Bio-inspired Sparse Representation of Images

Effrosyni DOUTSI[1,2], Lionel FILLATRE[1], Marc ANTONINI[1], Julien GAULMIN[2]

[1]Université Côte d'Azur, CNRS, I3S, France.

[2]4G-TECHNOLOGY, Mouans Sartoux - France.

doutsi@i3s.unice.fr, lionel.fillatre@i3s.unice.fr, am@i3s.unice.fr, julien.gaulmin@4g-technology.eu.

Résumé – Cet article présente une nouvelle représentation parcimonieuse des images inspirée de la rétine. Cette représentation consiste en un filtrage inspiré par le fonctionnement de la rétine qui imite la transformation du stimulus visuel en un signal spatio-temporel fortement redondant. Nous proposons d'utiliser un codage du type "Intègre-et-Tire" parfait pour réduire la redondance de cette transformation en mimant la génération des potentiels d'action neuronaux. L'encodage "Intègre-et-Tire" parfait est équivalent à un seuillage basé sur une zone morte dépendante du temps d'observation de l'image. Les résultats numériques montrent l'efficacité de cette représentation parcimonieuse qui fournit des résultats de reconstruction presque équivalents à ceux de la représentation redondante.

Abstract – This paper introduces a novel retina-inspired sparse representation which is applied to temporally constant 2D inputs. This architecture consists of the recently released retina-inspired filtering which mimics the transformation of the visual stimulus into current as it takes place in the retina. This transform is very redundant. As a result, we propose the Perfect Leaky Integrate and Fire (Perfect-LIF) as a model which sparsifies the over-complete retina-inspired decomposition mimicking the spike generation mechanisms of the neurons. The Perfect-LIF is a thresholding function based on a time-depended deadzone. Numerical results show the efficiency of our architecture which provides almost equivalent reconstruction results between the over-complete and the sparse representation of the input image.

1 Introduction

Compression has been one of the most challenging research fields over the last few decades. There are several reasons why the performance of compression algorithms needs to be improved. The most important of these reasons is the resolution of the multimedia devices which continues to dramatically grow (High Definition (HD), Ultra High Definition (UHD), 4K and 8K cameras, etc.) resulting in high spatiotemporal redundancy. As a result, the compression systems should be more efficient to be able to discard the repetitive data and encode only the informative ones.

The visual system is an efficient candidate to mimic because it selects very expeditiously the information of the visual stimulus which needs to be encoded and propagated to the brain. The encoding mechanism of neurons which are considered as analog to digital converters [1] generate a sparse code of spikes depending on the variations of their input stimulus. We are interested in proposing a novel retina-inspired coding principle for images which performs according to the visual system and the brain. Recently, we introduced the novel retina-inspired filtering which approximates the innest retina layer where the input stimulus is captured and transformed into electrical current. This filter is of a high redundancy which is mandatory to be eliminated. In our previous work, we have proposed an event-based code which is generated according to the behavior of neurons [2]. However, the Non-Leaky Integrate and Fire (NLIF) model is a very rough approximation of the neural mechanism. In this paper, we are interested in using a more reliable model, the Leaky Integrate and Fire (LIF) model to sparsify this redundant representation of an image because it describes in a better way how the neurons fire.

Section 2 is a brief introduction to the LIF model and the necessary assumptions to achieve a perfect reconstruction. Section 3 presents the Perfect-LIF which is the characteristic function that corresponds to the LIF model. We also introduce the sparse representation which adopts the Perfect-LIF. In section 4, we illustrate some numerical results and in the last section, we conclude this paper with a small discussion about the extension of this work.

