

- [7] G. Flitton, T. Breckon, and N. M. Bouallagu, "Object recognition using 3-D SIFT in complex CT volumes," in *Proc. British Mach. Vis. Conf.*, 2010, pp. 11.1–11.12.
- [8] M. Jenkin and J. K. Tsotsos, "Applying temporal constraints to the dynamic stereo problem," *Comput. Vis. Graph. Image Process.*, vol. 33, no. 1, pp. 16–32, Jan. 1986.
- [9] A. M. Waxman and J. H. Duncan, "Binocular image flows: Steps toward stereo-motion fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 715–729, Nov. 1986.
- [10] L. Zhang, B. Curless, and S. Seitz, "Spacetime stereo: Shape recovery for dynamic scenes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2003, pp. 367–374.
- [11] M. Gong, "Enforcing temporal consistency in real-time stereo estimation," in *Proc. Eur. Conf. Comput. Vis.*, May 2006, pp. 564–577.
- [12] F. Huguet and F. Devernay, "A variational method for scene flow estimation from stereo sequences," in *Proc. 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–7.
- [13] O. Williams, M. Isard, and J. MacCormick, "Estimating disparity and occlusions in stereo video sequences," in *Proc. Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2005, pp. 250–257.
- [14] T. Delbruck, B. Linares-Barranco, E. Culurciello, and C. Posch, "Activity-driven, event-based vision sensors," in *Proc. Int. Symp. Circuits Syst.*, Paris, France, May–Jun. 2010, pp. 2426–2429.
- [15] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 dB 15  $\mu$ s latency asynchronous temporal contrast vision sensor," *IEEE J. Solid State Circuits*, vol. 43, no. 2, pp. 566–576, Feb. 2008.
- [16] M. Mahowald and T. Delbruck, "Cooperative stereo matching using static and dynamic image features," in *Analog VLSI Implementation of Neural Systems*, C. M. Ismail, Ed. Boston, MA: Kluwer, 1989, pp. 213–238.
- [17] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, no. 4262, pp. 283–287, 1976.
- [18] M. Mahowald, *An Analog VLSI System for Stereoscopic Vision*. Boston, MA: Kluwer, 1994.
- [19] E. K. C. Tsang and B. E. Shi, "A neuromorphic multi-chip model of a disparity selective complex cell," in *Advances in Neural Information Processing Systems*, vol. 16. Cambridge, MA: MIT Press, 2004.
- [20] J. Kogler, C. Sulzbachner, and W. Kubinger, "Bio-inspired stereo vision system with silicon retina imagers," in *Proc. 7th Int. Conf. Comput. Vis. Syst.*, 2009, pp. 174–183.
- [21] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Seattle, WA, Jun. 1994, pp. 593–600.
- [22] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
- [23] A. Yuille and T. Poggio, "A generalized ordering constraint for stereo correspondence," Artificial Intelligence Laboratory, MIT, Cambridge, Memo Rep. 777, 1984.
- [24] R. Benosman, S.-H. Ieng, P. Rogister, and C. Posch, "Asynchronous event-based Hebbian epipolar geometry," *IEEE Trans. Neural Netw.*, vol. 22, no. 11, pp. 1723–1734, Nov. 2011.
- [25] J. V. Arthur and K. A. Boahen, "Synchrony in silicon: The gamma rhythm," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1815–1824, Nov. 2007.
- [26] G. Indiveri, E. Chicca, and R. Douglas, "A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 211–221, Jan. 2006.
- [27] G. Ermentrout, "Synchronization in a pool of mutually coupled oscillators with random frequencies," *J. Math. Biol.*, vol. 22, no. 1, pp. 1–9, 1985.
- [28] M. Abeles, "Role of the cortical neuron: Integrator or coincidence detector?" *Isr. J. Med. Sci.*, vol. 18, no. 1, pp. 83–92, Jan. 1982.
- [29] W. Softky and C. Koch, "The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs," *J. Neurosci.*, vol. 13, no. 1, pp. 334–350, Jan. 1993.
- [30] J. Bullier, "The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps," *Brain Res. Rev.*, vol. 36, nos. 2–3, pp. 96–107, Jan. 2001.

