procedure, though giving a good approximation of the stimulus, is still inaccurate. In this example, reconstruction quality is evaluated to 27 dB of PSNR.

## 3 INVERTING THE RETINA MODEL

In this section, we define an original and exact image reconstruction algorithm starting from the ROC. First, we introduce in Section 3.1 a low-pass scaling function in the analyzing filter bank. This modification will be shown to be necessary for the transform invertibility. Then, in Section 3.2, we give a matrix-based formalism for the transform and we use it to prove that our filter bank is a frame in Section 3.3. Finally, in Section 3.4, we show the exact reconstruction

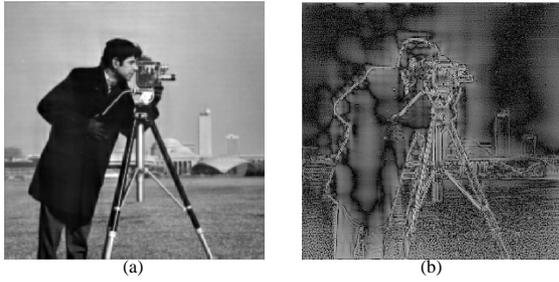


Figure 2: Result of the decoding procedure with the original approach using the totality of the retina cells. 2(a): The reconstructed image  $\tilde{f}_{N_s}$ . The PSNR quality measure of  $\tilde{f}_{N_s}$  yields 27 dB. 2(b): The error image shown in a logarithmic scale: high frequencies are the ones that are the most affected by this approach.

results using the dual frame and introduce an out-of-core algorithm to construct it.

### 3.1 Introduction of a Low-pass Scaling Function

We introduce a low-pass scaling function in the filter bank used for image analysis. This modification does not alter the ROC coder architecture and has both a mathematical and a biological justification.

Indeed, the Fourier transform of a Gaussian is another Gaussian, so that  $\mathcal{F}(DoG_k)$  is a difference of Gaussians. Therefore, with  $w^c = w^s = 1$  (cf. Equation (2)), we have:

$$\mathcal{F}(DoG_k) = 2\pi(\sigma_k^c)^2 G_{(\sigma_k^c)^{-1}} - 2\pi(\sigma_k^s)^2 G_{(\sigma_k^s)^{-1}}. \quad (6)$$

We can easily verify that for the central Fourier coefficient of each  $DoG_k$  filter we have:

$$\mathcal{F}(DoG_k)(u_0(0), u_0(0)) = 0, \forall k, \quad (7)$$

and that for the other coefficients we have:

$$\mathcal{F}(DoG_k)(i, j) > 0, \forall (i, j) \neq (u_0(0), u_0(0)). \quad (8)$$

This assertion is also verified on Figure 3(b). In order to cover up the centre of the spectrum, we propose to replace the  $DoG_0$  filter, with no change in the notation, by a Gaussian low-pass scaling function consisting in its central component, such that:

$$DoG_0(x, y) = w^c G_{\sigma_0^c}(x, y). \quad (9)$$

Figures 3(a) and 3(b) show the spectrum partitioning with the different  $DoG_k$  filters ( $k \geq 1$ , in blue) and the spectrum of the new scaling function  $DoG_0$  (in red dashed line) which covers low frequencies. With no scaling function, all constant images would be mapped into the null image 0 and this would make the transform be non-invertible. Here we overcome this

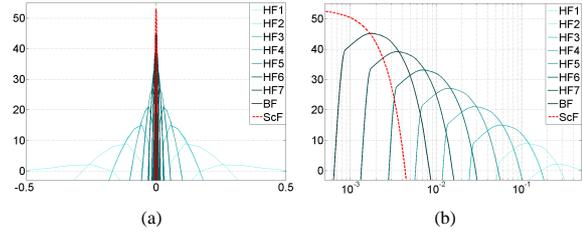


Figure 3: 3(a): Spectrum of the DoG filters. The abscissa represents the frequencies. The ordinate axis represents the different  $DoG_k$  filters gain in dB. 3(b): Half of the spectrum in 3(a) with the abscissa having a logarithmic step. The scaling function  $DoG_0$  is plotted in red dashed line.

problem as the central Fourier coefficient of  $DoG_0$  verifies:

$$\mathcal{F}(DoG_0)(u_0(0), u_0(0)) > 0. \quad (10)$$

The scaling function introduction is further justified by the actual retina behavior. Indeed, the surround component  $G_{\sigma_k^s}$  in Equation (2) appears progressively across time driving the filter passband from low frequencies to higher ones. So that, the Gaussian scaling function  $DoG_0$  represents the very early state of the retina cells.