2 Leaky Integrate and Fire (LIF) model

The LIF model [3] is a very well known model which has been widely used in literature [4, 5, 6, 7, 8, 9]. It approximates the neural spiking mechanism assuming that the cell is an electrical circuit given by :

$$I(t) = \frac{V(t)}{R} + C\frac{dV}{dt}, \tag{1}$$

where $I(t) = I_0 \mathbf{1}_{[0\leq t\leq T]}(t)$ is assumed to be a temporally constant input current, C the membrane capacitor of a neuron which is in parallel with the resistor R and $V(t)$ is the voltage across the resistor. If we multiply eq. (1) by R, we introduce the time constant $\tau_m = RC$ of the "leaky integrator". This leakage term is the main difference between the LIF and the NLIF which was used in our previous work [2]. This yields the standard form :

$$\tau_m \frac{dV}{dt} = V(t) + RI(t). \tag{2}$$

While the current $I(t)$ flows into the circuit, the capacitor has the ability to integrate the voltage $V(t)$ until the membrane potential reaches a threshold θ. At that moment, the neuron is excited and it emits a spike. As a result, the LIF model is a *time coder* which allows us to compute and encode the time arrival of the spikes. At time $t_k \in [0, T]$, when the k^{th} spike is emitted, the membrane potential is given by :

$$V_k(t) = RI_0 \left[1 - \exp(-\frac{t - t^k}{\tau_m})\right]. \tag{3}$$

We should remark that at time t^{k+1} the membrane potential is set to zero and the integration of the membrane potential starts all over again until the next spike $k+1$ will be emitted :

$$V_{k+1}(t^{k+1}) = V_{ref} = 0. \tag{4}$$

We have introduced in [10] that, under the assumption of the temporally constant input, we are able to compute the delay between two spikes as a function of voltage v and threshold θ which is given by :

$$d(v) = \begin{cases} +\infty, & \text{if} \quad v < \theta, \\ h(v;\theta) = -\tau_m \ln\left[1 - \dfrac{\theta}{v}\right], & \text{if} \quad v \geq \theta, \end{cases} \tag{5}$$

where $v = v(I_0) = RI_0$ is a function of the input constant in time current I_0.

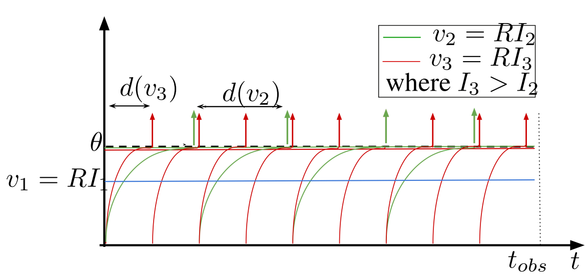

FIGURE 1 – Spike generation mechanism according to the LIF model.

Figure 1 shows the neural activity for different input values $I_1 < I_2 < I_3$. The amplitude of I_1 is too low resulting in $v_1 < \theta$. Consequently, no spike will be emitted. On the other hand, $v_2, v_3 \geq \theta$ meaning that there neuron will fire. The higher the value of I_0 is the smaller the delay $d(v)$ which is required for the spike arrival. Hence, the spike train which is generated due to I_3 is more dense comparing to the spike train of I_2.

If we know the delay, we are able to use the inverse function $h^{-1}(d(v);\theta)$ and perfectly reconstruct the input value :

$$\tilde{v} = \begin{cases} 0 & \text{if} \quad v < \theta, \\ v = h^{-1}(d(v);\theta) = \dfrac{\theta}{1 - \exp\left(-\dfrac{d(v)}{\tau_m}\right)} & \text{if} \quad v \geq \theta. \end{cases} \tag{6}$$

3 Proposed Image Coder

The neurons are able to dynamically encode their input signal and propagate the code of spikes to the visual cortex which is the analysis center. The neural code is not used to reconstruct the input signal but to learn and take decisions. However, our goal is to build a bio-inspired coding/decoding system for images. Figure 2 shows the proposed coding/decoding architecture we aim to build.

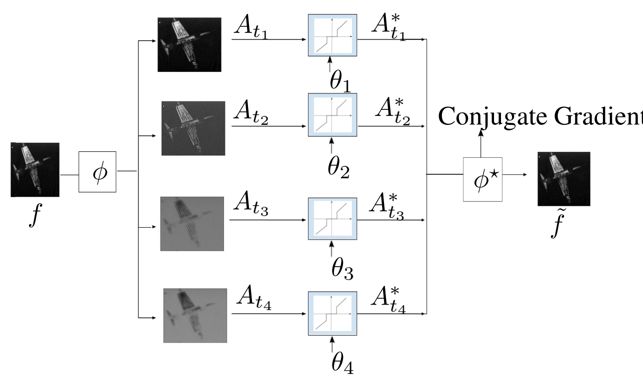

FIGURE 2 – Retina-inspired image coder. The figure describes the decomposition of the input image f using the retina-inspired filter ϕ into several layers A_{t_j} each one of which is sparsified by the LIF quantizer Q. Based on the quantized layers, one is able to reconstruct $\tilde{f}$ and compute the distortion comparing to f.

