Neural Networks 49 (2014) 51–58

Contents lists available at ScienceDirect

Neural Networks

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

A model for the receptive field of retinal ganglion cells

Myoung Won Cho [a,*], M.Y. Choi [b]

[a] Department of Global Medical Science, Sungshin Women's University, Seoul 142-732, Republic of Korea
[b] Department of Physics and Astronomy and Center for Theoretical Physics, Seoul National University, Seoul 151-747, Republic of Korea

ARTICLE INFO

Article history:
Received 8 March 2012
Received in revised form 2 September 2013
Accepted 18 September 2013

Keywords:
Receptive field model
Visual information processing
Sensory system

ABSTRACT

Most retina ganglion cells have center–surround receptive fields, where the center may be either ON or OFF while the surround is the opposite. We clarify the functional roles of the receptive field structure, on the basis of the modern theory of natural data processing. It is suggested that the retina shares the principal mechanism and performance of image processing with a video codec in computers, where the antagonism in spatial or temporal receptive fields originates from the orthogonality condition between linear filters for optimal coding of visual signals. We also reveal what visual information is multiplexed across the discharges of an ensemble of ganglion cells. Our theory makes it possible to predict the cross-correlations between ganglion cell spikes, which are optimized for LGN cells to respond accurately and quickly to their receptive fields.

1. Introduction

The retina transmits visual signals from a neural population of 10^8 photoreceptors into the lateral geniculate nucleus (LGN) via 10^6 optic nerve fibers of the retinal ganglion cells (RGCs) (Field & Chichilnisky, 2007; Meister & Berry, 1999; Troy & Shou, 2002). The visual responses of individual visual neurons have been studied in detail, which raises both our understanding and debating about how visual information is processed and conveyed from the eye to the brain.

An important issue in the early visual system is what the functional role of the center–surround cell receptive fields is. A common belief is that the receptive fields work as differential operators (Shanmugam, Dickey, & Green, 1979; Troy, 1993). The structure may allow ganglion cells to convey the information about discontinuities in the distribution of light falling on the retina; these often specify the edges of objects. The center–surround receptive fields are often modeled quantitatively as the difference of a Gaussian (DOG) function, which can have different functional characteristics depending on the ratio between parameters. The DOG function approximates well the Laplacian of a Gaussian (LOG) function, being near optimal for the task of revealing edges in images when the surround field has 1.6 times the spread of the center field. Note, however, that the data on RGC receptive fields give a factor larger than 1.6 (Marr, 1982; Marr & Hildreth, 1980). The idea of differential operators is also inconsistent with the considerable cell responses to spatially uniform visual stimuli, thus giving rise to different views on the belief. A second proposal is that the function of the surround is to pool signals from receptors over a reasonably wide area and to predict the local average luminance. The prediction would help to maximize the signal amplitude in relation to noise (Atick & Redlich, 1990; Srinivasan, Laughlin, & Dubs, 1982; Tsukamoto, Smith, & Sterling, 1990). A third proposal is that the center–surround receptive fields are designed to eliminate correlations between the messages carried by individual retinal ganglion cells (Atick & Redlick, 1992). The different types of ganglion or LGN cell also make the problem more confusing (Li, 1992). Although both P and M cells in monkeys (or X and Y cells in cats) have center–surround receptive fields; however, P (or X) cells respond approximately linearly to inputs while M (or Y) cells respond linearly only to stimuli of low spatial frequencies. Further, P cells are color-opponent while most M cells show little spectral selectivity.

Another important issue is that the correlations between two or more ganglion cell spikes are so strong. It was assumed that each ganglion cell responds to the stimulus within its receptive field and then transmits that information to the next visual stage independently of other ganglion cells; however, a series of experiments in various species show that nearby ganglion cells tend to fire together in synchrony over different time scales (Arnett, 1978; Mastronarde, 1989; Meister, Lagnado, & Baylor, 1995). It is often regarded that the redundancy in retinal coding may originate just from the natural correlations in impinged visual images or sharing electrical inputs from interneurons in the inner retina (Mukamel & Schnitzer, 2005; Ostijic, Brunel, & Hakim, 2009). Nevertheless, it seems that the internal circuitry in the retina is designed to enhance the synchrony between nearby ganglion cell firings more

---

[*] Corresponding author. Tel.: +82 2 920 7280; fax: +82 2 920 2091.
E-mail addresses: mwcho@sungshin.ac.kr (M.W. Cho), mychoi@snu.ac.kr (M.Y. Choi).

0893-6080/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neunet.2013.09.005

than the unavoidable factors do. The phenomenon is also contradictory to the 'redundancy reduction' hypothesis, which states that a visual neuron should remove correlations from an image to reduce redundancy in the spike train, thus increasing the efficiency of information coding (Barlow, 1961). A proposal about the functional role is that the synchrony targets to convey more information than the information available by treating them independently (Meister, 1996; Meister et al., 1995). For example, the activity of two ganglion cells, $R_1(t)$ and $R_2(t)$, can carry the three coefficients represented by the moving averages: $\langle R_1\rangle$–$\langle R_1R_2\rangle$, $\langle R_2\rangle$–$\langle R_1R_2\rangle$ and $\langle R_1R_2\rangle$. It is proposed that the multineuronal codes may be used to increase the spatial resolution of visual signals.

This paper addresses that the answer of the two issues could be entangled with each other. We interpret the functional role of the spatial and temporal receptive fields of visual neurons, on the basis of modern techniques of (visual) signal processing. It is argued that P/X cells may target to perform very efficient reduction of visual signals while M/Y cells may detach emergently irregular features in visual scenes based on the assumption that their receptive field follows the nature of low-pass and band-pass filters, respectively, in wavelet theory. Finally, we derive analytically the typical characteristics of the cross-correlations between RGC spikes based on the similarity between LGN and retina receptive fields and the properties of wavelet basis functions.