In order to define an inverse for the new transform, we demonstrate in the following that the set of DoG filters augmented with the  $DoG_0$  scaling function is a “frame”. So that, we will define the adequate notations in Section 3.2 and give the proofs in Section 3.3.

### 3.2 Matrix Notations for ROC

Unlike the implementations in (Van Rullen and Thorpe, 2001; Perrinet et al., 2004; Masmoudi et al., 2010), we use a matrix  $\Phi$  to compute the image transform through the modelled retina. The lines of  $\Phi$  are the different analyzing  $DoG_k$  filters. This yields an “undersampled Toeplitz-bloc” sparse matrix (see Figure 4(a)). Such an implementation allows fast computation of the multi-scale retinal transform through sparse matrix specific algorithms (Golub and Van Loan, 1996). This will in addition help us construct the dual frame of  $\Phi$ . The DoG transform is outlined in the following equation:

$$c = \Phi f. \quad (11)$$

Interestingly, the straightforward synthesis as defined in (5) amounts to the multiplication of the vector output by  $\Phi^*$  the Hermitian transpose of  $\Phi$  (see Figure 4(b)). The reconstruction procedure is then outlined in the following equation:

$$\tilde{f}_{N_s} = \Phi^* c. \quad (12)$$

Though small, the synthesis yields an error that limits the ability of the model to be used as a codec.

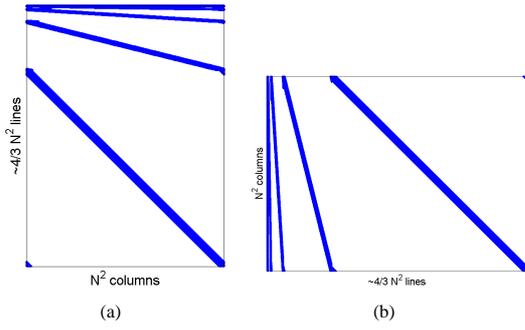


Figure 4: Template of the DoG analysis and synthesis matrices ( $\Phi$  and  $\Phi^*$ ). In this paper,  $\Phi$  and  $\Phi^*$  are represented as matrices where blue dots correspond to non-zero elements. Note here that  $\Phi$  and  $\Phi^*$  are highly sparse matrices.

### 3.3 The DoG Transform is a Frame Operator

Our aim is to prove that the bio-inspired retina image transform presented amounts to a projection of the input image  $f$  onto a frame of a vector space. The frame is a generalization of the idea of a basis to sets which may be linearly dependent (Duffin and Schaeffer, 1952; Kovacevic and Chebira, 2008). These frames allow a redundant signal representation which, for instance, can be employed for coding with error resilience. By proving the frame nature of this transform, we will be able to achieve an exact reverse transform through the construction of a dual frame.

A set of vectors is a frame if it verifies the so-called ‘‘frame condition’’ (Duffin and Schaeffer, 1952; Kovacevic and Chebira, 2008) which states that  $\exists \beta \geq \alpha > 0$  such that :

$$\alpha \|f\|^2 \leq \sum_{kij \in \Gamma} (c_{kij})^2 \leq \beta \|f\|^2, \forall f. \quad (13)$$

**Positioning with Respect to the State of the Art.** Pyramid architectures are very common in signal processing and involve a wide range of filters (Van Rullen and Thorpe, 2001; Do and Vetterli, 2003; Rakshit and Anderson, 1995). Some works in the literature made the connection between the pyramidal architecture and frames. For example, in (Rakshit and Anderson, 1995) the authors proved experimentally that the classical Laplacian pyramid is a frame. However, in our case, we prove that the pyramid introduced in (Van Rullen and Thorpe, 2001) -which is not Laplacian- is a frame. We showed this mathematically through an original demonstration. Also, in (Do and Vetterli, 2003) the authors proposed the design of a set of orthogonal vectors inspired from the Laplacian pyramid to conceive a new orthogonal and tight ( $\alpha = \beta$  cf. Eq. (13)) frame. The filter bank defined

from (Van Rullen and Thorpe, 2001) form a frame that is neither orthogonal nor tight.