The input image f is filtered by the retina-inspired filtering $\phi(x, t)$ that we proposed in [11]. This filter is a novel Weighted Difference of Gaussian (WDoG) kernel that mimics the behavior of the retina transform :

$$\phi(x, t) = a(t) G_{\sigma_C}(x) - b(t) G_{\sigma_S}(x), \tag{7}$$

where $a(t)$ and $b(t)$ are two time-varying weights which tune the shape of the DoG, σ_c and σ_c are the standard deviations of the center and the surround Gaussians respectively with $\sigma_c < \sigma_s$. The retina-inspired filtering, which is a frame, is applied to temporally constant input signals $f(x, t) = f(x)\mathbf{1}_{[0 \leq t \leq T]}(t)$ resulting in high redundancy :

$$A(x, t) = \phi(x, t) \overset{x}{*} f(x), \tag{8}$$

where $\overset{x}{*}$ is a spatial convolution. For each time instant t_j there is a different decomposition layer $A_{t_j} = A(x, t_j)$. This redundancy is sufficient to perfectly reconstruct the input signal $\tilde{f}$ using the dual frame ϕ^*. We are interested in reducing this redundancy and discard all the coefficients of low energy keeping only the most informative ones for the reconstruction. To achieve this goal we propose in this section a thresholding function which is inspired by the LIF model. This function is called Perfect-LIF and its output is a sparse decomposition layer $A^*_{t_j}$ that consists of some zero and non-zero values. The characteristic function which describes the LIF encoder/decoder is illustrated in Fig. 3 (black solid line). This is a thresholding function which is called Perfect-LIF. It is able to discard all the input values $v < \theta$ while it perfectly restores the values $v \geq \theta$. The deadzone of the Perfect-LIF is of length 2θ.

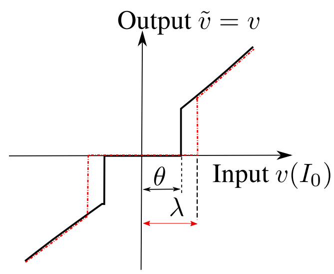

FIGURE 3 – Perfect-LIF characteristic function with and without any time constraint (black solid and red dash-dot curves respectively).

3.1 Constrained Perfect-LIF

Based on the initial assumption that the current is constant for a given time T, we impose a temporal constraint that discards the group of values which excite the neuron too late. Within this framework, the observation window is bounded by $t_{obs} = T$. Figures 3 (red dash-dot line) and 4 illustrate what is the impact of the temporal constraint with respect to the length of the deadzone. It is obvious that the delay $d(v)$ for all the input values $v < \theta$ is infinite. The observation window t_{obs} imposes a novel threshold $\lambda \geq \theta$ eliminating all the values with $d(v) > t_{obs}$:

$$\lambda = \lambda(t_{obs}) = \frac{\theta}{1 - \exp\left(-\dfrac{t_{obs}}{\tau_m}\right)}. \quad (9)$$

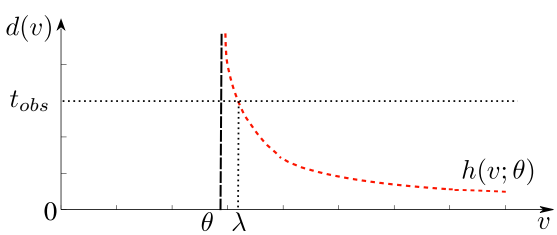

FIGURE 4 – The time constraint t_{obs} imposes a novel threshold $\lambda > \theta$ to discard some input values v.

4 Results

The Perfect-LIF model approximates the LIF model and the way the neurons generate the spike trains. We have filtered images using the retina-inspired filter [11] and we apply the Perfect-LIF to encode each decomposition layer. Of course, we are aware of the high bitrate cost that the Perfect-LIF provides. This is due to the fact that the delay values belong to an infinite set of real values. As a result, we discuss the efficiency of such a model with respect to the reconstruction error measured by two different metrics ; Peak Signal to Noise Ratio (PSNR) and Structure SIMilarity (SSIM).

