ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

A neural model for binocular vergence control without explicit calculation of disparity

Agostino Gibaldi, Manuela Chessa, Andrea Canessa,
Silvio P. Sabatini, Fabio Solari

Department of Biophysical and Electronic Engineering, University of Genoa
Via all'Opera Pia 11/A - 16145 Genova-ITALY

Abstract. A computational model for the control of horizontal vergence, based on a population of disparity tuned complex cells, is presented. The model directly extracts the disparity-vergence response by combining the outputs of the disparity detectors without explicit calculation of the disparity map. The resulting vergence control yields to stable fixation and has small response time to a wide range of disparities. Experimental simulations with synthetic stimuli in depth validated the approach.

1 Introduction

The retinal binocular disparity is used by the brain as a source of information for the creation of the disparity map to achieve depth perception, and for controlling the movement of the eyes in order to actively get a better perception of the scene. Experimental evidences show how the complex cells of the primary visual cortex (V1) are at the base of both the perception of disparity [1], and the control of vergence eye movements [2], and how these two tasks are carried out by two separate cortical mechanisms. Previous vergence models that are based on a distributed population of disparity detectors, require first the computation of the disparity map for the extraction of the control signals [3] [4], thus limiting the functionality of the vergence system within their fusibility range (FR). In this paper, mimicking the behaviour of the cells of the Medial Superior Temporal (MST) area [5], we present a model that, by combining the response of the population of complex cells, does not take a decision on the disparity values, but extracts disparity-vergence responses that allow us to nullify the disparity in the fovea, even when the stimulus disparities are far beyond the FR. Furthermore, on the basis of the Dual Mode theory [6], the model provides two distinct vergence control mechanisms: a "fast" mode enabled in the presence of large disparities, and a "slow" mode enabled in the presence of small disparities. An additional signal provides the switch between the two modes, according to the stimulus disparities.

2 Model description

2.1 Distributed representation of binocular disparity

Disparity information is extracted from a sequence of stereo image pairs by using a distributed cortical architecture that resorts to a population of simple

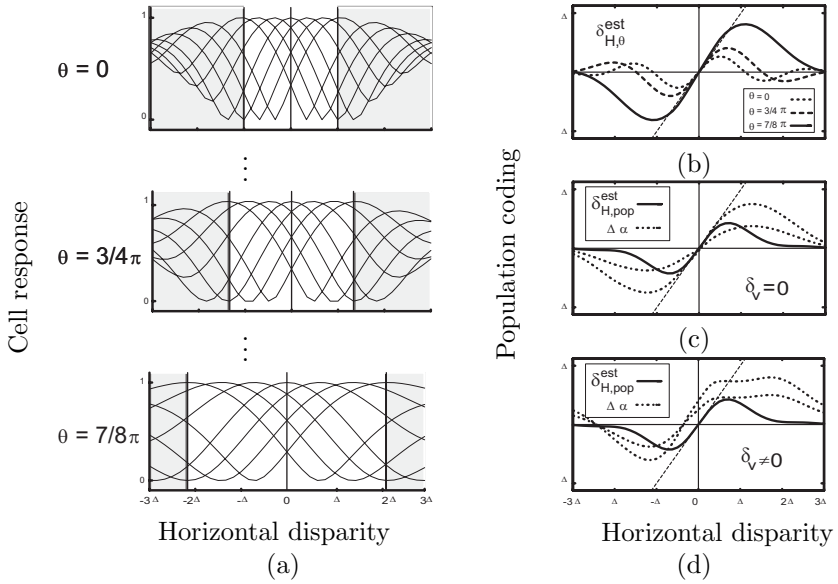


Fig. 1: (a) Tuning curves of complex cells at different orientations. (b) Estimated horizontal disparity using single orientation channels. (c-d) Disparity vergence responses for $\delta_V = 0$ and $\delta_V \neq 0$ (see text).

