Vision Research 50 (2010) 2200–2212

Contents lists available at ScienceDirect

Vision Research

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

Encoding natural scenes with neural circuits with random thresholds

Aurel A. Lazar *, Eftychios A. Pnevmatikakis, Yiyin Zhou

Department of Electrical Engineering, Columbia University, New York, NY 10027, USA

A R T I C L E I N F O

Article history:
Received 16 October 2009
Received in revised form 20 March 2010

Keywords:
Neural encoding of natural scenes
Receptive fields
Neural circuits with random thresholds
Reconstruction of visual stimuli
Reproducing Kernel Hilbert Spaces
Regularization

A B S T R A C T

We present a general framework for the reconstruction of natural video scenes encoded with a population of spiking neural circuits with random thresholds. The natural scenes are modeled as space-time functions that belong to a space of trigonometric polynomials. The visual encoding system consists of a bank of filters, modeling the visual receptive fields, in cascade with a population of neural circuits, modeling encoding in the early visual system. The neuron models considered include integrate-and-fire neurons and ON–OFF neuron pairs with threshold-and-fire spiking mechanisms. All thresholds are assumed to be random. We demonstrate that neural spiking is akin to taking noisy measurements on the stimulus both for time-varying and space-time-varying stimuli. We formulate the reconstruction problem as the minimization of a suitable cost functional in a finite-dimensional vector space and provide an explicit algorithm for stimulus recovery. We also present a general solution using the theory of smoothing splines in Reproducing Kernel Hilbert Spaces. We provide examples of both synthetic video as well as for natural scenes and demonstrate that the quality of the reconstruction degrades gracefully as the threshold variability of the neurons increases.

1. Introduction

In the recent years the increasing availability of multi-electrode recordings and functional imaging methods has led to the application of neural decoding techniques to the recovery of complex stimuli such as natural video scenes. An algorithm based on the optimal linear decoder derived in Warland, Reinagel, and Meister (1997) for a rate model was presented in Stanley, Li, and Dan (1999) for the reconstruction of natural video scenes with recognizable moving objects from recordings of a neural population of the cat's lateral geniculate nucleus (LGN). Visual image reconstruction from fMRI data was examined in Miyawaki et al. (2008), whereas in Kay, Naselaris, Prenger, and Gallant (2008) fMRI data was used to identify natural images. The above works suggest that the visual information is preserved along the different layers of the visual system and call for the development of novel algorithms for neural decoding algorithms that are based on spike times.

In this paper we present a formal mathematical, model based approach, for coding and reconstruction in the early visual system. Our neural architecture consists of a population of N spatial filters that model the classical receptive fields, in cascade with an equal number of spiking neural circuits. The neural circuits considered are either integrate-and-fire neurons or ON–OFF neuron pairs with thresholding and feedback. In our architecture the neuronal variability is not attributed to a probabilistic code (Ma, Beck, Latham, & Pouget, 2006); rather the neural circuits are assumed to have random thresholds with known a priori distribution. Neurons with random thresholds have been used to model the observed spike variability of biological neurons of the fly visual system (Gestri, Mastebroek, & Zaagman, 1980), as well as neurons in the early visual system of the cat (Reich, Victor, Knight, Ozaki, & Kaplan, 1997).

We show that neural spiking with these neural circuits represents noisy and independent (Knight, 1972) (generalized) measurements of the input visual stimulus. Based on these measurements, we construct regularized cost functionals and identify the reconstructed stimulus as its minimizer. For simplicity, we assume that the input visual space belongs to a finite-dimensional Hilbert space and use standard optimization techniques to find the reconstructed stimulus. However, as it will be discussed, the results can be directly extended to infinite-dimensional spaces, using the theory of smoothing splines (Wahba, 1990) in Reproducing Kernel Hilbert Spaces (Berlinet & Thomas-Agnan, 2004).

The work presented here builds and extends upon previous work on the representation of stimuli with *deterministic* spiking neurons. Assuming that the input signal is bandlimited and the bandwidth is known, a perfect recovery of the stimulus based upon the spike times can be achieved provided that the spike density is above the Nyquist rate of the stimulus. These results hold for a wide variety of sensory stimuli, including audio (Lazar & Pnevmatikakis, 2008b) and video streams (Lazar & Pnevmatikakis, 2008a; Lazar & Pnevmatikakis, submitted for publication) encoded with a population of spiking neurons. The model of stimuli considered

* Corresponding author. Fax: +1 212 932 9421.
E-mail address: aurel@ee.columbia.edu (A.A. Lazar).

0042-6989/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2010.03.015

in this paper are defined on a discretized version of a band-limited signal space, known as the space of *trigonometric polynomials*. Such spaces are suitable for modeling since they have all the desirable properties of band-limited signal spaces with the added benefit of being finite-dimensional and thus numerically tractable (Lazar, Simonyi, & Tóth, 2008). Moreover, as it will be demonstrated, the finite-dimensionality of the space determines to a first order the complexity of the reconstruction algorithm. Consequently, data recorded from additional neurons can be included into the recovery algorithm at a very moderate computational cost.

Since the encoding neural circuits have random thresholds, a perfect recovery of the input stimulus is not possible. In order to derive an optimal recovery algorithm, we setup the stimulus recovery as a regularized optimization problem. Signal representation using regularization techniques has been discussed in the computational vision (Poggio, Torre, & Koch, 1985) and neural networks (Girosi, Jones, & Poggio, 1995) literature. In this paper we present a formal model for stimulus reconstruction from spike timing using a method of regularization that, as we will show, can approximate complex visual streams, such as natural scenes, in a very efficient way. Using regularization to reconstruct signals encoded with neurons with random thresholds was first presented in Lazar and Pnevmatikakis (2009) in the context of time-varying stimuli belonging to Sobolev spaces encoded with a population of leaky integrate-and-fire neurons.