## Frames for Exact Inversion of the Rank Order Coder

Khaled Masmoudi, *Student Member, IEEE*,  
 Marc Antonini, *Member, IEEE*, and Pierre Kornprobst

**Abstract**—Our goal is to revisit rank order coding by proposing an original exact decoding procedure for it. Rank order coding was proposed by Thorpe *et al.* who stated that the order in which the retina cells are activated encodes for the visual stimulus. Based on this idea, the authors proposed in [1] a rank order coder/decoder associated to a retinal model. Though, it appeared that the decoding procedure employed yields reconstruction errors that limit the model bit-cost/quality performances when used as an image codec. The attempts made in the literature to overcome this issue are time consuming and alter the coding procedure, or are lacking mathematical support and feasibility for standard size images. 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 twofold. First, 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. Second, to deal with the problem of memory overhead, we design a 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 defines a simple and exact reverse transform for it with over than 265 dB of increase in the peak signal-to-noise ratio quality compared to [1]. Furthermore, the framework presented here can be extended to several models of the visual cortical areas using redundant representations.

**Index Terms**—Bio-inspired image coding, frames theory, out-of-core, rank order code, scalability.

### I. INTRODUCTION

Neurophysiologists made substantial progress in better understanding the early processing of the 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 [2]–[4]. Thus, our aim is to use the computational neuroscience models that mimic the retina behavior to design novel lossy coders for static images. In this brief, we assume that the retina encodes the visual information by the order in which its ganglion cells react to the stimulus - recalling that these cells react through the emission of electrical impulses (the spikes). This choice was motivated by Thorpe *et al.* neurophysiologic results on ultrarapid stimulus categorization [2], [5]. The authors showed that still image classification can be achieved by the visual cortex within very short latencies of about 150 ms. 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.

Manuscript received May 20, 2011; accepted December 3, 2011. Date of publication December 27, 2011; date of current version February 8, 2012.

K. Masmoudi and M. Antonini are with I3S Laboratory-CNRS-UNS, Sophia-Antipolis 06903, France (e-mail: kmasmoud@i3s.unice.fr; am@i3s.unice.fr).

P. Kornprobst is with NeuroMathComp Team Project, INRIA, Sophia-Antipolis 06902, France (e-mail: pierre.kornprobst@sophia.inria.fr).

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

Digital Object Identifier 10.1109/TNNLS.2011.2179557

This code, termed as rank order code (ROC), is at the origin of a classical bio-inspired retina model [1].

However, one major limitation of the ROC coder defined in [1] prevents its use for the design of image codecs. It is the inaccuracy of the proposed decoding procedure. Indeed, the 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 bit-cost/quality performance [6]–[8]. Efforts to correct this issue followed two main approaches. A first one consisted in inverting directly the transform operator to obtain a reverse filter bank as in [7]. The method presented is based on a pseudo-inversion. Though interesting, we will show that this method lacks mathematical support. Besides the procedure used deals with a high dimension matrix and thus is infeasible, as such, for standard size images. A second approach relies on matching pursuit (MP) algorithms as in [6] and [7]. These methods are time consuming and alter the coding procedure. In addition, the MP approach depends on the order in which the “match and update” mechanism is performed, and this makes the coding procedure depend on the stimulus itself.

In this brief, we give an original solution relying on the mathematical concept of “frames” [9]. Frames extend the notion of basis for sets of filters which are linearly dependent. Their use for redundant signal analysis and reconstruction has been experienced in the signal processing literature. For example, authors in [10] proved experimentally that the filters of the classical Laplacian pyramid [11] form a non-orthogonal frame of filters and proposed an algorithm for partial error removal from the reconstructed signal. Also, the authors in [12] designed a new frame of orthogonal filters for signal analysis that is inspired from the Laplacian pyramid too. Strong design constraints imposed on the filters enabled errorless reconstruction.

Though the retinal model defined in [1] is based on a different set of filters (that in addition are not orthogonal), the cited works inspired a solution to our specific case. This brief brings two main contributions. 1) We add to the original retinal filter bank an adequate scaling function, and we provide an original mathematical demonstration that we get a frame. Then we propose an algorithm for errorless reconstruction through the construction of a so-called “dual frame.” 2) We solve the technical issue related to memory overhead that prevented the use of frames for high dimension spaces, with a novel out-of-core algorithm for the computation of the dual frame. Thanks to our new approach, we show that the image that we reconstruct is equal to the original stimulus.