Figure 5 shows the reconstruction results of a still image which has been encoded/decoded with the Perfect-LIF for different values of threshold λ. The threshold λ has been tuned in such a way that it corresponds to a percentage p of the amount of coefficients which are available at t_{obs} i.e. when $p = 100\%$ all the coefficient are processed. The lower the threshold, the higher the percentage of neurons which spike. The percentage p is related to the statistical distribution of each decomposition layer. The observation window that each subband is allowed to be encoded is $t_{obs} \in \{0.3, 1, 10\}$ ms.

It is proven in [11] that the spectrum of each layer is time-varying starting from low-frequency layers which turn into high-frequency layers. We have also proven in [12] that the degradation of the image depends on λ when the retina-inspired decomposition is a frame. This is the case when $t_{obs} = T = 150$ ms which means that all the layers participate to the reconstruction. However, in these experiments, λ is applied to a group of decomposition layers which appear progressively in time. When $t_{obs} = 0.3$ ms the threshold λ concerns only the low frequency layers which are available. While time increase the threshold λ is imposed to a larger group of layers but still not to the whole frame. This is the reason why the degradation does not decreases while the percentage of active neurons increases.

5 Conclusion

In this paper, we have explored the similarities between a thresholding function, which is called Perfect-LIF, and the way the neurons generate the code of spikes. The LIF model allows us to encode the time each spike arises. We show that we are able to perfectly reconstruct the input signal if the value is higher than a threshold value which is imposed by a temporal constraint.

In the future, we would like to improve the performance of this model and provide a more efficient compression with a reasonable bitrate. We are also interested in using this Perfect-LIF to improve and/or replace the motion estimation which is used in standards taking advantage of the neurons sensitivity.

Références

[1] K. Masmoudi, M. Antonini, and P. Kornprobst, “Frames for exact inversion of the rank order coder,” *IEEE Transaction on Neural Networks*, vol. 23, no. 2, pp. 353–359, 2012.

[2] E. Doutsi, L. Fillatre, M. Antonini, and J. Gaulmin, “Signal reconstruction from a bio-inspired event-based code,” *Gretsi*, 2015.