**Upper Bounding.** Let us prove the upper bounding in Equation (13). We have:

$$\sum_{kij \in \Gamma} (c_{kij})^2 = \sum_{k=0}^{K-1} \|B_k\|^2, \quad (14)$$

where  $B_k$  is the subband of scale  $k$  generated by the image transform with:

$$B_k(i, j) = \sum_{x=u_k(i)-M_k, y=u_k(j)-M_k}^{x=u_k(i)+M_k, y=u_k(j)+M_k} DoG_k(u_k(i) - x, u_k(j) - y) f(x, y).$$

If we denote by  $U_k$  the undersampling operator corresponding to the function  $u_k$  (cf. Equation (3)), we can write the following:

$$B_k = U_k(DoG_k * f). \quad (15)$$

Then, we have the following obvious inequalities:

$$\begin{aligned} \|B_k\| &= \|U_k(DoG_k * f)\| \leq \|U_k(|DoG_k| * |f|)\| \\ &\leq \| |DoG_k| * |f| \| \\ &\leq \|DoG_k\| \|f\|. \end{aligned}$$

Then, we get back to (14) and infer the following bounding:

$$\begin{aligned} \sum_{kij \in \Gamma} (c_{kij})^2 &= \sum_{k=0}^{K-1} \|B_k\|^2 \leq \left( \sum_{k=0}^{K-1} \|DoG_k\|^2 \right) \|f\|^2 \\ &= \beta \|f\|^2. \end{aligned} \quad (16)$$

**Lower Bounding.** Now, let us demonstrate the lower bounding in Equation (13). We start from the fact that:

$$\sum_{k=0}^{K-1} \|B_k\|^2 \geq \|B_{K-1}\|^2 + \|B_0\|^2, \quad (17)$$

which amounts to write the following inequalities:

$$\begin{aligned} &\sum_{kij \in \Gamma} (c_{kij})^2 \\ &= \sum_{k=0}^{K-1} \|B_k\|^2 \\ &\geq \|DoG_{K-1} * f\|^2 \\ &+ \|(DoG_0 * f)(u_0(0), u_0(0))\|^2 \\ &= \|\mathcal{F}(DoG_{K-1}) \mathcal{F}(f)\|^2 \\ &+ \|\mathcal{F}(DoG_0) \mathcal{F}(f)(u_0(0), u_0(0))\|^2 \\ &= \sum_{i,j=0}^{N-1} (\mathcal{F}(DoG_{K-1})(i, j) \mathcal{F}(f)(i, j))^2 \\ &+ \|\mathcal{F}(DoG_0)(u_0(0), u_0(0)) \mathcal{F}(f)(u_0(0), u_0(0))\|^2, \end{aligned}$$

where  $\mathcal{F}$  designates the discrete Fourier transform. We know that  $\mathcal{F}(DoG_{K-1})(i, j) > 0, \forall (i, j) \neq (u_0(0), u_0(0))$  and that  $\mathcal{F}(DoG_{K-1})(u_0(0), u_0(0)) = 0$ . We also have  $\mathcal{F}(DoG_0)(u_0(0), u_0(0)) > 0$  (cf. Section 2.1). So, if we define a set  $S_{K-1}$  by  $S_{K-1} = \llbracket 0, N-1 \rrbracket^2 \setminus (u_0(0), u_0(0))$  and  $\alpha$  by:

$$\alpha = \min \left\{ \mathcal{F}(DoG_0)^2(u_0(0), u_0(0)), \left\{ \mathcal{F}(DoG_{K-1})^2(i, j), (i, j) \in S_{K-1} \right\} \right\} > 0,$$

then we get the following:

$$\begin{aligned} & \sum_{i,j=0}^{N-1} (\mathcal{F}(DoG_{K-1})(i, j) \mathcal{F}(f)(i, j))^2 \\ & + \|\mathcal{F}(DoG_0)(u_0(0), u_0(0)) \mathcal{F}(f)(u_0(0), u_0(0))\|^2 \\ = & \sum_{i,j \in S_{K-1}} (\mathcal{F}(DoG_{K-1})(i, j) \mathcal{F}(f)(i, j))^2 \\ & + \|\mathcal{F}(DoG_0)(u_0(0), u_0(0)) \mathcal{F}(f)(u_0(0), u_0(0))\|^2 \\ \geq & \alpha \sum_{i,j \in \llbracket 0, N-1 \rrbracket^2} (\mathcal{F}(f)(i, j))^2 \\ = & \alpha \|f\|^2, \end{aligned}$$

and finally,  $\sum_{kij \in \Gamma} (c_{kij})^2 \geq \alpha \|f\|^2$ . Thus, the set of DoG filters satisfies the frame condition (13).