This brief is organized as follows. In Section II, we present the three stages of the rank order coding/decoding method. Then in Section III, we define an exact decoding scheme through the construction of a dual frame. Finally in Section IV, we show the gain that we obtain in terms of bit-cost/quality tradeoff.

## II. RANK ORDER CODEC: THREE STAGES

This section summarizes the three stages of the ROC coding/decoding procedure as defined in [1]. First we present

in Section II-A the image transform as performed by a bio-inspired retina model. We then give the specification of the subsequent rank ordering and decoding procedures in Sections II-B and C.

### A. Image Transform: A 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 have proven that the outer cells processing can be approximated by a linear filtering. In particular, the authors in [13] proposed the largely adopted difference of Gaussian (DoG) filter which is a weighted difference of spatial Gaussians that is defined as

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

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 Fig. 1(b) [1], [6], [8]. As in the retina topology, the cells density and scale are inversely proportional. This keeps the model strongly inspired from the mammals retina, though the authors in [1] 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

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

where  $\sigma_{k+1}^c = (1/2)\sigma_k^c$  and  $\sigma_{k+1}^s = (1/2)\sigma_k^s$ . Each  $DoG_k$  filter has a size of  $(2M_k + 1)^2$ , with  $M_k = 3\sigma_k^s$ . Authors in [1] chose the biologically plausible parameters as estimated in [13]  $w^c = w^s = 1$ ,  $\sigma_k^c = (1/3)\sigma_k^s \forall k$ , 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 step 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:

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

$u_k$  is an undersampling function. We notice that  $u_{K-1}(i) = i$ , and that  $(u_k)_{k \in \llbracket 0, K-2 \rrbracket}$  are undersampled versions of  $u_{K-1}$ .  $c_{kij}$  is then computed as follows:

$$c_{kij} = \sum_{x=u_k(i)-M_k, y=u_k(j)-M_k}^{x=u_k(i)+M_k, y=u_k(j)+M_k} DoG_k(u_k(i) - x, u_k(j) - y) f(x, y). \quad (3)$$

This transform generates a vector  $c$  of  $(4/3N^2 - 1)$  coefficients  $c_{kij}$  for an  $N^2$ -sized image (if  $N$  is a power of 2). This architecture is similar to a Laplacian pyramid [11]. An example of such a transform performed on the *cameraman* test image is shown in Fig. 1.

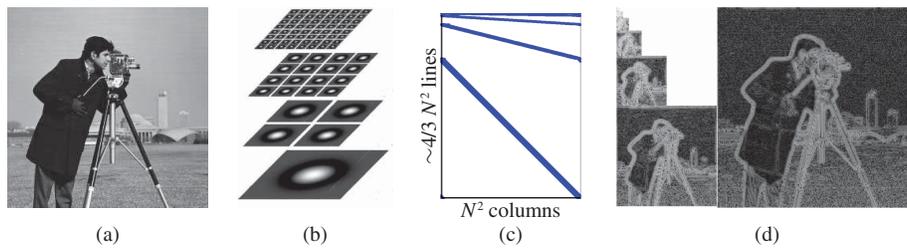


Fig. 1. (a) Illustration of the retinal image transform applied on *cameraman*. Image size is  $257 \times 257$  pixels. (b) Example of a dyadic grid of DoG's used for the image analysis (from [1]). (c) Template of the DoG transform matrix: In this brief, the transform  $\Phi$  is represented as a matrix where blue dots correspond to non-zero elements. (d) Transform result showing the generated subbands in log scale.

### B. Sorting: The Generation of the ROC

Thorpe *et al.* [2], [5] proposed that the order in which the spikes are emitted encodes for the stimulus. This yielded the ROC which relies on the following simplifying assumptions: 1) from stimulus onset, only the first spike emitted by each cell is considered; 2) the time to fire of each cell is proportional to its degree of activation; and 3) only the order of firing encodes for the stimulus.

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 [5]. So, the neurons responses  $(c_{kij})_{kij \in \Gamma}$  defined in (3) are sorted in the decreasing order of their amplitude  $|c_{kij}|$ .