2. Preliminaries

The linear–nonlinear (LN) model provides a convenient scheme to predict the responses of RGCs to visual stimuli or the reverse (Karklin & Simoncelli, 2011; Rodieck, 1965; Sakuranaga & Naka, 1985; Victor, 1987). In the model, discharges of a single ganglion cell are born of a nonlinear transformation for the convolution of visual stimuli through a linear filter. We consider the convolution process to associate with the transform coding, which is a type of data compression for natural data such as video signals, audio signals, or photographic images. The key idea of the lossy compression algorithm is to ignore less important coefficients after transforming data into another coordinate by convolving them through linear filters. While audio and still images can usually be compressed at the ratio 10:1 with imperceptible quality loss, the compression ratio of a lossy video codec, taking the value, e.g., 100:1, is almost always far superior to that for the audio and still-image equivalents (Mallat, 1999). It is noticeable that the compression ratio of a video codec is similar to the ratio of the number of photoreceptors to that of optic nerve fibers.

The properties of a transform coding are determined mainly by its filters employed. In the original JPEG standard, an example of the transform coding, the two-dimensional discrete cosine transform (DCT) is used (Wallace, 1991). Most principal components take after the cosine basis functions with lower frequencies when a natural image is partitioned by rectangular window functions. If the image is partitioned by Gaussian functions, the principal components would follow the Garbor functions, used to model the receptive fields of simple cells in V1 (Daugman, 1980; Marcelja, 1980).

Meanwhile, some formats such as the JPEG-2000 standard use the discrete wavelet transform (DWT) for processing natural data (Skodras, Christopoulos, & Ebrahimi, 2001). The DWT possesses excellent signal compaction properties for many classes of real-world signal while being computationally very efficient, and has been applied to various technical fields including image compression, image denoising, image enhancement, and pattern recognition (Mallat, 1999; Mertins, 1999). It is thus conceivable that the signal processing in the early visual areas may also relate to such a wavelet-based mechanism. Indeed the receptive field structure of a visual cell is often modeled as the LOG wavelet function; the hierarchical processing in the early visual system is also reminiscent of a multiresolution analysis (MRA) of visual signals, based on DWT. Such use of operators of different sizes is necessary for detecting properly detailed features or edges because intensity changes occur on different scales in an image.

Specifically, there are two different types of filter in the DWT, which are called scaling functions and wavelets, respectively. A scaling function and its corresponding wavelet(s) have the ability to decompose an empirical signal into different quantities. While the convolution with wavelets tends to detach irregularities, discontinuities, or fluctuations in the signal, that with scaling functions captures the remains which are regular or smoothly varying components in the signal. Scaling functions are sometimes referred to as averaging filters because of their low-pass nature in the frequency domain.

Fig. 1 illustrates the Haar scaling function and wavelet, the simplest wavelet bases. It is shown that the product of the Haar scaling function (or wavelet) and a step function results in finite values during on timings (or at onset and offset timings) owing to its nature of the averaging (or differential) operator. Usually, wavelet bases yield similar convolution results for the step function. From this, we presume that the temporal filter of a P/X (or M/Y) cell may be modeled by a scaling function (or wavelet) (Enroth-Cugell & Robson, 1966).

However often the spatial receptive fields of RGCs may be regarded as edge sharpening filters, i.e., wavelets, it is difficult to clarify their attributes based simply on their appearances, for the center–surround opposite structure is a common characteristic of scaling functions and wavelets. While Gaussian-like functions were used as low-pass filters for image processing in engineering (Adelson, Anderson, Bergen, Burg, & Ogden, 1984), DOG-like functions are adopted as more efficient ones, with the orthogonality condition considered. The inner product of filters with different centers may vanish when they have non-overlapping concerns like the translated versions of the Haar scaling (or a rectangular) function or have opposite signs in the center and surrounds. Such center–surround structure, which is found also in many wavelets, e.g., the LOG wavelet (cf. Fig. 3(B)), has its origin in the nature of differential operators. A wavelet ψ should have p vanishing moments, i.e., $\int dt\, t^n \psi(t) = 0$ for $0 \le n < p$ when it operates as differentiation of order p. Note that the center–surround structure provides a potential solution of $\psi(t)$ when it has vanishing zeroth and first moments. Here the number of vanishing moments is one of the essential characteristics of wavelets. For example, primal and dual wavelets of Cohen–Daubechies–Feauveau (CDF) 9/7, used in the JPEG-2000 standard, possess four vanishing moments. Both the CDF 9/7 scaling function and its wavelet also exhibit the center–surround structure, as shown in Fig. 2(A) and (B).

3. Receptive field model

Suppose the response of retina P cell a at position $\mathbf{r}'_a$ and time t'_a is described by the LN model:

$$R(\mathbf{r}'_a, t'_a) = \bar{R} + g\big(c\,(\mathbf{r}_a, t_a)\big) \tag{1}$$

with baseline firing rate $\bar{R}$ and gain function g, as in the leaky integrate-and-fire model. Here $\mathbf{r}_a = (x_a, y_a)$ specifies the receptive field center of cell a, and t_a the reference timing. The difference between the firing timing t'_a and the reference timing t_a is assumed to be the same. Under external (or target) stimulus $S(\mathbf{r}, t)$ to the photoreceptor cell at position $\mathbf{r} = (x, y)$ and time t, in addition to the base (or non-target) stimulus $\bar{S}$, the coefficient takes the form

$$c\,(\mathbf{r}_a, t_a) = \int d\mathbf{r} \int dt\, A(\mathbf{r} - \mathbf{r}_a) B(t - t_a) \big[S(\mathbf{r}, t) - \bar{S}\,\big], \tag{2}$$