### 3.4 Synthesis using the Dual DoG Frame

We introduce in this section a correction means for the reconstruction error in the retina model presented through the frame theory.

The straightforward analysis/synthesis procedure can be outlined in the relation between the input image and the reconstruction estimate:

$$\tilde{f}_{N_s} = \Phi^* \Phi f. \quad (18)$$

As we already demonstrated that the DoG transform is a frame,  $\Phi^* \Phi$  is said to be the frame operator. To have an exact reconstruction of  $f$ , one must construct the dual DoG vectors. A preliminary step is to compute  $(\Phi^* \Phi)^{-1}$ , the inverse frame operator. We then get a corrected reconstruction  $f_{N_s}^*$ , defined by:  $f_{N_s}^* = (\Phi^* \Phi)^{-1} \tilde{f}_{N_s}$ . If  $N_s$  is the total number of the retina model cells, we have:

$$\begin{aligned} f_{N_s}^* &= (\Phi^* \Phi)^{-1} \tilde{f}_{N_s} \\ &= (\Phi^* \Phi)^{-1} \Phi^* c \\ &= (\Phi^* \Phi)^{-1} \Phi^* \Phi f \\ f_{N_s}^* &= f. \end{aligned} \quad (19)$$

As made clear through Equation (19), the dual vectors are the lines of  $(\Phi^* \Phi)^{-1} \Phi^*$ . Besides,  $(\Phi^* \Phi)$  is a square, definite positive invertible matrix (Kovacevic and Chebira, 2008). This is a crucial issue as it ensures the exactness of the reverse frame operator.

We compute the reverse frame operator  $(\Phi^* \Phi)^{-1}$  and get the results shown in Figure 5. The reconstruction obtained by the means of the dual frame operator is accurate and requires only a simple matrix multiplication. In this example, reconstruction quality is evaluated to 295 dB of PSNR.

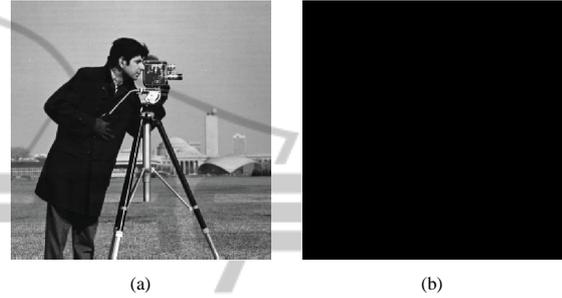


Figure 5: Result of the decoding procedure with the dual DoG frame using the totality of the retina cells. 5(a): The reconstructed image  $f_{N_s}^*$ . The PSNR quality measure of  $f_{N_s}^*$  yields 295 dB. 5(b): The error image in logarithmic scale. This shows that the reconstruction using the dual frame is very precise.

Dual vectors resemble the DoG analyzing filters. This is obvious as the straightforward image reconstruction  $\tilde{f}_{N_s}$  is already close to  $f$ , which means that  $\Phi^* \Phi$  is close to identity. However, the dual filters lose the symmetry property of the primal ones. An example of dual vectors constructed as the rows of  $(\Phi^* \Phi)^{-1} \Phi^*$  is shown in Figure 3.4. Figure 3.4 shows also that the exact reconstruction of  $f$  is obtained by a slight relaxation in the symmetry constraint of the DoG filters.

#### 3.4.1 The Recursive Out-of-core Blockwise Inversion Algorithm

Though the mathematical fundamentals underlying this work are simple, the implementation of such a process is a hard problem. In spite of the sparsity of  $\Phi$  and  $\Phi^*$ , the frame operator  $\Phi^* \Phi$  is an  $N^4$ -sized dense matrix for an  $N^2$ -sized image  $f$ . For instance, if  $N = 257$ ,  $\Phi^* \Phi$  holds in 16 Gbytes, and 258 Gbytes if  $N = 513$ . As noticed in (Do & Vetterli 2003): *A key observation is that one should use the dual frame operator for the reconstruction. While this seems a somewhat trivial observation, it has not been used in practice, probably because the usual reconstruction, while suboptimal, is very simple.* Indeed due to its

