



An end-to-end spike-based image compression architecture

Effrosyni Doutsis, Marc Antonini, Panagiotis Tsakalides

► **To cite this version:**

Effrosyni Doutsis, Marc Antonini, Panagiotis Tsakalides. An end-to-end spike-based image compression architecture. 2020. hal-02908882

HAL Id: hal-02908882

<https://hal.archives-ouvertes.fr/hal-02908882>

Preprint submitted on 29 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An end-to-end spike-based image compression architecture

Effrosyni Doutsis¹, Marc Antonini², Panagiotis Tsakalides^{1,3}

¹ Foundation for Research and Technology - Hellas, Institute of Computer Science

² Université Côte d'Azur, I3S lab, CNRS, ³ University of Crete, Computer Science Department

Emails: edoutsis@ics.forth.gr, am@i3s.unice.fr, tsakalid@ics.forth.gr

Abstract—In this paper, we discover the special properties of neurons in terms of compression. Neurons are able to transform a visual stimulus into a sequence of discrete biphasic events, called spikes trains, forming the neural code. The neural spike generation properties are beneficial to image processing community as the neural code is very compact, yet informative enough, to be used in the input stimulus recovery. We show that the spike-based compression enables to improve the reconstruction quality in time which is a completely novel feature compared to compression standards. In addition, we mathematically prove that the proposed neuro-inspired mechanism behave either as a uniform or a non-uniform quantizer depending on its parameter. Last but not least, we build an end-to-end spike-based coding/decoding architecture that first transforms an image with a DCT filter and then, it generates spikes to compress the transformed coefficients. Based on these spike trains we reconstruct the input image and we estimate the rate-distortion performance of the whole system. Last but not least, we provide numerical results that confirm that our proposed architecture is much more efficient compared to the JPEG standards.

Index Terms—spikes, leaky integrate-and-fire (LIF) neurons, quantization, image coding/ decoding.

I. INTRODUCTION

The human visual system (HVS) has gaining an increasing interest as a powerful and intelligent mechanism that operates numerous complex tasks during the reception, transformation and propagation of the visual stimulus to the visual cortex, such as object detection and identification, distance assessment to and between objects, motion perception, and color vision, just to name a few. The signal processing society, being inspired by the structure and functionalities of the HVS, has built several neuromimetic systems such as neuromorphic chips [1], neuromorphic vision sensors [2], [3], deep learning [4], virtual retina [5], neuro-inspired image and video coding/decoding mechanisms [6]–[9], etc.

This work is motivated by the capacity of the HVS to transform the visual world into a sequence of discrete biphasic events, called spike trains, forming a highly compact yet informative code that flows through the visual system processing layers and carries every detail concerning the input stimulus to the visual cortex. It is proven in [9] that a spike train generated according to the leaky integrate-and-fire (LIF) neurons; a very

This research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under HFRI faculty grant no. 1725, and by the Stavros Niarchos Foundation within the framework of the project ARCHERS.

well-known neuroscience model that approximates the spike generation mechanism, could be used to code natural images in a lossless manner. Here, we are interested in studying the compression properties of neurons as the compactness of the neural code seems to be groundbreaking comparing to the state-of-the-art compression standards. We propose a *neuro-inspired lossy coding/decoding method* while discovering the impact of each LIF parameter on the uniform or non-uniform behavior of the proposed mechanism.

II. LIF BACKGROUND

The well-known LIF model approximates the neuronal encoding by a first order differential equation derived from a resistor-capacitor circuit

$$I(t) = \frac{u(t)}{R} + C \frac{du}{dt}(t), \quad (1)$$

where $I(t)$ is the input current standing for the intensity of light running through the circuit, C is the capacitance, R is the resistance and $u(t)$ is the voltage across the resistor modeling the membrane potential of neuron [10]. If voltage value $u = RI$ is higher than the membrane threshold θ , the neuron emits a spike, otherwise it remains silent.