The final output of this stage, the ROC, is then a sorted list of  $N_s$  couples  $(p, c_p)$  such that  $|c_p| \geq |c_{p+1}|$ , with  $p$  being the index of the cell defined by  $p(k, i, j) = k N_k^2 + i N_k + j$  and  $N_k^2$  being the number of cells in the subband  $B_k$ . Here, the generated series  $(p, c_p)_{0 \leq p < N_s}$  is the only data that the coder transmits to the decoder. Note that in some implementations as [1] and [6], 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 brief).

### C. Decoding Procedure of the ROC

We consider the set of the first  $N_s$  highest cell responses forming the ROC of a given image  $f$ . In [1], the authors defined  $\tilde{f}_{N_s}$ , the decoded estimation of  $f$  by

$$\tilde{f}_{N_s}(x, y) = \sum_{p(k,i,j)=0}^{N_s-1} c_p \text{DoG}_k(u_k(i) - x, u_k(j) - y). \quad (4)$$

Equation (4) defines 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 cells of the retina. This feature makes the coder scalable [8].

An example of such a reconstruction is given in Fig. 2(a), with all the retina cells taken into account. Fig. 2(b) 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 peak signal-to-noise ratio (PSNR). This is due to the fact that the DoG filters are almost but not exactly orthonormal. We detail in the next section our original method to reconstruct exactly the input  $f$ .

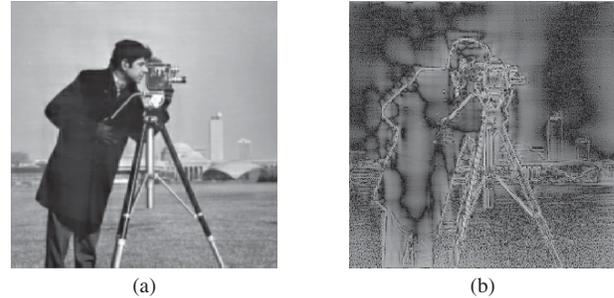


Fig. 2. Result of the decoding procedure with the original approach using all of the retina cells responses. (a) Reconstructed image. The PSNR of  $\tilde{f}_{N_s}$  is 27 dB. (b) Error image in a log-scale.

## III. INVERTING THE BIO-INSPIRED RETINA MODEL

In this section, we define an original and exact image reconstruction algorithm starting from the ROC. First, we introduce in Section III-A a low-pass scaling function in the analyzing filter bank. This modification will be shown to be necessary for the transform reversibility. Then, in Section III-B, we give a matrix-based formalism for the transform and we use it to prove that our filter bank is a frame in Section III-C. Finally, in Section III-D, we show the exact reconstruction results using the dual frame and introduce an out-of-core algorithm to construct it.

### A. 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  $\mathcal{F}(\text{DoG}_k)$  is a DoGs. Therefore, with  $w^c = w^s = 1$  [see (1)], we have

$$\mathcal{F}(\text{DoG}_k) = 2\pi (\sigma_k^c)^2 G_{(\sigma_k^c)^{-1}} - 2\pi (\sigma_k^s)^2 G_{(\sigma_k^s)^{-1}}. \quad (5)$$

We can easily verify that the central Fourier coefficient  $\mathcal{F}(\text{DoG}_k)(u_0(0), u_0(0)) = 0 \forall k$ , and that  $\mathcal{F}(\text{DoG}_k)(i, j) > 0 \forall (i, j) \neq (u_0(0), u_0(0))$ .

In order to cover up the center of the spectrum, we propose to replace the  $\text{DoG}_0$  filter, with no change in the notation, by a Gaussian low-pass scaling function consisting in its central component, such that

$$\text{DoG}_0(x, y) = w^c G_{\sigma_0^c}(x, y). \quad (6)$$

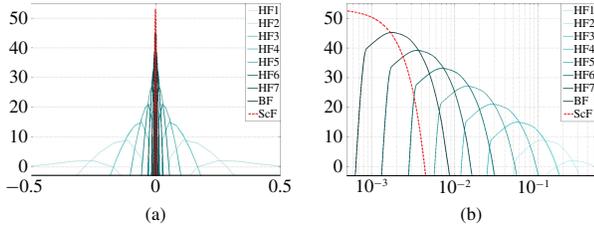


Fig. 3. (a) Spectrum of the DoG filters. The abscissa represents the frequencies. The ordinate represents the  $DoG_k$  filters gain in dB. (b) Half of the spectrum in (a) with the abscissa having a log-step. The scaling function  $DoG_0$  is plotted in red dashed line.

