Neural Networks 46 (2013) 32–39

Contents lists available at ScienceDirect

Neural Networks

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

Neural networks letter

Local stimulus disambiguation with global motion filters predicts adaptive surround modulation

Babette Dellen *, Carme Torras

Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Llorens i Artigas 4-6, 08028 Barcelona, Spain

ARTICLE INFO

Article history:
Received 6 August 2012
Received in revised form 17 April 2013
Accepted 19 April 2013

Keywords:
Stimulus disambiguation
Motion processing models
Fourier analysis
Global filters
Aperture problem
Motion repulsion
Constructive interference

ABSTRACT

Humans have no problem segmenting different motion stimuli despite the ambiguity of local motion signals. Adaptive surround modulation, i.e., the apparent switching between integrative and antagonistic modes, is assumed to play a crucial role in this process. However, so far motion processing models based on local integration have not been able to provide a unifying explanation for this phenomenon. This motivated us to investigate the problem of local stimulus disambiguation in an alternative and fundamentally distinct motion-processing model which uses global motion filters for velocity computation. Local information is reconstructed at the end of the processing stream through the constructive interference of global signals, i.e., inverse transformations. We show that in this model local stimulus disambiguation can be achieved by means of a novel filter embedded in this architecture. This gives rise to both integrative and antagonistic effects which are in agreement with those observed in psychophysical experiments with humans, providing a functional explanation for effects of motion repulsion.

1. Introduction

The classical receptive field (CRF) represents one of the most important concepts in neuroscience and has received large support since the work of Hubel and Wiesel (1968), who studied the response properties of neurons in the striate cortex of monkey to visual stimuli. The CRF has been defined as the region of the visual space within which a stimulus evokes neuronal activity. Neurons in the primary visual cortex and area MT, which are considered to belong to the main motion processing pathway in most mammals, have been shown to have localized CRFs, and based on these findings, it was assumed that neural integration in the CRF is rather local than global, e.g. Livingstone and Hubel (1987).

However, the study of waves tells us that local information can also be represented in a distributed fashion by global signals. Fourier transformations use this principle to change from a space–time representation of the world to the frequency space, and vice versa. Without doubt, the principle of constructive interference[1] provides an equally plausible computational explanation for locality.

While both mechanisms can be used to predict local responses, using either one or the other necessarily results in two distinct, basic processing schemes, which are sketched in Fig. 1(A)–(B). The two approaches incorporate global information differently. Hence, the study of locally ambiguous stimuli might shed some light on the important question of which processing architecture has been realized in the human visual system.

When characterizing the integrations used in classical models, the term *local* is used in the sense of *nearby*, meaning that integration is performed for nearby signals only, i.e., within a small area of the visual space. However, the term *local* is also used to describe a *specific location* in space. The superposition of global signals leads to the reconstruction of local image properties, despite the global nature of the preceding computations.

The aperture problem exemplifies how local integration limits the correct determination of stimulus velocity: when a moving line is seen through a small window such that the end points of the line are not visible, the motion of the line is ambiguous, because the observed temporal change in the window could have been caused by many different motions. Classical models attempt to disambiguate the stimulus using results obtained at neighboring locations (Fukushima, 1980; Huang, Jiao, & Jia, 2008), assuming some locality-preserving interaction between local signals (see Fig. 1(A)). However, in the case of multi-object stimuli, this strategy contains the risk that signals belonging to different entities are falsely combined, leading to a paradox: before having disambiguated the stimuli, we cannot know which local signals we can combine, and vice versa. This might explain why classical approaches so far have not

* Corresponding author. Tel.: +34 934015786.
E-mail addresses: bdellen@iri.upc.edu (B. Dellen), torras@iri.upc.edu (C. Torras).

[1] The principle of superposition of waves states that when two or more waves are incident on the same point, the total amplitude at that point is equal to the sum the amplitudes of the individual waves. Constructive interference occurs when a crest of a wave meets a crest of another wave. Constructive interference is achieved in this work through inverse transformations, and both terms can be used synonymously here.

0893-6080/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neunet.2013.04.007

been able to predict human perceptual responses to the stimuli shown in Fig. 4(A) on the basis of a single mechanism (Amano, Edwards, Badcock, & Nishida, 2009; Huang, Albright, & Stoner, 2008).

In this work, the problem of stimulus disambiguation is investigated in the context of a framework based on constructive interference of global Fourier components, and a genuine solution to the problem is developed, made possible by a novel filter implemented in this architecture.

The paper is structured as follows. In Section 2, the underlying processing architecture based on constructive interference is introduced. Then, the results are presented, starting with a problem statement (Section 3), illustrating the problem of stimulus ambiguity using a typical example, and a solution is proposed (Section 4). Simulation results showing predictions of perceptual responses, including the motion-repulsion illusion are presented in Section 5. Finally, the results are discussed with respect to classical motion processing models (Adelson & Bergen, 1985; Fukushima, 1980; Huang, Jiao et al., 2008) based on local integration in Section 6.

2. Methods I: mathematical model

We build upon a recent method and represent the visual stimulus through global Fourier components (Dellen & Wörgötter, 2011). This algorithm follows a processing structure based on the principle of constructive interference (see Fig. 1(B)) and thus provides an adequate ground for this case study.

Let $I(x, y, t)$ be a visual stimulus, where x and y define the spatial positions and t is the time. The spatial Fourier transform of $I(x, y, t)$ is then represented by

$$F(\mathbf{k}, t) = \iint_{-\infty}^{+\infty} \exp(i\mathbf{k}\mathbf{x}')I(\mathbf{x}', t)d\mathbf{x}', \tag{1}$$

where $\mathbf{k} = (k_x, k_y)$ is the spatial frequency vector, and $\mathbf{x}' = (x', y')$. Accordingly, the spatio-temporal Fourier transform is defined as

$$F(\mathbf{k}, \omega) = \int_{-\infty}^{+\infty} F(\mathbf{k}, t') \exp(i\omega t')dt', \tag{2}$$

where ω is a temporal frequency.

