Pattern Recognition Letters 32 (2011) 1223–1239

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Linear-nonlinear neuronal model for shape from shading

José R.A. Torreão *, João L. Fernandes

Instituto de Computação, Universidade Federal Fluminense, 24210-240 Niterói RJ, Brazil

ARTICLE INFO

Article history:
Received 4 August 2009
Available online 13 April 2011

Keywords:
Biological vision
Computer vision
Linear-nonlinear model
Shape from shading

ABSTRACT

The goal of shape from shading (SFS) is to recover a relative depth map from the variations of image intensity associated to changes in surface shape. There have been very few attempts at developing biologically plausible solutions to this problem, and a sound neurophysiological basis is still missing. Here we present a biologically inspired approach to SFS, formulated in terms of the well-known linear-nonlinear model of neuronal responses. Without resorting to the image irradiance equation, which is at the heart of the traditional SFS algorithms, we submit the input image to a linear filter followed by nonlinear transformations modelled on the tuning curves of the disparity-selective binocular neurons. This yields plausible shape estimates, without requiring information regarding surface reflectance or illumination.

1. Introduction

The shape from shading technique (SFS) has a long tradition in computational vision. Given a monocular image, its goal is to recover a depth map from the brightness pattern associated with the spatially varying orientation of the imaged surfaces, usually assumed smooth, uniform, and distantly illuminated. Several different approaches have been proposed for shape from shading (Zhang et al., 1999; Durou et al., 2008), all of which try to solve some version of the image irradiance equation, which relates image intensities to the photometric and geometric properties of the scene, embodied in the reflectance map function (Horn, 1986). Very few attempts have been made at developing biologically plausible solutions to SFS (Lehky and Sejnowski, 1996; Pentland, 1989), and a sound neurophysiological basis for the process is still missing. Here we introduce an SFS algorithm based on a commonly used model of neuronal responses, the linear–nonlinear (LN) model (Movshon et al., 1978). Without having recourse to the image irradiance equation, we compute depth estimates by submitting the input image to a linear filter followed by nonlinear transformations modelled on the tuning curves of the disparity-selective binocular neurons of the visual cortex (Poggio and Talbot, 1981). This choice of nonlinear functions has been motivated, in part, by the goal to formulate SFS in similar terms as stereoscopy, since both visual processes deal with depth estimation (relative depth, in the case of SFS). More importantly, though, we have also found that, under imaging conditions normally assumed to hold in shape from shading, a parallel may be drawn between the input to the LN model's nonlinear stage and a defocus measure. This establishes a relation between our approach and depth from defocus estimation; the latter, on the other hand, has already been proven formally equivalent to stereoscopy (Schechner and Kiryati, 2000).

As shown by the experimental analysis reported here, the proposed SFS model is able to yield faithful surface reconstructions requiring only the input images, with no need for information regarding surface reflectance or illumination.

2. A linear–nonlinear SFS model

The response of some visual neurons has been described by the so-called linear-nonlinear model: the neuron input passes through a cascade of a linear filter and a nonlinearity, to yield the average rate that will drive a (usually Poisson) spike generator (Movshon et al., 1978). The estimated rate can be expressed as $r_{est} = F(L)$, where L is the output of the linear filter and F is the nonlinear function, commonly chosen under different guises, such as a thresholding, a rectifying, or a sigmoidal function.

Here we investigate the use of an LN model to compute shape from shading (Fig. 1a). At each image location, linear and nonlinear stages are assumed, and the depth value is obtained as the estimated LN response.

2.1. Linear stage

In Lehky and Sejnowski (1996), a neural network trained to perform shape from shading developed receptive fields similar to those of the simple cells of the visual cortex (Hubel and Wiesel,

* Corresponding author. Tel.: +55 21 2629 5683.
E-mail addresses: jrat@ic.uff.br (J.R.A. Torreão), ccmjlf@ic.uff.br (J.L. Fernandes).