Fig. 3(a) and (b) shows 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 problem as the central Fourier coefficient  $\mathcal{F}(DoG_0)(u_0(0), u_0(0)) > 0$ .

The scaling function introduction is further justified by the actual retina behavior. Indeed, the surround  $G_{\sigma_k^s}$  in (1) appears progressively in time driving the filter passband from low to high frequencies. So that,  $DoG_0$  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.”

### B. Matrix Notations for ROC

Unlike the implementations in [1], [6], and [8], let us introduce the matrix  $\Phi$  to compute the image transform through the retina model. The rows of  $\Phi$  are the different  $DoG_k$  filters. This yields an “undersampled Toeplitz-block” sparse matrix [Fig. 1(c)]. Such an implementation allows fast computation of the multi-scale retinal transform through sparse matrix specific algorithms. This will in addition help us to construct the dual frame of  $\Phi$ .

With this notation, the DoG transform is rewritten

$$c = \Phi f. \quad (7)$$

Interestingly, the straightforward synthesis as defined in (4) amounts to the multiplication of the vector output  $c$  by  $\Phi^*$  the Hermitian transpose of  $\Phi$ . Then, the reconstruction procedure defined in (4) is rewritten

$$\tilde{f}_{N_s} = \Phi^* c. \quad (8)$$

### C. DoG Transform is a Frame Operator

Our aim is to prove that the bio-inspired image transform presented amounts to a projection of the input  $f$  on a frame of a vector space. The frame is a generalization of the idea of a basis to sets which may be linearly dependent [9]. 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.

According to [9], to prove that our transform is a frame we need to show that  $\exists 0 < \alpha \leq \beta < \infty$ , such that

$$\alpha \|f\|^2 \leq \sum_{kij \in \Gamma} (c_{kij})^2 \leq \beta \|f\|^2 \quad \forall f. \quad (9)$$

1) *Positioning With Respect to the State-of-the-Art:* Pyramid architectures are very common in signal processing and involve a wide range of filters, some of which are loosely referred to as DoG. These are of different sorts. For instance, [1], [12], and [10] are three different implementations of pyramids dealing with different types of DoG’s. Note that the term “DoG” may lead to some confusion. Indeed, according to the respective weights and standard deviations, an infinity of filters with different properties may be implemented as the difference of two Gaussians. For example, in [10] the authors proved experimentally that the classical Laplacian pyramid is a frame. However, in our case, we prove that the pyramid introduced in [1] -which is not Laplacian- is a frame. We showed this mathematically through an original demonstration. Also, in [12] the authors proposed the design of a set of orthogonal vectors inspired from the Laplacian pyramid to conceive a new orthogonal and tight [ $\alpha = \beta$  see (9)] frame. The filter bank defined from [1] form a frame that is neither orthogonal nor tight.

*Proposition 3.1:* Let  $\Psi(\phi) = \{\phi_{kij}, (k, i, j) \in \Gamma\}$  be the set of vectors  $\phi_{kij}$ , such that  $\phi_{kij}(x, y) = DoG_k(u_k(i) - x, u_k(j) - y)$  as defined in (1) for  $k > 0$ , and in (6) for  $k = 0$ . Then  $\Psi(\phi)$  is a frame of the  $N \times N$  images vector space.

*Proof:*

*Upper bounding:* We have

$$\sum_{kij \in \Gamma} (c_{kij})^2 = \sum_{k=0}^{K-1} \|B_k\|^2 \quad (10)$$

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

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

If we denote by  $U_k$  the undersampling operator corresponding to the function  $u_k$  [see (2)], we can write

$$B_k = U_k(DoG_k * f). \quad (11)$$

Then, we have the following inequalities:

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

So, with (10) we infer the following bounding:

$$\begin{aligned} \sum_{kij \in \Gamma} (c_{kij})^2 &= \sum_{k=0}^{K-1} \|B_k\|^2 \leq \left( \sum_{k=0}^{K-1} \|DoG_k\|^2 \right) \|f\|^2 \\ &= \beta \|f\|^2 \end{aligned} \quad (12)$$



Fig. 4. Result of the decoding procedure with the dual DoG frame using all of the retina cells responses. (a) Reconstructed image. The PSNR of  $f_{N_s}^*$  is 296 dB. (b) Error image in log-scale.

which shows the first inequality.

