ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 313-318

Sparse Image Coding Using an Asynchronous Spiking Neural Network

Laurent Perrinet,* Manuel Samuelides
ONERA/DTIM, 2, av. Belin, 31055 Toulouse, France

Abstract.
In order to explore coding strategies in the retina, we use a wavelet-like transform which output is sparse, as is observed in biological retinas [4]. This transform is defined in the context of a one-pass feed-forward spiking neural network, and the output is the list of its neurons' spikes: it is recursively constructed using a greedy matching pursuit scheme which first selects higher contrast energy values. As in [7], we find invariants in the output for some classes of images, allowing to code the absolute contrast value solely by its rank in the spike list. An application to image compression is shown which is comparable to other techniques such as JPEG at low bit compression.

1 Introduction

1.1 What is the goal of retinal coding?

Despite the intensive research in neuroscience on the retina, image processing strategies haven't seen any major revolutions that permitted for artificial retinas to compare with their biological counterparts. So far, it is assumed that the goal of the retinal coding strategy is to transmit as much information from the 10^8 photoreceptors to the brain through the 'informational bottleneck' of the 10^6 of axons of the ganglion cells (GC) which form the fibers of the optical nerve.
Among proposed strategies, dimension reduction (PCA), blind source separation (e.g. ICA) and sparse coding [4] are the most successful. This last method suggests that the code could consist of a relatively small number of active spiking GC if they form an *overcomplete basis*, but it fails so far to use the temporal aspect of retinal processing. In fact, this aspect seems crucial and recently, ultra rapid categorization [6] was shown in humans and monkeys and urged the computational neuroscience community to explore new coding strategies accounting for the consequences of those experiences. To gain advantage over the speed of retinal processing, the code should convert the analog intensities into a 'wave' of spikes in less than $20\ ms$, the most 'important' spikes

*Corresponding author, e-mail: perrinet@cert.fr

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 313-318

being fired first. This defines a new goal for the retinal code: the analog image should be temporally transformed so that the spike list transmits progressively the most information with the shortest latency.

1.2 Retinal image processing

Although the retinal neural network is a complex architecture of several layers, the processes from the light influx detected by the photoreceptors to the potential at the soma of the GCs are chemical and no spike occur. Therefore, the transform is essentially linear and a typology may be drawn from the response of GCs to light stimulation [5]. Grossly, we model the linear layer of the retina by a layer of localized, bandpass and non-oriented linear neurons. Their receptive fields are overlapping and their scales are in general distributed non-uniformly over the retina.
Each GC integrates at his soma this information until it reaches a threshold: it then emits a spike. This forms the non-linear layer which is generally modelized by an Integrate-and-fire model where the strongest responses are fired first; the spikes are then transmitted via the optic nerve: the image code consists exactly on the spiking times (or *latencies*) for the different fibers (i.e. GC's axons).
A possible descriptive algorithm is a wavelet-like transform [2] which output coefficients would be temporally transformed to a pattern of spikes. Such an algorithm was successfully proposed by Van Rullen and Thorpe [7] with a nearly orthogonal basis on a dyadic scale. However, the orthogonality criteria is sometimes hard to fulfill and means that the algorithm may be sub-optimal for the sparsity of the coefficients and the separation of the components of the image. Moreover, a dyadic scale is biologically not plausible and the resulting code is not well suited to group transforms like translation, rotation and scaling.

1.3 Definition of the article's framework

Following Olshausen and Field [4], we assume that an approximation I_f of an image I may be calculated as the linear sum of 'patches' of different sizes and localizations chosen from a given 'dictionary' $\mathcal{D}$ of GC characteristic spatial weighting functions ϕ_i:

$$I_f = \sum_{i \in \mathcal{D}} a_i \phi_i \tag{1}$$

Minimizing $\|I - I_f\|$ under informational constraints leads to a combinatorial explosion of the freedom of choice of the subset of $\mathcal{D}$ and of the values a_i (it is a *NP hard* problem [3]). In this article, we derive an algorithm from a spiking neural network model of the retina using a greedy matching pursuit algorithm and we apply it to the framework of the experiences of Thorpe et al. [6]. Our goal is then to study the influence of the tuning of the GC norm and also the invariants in the value of the coefficients according to their spiking rank.
Finally, to illustrate the performance of our coding strategy, we apply our results to a compression application : after building a simple reconstruction scheme, the code may be compared to other models -as the popular JPEG format- by its reconstruction quality in comparison with the compression rate.