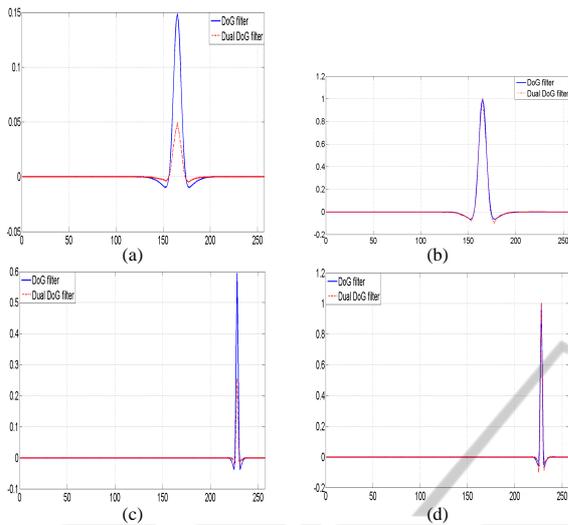


Figure 6: 6(a)- 6(c): Comparison between the DoG filters (in blue line) and their duals (in red dashed line). 6(b)-6(d): Same as previous with the highest value of each filter normalized to 1. Though close in shape to the DoG filters, the dual DoG filters are asymmetric.

technical difficulty, there is no solution in the literature that computes explicitly the dual of a frame in a general case like ours. In our work, we tackled this technical issue and resolved it with success by designing an original “out-of-core” inversion algorithm.

The frame operator  $\Phi^*\Phi$  is constructed bloc by bloc, and each bloc is stored separately on disk. The inversion is then performed using a recursive algorithm that relies on the blockwise matrix inversion formula that follows:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BQ^{-1}CA^{-1} & -A^{-1}BQ^{-1} \\ -Q^{-1}CA^{-1} & Q^{-1} \end{pmatrix}, \quad (20)$$

where  $Q$  is the Schur complement of  $A$ , such that:

$$Q = D - CA^{-1}B.$$

Thus, inverting a matrix amounts to the inversion of two matrices,  $A$  and  $D$  (cf. Equation (20)), that are 4 times smaller. The inversion consists then in subdividing the problem by a factor 4 at each recursion level until we reach a single bloc problem. Thereby, we invert the single bloc remaining matrix using a classical Gauss-Jordan elimination procedure. Obviously, this algorithm requires out-of-core blockwise matrix routines for multiplication, subtraction and addition, that we implemented in a “multi-threaded” fashion to accelerate the computation.

**Advantages of our Approach.**  $(\Phi^*\Phi)$  is a square, definite positive, and invertible matrix (Kovacevic and Chebira, 2008). Thus  $(\Phi^*\Phi)^{-1}$  exists and obviously the exact reverse transform of  $\Phi$  too. Another advantage of our method is that  $(\Phi^*\Phi)$  is well conditioned, with a conditioning number estimated to around 16, so that its inversion is stable. This is a crucial issue as previous work aimed at conceiving the DoG reverse transform tried to invert the original filter bank with no scaling function  $DoG_0$  (Bhattacharya and Furber, 2007; Bhattacharya and Furber, 2010). This is obviously mathematically incorrect as the filter bank thus defined is not a frame and thus its pseudo inverse  $(\Phi^*\Phi)^{-1}\Phi^*$  does not exist. The solution proposed by the authors of (Bhattacharya and Furber, 2007; Bhattacharya and Furber, 2010) gives only a least squares solution to an ill-conditioned problem. Our method instead is stable. Besides through the out-of-core algorithm that we designed we can invert  $(\Phi^*\Phi)$  even for large images whereas authors in (Bhattacharya and Furber, 2007; Bhattacharya and Furber, 2010) are restricted to a maximum size of  $32 \times 32$ . Indeed, we were able to reconstruct  $257 \times 257$  and  $513 \times 513$  images through our new approach.

The exactness of our decoding schema is confirmed when applied on several classical test images. For example *cameraman* reconstruction quality increases from 27 dB with the classical decoder to 296 dB with ours (see Figure 7). The same test on *Lena* leads to an increase from 31 to 300 dB of PSNR (see Figure 8). Numerically, the only limitation comes from the machine precision which is finite. Given this fact, and with a 64-bit machine precision, a value of PSNR that is greater than 290 dB means that the reconstruction is numerically exact. We also confirm these results by using quality metrics that are more consistent than PSNR with the human eye perception. Here we show the mean structural similarity measure (SSIM) (Wang et al., 2004) (an index between 0 and 1) which also confirms the precision of our new decoder with an increase in quality from 0.9 to 1 when all the retina cells have fired (Figures 7 and 8 captions).

