Emergence of filters from natural scenes in a sparse spike coding scheme

Laurent Perrinet

INPC-CNRS, 31, ch. J. Aiguier, 13402 Marseille, France

Abstract

As an alternative to classical representations in machine learning algorithms, we explore coding strategies using events as is observed for spiking neurons in the central nervous system. Focusing on visual processing, we have previously shown that we may define a sparse spike coding scheme by implementing accordingly lateral interactions [6]. This class of algorithms is both compatible with biological constraints and also to neurophysiological observations and yields a performant algorithm of *computing by events*. We explore here learning mechanisms to unsupervisely derive an optimal overcomplete set of filters based on previous work of Olshausen and Field [3] and show its biological relevance.

Key words: Vision, sparse spike coding, unsupervised learning.

1 Toward sparse spike coding

The central nervous system forms a highly parallel and interconnected network of neurons that interact with a wide range of signals. However, a striking structure from this signal is the presence of brief electrical pulses, *spikes*, that are mostly identitical across a nervous system but also across individuals. But the hypothetical "language" that neurons could share using these spikes is still unknown and most models rely on the assumption that information is conveyed from neuron to neuron through the *frequency* of the spikes. However, a this approach is incompatible with biological constraints on the speed of visual processing [7] and urge us to study alternatives such as temporal coding.

We are here especially concerned with visual processing and we applied these models of spike coding to a model of information processing, transmission and *adaptation* from one neuronal layer to another and especially in the retina and the primary visual area (V1).

1.1 Coding in a neuronal layer: from analog to spike and back to analog

Let us first define a model neuronal layer by a set of neurons i receiving in parallel an image I of synaptic currents. The dendrite of a neuron may be characterized by its synaptic weight vector ϕ_i over these positions $\vec{l}$ in the image so that these currents add up linearly to form a driving current $C_i =< I, \phi_i >= \sum_{\vec{l}} I(\vec{l}).\phi_i(\vec{l})$ at the soma of the neuron. In the case of the retina, we would choose the output of the photo-receptors (for image processing the luminance of a pixel) as the input. The filters could be defined as dilated, translated and sampled Difference of Gaussian filters (see [2, pp. 77]) to yield measures of the multi-scale contrast. These form thus a wavelet representation of the input which is driving the output of the retina.

As in most biological neurons, our model integrates the driving current as a hyper-polarization of its membrane. When this potential reaches a given threshold from below, it obeys a non-linear mechanism which results in a spike and the potential abruptly decreases to a resetting potential. Using the

IF model, the integration is linear and the spike is idealized by an impulse event [4]. The frequency of firing is therefore proportional to the driving current but analogically the more the neuron is activated, the more the spike is fired rapidly: the latency is inversely linked to the corrected activity $|C_i|$.

This model thus converts an analog image of synaptic input into a spatio-temporal pattern of spikes. For instance, if we present to the model retina an image, the analog luminance image is transformed in a spatio-temporal pattern of spikes whose instantaneous frequency may constitute the spike code. But alternatively, the code is carried by the exact spiking time (or *latency*) of the first spike for each neuron. This algorithm therefore defines a coding scheme from an analog matrix to a spike 'wave front' which will travel along the optic nerve, the spikes corresponding to higher activities being emitted quicker. In this paper, we will focus on this transient part of the spiking information.

As a case study for a visual code and even if this situation is biologically highly unrealistic, we studied the quality of the image reconstruction by the wave front as an upper bound of information transmission quality. Thanks to a property of the orthogonal wavelet transform, the image may be reconstructed back from the coefficients' values. Moreover, Van Rullen and Thorpe [8] have shown that these values observed regularities across natural images as they were rectified and ordered from the largest to the lowest. A solution is therefore to decode back the analog values from their rank : this reconstructed image converges to I, forming a *compact* transmission code of the image [5]. Though the condition on the filters for a compact reconstruction is a strong constraint on the architecture and is achieved only approximatively with the model presented in [8], this scheme builds a complete code for the optic nerve from the retina (analog to spike coding) to the brain (spike to analog coding) using solely the rank of the spikes in the wave front.

1.2 Overcomplete representations in the cortex

However, in biological neuronal layers, neurons have correlated sensitivities, a condition which is necessary to implement a continuous representation relative to frequent transformations such as translations or rotations and it is therefore too restrictive in order to build a biologically inspired model of visual processing in V1. Instead, we may want to use an *overcomplete representation* of the image, *i.e.* for which the dictionary $\mathcal{D}$ of filters is far greater then in the previous model as with a dictionary of oriented Gabor filters. But using the previous coding model would yield a highly redundant spike code.