*Lower bounding:* We start from the fact that

$$\sum_{k=0}^{K-1} \|B_k\|^2 \geq \|B_{K-1}\|^2 + \|B_0\|^2 \quad (13)$$

which amounts to write the following inequalities:

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

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  $\mathcal{F}(DoG_0)(u_0(0), u_0(0)) > 0$  (see Section III-A). 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:

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

so that,  $\sum_{kij \in \Gamma} (c_{kij})^2 \geq \alpha \|f\|^2$ . ■

#### D. Synthesis Using the Dual DoG Frame

Based on the previous proof, we detail in this section an exact reconstruction procedure for the DoG transform.

The straightforward analysis/synthesis procedure can be outlined in the relation between the input image and the reconstruction estimate:  $\tilde{f}_{N_s} = \Phi^* \Phi f$ . As we already demonstrated that the DoG transform is a projection on 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

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

As made clear through the equation above, the dual vectors are the rows of  $(\Phi^* \Phi)^{-1} \Phi^*$ . If we reconstruct  $f$  starting from the ROC output  $c$  and using the dual frame vectors, we get the results shown in Fig. 4. The reconstruction obtained is accurate and requires only a simple matrix multiplication. In this example, reconstruction quality is evaluated to 296 dB of PSNR.

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 [14].

1) *Recursive Out-of-Core Blockwise Inversion Algorithm:* Though the mathematical fundamentals underlying this brief are simple, the implementation of such a process is a hard problem. Indeed, 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  $N = 513$ . The solution is to recourse to the out-of-core algorithms [15].

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

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1} B Q^{-1} C A^{-1} & -A^{-1} B Q^{-1} \\ -Q^{-1} C A^{-1} & Q^{-1} \end{pmatrix}$$

where  $Q$  is the Schur complement of  $A$ , such that  $Q = D - C A^{-1} B$ . Thus, inverting a matrix amounts to the inversion of two matrices that are four times smaller. The inversion consists then in subdividing the problem by a factor four at each recursion level until we reach a single block problem. 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.

2) *Advantages of Our Approach:*  $(\Phi^* \Phi)^{-1}$  is a square, definite positive, and invertible matrix [9]. Another advantage of our method is that  $(\Phi^* \Phi)^{-1}$  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



Fig. 5. Reconstruction of the *cameraman* image  $f$  using different percentages of the highest  $c_{kij}$ . (a) Straightforward progressive synthesis  $\tilde{f}_{N_s}$ . (b) Corrected reconstruction  $f^*_{N_s}$  using the dual frame. PSNR for the upper/lower image is from left to right: (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 left to right: (0.49 /0.57), (0.56 /0.61), (0.72 /0.76), (0.77 /0.81), and (0.91 /1).

aimed at conceiving the DoG reverse transform tried to invert the original filter bank with no scaling function  $DoG_0$  [7]. 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 in [7] 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  $(\Phi^*\Phi)$  for large images whereas [7] are restricted to a maximum size of  $32 \times 32$ .

Furthermore, correcting the reconstruction errors using the adequate dual frame does not alter the coding procedure. Indeed, methods introduced in [6] and [7] are based on MP. MP mainly relies on the principle of progressive elimination of the filters overlap with a “match and update” routine. Thus MP depends on the order in which the “match and update” mechanism is performed. This amounts to have a specific “reconstruction basis” for each specific image. However, our algorithm provides us with a single frame of vectors (independent of the stimulus) that guarantees the exactness of the reconstruction for all possible images. Our decoding scheme does not implement an *a posteriori* lateral inhibition mechanism but rather supposes the cells to be independent. Our work: 1) points out the reason that makes the retina model in [1] non-invertible as it is and resolve it with the introduction of a low-pass scaling function; and 2) offers an alternative to MP that has the advantage of being mathematically exact for all images and reproducible for other representations through the concept of “frames.” Besides MP, though interesting, remains time consuming. Yet our method keeps the coding procedure straightforward, order-independent, and multi-threadable.