To compute image motion, we can filter the Fourier components temporally with

$$b(\mathbf{k}, \omega, \mathbf{v}_p, t) = a(\mathbf{k}, \omega, \mathbf{v}_p) \exp(i\omega t), \tag{3}$$

where

$$a(\mathbf{k}, \omega, \mathbf{v}_p) = \exp[-(\omega - \mathbf{k}\mathbf{v}_p)^2/(|\mathbf{k}|^2\sigma^2)]. \tag{4}$$

Here, $\mathbf{v}_p = (v_p \cos\Theta, v_p \sin\Theta)$ is the preferred (or search) velocity (Dellen & Wörgötter, 2011), defined through a preferred motion direction Θ and a preferred speed v_p. The parameter σ determines how strongly the motion-constraint equation is enforced. We use $\sigma = 0.2$ pixels/frame. With a larger σ, a wider range of values is sampled.

Applying the inverse Fourier transform to the filtered Fourier space allows us to return in real space and reconstruct those parts of the image that move with a velocity $\mathbf{v}_p$, yielding a local response

$$r_L(\mathbf{x}, t, v_p, \Theta) = \iiiint_{-\infty}^{+\infty} \alpha b(\mathbf{k}, \omega, \mathbf{v}_p, t - t') \times F(\mathbf{k}, t')d\mathbf{k}d\omega dt' \tag{5}$$

$$= \iiint_{-\infty}^{+\infty} \alpha\beta a(\mathbf{k}, \omega, \mathbf{v}_p)F(\mathbf{k}, \omega)d\mathbf{k}d\omega \tag{6}$$

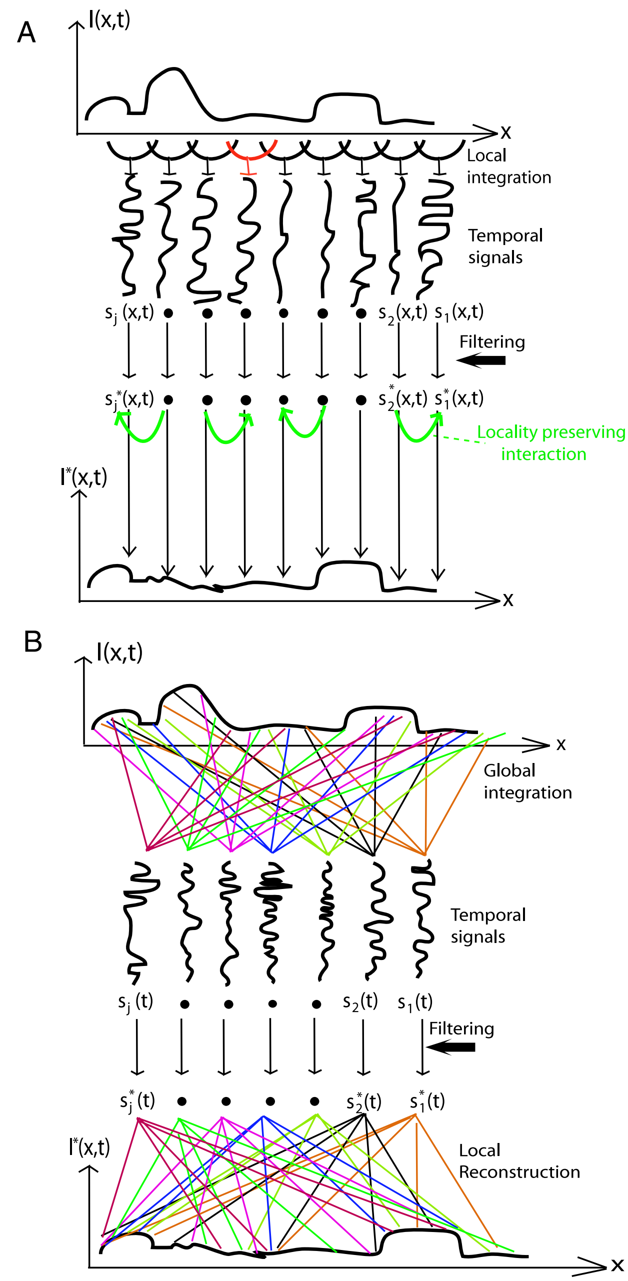

Fig. 1. (A) Classical models assume a processing architecture using local integration inside small patches of the stimulus $I(x, t)$ to compute local stimulus properties. The resulting signals $s_i(x, t)$ depend thus on the location of the respective patch and encode relevant local stimulus features, as for example motion. Global information is incorporated by enabling a locality-preserving interaction between local signals. The final responses $s_i^*(x, t)$ then map the input space in a trivial way. (B) The principle of constructive interference defines an alternative computational strategy. Global integrations generate structured, temporal signals $s_i(t)$, which, taken individually, do not provide information about location. When superimposing the signals in an appropriate manner, spatial information can be reconstructed fully. Specific stimulus features can be extracted through temporal filtering of the global signals before reconstruction.

with the inverse transformation factors $\alpha = \exp(-i\mathbf{k}\mathbf{x})$ and $\beta = \exp(-i\omega t)$. Using the velocity filter, a global response can be defined in an analog manner as

$$r_G(v_p, \Theta) = \iiint_{-\infty}^{+\infty} |a(\mathbf{k}, \omega, \mathbf{v}_p)F(\mathbf{k}, \omega)|d\mathbf{k}d\omega. \tag{7}$$