To mimic biological efficiency, we need to obtain a *sparse code*, that is for which only relatively a few neurons are active. Following the framework presented in Olshausen and Field [3], the image is modeled as the probabilistic realization of a Linear Generative Model (LGM, *i.e.* where the input image is approached as a linear sum of neuronal filters, $I = \sum_i a_i.\phi_i$). This leads theoretically to a combinatorial explosion (it is a *NP-hard* problem [2]). The solution of [3] is to choose the best representation as a trade-off between precision and the sparsity of the chosen filters' coefficients and is implemented using a conjugate gradient descent.

In the framework of event-based computation inspired by spike coding, another strategy is to use a *Matching Pursuit* (MP) [2, pp.412–9] algorithm. Under the same assumption as in [3] we optimize the representation one filter after the other by choosing at every step the most probable filter for representing the image (hence the name of "greedy pursuit") knowing the previously chosen filters and coefficients. As described in [6], at the step t (and which corresponds in a one to one correspondence to an exact spiking latency) this is the filter of index $i^t = \text{ArgMax}_{i \in \mathcal{D}}(|C_i^t|)$, that is corresponding to the quicker spike. The associated coefficient minimizing the error image is defined as the projection of the image along the chosen filter, and the algorithm iterates then on the successive image residuals (*i.e.* using the probability of the image knowing what filters and coefficients were already chosen), that is $I^{t+1} = I^t - \frac{C_{i^t}^t}{N_{i^t}{}^2}.\phi_{i^t}$, where $N_i = \|\phi_i\|$. In particular, the current of a winning neuron becomes $C_{i^t}^{t+1} = 0$. In comparison with [3], we don't define *a priori* an objective sparsity function, but minimize the coefficients and a less "greedy" option is either to cancel only a fraction of the projection or to follow a given sparsity function. As for the retina, I observed the behavior of the absolute coefficients' values in function of the rank of propagation. Similarly as the previous model, I observed regularities across natural images but also that due to its adaptive behavior (the error is propagated in the residuals), the model showed up to be more stable and faster decrease of both the coefficients and the reconstruction error, and particularly for a model of V1[5]. This algorithm thus forms an efficient *sparse spike coding* of visual information.

2 Emergence of spatial filters

Due to the similarity of the algorithm underlying the sparse spike coding scheme to the algorithm presented in [3] and which was further studied e. g. in [1], it was clearly obvious that we could use the same formalism for learning the dictionary of filters. We already saw that the filters and their corresponding coefficients were chosen in an inner loop using an incremental method and corresponded to inferring the best representation of the image given the observation and the dictionary. The idea is then to adapt the dictionary in an outer loop to the set of image (here a set of natural images), by looking for the maximum of the "goodness" of the representation of this set using the given dictionary. This "goodness" is expressed in this probabilistic framework as the average log-probability of the images given the dictionary, which will correspond to the minimization of a representation cost. This is maximized using a gradient descent and may be derived analytically.

However, the situation is simpler for the sparse spike coding scheme since we have at every time step a measure of the representation error by the residual image. We may thus derive the learning rule as a gradient descent on the representation cost. In our framework, this may be implemented by adding in the same loop after the coding sweep a learning step after each selection of a best match i^t as $\Delta\phi_{i^t} = \gamma.S(t).I^{t+1}$. Here γ is the learning factor, $S(t) = \frac{C_{i^t}}{N_{i^t}^2}$ is the coefficient corresponding to ϕ_{i^t} and I^{t+1} is the residual image, *i.e.* the error at this time step. In addition, since an accurate learning is proportionally more probable with the goodness of the match, γ is taken proportionally to the activity. This corresponds to a measure of the prior probability that the learned filter corresponds to an accurate representation and similarly as in [1], this leads to a more stable learning convergence.

To compare both methods, I used the same protocol as [3] in terms of the image database, the size (8×8) and number (64) of the filters, and the learning parameters. Only the competition between neurons was controlled by a a different and simpler rule based on reinforcement. After approximately 1000 learning steps, filters converge to direction selective filters (see Fig. 1, Left) similar to the results of the Sparsenet algorithm [3] and suggest that our algorithm extracts independent components of the image. In fact, the spike