#### IV. COMPARISON TO THE ORIGINAL ROC CODEC

We experiment our new decoder in the context of scalable image decoding. We reconstruct the *cameraman* test image using an increasing number  $N_s$  of highest responses taken into account [see (4) and (8)]. We compare the results when using

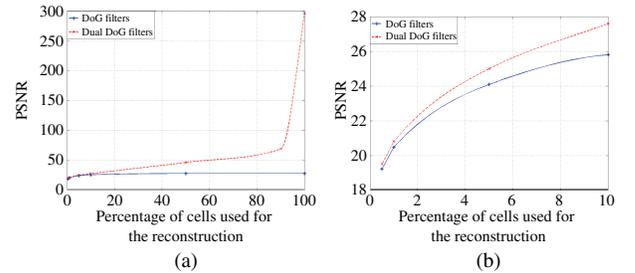


Fig. 6. PSNR quality of the reconstruction using the DoG filters [1] (in blue solid line) and the dual DoG filters (in red dashed line) as a function of the percentage of cells taken into consideration. (a) Results shown for percentages between 0% and 100%. (b) Results shown for percentages between 0% and 10%.

the original DoG filters in  $\Phi^*$  and the dual DoG filters in  $(\Phi^*\Phi)^{-1}\Phi^*$  for the decoding procedure. Fig. 5 summarizes the results obtained, with the upper line showing the progressive straightforward reconstruction  $\tilde{f}_{N_s}$  and the bottom line showing the corrected progressive reconstruction  $f^*_{N_s}$ . The increase in quality measured in PSNR is significant for all bit-rates and goes up with  $N_s$ . For low bit-rates, the PSNR increase is around 0.3 dB. For higher bit-rates, the increase of PSNR reaches almost 270 dB (from 27 to 296 dB) when all the retina cells have fired. Though we do not show it, the exactness of our decoding schema is confirmed when applied on several classical test images. For example *Lena* reconstruction quality increases from 31 dB with the classical decoder to 300 dB with ours. We also confirm these results by using other quality metrics that are more consistent than PSNR with the human eye perception: we show the mean structural similarity measure (SSIM) [16] (an index between 0 and 1) which confirms the precision of our new decoder with an increase in quality from 0.9 to 1 when all the retina cells have fired (Fig. 5 caption).

Fig. 6 compares the bit-cost/quality curves of the two methods and shows the high improvement we obtain. Besides, our method does not alter the coding procedure and keeps it straightforward, fast and order-independent.

## REFERENCES

- [1] R. Van Rullen and S. Thorpe, "Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex," *Neural Comput.*, vol. 13, no. 6, pp. 1255–1283, Jun. 2001.
- [2] S. Thorpe, "Spike arrival times: A highly efficient coding scheme for neural networks," in *Parallel Processing in Neural Systems and Computers*, R. Eckmiller, G. Hartmann, and G. Hauske, Eds. Amsterdam, The Netherlands: Elsevier, 1990, pp. 91–94.
- [3] M. Meister and M. Berry, "The neural code of the retina," *Neuron*, vol. 22, no. 3, pp. 435–450, Mar. 1999.
- [4] T. Gollisch and M. Meister, "Rapid neural coding in the retina with relative spike latencies," *Science*, vol. 319, no. 5866, pp. 1108–1111, Feb. 2008.
- [5] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, pp. 520–522, Jun. 1996.
- [6] L. Perrinet, M. Samuelides, and S. Thorpe, "Coding static natural images using spiking event times: Do neurons cooperate?" *IEEE Trans. Neural Netw.*, vol. 15, no. 5, pp. 1164–1175, Sep. 2004.
- [7] B. S. Bhattacharya and S. Furber, "Biologically inspired means for rank-order encoding images: A quantitative analysis," *IEEE Trans. Neural Netw.*, vol. 21, no. 7, pp. 1087–1099, Jul. 2010.
- [8] K. Masmoudi, M. Antonini, P. Kornprobst, and L. Perrinet, "A novel bio-inspired static image compression scheme for noisy data transmission over low-bandwidth channels," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Dallas, TX, Mar. 2010, pp. 3506–3509.
- [9] J. Kovacevic and A. Chebira, "An introduction to frames," *Foundations and Trends in Signal Processing*. Hanover, MA: Now Publishers, 2008.
- [10] S. Rakshit and C. Anderson, "Error correction with frames: The Burt-Laplacian pyramid," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 2091–2093, Nov. 1995.
- [11] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.
- [12] M. Do and M. Vetterli, "Framing pyramids," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2329–2342, Sep. 2003.
- [13] D. Field, "What is the goal of sensory coding?" *Neural Comput.*, vol. 6, no. 4, pp. 559–601, 1994.
- [14] K. Masmoudi, M. Antonini, and P. Kornprobst, "Frames for exact inversion of the rank order code," INRIA, Le Chesnay, France, Res. Rep. RR-7744, 2011.
- [15] S. Toledo, "A survey of out-of-core algorithms in numerical linear algebra," in *External Memory Algorithms and Visualization*, J. Abello and J. S. Vitter, Eds. Providence, RI: AMS, 1999, pp. 161–180.
- [16] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