For a temporally constant input signal during a given duration $T > 0$ the LIF model offers a precise calculation of the exact moment t^{k+1} when the $k + 1$ th spike is emitted just after the emission of k th spike. In this case, the interspike delay $d(I) = t^{k+1} - t^k$ is constant (see Fig. 1) given by

$$d(I) = \begin{cases} +\infty, & \text{if } RI < \theta, \\ h(I) = -\tau \ln \left[1 - \frac{\theta}{RI} \right], & \text{if } RI > \theta. \end{cases} \quad (2)$$

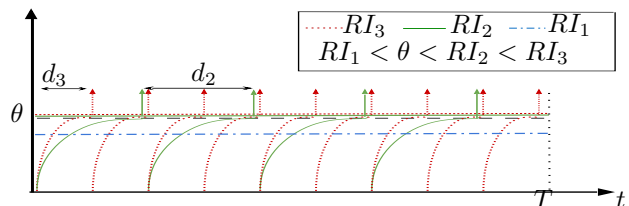


Fig. 1. LIF model with observation window T and threshold θ . If the intensity I satisfies $RI > \theta$, the neuron spikes (case $I \in \{I_2, I_3\}$), otherwise it remains silent (case $I = I_1$).

It was proposed in [11] that if one (i) transforms the pixel intensities of an image into a sequence of spikes and (ii) encodes for each intensity the exact time when the first spike appears, it is possible to perfectly recover the initial image. The authors proposed in [9] that the duration of time T during which the input signal remains constant is the upper bound of the interspike delays that can participate to reconstruction of the input signal. In addition, instead of encoding the interspike delays, one can “listen” to the neural firing response and describe each pixel intensity by the number of spikes emitted during the observation window T .

$$N(I) = \begin{cases} 0, & \text{if } RI < \theta, \\ \frac{T}{d(I)}, & \text{if } RI > \theta. \end{cases} \quad (3)$$

III. IMPACT OF LIF PARAMETERS

A mechanism that encodes pixel intensities based on the number of emitted spikes behaves as a quantizer. The authors proposed in [12] that counting the number of spikes is more efficient in terms of compression than quantizing the interspike delays. The properties of this neuro-inspired quantizer (NQ) highly depends on LIF parameters which tune the frequency of spike emission. In this work we extensively study the impact of parameters R and T on the behavior of the NQ. In addition, we build an end-to-end compression algorithm that (i) transforms the pixel intensities with the Discrete Cosine Transform (DCT) and (ii) use LIF-based quantizer to reduce the spatial redundancy. We compare the rate-distortion performance of the end-to-end NQ to JPEG standard.

A. Resistance R

It has been mathematically proven in [12] that the resistance R affects the behavior of the NQ that performs as uniform or non-uniform. When R is small the NQ is non-uniform but when R is arbitrarily large its behavior approximates a uniform quantizer with a constant quantization step size ℓ (see Fig. 2).

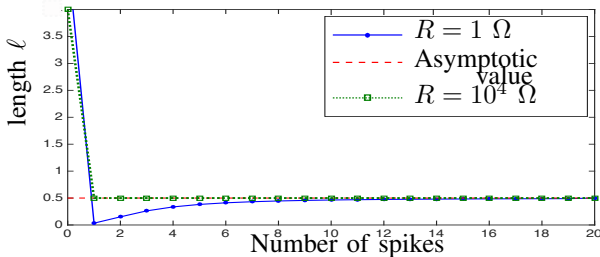


Fig. 2. Performance of the NQ as (i) non-uniform (small R) and (ii) uniform (R arbitrarily large). The number of spikes varies in function of the intensity I (set of parameters: $\theta = 5$ V, $C = 10$ F and $T = 100$ ms).

B. Observation Window T

It is easy to note in Fig. 2 that as the number of spikes increases, the behavior of the NQ is always uniform. A high number of spikes might be caused either by high pixel intensity values (as in Fig. 2) or because of an arbitrarily large

observation window T . Here, we are studying whether the size of the observation window could also lead to a uniform or a non-uniform behavior of the NQ.

Proposition 1: Given that the input intensity I , the resistance R and the membrane threshold θ are constant and the observation window T is the only parameter that influences number of emitted spikes k , the quantization length ℓ_k converges to an asymptotic value

$$\lim_{k \rightarrow \infty} \ell_k = 0. \quad (4)$$

Proof 1: The proof is omitted due to lack of space.