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 313-318

2 Description of the retinal model

2.1 Response of the ganglion cells

For simplicity and given our application's goal, we will define different scales σ and distribute the neurons' localizations $\vec{\lambda}$ uniformly on a rectangular grid for each scale. The grid's spacing depends proportionally on the scale so that the neurons fill the spatial/frequency scale.
Generally, we write as in [1] the output of the linear layer:

$$C_{\sigma,\vec{\lambda}} =< I, \phi_{\sigma,\vec{\lambda}} >= \sum\nolimits_{\vec{l}\in\mathcal{R}_{\sigma,\vec{\lambda}}} I(\vec{l}).\phi_{\sigma,\vec{\lambda}}(\vec{l}) \tag{2}$$

where $I(\vec{l})$ is the light intensity at $\vec{l}$ and $\mathcal{R}_{\sigma,\vec{\lambda}}$ is the receptive field (i.e support) of $\phi_{\sigma,\vec{\lambda}}$. Also, we deduce the weight vectors $\phi_{\sigma,\vec{\lambda}}$ of the GCs as a dilated, translated and sampled *Mexican Hat* (see [2, p. 77]) which itself is defined as the normalized laplacian of the gaussian $\mathcal{G}$: $\Delta\mathcal{G}(\vec{r}) = \frac{2}{\sqrt{3\sigma\sqrt{\pi}}}(2 - \|\vec{r}\|^2)\exp(-\frac{\|\vec{r}\|^2}{2})$. The norm of the vectors $\phi_{\sigma,\vec{\lambda}}$ are controlled with respect to the scale by $N_\sigma = \|\phi_{\sigma,\vec{\lambda}}\|$:

$$\phi_{\sigma,\vec{\lambda}}(\vec{r}) = N_\sigma \Delta\mathcal{G}(\frac{\vec{r}}{\sigma}) * \delta_{\vec{\lambda}} \tag{3}$$

where $*\delta_{\vec{\lambda}}$ denotes the translation by $\vec{\lambda}$. At last, the spike latencies are inversely proportional to the integration activity of the neuron $|C_{\sigma,\vec{\lambda}}|$.

2.2 Propagation algorithm

First, we determine the first GC cell in the retina to fire:

$$(\sigma^0, \vec{\lambda}^0) = \text{ArgMax}_{\sigma,\vec{\lambda}}(|C_{\sigma,\vec{\lambda}}|) \tag{4}$$

and for this index $(\sigma^0, \vec{\lambda}^0)$, we define the extremal contrast value $C_{\sigma^0,\vec{\lambda}^0}(\vec{\lambda}^0)$. Actually, we found the best match in the sense of the projection on our basis and we therefore use a projection pursuit scheme, i.e. we subtract the projection from $I^0 := I$.

$$I^1 = I^0 - \frac{< I^0, \phi_{\sigma^0,\vec{\lambda}^0} > \phi_{\sigma^0,\vec{\lambda}^0}}{\|\phi_{\sigma^0,\vec{\lambda}^0}\|^2} = I^0 - C_{\sigma^0,\vec{\lambda}^0}\frac{\phi_{\sigma^0,\vec{\lambda}^0}}{{N_{\sigma^0}}^2} \tag{5}$$

The contrast becomes if we set $C^0_{\sigma,\vec{\lambda}} := C_{\sigma,\vec{\lambda}}$,

$$C^1_{\sigma,\vec{\lambda}} =< I^1, \phi_{\sigma,\vec{\lambda}} >= C^0_{\sigma,\vec{\lambda}} - C^0_{\sigma^0,\vec{\lambda}^0}\frac{< \phi_{\sigma^0,\vec{\lambda}^0}, \phi_{\sigma,\vec{\lambda}} >}{{N_{\sigma^0}}^2} \tag{6}$$