### A Globally Convergent MC Algorithm With an Adaptive Learning Rate

Dezhong Peng, *Member, IEEE*, Zhang Yi, *Senior Member, IEEE*,  
Yong Xiang, and Haixian Zhang, *Member, IEEE*

**Abstract**—This brief deals with the problem of minor component analysis (MCA). Artificial neural networks can be exploited to achieve the task of MCA. Recent research works show that

Manuscript received December 30, 2010; revised November 10, 2011; accepted December 3, 2011. Date of publication January 3, 2012; date of current version February 8, 2012. This work was supported in part by the National Basic Research Program of China 973 Program, under Grant 2011CB302201, the National Natural Science Foundation of China, under Grant 60970013 and Grant 61172180, and the Australian Research Council, under Grant DP110102076.

D. Peng, Z. Yi, and H. Zhang are with the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: dezhongpeng@gmail.com; zhangyi@scu.edu.cn; hxzhang530@gmail.com).

Y. Xiang is with the School of Information Technology, Deakin University, Melbourne, VIC 3125, Australia (e-mail: yxiang@deakin.edu.au).

Digital Object Identifier 10.1109/TNNLS.2011.2179310

convergence of neural networks based MCA algorithms can be guaranteed if the learning rates are less than certain thresholds. However, the computation of these thresholds needs information about the eigenvalues of the autocorrelation matrix of data set, which is unavailable in online extraction of minor component from input data stream. In this correspondence, we introduce an adaptive learning rate into the OJAN MCA algorithm, such that its convergence condition does not depend on any unobtainable information, and can be easily satisfied in practical applications.

**Index Terms**—Deterministic discrete time system, eigenvalue, eigenvector, minor component analysis, neural networks.

## I. INTRODUCTION

The eigenvector associated with the smallest eigenvalue of the autocorrelation matrix of data set is called minor component (MC). As an effective method for data analysis, minor component analysis (MCA) is aimed at extracting MC from data set and has many important applications [1]–[3], such as curve and surface fitting, digital beamforming, frequency estimation, moving target indication, and clutter cancellation. Recently, neural network based MCA algorithms have received considerable research interests [11]–[13]. Although the batch MCA methods, e.g., power algorithms, which usually depend on the computation of the correlation matrix of inputs, are more efficient and can achieve better performance than the online neural networks algorithms [20], the online algorithms do not need to compute and save the correlation matrix and only deal with the computation of vectors and scalars, which results in that the online neural networks algorithms have the lower storage requirement than the batch algorithms.

So far, numerous neural networks learning algorithms have been developed to solve the MCA problem [6]–[14]. For these MCA algorithms, convergence is essential and considerable research has been conducted to analyze their convergence properties. Since direct convergence analysis is rather difficult for stochastic discrete time (SDT) systems describing the MCA algorithms, some indirect dynamics study methods are referable. In the existing literature, deterministic continuous time (DCT) method and deterministic discrete time (DDT) method are two widely-used analysis approaches for the stochastic learning algorithms. The DCT method transforms the original SDT systems into their corresponding DCT systems, based on the stochastic approximation theory [4]. Then the convergence properties of MCA neural networks algorithms are indirectly attained by studying the dynamics of these DCT systems [5]–[9]. The DDT method exploits the conditional expectation to transform the original SDT system describing the learning algorithm into a corresponding DDT system. Through studying dynamics of the obtained DDT system, one can indirectly investigate the convergence properties of the original SDT system. Both the DCT method and the DDT method indirectly shed some light on the convergence characteristics of the original stochastic algorithms.

Both the DDT method and the DCT method show that the selection of the learning rate plays an essential role in the dynamical behaviors of algorithms. In order to guarantee convergence of algorithms, the learning rate is usually required