Neural Networks 44 (2013) 22–35

Contents lists available at SciVerse ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

The power of connectivity: Identity preserving transformations on visual streams in the spike domain

Aurel A. Lazar [a,*,1], Eftychios A. Pnevmatikakis [b,1], Yiyin Zhou [a,1]

[a] Department of Electrical Engineering, Columbia University, New York, NY, USA

[b] Department of Statistics and Center for Theoretical Neuroscience, Columbia University, New York, NY, USA

ARTICLE INFO

Article history:
Received 9 July 2012
Received in revised form 7 February 2013
Accepted 27 February 2013

Keywords:
Identity preserving transformations
Invariant representations
Spiking neurons
Time encoding machines
Group theory
Connectivity

ABSTRACT

We investigate neural architectures for identity preserving transformations (IPTs) on visual stimuli in the spike domain. The stimuli are encoded with a population of spiking neurons; the resulting spikes are processed and finally decoded. A number of IPTs are demonstrated including faithful stimulus recovery, as well as simple transformations on the original visual stimulus such as translations, rotations and zoomings. We show that if the set of receptive fields satisfies certain symmetry properties, then IPTs can easily be realized and additionally, the same basic stimulus decoding algorithm can be employed to recover the transformed input stimulus. Using group theoretic methods we advance two different neural encoding architectures and discuss the realization of exact and approximate IPTs. These are realized in the spike domain processing block by a "switching matrix" that regulates the input/output connectivity between the stimulus encoding and decoding blocks. For example, for a particular connectivity setting of the switching matrix, the original stimulus is faithfully recovered. For other settings, translations, rotations and dilations (or combinations of these operations) of the original video stream are obtained. We evaluate our theoretical derivations through extensive simulations on natural video scenes, and discuss implications of our results on the problem of invariant object recognition in the spike domain.

1. Introduction

The brain must be capable of forming object representations that are invariant with respect to the large number of fluctuations occurring on the retina (DiCarlo & Cox, 2007). These include object position, scale, pose and illumination, and the presence of clutter. In a simple model of the visual system in primates, the incoming visual stimulus is first represented in the responses of the retinal ganglion cells (RGCs). Subsequently, the stimulus is re-represented at each neural layer starting with the first relay center (LGN) and followed by the visual cortex (V1, V2, V4 and IT cortex). Each of these representations can be modeled as an Identity Preserving Transformation (IPT). At the final stage, the visual objects are represented in a way that is amenable to an efficient comparison with an internal (memory) representation of the object. Since spike trains are the language of the brain, the latter representation is in the form of a neural population activity. Consequently, the decision whether the object is present or absent takes place in the spike domain (Logothetis & Sheinberg, 1996).

What are some plausible computational or neural mechanisms by which invariance could be achieved? An early pioneering work (Olshausen, Anderson, & Essen, 1993) provides a model mechanism for shifting and rescaling the representation of an object from its retinal reference frame into an object-centered reference frame (see also Anderson & Essen, 1987). In one class of models used in the invariant recognition literature, transformations of the incoming visual signal are matched with an existing stored version of the image (Bülthoff & Edelman, 1992). More formally, let I be a visual sensory object (stimulus). An IPT acting on I is modeled as an invertible transformation $\mathcal{T}$ that, in turn, consists of a composition of a set of elementary operators (e.g., rotation, dilation, translation, etc.). The set of all spike trains produced by $\mathcal{T}(I)$ for all possible IPTs $\mathcal{T}$ defines the object-manifold. For identifying the instantiation of a stored object in the incoming object-manifold, the algorithm presented in Arathorn (2002) calls for the identification of the operator $\mathcal{T}$ (and its inverse). More recent research focuses on routing/connectivity operators in support of information delivery (e.g., sensory information) to higher brain centers (Wolfrum & von der Malsburg, 2007).

In this paper we focus on the *realization* of IPTs in the *spike domain*. The spike domain is a non-linear, stimulus-dependent representation space. The non-linear nature of the stimulus

* Corresponding author. Tel.: +1 212 854 1747; fax: +1 212 932 9421.
E-mail addresses: aurel@ee.columbia.edu (A.A. Lazar), eftychios@stat.columbia.edu (E.A. Pnevmatikakis), yiyin@ee.columbia.edu (Y. Zhou).

[1] The author's names are alphabetically listed.

0893-6080/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neunet.2013.02.013

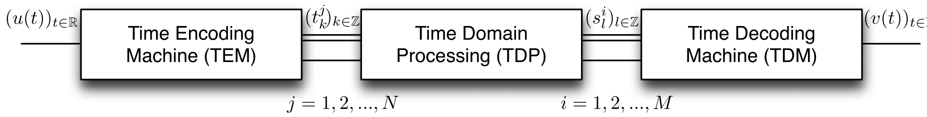

Fig. 1. General signal processing chain with a Time Domain core.

representation has proven to be a major challenge for spike domain computation. Our goal here is to put forth the first efficient rigorous computational model that allows formal reasoning in the spike domain while at the same time it is biologically relevant. Our model of computation can briefly be summarized in block diagram form in Fig. 1. The input visual stimulus is encoded in the time domain by an instantiation of a Video Time Encoding Machine (Video TEM) (Lazar & Pnevmatikakis, 2011b). Video TEMs are spatio-temporal models of neural encoding that are realized with receptive fields in cascade with a population of spiking neurons (Lazar, Pnevmatikakis, & Zhou, 2010). The encoded stimulus (in the form of spikes) is first processed in the Time Domain Processing (TDP) block and then decoded by a Time Decoding Machine (TDM). The output of the TDM is again an analog signal. Our time (spike) domain computation chain resembles the traditional digital signal processing (Oppenheim, Schafer, & Buck, 1999) chain where an analog signal is converted into a digital signal using an analog-to-digital converter, then processed with a digital signal processor and finally converted back to an analog signal with a digital-to-analog converter.