Furthermore, correcting the reconstruction errors using the adequate dual frame does not alter the coding procedure. Indeed, methods introduced in (Perinet et al., 2004; Bhattacharya and Furber, 2010) are based on the matching pursuit (MP) algorithm. MP is time consuming and depends on the order in which the “match and update” mechanism is performed. Our method keeps the coding procedure straightforward, multi-threadable and order-independent.

Besides, the results presented here are also beneficial to the image coders that we previously designed



Figure 7: Reconstruction of the cameraman image  $f$  using different percentages of significant coefficients. To get this reconstruction coefficients are set to 0 if under a descending threshold. The left column shows the progressive  $\tilde{f}_{N_s}$  synthesis. The right column shows  $f_{N_s}^*$ . PSNR for the left/right image is from top to bottom: (19.2 dB/19.5 dB), (20.4 dB/20.8 dB), (24.08 dB/25 dB), (25.8 dB /27.5 dB), and (27.9 dB/296 dB). The mean SSIM for the upper/lower image is from top to bottom: (0.49 /0.57), (0.56 /0.61), (0.72 /0.76), (0.77 /0.81), and (0.91 /1).

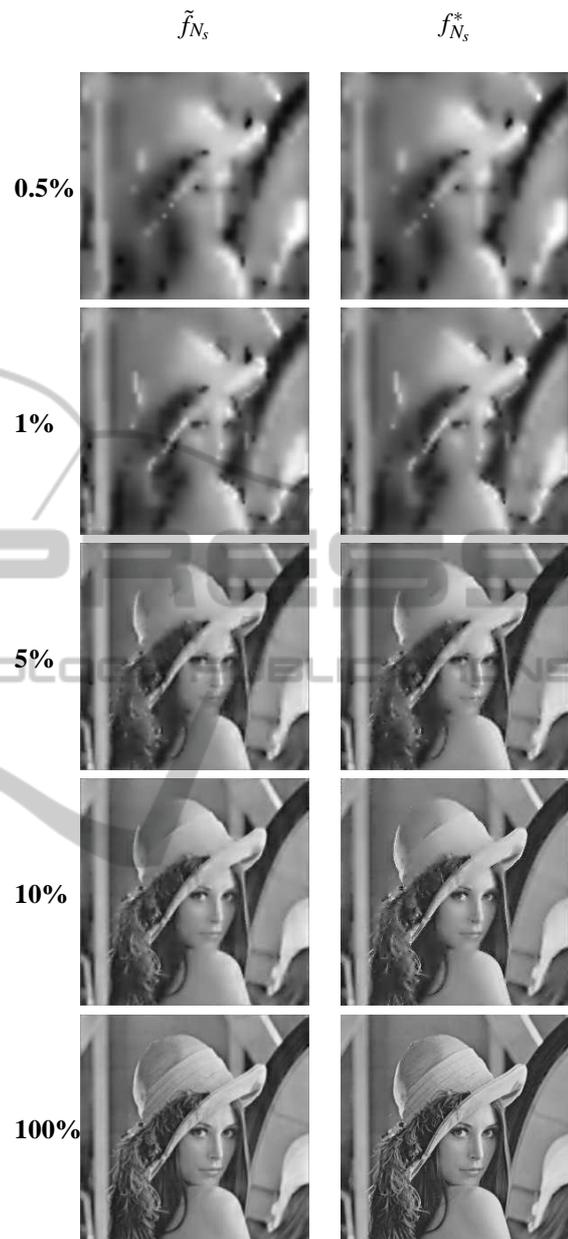


Figure 8: Reconstruction of the Lena image  $f$  using different percentages of significant coefficients. To get this reconstruction coefficients are set to 0 if under a descending threshold. The left column shows the progressive  $\tilde{f}_{N_s}$  synthesis. The right column shows  $f_{N_s}^*$ . PSNR for the left/right image is from top to bottom: (20.06 dB/20.17 dB), (21.73 dB/21.83 dB), (25.93 dB/26.65 dB), (27.97 dB /28.84 dB), and (31.23 dB/299.60 dB). The mean SSIM for the upper/lower image is from top to bottom: (0.38 /0.38), (0.46 /0.46), (0.69 /0.72), (0.77 /0.81), and (0.90 /1).

and that are based on this same retina model (Masmoudi et al., 2010; Masmoudi et al., 2011).