Iterating this steps in time, our algorithm is simply for $t \geq 0$, given the initialization:

$$(\sigma^t, \vec{\lambda}^t) = \text{ArgMax}_{\sigma,\vec{\lambda}}(|C^t_{\sigma,\vec{\lambda}}|) \tag{7}$$

$$\text{and} \quad C^{t+1}_{\sigma,\vec{\lambda}} = C^t_{\sigma,\vec{\lambda}} - C^t_{\sigma^t,\vec{\lambda}^t}\frac{< \phi_{\sigma^t,\vec{\lambda}^t}, \phi_{\sigma,\vec{\lambda}} >}{{N_{\sigma^t}}^2} \tag{8}$$

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 313-318

This algorithm is exactly similar to the work of Mallat and Zhang [3] (which is reviewed in [2, pp.412–419]) for normalized filters ($N_\sigma = 1$) under the term *Matching Pursuit* (MP). The image code is then given by the spike list $(\sigma^t, \vec{\lambda}^t)$ with the corresponding value $C^t_{\sigma^t,\vec{\lambda}^t}$ of the extremal contrast value. Finally the algorithm is stopped at time t_f when the contrast is less than a given threshold.
To reconstruct the image we derive immediately

$$I = \sum_{t=1}^{t_f} C^t_{\sigma^t,\vec{\lambda}^t} \frac{\phi_{\sigma^t,\vec{\lambda}^t}}{{N_{\sigma^t}}^2} + I^{t_f} \tag{9}$$

which was our goal in (Eq. 1).

2.3 Sparse coding

This coding strategy provides a sparse representation of the signal: in comparison with a dyadic decomposition [7], the matches are more precise and the choice of a match is fed back to the coefficients as a lateral interaction proportional to $C^t_{\sigma^t,\vec{\lambda}^t}$ and the correlation weight $< \phi_{\sigma^t,\vec{\lambda}^t}, \phi_{\sigma,\vec{\lambda}} > / {N_{\sigma^t}}^2$.
Moreover, since the residual image I^{t+1} is orthogonal to $\phi_{\sigma^t,\vec{\lambda}^t}$, then a convergence theorem is available, namely that its norm converges exponentially to 0 when t tends to infinity(see proof in [2, p.414] and Fig. 2 - B).
This result is essential in the context of retinal coding since it means that all the information will be stored in a few active spiking neurons. This code, which transmits most active spikes first, will therefore lead to very good information transfer rate even on very short latencies.

3 Results

3.1 Image reconstruction using the matching pursuit (MP) algorithm

We apply our method to different 128x128 images and, as in [3], we define $N_\sigma = \|\phi_{(\sigma,\vec{\lambda})}\| = 1$ for each scale. The scale grows geometrically with a factor $\rho = \sqrt[4]{2}$ (i.e. 4 layers per octave) on 24 scales. As in [7], the recoded image is recognizable after only a few spikes (see Fig. 1 - B) and convergence is fast as of the Mean Square Error (MSE) (see Fig. 2 - B-*a*).
We may easily compute the maximum number of bytes necessary to describe the spike list as $\mathcal{I} = n_{spike} \times \log_2(n_{neurons} \times n_{quantize})/8$ where $n_{quantize}$ is the number of quantization levels for the contrast. Moreover, if n_{pixel} is the number of pixels in the image, $n_{neurons} = \frac{n_{pixels}}{1-\rho^{-2}}$. Numerically, the information rate is ~ 3.19 byte/spike.

3.2 Weighted matching pursuit

Experimental data on natural images show that their spatial power spectrum obey to some invariants [1] and that we may find *a priori* a whitening filter so that the filters'

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 313-318

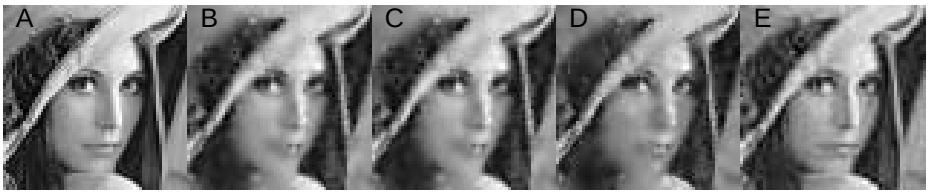

Figure 1: Different reconstruction of *Lena* for a given file size of 2000 bytes. (A) Detail of the original 128x128 image; reconstruction using (B) MP; (C) MP with Look-Up-Table; (D) Weighted MP; and in comparison with (E) JPEG (Quality 32)