We explore the recovery of natural scenes and synthetic video streams as a function of the variability of the random thresholds. Variability is quantified as the ratio between the variance and the mean of the threshold. We also explore the modeling of natural scenes with the sample functions that are defined in the space of trigonometric functions. Finally, we present for the first time video sequences of visual stimuli encoded with neural circuit architectures based on neurons with random thresholds. We evaluate the recovery using both traditional measures of signal-to-noise ratio (SNR) as well structural similarity index (SSIM) (Wang, Bovik, Sheikh, & Simoncelli, 2004). The latter more closely relates to perceptual quality of visual stimuli. Rather than focusing on modeling a specific region of the early visual system, we show that the methodology presented here is general and can be applied to arbitrary combinations of receptive fields and neural spiking mechanisms. These include classic models of the early visual pathway (retina, LGN and V1).

The paper is organized as follows. Section 2 deals with the problem of encoding and reconstruction of time-varying stimuli. In Section 2.1 we give a short overview of the spaces of trigonometric functions and discuss how these constitute a natural discretization of spaces of bandlimited functions. In Section 2.2 we present how time-varying stimuli can be encoded with ON–OFF neuron pairs with random thresholds and present their reconstruction by finding the minimizer of an appropriate quadratic cost functional. In Section 2.3 integrate-and-fire neurons with random thresholds encode time-varying stimuli; their recovery is presented in the same section. Examples are given in Section 2.4 that explore the quality of the reconstruction as a function of threshold variability. In Section 3 we introduce the full model for video encoding and reconstruction with a population of spiking neurons with random thresholds. We discuss how video streams can be modeled as space-time trigonometric polynomials and discuss their representation and reconstruction based on this working assumption. Section 4 presents examples of both synthetic and natural video scenes, encoded with neural circuits build with classic models of receptive fields and spiking neurons arising in the retina, LGN and V1. The examples demonstrate the effectiveness of our algorithm by measuring various different quality metrics (Peak SNR, and SSIM) for two different choices of random threshold (Gaussian, Gamma). Actual videos can be found in the Supplementary material. Section 5 discusses various extensions of our work to the recovery of infinite-dimensional stimuli. Finally, Section 6 provides the context for our research and its relation to Bayesian estimates, as well as approaches to globally optimal reconstructions. Section 7 concludes our work and discusses potential future directions.

2. Representation and recovery of time-varying stimuli

Encoding of space-time visual stimuli with neural circuits leads to a fairly complex neural architecture. Since our goal is to present in this paper a rigorous framework for both representation and recovery of visual information, we will first introduce the simpler case of encoding time-varying signals. In this way the reader can develop the needed intuition to deal with the more general encoding of space-time stimuli. As will be clear in Section 3, the key neural building blocks of the encoding architecture for visual stimuli require the careful treatment described below.

Following a short introduction to the space of trigonometric functions, we present a general framework for the representation and recovery of time-varying functions with spiking neuron models. The neuron models considered are of integrate-and-fire and threshold-and-fire type and arise as spiking neuron models in early vision.

2.1. Modeling stimuli as trigonometric functions

In this section we briefly introduce the spaces of trigonometric polynomials and discuss how they can be used for modeling sensory stimuli of interest. We show that trigonometric polynomials are natural discretizations of bandlimited functions, suitable for applications.

In the univariate case, the space of trigonometric polynomials consists of functions that are simultaneously bandlimited with bandwidth Ω (in rad/sec) and periodic with period T. The period and bandwidth are related with each other by the relation

$$T = \frac{2\pi M}{\Omega}, \tag{1}$$

where M is a positive integer that denotes the order of the space. Let $\mathcal{H}$ denote this space. Then $\mathcal{H}$ consists of all the functions $u = u(t), t \in \mathbb{R}$, of the form

$$u(t) = \sum_{m=-M}^{M} a_m \exp(jm\omega_M t), \tag{2}$$

where $\omega_M = \Omega/M$. Note that the space of trigonometric polynomials of order M is a natural discretization of the space of bandlimited functions. The discretization is best viewed in the frequency domain. The exponentials in (2) have a line Fourier spectrum at the points $m\omega_M$ with $m = -M, \dots, M$. By letting $M \to \infty$, this spectrum becomes $[-\Omega, \Omega]$.

Remark 1. The stimuli defined in (2) are in general complex valued functions. To obtain real valued functions, we require $u = \bar{u} \Rightarrow a_0 \in \mathbb{R}$ and $a_{-m} = \overline{a_m}, m = 1, \dots, M$, where $\bar{u}$ denotes the complex conjugate of u.

The sesquilinear form $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \mapsto \mathbb{C}$ defined by

$$\langle u, v \rangle = \int_{-T/2}^{T/2} u(s)\overline{v(s)}ds, \tag{3}$$

is an inner product for $\mathcal{H}$ and thus the space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is a well defined Hilbert space. It is easy to see that under the inner product (3), the set of functions $(e_m), m = -M, \dots, M$ defined as

$$e_m(t) = \frac{1}{\sqrt{T}} \exp(jm\omega_M t), \tag{4}$$

constitutes an orthonormal basis for $\mathcal{H}$.