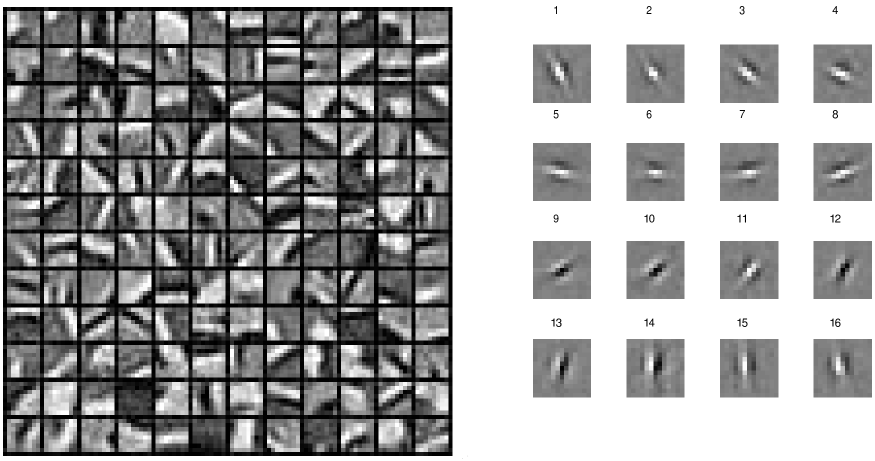

Fig. 1. **Emergence of filters using an adaptive scheme**. *(Left)* Using the same parameters as Olshausen and Field [3], I simulated the propagation of the visual information through a neuronal layer which filters are initially randomly set. By using the learning rule, neurons selective to orientation emerge, as may be observed for V1 simple cells. *(Right)* Introducing lateral connections between filters from a model cortical column (see text) in favor of neighboring activity, we observe the emergence of topological relations between neighboring filters (cyclically numbered from 1 to 16) similar to the neurophysiologically observed *pinwheels*.

model is an implementation of Independent Components Analysis (ICA) : the distribution of the data for natural images is highly non-uniform and is characterized by components with high kurtosis along some axes of the representation space. By adaptively finding the filters, *i.e.* the axes in the representation space, which yield a sparse representation of the signal, [3, 1] proved that this representation relates to ICA since it uses the same informational optimizing principle. Our results provides therefore a simple spike coding model which could be a facet of an underlying principle of the reverse engineering processes that may take place in the brain.

2.1 *Extension to a model of a cortical column*

From simulations on larger image patches (128×128), we easily verify that these results are independent of arbitrary translations of the image, since these patches are chosen at random positions in images. Since we also observed that most neurons are local, we may therefore simplify the algorithm by implementing a model of cortical columns. Mimicking the architecture of V1, it consists in defining at every pixel for a large image a set of filters (a column) which is repeated identically. Even though it seems less biological, this assumption permits to calculate activities as simple convolution with small kernels and therefore to simulate the model on more realistic images. Applying the learning rule in this architecture we may predict that if two filters are similar with respect to a small translation, they will be in parallel competition and only the best one will be chosen. It results that the filters obtained using this representation provided no doubles regarding small translations [4].

A further step in this model that may be particularly efficient for image representation is to topologically associate activities by favoring neurons that were close to the neurons which recently fired. Since smooth borders (lines and edges of solids) are more probable in natural images, filters will associate by similarity of contour direction. This scheme provides therefore an interdependence similar to those found in *Kohonen maps* which in our simulations led to the emergence of a "pin-wheel" (see Fig. 1, Right) and could be a the origin of the emergence of hyper-columns as a basic "brick" of V1.

Conclusion : toward a dynamical model of neural representation

Extending the sparse spike coding algorithm, this learning strategy builds an efficient dictionary of spatial filters which are particularly adapted to the processing of images (all scripts to reproduce the experiments may be found at `http://laurent.perrinet.free.fr/app/`). It leads to the emergence of independent filters which are direction selective filters, as is observed in the primary visual cortex (V1). This strategy is improved by introducing topological constraints and further investigations leads to the formation of spatio-temporal filters in the visual flow which may be fitting the response of neurons in V1.

From its event-based aspect, this adaptive sparse spike coding algorithm differs from classical algorithms and fits particularly well to efficient algorithms that may be implemented in the central nervous system. In particular, its extension to the time domain fits well a to representation of visual flows. As is proved by simple visual illusions as the aperture problem, a challenge is then to achieve a link between the local features given by the neurons to model the neuronal integration of *local* information which give rise to the single and *global* perception as we experience it.

References

[1] M. S. Lewicki and T. J. Sejnowski. Learning overcomplete representations. *Neural Computation*, 12(2):337–365, 2000. URL `http://citeseer.nj.nec.com/lewicki98learning.html`.

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

[3] B. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–25, 1998.

[4] L. Perrinet and M. Samuelides. Coherence detection in a spiking neuron via hebbian learning. *Neurocomputing*, 44–6(C):133–9, 2002.

[5] L. Perrinet, M. Samuelides, and S. Thorpe. Coding static natural images using spiking event times : do neurons cooperate? *Submitted*, 2002. URL `http://laurent.perrinet.free.fr/publi/perrinet03ieee.pdf`.

[6] L. Perrinet, M. Samuelides, and S. Thorpe. Sparse spike coding in an asynchronous feed-forward multi-layer neural network using matching pursuit. *Neurocomputing, in press*, 2002. URL `http://laurent.perrinet.free.fr/publi/perrinet02sparse.pdf`.

[7] S. J. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381:520–2, 1996.

[8] 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):1255–83, 2001.