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.03.017

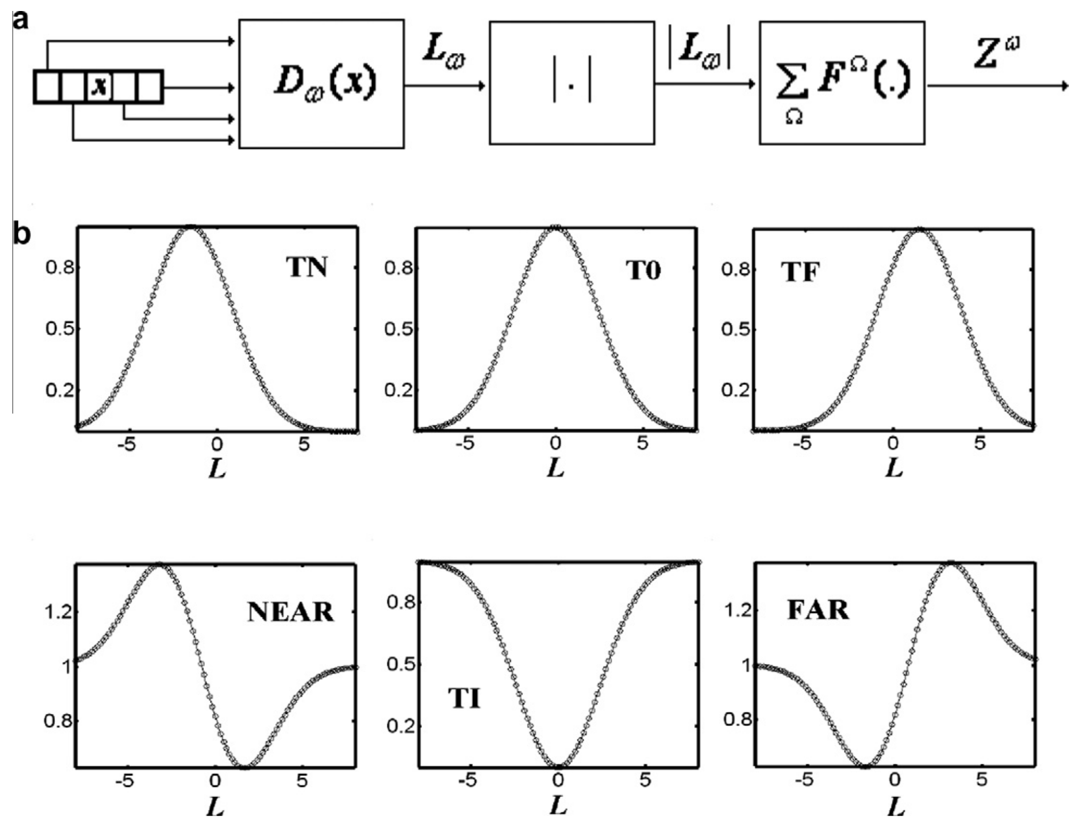

Fig. 1. Linear-nonlinear system for SFS estimation. (a) A window of fixed length W is centered at each image pixel, and a linear measure, L_ω (Eq. (2)), is obtained by filtering the windowed intensities by a complex Gabor function, $D_\omega(x)$, of frequency ω. The first nonlinear stage computes the magnitude of the transformed intensities. This is then submitted to a series of nonlinear transformations, modelled on the disparity tuning profiles of the binocular neurons (b), to yield the depth estimate at frequency ω, Z^ω. The nonlinear transformations depend on the linear outputs, L_Ω, at all frequencies Ω, and are combined into the function F^Ω (Eq. (6)), which adds tuned-inverse (TI), tuned-zero (T0), tuned-near (TN), and tuned-far (TF) profiles, modelled by gaussians, in such a way that a near (NE) and a far (FA) profiles also result (Eq. (7)).

1962). Since such cells are well described by Gabor functions (Marcelja, 1980), we chose these as the linear kernels of our model. For simplicity, a one-dimensional model was assumed, and, at a given position x_0, we defined the linear filter as

$$D_\omega(x) = e^{i\omega x} e^{-\frac{(x-x_0)^2}{2\sigma^2}} \tag{1}$$

where σ is a free parameter, kept fixed over the whole image. The local linear output at the site (x_0, y) becomes

$$L_\omega = \int dx e^{i\omega x} e^{-\frac{(x-x_0)^2}{2\sigma^2}} I(x, y) \tag{2}$$

meaning that, at each image pixel, the linear response is given by a Gaussian-windowed Fourier transform (i.e., a Gabor transform) of the input image. Since L_ω is a complex response, in neurophysiological terms this means that we would be dealing with a quadrature pair of simple cells at each image site. Such pairs have already been found in the cat's visual cortex (Pollen and Ronner, 1981).

2.2. Nonlinear stages

2.2.1. First nonlinear stage

The linear filter output is next submitted to nonlinear transformations, in order to yield the shape estimate. We assumed a cascade of two nonlinear stages, the first just computing the magnitude of the transform in Eq. (2), such as to yield a quantity that can be interpreted as a defocus measure (see the Appendix). On the other hand, in terms of the parallel with a stereoscopic system (see below), L_ω will amount to a disparity, and thus, by working only with its magnitude, we are restricting our model to an *uncrossed* stereo configuration – that is to say, the whole imaged surface is assumed to lie farther away than the horopter, the locus of the scene points which yield zero disparity (Poggio, 1995).

2.2.2. Second nonlinear stage

The second nonlinear stage was so chosen as to establish a parallel between our approach and biological stereoscopic processing. Stereoscopy, similarly as shape from shading, deals with depth estimation (relative depth, in SFS), and it seems reasonable to look for a common biological basis for both. Moreover, under imaging conditions generally assumed to hold in SFS, the Fourier transform which is the input to the second nonlinear stage (Eq. (2)) can be related to a defocus measure (see the Appendix); the theoretical equivalence between shape from defocus and stereoscopy, on the other hand, has already been established (Schechner and Kiryati, 2000). Considering this, we chose, for the computation on the second nonlinear stage of our SFS algorithm, a combination of functions modelled on the disparity-tuning curves of the binocular cortical neurons (Poggio and Talbot, 1981). A tuning curve is any mapping of a neuron's average firing rate as a function of a given stimulus, and it can often be well approximated by a Gaussian. Here we will be working with the tuning curves of binocular neurons, which are sensitive to stereoscopic disparity.

2.2.2.1. Binocular neurons. Binocular neurons are first found in area V1 of the visual cortex. They are sensitive to light presented to either eye, and play a fundamental role in biological stereo processing, by giving different responses to stimuli comprising different binocular disparities. A binocular disparity arises from the