We verify numerically proposition 1 and we present in Fig 4 the asymptotic behavior the NQ for 4 different intensity values when the observation window $T \in [1, 200]$ ms. As expected when the number of spikes increases due to the size of the observation window the length of the quantization step converges to the zero value. We also show in Fig. 3 that by increasing the observation window T there is no impact on the behavior of the NQ. Rather, it is only the parameter R that cause this effect as for large R values, either small or large T values result in a uniform behavior (see Fig. 4 (a)), while for small R values the quantizer follows a non-uniform behavior (see Fig. 4 (b)).

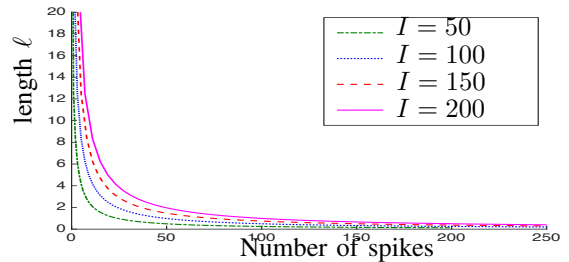


Fig. 3. Performance of the NQ when the number of spikes varies in function of the observation window T (set of parameters: $\theta = 5$ V, $C = 10$ F $R = 1\Omega$ and $T \in [1, 200]$ ms).

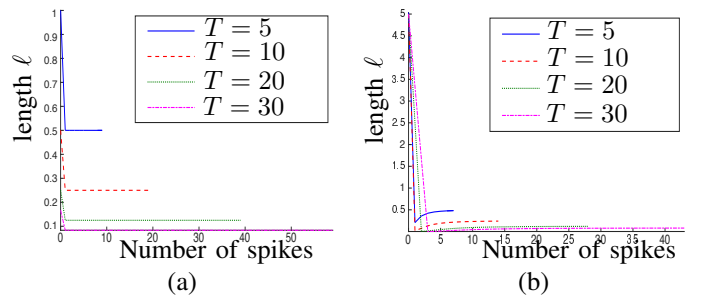


Fig. 4. (a) Uniform NQ for different T values when the parameter $R = 10^4$. (b) Non-uniform NQ for different T values when the parameter $R = 1$.

The size of the observation window has also a strong impact on the quality of the reconstruction. Figure 5 depicts that reconstruction is poor when T is small because most of the small intensities will be represented by one or none spikes. On the other hand, when T is large the reconstruction quality is improved. This progressive reconstruction has yet to be addressed in the state-of-the-art.

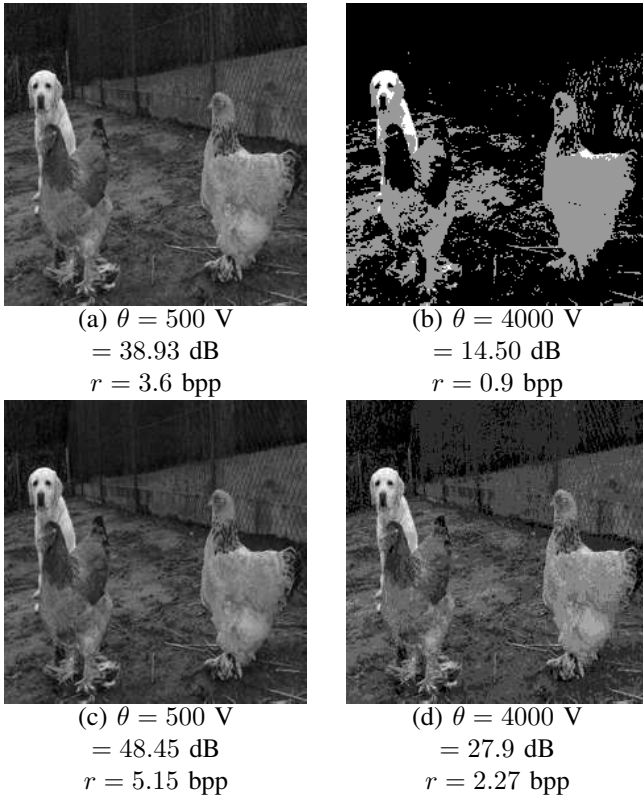


Fig. 5. Progressive reconstruction for $T = 50$ ms (a)-(b) and $T = 150$ ms (c)-(d) (set of parameters: $R = 10^3 \Omega$ and $C = 1$ F).