## 4 RESULTS

We experiment our new decoder in the context of scalable image decoding. We reconstruct the test image using an increasing number  $N_s$  of significant coefficients (cf. Equation (5)). We then compare the results when using the original DoG filters in  $\Phi$  and their dual DoG filters in  $(\Phi^*\Phi)^{-1}\Phi^*$  for the decoding procedure.

Figures 7 and 8 show two example results obtained for *cameraman* and *Lena*. In both figures the left column shows the progressive straightforward reconstruction  $\tilde{f}_{N_s}$  and the right column shows the corrected progressive reconstruction  $f_{N_s}^*$  using the dual frame.

Qualitatively speaking, the high frequencies which were the most altered by the straightforward synthesis are now well rendered. This is obvious even for low rates and we can verify it in the camera contours (Figure 7) or in *Lena* hair details (Figure 8). This is an important issue as contours and salient points are the most important features used for categorization tasks. Bearing in mind that the retina model under study was first designed for fast categorization, our results become crucial.

Quantitatively speaking, the gain in PSNR is significant for low rates (around 0.2 dB for both cases) and very high when we consider the totality of the retina cells in the reconstruction (over 260 dB for both cases). Figure 9 compares the Rate/quality curves of the two methods and shows the high improvement we obtain. Here the rate is implicitly related to the number of neurons taken into account for the image reconstruction. In all the curves shown the abscissa represents the percentage of the highest responses used for the reconstruction and the ordinate represents the reconstruction quality in terms of PSNR. This figure shows that the PSNR gain grows exponentially with the number of neurons taken into account for the reconstruction. This means that the dual frame correction, though already significant for low rates, becomes extremely important for high rates.

## 5 DISCUSSION

We proposed in this work an original exact decoding procedure for the classical rank order coder defined in (Van Rullen and Thorpe, 2001). The authors has then designed a bio-inspired retina model for the image transform and reconstruction. Our contribution encompasses a theoretical and a technical aspect.

Regarding the theoretical aspect, (i) we proved that the bio-inspired transform used to model the

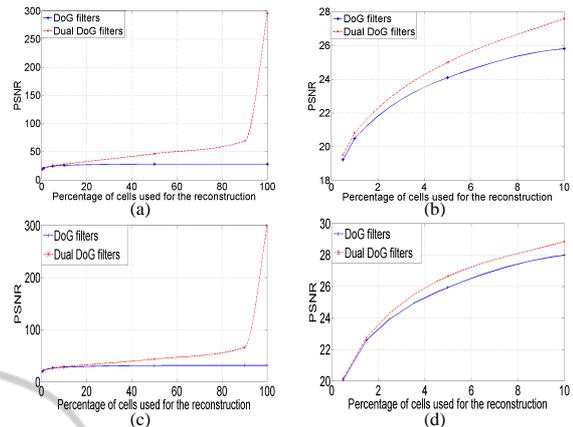


Figure 9: PSNR quality of the reconstruction using the DoG filters (Van Rullen and Thorpe, 2001) (in blue solid line) and the dual DoG filters (in red dashed line). Results for *cameraman* are shown in the first line and results for *Lena* are shown in the second line. The abscissa represents the percentage of the highest responses used for the reconstruction. The ordinate represents the reconstruction quality in terms of PSNR. 9(a)- 9(c): Results shown for percentages between 0% and 100%. 9(b)- 9(d): Results shown for percentages between 0% and 10%.

retina in (Van Rullen & Thorpe 2001) is non-invertible as it is, and (ii) we gave an original mathematical proof that this transform if augmented with an adequate scaling function is a frame. We also showed that the scaling function besides its mathematical justification has a biological one. We then defined the corresponding dual frame that is necessary for the exact image reconstruction.

Regarding the technical aspect, we overcame the problem of memory overhead encountered while computing this dual frame. Up to our knowledge, no work in the literature concerned with high dimensionality frame inversion tackled explicitly this problem (Rakshit and Anderson, 1995; Do and Vetterli, 2003). Indeed usual reconstruction algorithms avoid such a calculation by using inaccurate, though very simple, methods. Thus we designed an original recursive out-of-core blockwise algorithm. Our algorithm is general and could be used in a variety of applications requiring a high dimension matrix inversion.