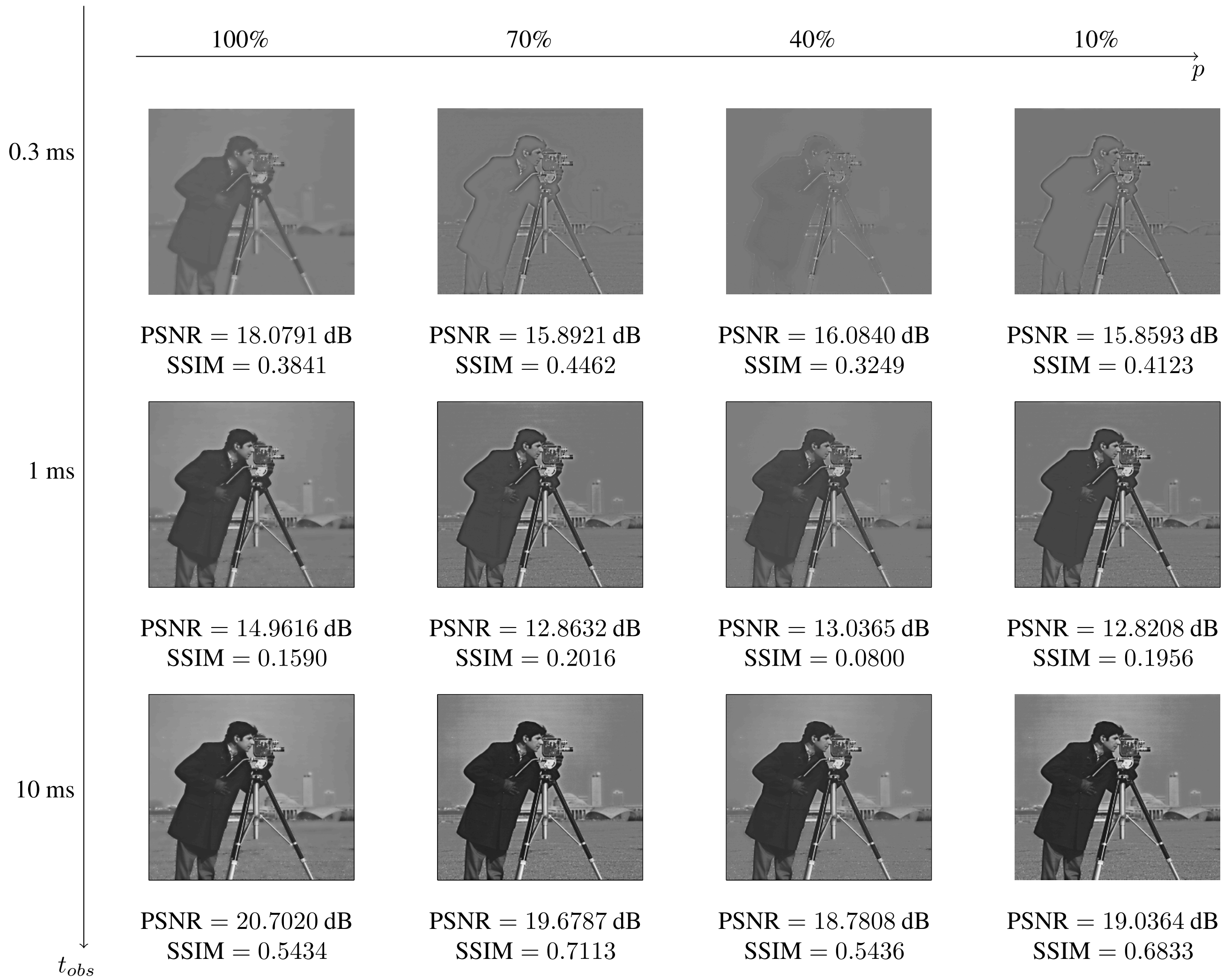

FIGURE 5 – Decoding a sparse representation of an image using the Perfect-LIF. The figure illustrates reconstruction results for different values of p and t_{obs}. All the tuning parameter with respect to the retina-inspired filter are given in [11].

[3] Wulfram Gerstner and Werner Kistler, *Spiking Neuron Models : An Introduction*, Cambridge University Press, New York, NY, USA, 2002.

[4] F. Rieke, D. Warland, R. R. Steveninck, and W. Bialek, *Spikes : Exploring the neural code*, Computational Neuroscience. MIT Press, 1999.

[5] A. Wohrer, P. Kornprobst, and M. Antonini, "Retinal filtering and image reconstruction," Tech. Rep., Inria Research ReportRR- 6960, 2009.

[6] K. Masmoudi, M. Antonini, and P. Kornprobst, "Streaming an image through the eye : The retina seen as a dithered scalable image coder," *Signal processing Image Communication*, vol. 28, no. 8, pp. 856–869, 2013.

[7] Aurel A. Lazar and A. Pnevmatikakis, "Video time encoding machines," *IEEE Transaction on Neural Networks*, vol. 22, no. 3, pp. 461–473, 2011.

[8] R. Jolivet, T. J. Lewis, and W. Gerstner, "Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy," *Journal of Neurophysiology*, vol. 92, pp. 959–976, 2004.

[9] G. C. Cardarilli, A. Cristini, L. D. Nunzio, M. Re, M. Salerno, and G. Susi, "Spiking neural networks based on LIF with latency : Simulation and synchronization effects," *Asilomar Conference on Signals, Systems and Computers*, 2013.

[10] E. Doutsi, L. Fillatre, M. Antonini, and J. Gaulmin, "Retina-inspired video codec," *Picture Coding Symposium (PCS)*, 2016.

[11] E. Doutsi, L. Fillatre, M. Antonini, and J. Gaulmin, "Retina-inspired filtering," *IEEE Transactions on Image Processing*, (submitted) 2016.

[12] E. Doutsi, L. Fillatre, M. Antonini, and J. Gaulmin, "Retina-inspired filtering for dynamic image coding," *IEEE International Conference in Image Processing (ICIP)*, pp. 3505–3509, 2015.