IV. EXPERIMENTAL RESULTS

To evaluate the performance of the NQ we built an end-to-end coding/decoding mechanism that first transforms an input image using the DCT transform and then encodes using spikes the transformed coefficients. Figure 6 shows the average response when 100 images have been compressed either by the proposed architecture or the JPEG standard. According to two different image quality metrics; the peak signal to noise ratio (PSNR) and the structure similarity index (SSIM) the proposed architecture outperforms the state-of-the-art.

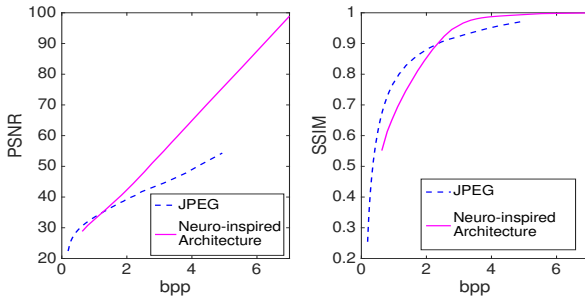


Fig. 6. Comparison between the proposed neuro-inspired end-to-end architecture and the JPEG standard.

REFERENCES

[1] C. S. Thakur, J. L. Molin, G. Cauwenberghs, G. Indiveri, K. Kumar, N. Qiao, J. Schemmel, R. Wang, E. Chicca, J. Olson H., J. Seo,

S. Yu, Y. Cao, A. van Schaik, and R. Etienne-Cummings, "Large-scale neuromorphic spiking array processors: A quest to mimic the brain," *Frontiers in Neuroscience*, vol. 12, pp. 891, 2018.

[2] Jae-Yeon Y. Won, Hyunsurk Ryu, T. Delbruck, Jun Haeng Lee, Jiang Hu, J. H. Lee, and Jiang Hu, "Proximity Sensing Based on a Dynamic Vision Sensor for Mobile Devices," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 536–544, 2015.

[3] Z. Bi, S. Dong, Y. Tian, and T. Huang, "Spike coding for dynamic vision sensors," in *2018 Data Compression Conference*, 2018, pp. 117–126.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," in *Nature*, 2015, vol. 521, pp. 436–444.

[5] A. Wohrer and P. Kornprobst, "Virtual Retina: A biological retina model and simulator, with contrast gain control," *Journal of Computational Neuroscience*, vol. 26, no. 2, pp. 219–249, 2009.

[6] R. VanRullen and S. J. Thorpe, "Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex," *Neural Computation*, vol. 13, no. 6, pp. 1255–1283, 2001.

[7] K. Masmoudi, M. Antonini, and P. Kornprobst, "Frames for exact inversion of the rank order coder," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 2, pp. 353–359, 2012.

[8] Aurel A. Lazar, A. Pnevmatikakis, and Eftychios A. Pnevmatikakis, "Video Time Encoding Machines," *IEEE Transactions on Neural Networks*, vol. 22, no. 3, pp. 461–473, 2011.

[9] E. Doutsis, L. Fillatre, M. Antonini, and J. Gaulmin, "Neuro-inspired Quantization," in *IEEE International Conference on Image Processing*, Athens, 2018.

[10] W. Gerstner and W. Kistler, *Spiking neuron models: Single Neurons, Populations, Plasticity*, Cambridge University Press, 2002.

[11] E. Doutsis, L. Fillatre, M. Antonini, and J. Gaulmin, "Bio-inspired Sparse Representation of Images," in *Groupe d'Etudes du Traitement du Signal et des Images (Gretsi)*, 2017.

[12] E. Doutsis, G. Tzagkarakis, and P. Tsakalides, "Neuro-inspired compression of rgb images," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.