output would be sphered accross scales. Using the work of Olshausen and Field [4] we propose that:

$$N_\sigma = \frac{1}{\sigma} \exp(-(\sigma.f_0)^{-1.4}) \tag{10}$$

where experimentally $f_0 = 200$ cycles/picture.
This alternative to the case where the filters are normalized shows that we may tune each GC so that the propagation doesn't favor any scale in particular in time. In our case, we observe that the reconstruction is similar (see Fig. 1 - D) but that less low spatial frequency spikes were sent first, the spike list is less predictale and its entropy is enhanced. A face recognition application should therefore use this tuning.

3.3 Invariance of the contrast value: Rank Order Coding

As in [7], we observe a relative invariance in the absolute value of the contrast in function of the rank of the spikes across natural images. This leads to a even better strategy of compression: the rank and the polarity (ON or OFF) of the contrast are enough to code the contrast value as it is given by a 'look up table' (see Fig. 2 - A). Therefore $n_{quantize} = 2$ and numerically, the information rate shrinks to ~ 2.09 byte/spike.
The reconstruction is very similar (see Fig. 1 - C) and the MSE converges similarly as the case where we use the exact value of the contrast (see Fig. 2 - B- curve *a*).

Conclusion

We have proved that we may define a code based on a dictionary of ganglion weight vectors and that this code is efficient and sparse as is observed in the retina. We've also shown that tuning the scale sensibility according to the statistics of the natural images and using rank order coding led to a very efficient coding strategy which compares to image processing standards like JPEG. Therefore, this may be a strategy used by the retina especially for low bit compression and fast image transmission.
Its biological plausibility and high performance demonstrate that such temporal processes may be crucial in the retina. Finally, it advocates for the use of Support Vector

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 313-318

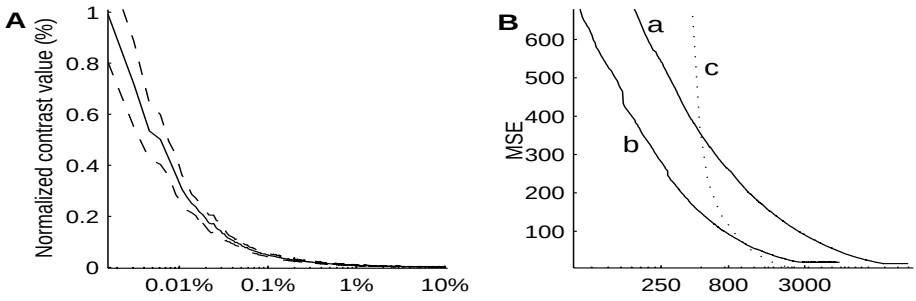

Figure 2: (A) Look-Up-Table : mean and variance of the absolute contrast in function of the relative rank for a database of 100 natural images (B) Comparison of MSE in function of the file size (in $bits$) for the different methods : (*a*) MP; (*b*) MP with Look-Up-Table; (*c*) JPEG at different qualities.

Machine algorithms and for a hierarchical feed-forward architecture. This work is therefore a first step before the implementation of the algorithm to a layer of orientation selective neurons (layers V1 and V2) and then to layers V4 and MT with more 'abstract' dictionaries.

Online simulations

`http://laurent.perrinet.free.fr/code/retina.html`

References

[1] J. J. Atick and A. N. Redlich. What does the retina know about natural scenes? *Neural Computation*, 4(2):196–210, 1992.

[2] S. Mallat. *A Wavelet Tour of signal Processing*. Academic Press, 1998.

[3] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3414, 1993.

[4] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37:3311–3325, 1998.

[5] R. W. Rodieck. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5:583–601, 1965.

[6] S. J. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *381*, pages 520–522, 1996.

[7] R. Van Rullen and S. J. Thorpe. Rate coding versus temporal order coding: What the retina ganglion cells tell the visual cortex. *Neural Computation*, 13(6):1255–1283, 2001.