Furthermore, the method presented in this paper does not alter the coding procedure and keeps it straightforward unlike the stimulus-dependent MP methods in (Perrinet et al., 2004; Bhattacharya and Furber, 2010). In fact, MP-like algorithms require the computation of a specific reconstruction filter bank for each specific image. On the contrary, in our case the dual vectors used for the reconstruction (i) are computed once for all and (ii) are the same for all images. This keeps the decoding procedure committed to the rank order coding philosophy. Yet, the rank or-

dering supposes that the retina cells are independent and fire asynchronously.

One last major advantage is that our algorithm is multi-threadable. Indeed, there is no possible data hazard in the decoding procedure. Each value of  $f_{N_s}^*$  is independent from the others and the concurrent reading in  $\Phi^* \Phi$  does not alter the data.

Though it is to be noted that, in some implementations, the rank order decoder inaccuracy is enhanced for a supplemental reason: the inexactness of the look-up-table that might be used to re-generate the transform coefficients  $c_p$ . In this work and for a sake of clarity, we considered only the filters overlap as a source of error. Otherwise the reader could not distinguish the part of error due to the filters overlap and the other part that is due to the look-up-table. In our case, the decoder is supposed to be provided with an optimal look-up table. Still the approach presented remains (i) relevant because the inaccuracy of any look-up-table that might be used will affect both the "classical" reconstruction and the "dual frame", (ii) novel through the introduction of the frames theory and (iii) general and thus could be extended to several models of cortical areas using redundant representations.

## REFERENCES

- Bhattacharya, B. S. and Furber, S. (2007). Maximising information recovery from rank-order codes. In *Proceedings of SPIE*, volume 6570, Orlando, FL, U.S.A.
- Bhattacharya, B. S. and Furber, S. (2010). Biologically inspired means for rank-order encoding images: A quantitative analysis. *IEEE Trans. Neural Netw.*, 21(7):1087–1099.
- Burt, P. and Adelson, E. (1983). The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.*, 31(4):532–540.
- Do, M. and Vetterli, M. (2003). Framing pyramids. *IEEE Trans. Signal Process.*, 51(9):2329–2342.
- Duffin, R. J. and Schaeffer, A. C. (1952). A class of non-harmonic fourier series. *Transactions of the American Mathematical Society*, 72(2):pp. 341–366.
- Field, D. (1994). What is the goal of sensory coding? *Neural Comput.*, 6(4):559–601.
- Gollisch, T. and Meister, M. (2008). Rapid neural coding in the retina with relative spike latencies. *Science*, 319(5866):1108–1111.
- Golub, G. and Van Loan, C. (1996). *Matrix computations*. Johns Hopkins Univ Pr.
- Kovacevic, J. and Chebira, A. (2008). *An introduction to frames*. Foundations and Trends in Signal Processing, Now Pub.
- Masmoudi, K., Antonini, M., Kornprobst, P., and Perrinet, L. (2010). A novel bio-inspired static image compression scheme for noisy data transmission over low-bandwidth channels. In *Proceedings of ICASSP*, pages 3506–3509. IEEE.
- Masmoudi, K., Antonini, M., Kornprobst, P., and Perrinet, L. (2011). A bio-inspired image coder with temporal scalability. In *Proceedings of ACIVS*, pages 447–458. Springer.
- Meister, M. and Berry, M. (1999). The neural code of the retina. *Neuron*, pages 435–450.
- Perrinet, L., Samuelides, M., and Thorpe, S. (2004). Coding static natural images using spiking event times: do neurons cooperate? *IEEE Trans. Neural Netw.*, 15(5):1164–1175.
- Rakshit, S. and Anderson, C. (1995). Error correction with frames: the Burt-Laplacian pyramid. *IEEE Trans. Inf. Theory*, 41(6):2091–2093.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1997). *Spikes: Exploring the Neural Code*. The MIT Press, Cambridge, MA, USA.
- Rodieck, R. (1965). Quantitative analysis of the cat retinal ganglion cells response to visual stimuli. *Vision Research*, 5(11):583–601.
- Thorpe, S. (1990). Spike arrival times: A highly efficient coding scheme for neural networks. In Eckmiller, R., Hartmann, G., and Hauske, G., editors, *Parallel Processing in Neural Systems and Computers*, pages 91–94. Elsevier.
- Thorpe, S. (2010). Ultra-rapid scene categorization with a wave of spikes. In *Biologically Motivated Computer Vision*, pages 335–351. Springer.
- Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381:520–522.
- Van Rullen, R. and Thorpe, S. (2001). Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex. *Neural Comput.*, 13:1255–1283.
- Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.