An example of processing in the time domain appeared in Lazar (2006) where it was demonstrated how an arbitrary linear filter can be implemented in the time domain, using neural components. By building upon these results, any IPT acting on the input stimulus can be realized in the time domain. However, the setup of Lazar (2006) is rather complex as it requires a different TDM for any desired transformation of the sensory stimulus.

There are two types of operators that are used for encoding of stimuli with TEMs: linear operators (receptive fields) followed by non-linear operators (spiking neural circuits). These operators are cascaded. The efficient realizability of IPTs presented here is primarily due to the structure of the receptive fields of the Video TEM. These are required to form an overcomplete spatial (or spatio-temporal) filterbank. Furthermore the set of receptive fields has to exhibit certain symmetry properties (in group theoretic sense). If the receptive fields (linear filters) have a group structure transformations on the stimulus can be realized via transformations on the filters. However, these group operations cannot be, in general, "propagated" through the neural encoding circuits (formally non-linear operators). Surprisingly, however, under certain conditions described in the paper, rotations, scaling and translations can be efficiently executed in the spike domain.

We show that a large class of IPTs can be efficiently realized by making connectivity changes in the TDP block while the TDM block remains the same. The TDP block consists of a "switching matrix" that simply regulates the connectivity between the TEM and TDM blocks. We will show that different IPTs can be realized with different connectivity settings of the switching matrix. For example, for a particular setting of the switching matrix, the original stimulus is faithfully recovered. For other settings, translations, rotations and dilations (or combinations of these transformations) of the original video stream are obtained. We will also show that IPTs can be computed in parallel.

Our model can be viewed as a generalization of the shifting and rescaling mechanisms proposed in Olshausen et al. (1993). We extend these operations to include rotations and show how to efficiently implement them in the spike domain. We also discuss the constraints that the finite size of the neural population imposes on the set of achievable transformations. By starting from the continuous group on the plane characterizing all the possible IPTs, we advance two different encoding architectures whose receptive fields are defined on two different discrete grids. The first is a log-polar grid, similar to the ones used in models of foveated vision (Nattel & Yeshurun, 2000; Weber & Triesch, 2009; Wohrer & Kornprobst, 2009). On the log-polar grid the switching matrix can realize combinations of rotations and dilations in a lossless manner in the spike domain. The second is a Cartesian grid (Field & Chichilnisky, 2007; Lee, 1996). On the latter grid the switching matrix can realize combinations of dilations and translations in a lossless manner in the spike domain as well. Finally, we discuss how discrete approximations of the continuous symmetry group can be used to perform arbitrary but approximate IPTs in the spike domain. Examples are given that intuitively demonstrate our methodology.

2. Methods

2.1. The architecture of the model of computation

An illustration of a general switching ("rewiring") architecture for encoding, processing and decoding video streams is shown in Fig. 2. Our architecture follows the general one depicted in Fig. 1.

The input signal is an analog video stream and is encoded by a canonical Video Time Encoding Machine (see Fig. 2). A more formal overview of Video TEMs is available in Appendix A.1. Briefly, the Video TEM consists of a bank of linear filters/receptive fields $D^j(x, y, t)$, $j = 1, \dots, N$, in cascade with non-linear spiking circuits (e.g., neural circuits realized with Integrate-and-Fire neurons). Hence, a Video TEM maps an input visual stimulus into a vector of spike trains. The spiking activity of the neural circuits can be interpreted as signal dependent sampling. This sampling operation, defined as the t-transform, is expressed by the following set of equations

$$\left\{\langle I, \phi_k^j\rangle = q_k^j,\ k \in \mathbb{Z},\ j = 1, \dots, N\right\}, \tag{1}$$

where $I = I(x, y, t)$ is the input visual stimulus that belongs to a Hilbert space, ϕ_k^j is the sampling function associated with k-th spike of neuron j and q_k^j is its measurement (the projection of I onto the sampling function). The sampling functions are determined by the linear receptive fields, the spike times and the parameters of the neural circuits, whereas the outcomes of these projections depend on the spike times and the parameters of the neural circuits. Although the left-hand-side of Eq. (1) is an inner product, the sampling by the neural circuits is highly non-linear, and the sampling functions are, through the spike times, stimulus dependent.

The TDM block implements decoding algorithms for the canonical Video TEM (see Fig. 2). A more formal overview of Video TDMs is available in Appendix A.2. Under certain conditions the Video TEMs can faithfully encode the input video stream as a multidimensional sequence of spike trains. The TDM architecture implements a perfect decoding algorithm of the input video stream (see Appendix A.2). Briefly, the faithful representation condition ensures that (i) the set of linear receptive fields does not filter out any spatial information contained in the input stimulus (Lazar & Pnevmatikakis, 2011a) and (ii) the spiking frequency of the neurons is high enough so that it can represent the temporal information of the stimulus (Lazar & Pnevmatikakis, 2011b).