and complex cells. The population is composed of cells sensitive to $N_p \times N_o$ vector disparities $\boldsymbol{\delta} = (\delta_H, \delta_V)$ with N_p magnitude values distributed in the range $[-\Delta, \Delta]$ pixels and along N_o orientations uniformly distributed between 0 and π. Each simple cell has a binocular receptive field $g_L(x,y)+g_R(x,y)$ defined by a pair of Gabor functions:

$$g(x,y;\psi,\theta) = \exp\left(-\frac{1}{2\sigma^2}(x_\theta^2 + y_\theta^2)\right)\cos(2\pi k_0 x_\theta + \psi) \tag{1}$$

positioned in corresponding points $\mathbf{x} = (x,y)$ of the left and the right images, rotated by the same angle θ with respect to the horizontal axis, and characterized by the same peak frequency k_0 and spatial envelope σ, and by a proper binocular phase shift ($\Delta\psi = \psi_L - \psi_R$), along the rotated axis x_θ, which confers to the cell its specific tuning to a disparity $\delta_{pref} = \Delta\psi/2\pi k_0$, along the direction orthogonal to θ. Formally, given $I_L(\mathbf{x})$ and $I_R(\mathbf{x})$ the left and the right images and $\boldsymbol{\delta}(\mathbf{x})$ the image disparities so that $I_L(\mathbf{x}) = I_R(\mathbf{x} + \boldsymbol{\delta}(\mathbf{x}))$, for every image position $\mathbf{x}$, the response of a simple cell r_s is given by:

$$r_s(\boldsymbol{\delta}(\mathbf{x});\theta,\Delta\psi) = \int\int g_L(\mathbf{x}'-\mathbf{x})I_R(\mathbf{x}'+\boldsymbol{\delta}(\mathbf{x}')) + g_R(\mathbf{x}'-\mathbf{x})I_R(\mathbf{x}')d\mathbf{x}' \tag{2}$$

The response of a complex cell r_c is modeled by the sum of the squared response of a quadrature pair of simple cells, and its response is given by [7]:

$$r_c(\boldsymbol{\delta}(\mathbf{x});\theta,\Delta\psi) = r_s^2(\boldsymbol{\delta};\theta,\Delta\psi) + r_s^2(\boldsymbol{\delta};\theta,\Delta\psi+\pi/2) \tag{3}$$

Accordingly, r_c has its maximum when the product of the magnitude of stimulus disparity and the spatial peak frequency equals the phase difference in the binocular receptive field ($\Delta\psi = |\boldsymbol{\delta}|2\pi k_0$) [8]. The maximum disparity detectable by the population is given by $\Delta = \frac{\Delta\psi_{max}}{2\pi k_0}$. It is worth noting that the population cells are, by construction, jointly tuned to horizontal (δ_H) and vertical disparities (δ_V). Each orientation channel has a sensitivity for the horizontal disparity and, assuming $\delta_V = 0$, we can obtain its estimate δ_H^θ by the projection of the oriented phase difference on the (horizontal) epipolar line in the following way: $\delta_H^\theta = \frac{\Delta\psi}{2\pi k_0 cos\theta}$. To decode the horizontal disparity at a specific image point, the whole activity of the population of cells, with receptive fields centered in that location, is considered. By using a center-of-mass decoding strategy, the estimated horizontal disparity is obtained by:

$$\delta_H^{est} = \frac{\sum_{i=1}^{N_p}\sum_{j=1}^{N_o}\frac{\Delta\psi_i}{2\pi k_0 \cos\theta_j} r_c^{ij}}{\sum_{i=1}^{N_p}\sum_{j=1}^{N_o} r_c^{ij}} \tag{4}$$

where r_c^{ij} denotes the response of the complex cell characterized by the i-th phase difference and by the j-th orientation. Fig.1a shows the horizontal disparity tuning curves of the population for the different orientations. Fig.1b and Fig.1c-d (solid thick line) show the estimated disparity curves obtained by using single orientation channels and the whole population, respectively. In any case, the estimate of the disparity is correct in a range around $[-\Delta, \Delta]$, only. The increase of the sensitivity range as the orientation of the receptive fields deviates from the vertical, comes at the price of a reduced reliability of the measure. As for the control of vergence, larger disparities have to be discriminated while keeping a good accuracy around the fixation point for allowing finer refinement and achieving stable fixations, alternative strategies might be employed to gain effective vergence signals directly from the complex cell population responses, without explicit computation of the disparity map.

2.2 Control signal extraction

A desired feature of the disparity tuning curves for vergence is an odd symmetry with a linear segment passing smoothly through zero disparity (cf. dotted lines in Fig.1c), which defines critical servo range over which changes in the stimulus horizontal disparity elicit roughly proportional changes in the amount of horizontal vergence eye movements, $\Delta\alpha = p\delta_H$ where α is the vergence angle. We demonstrate that such disparity vergence response can be approximated by proper weighting of the population cell responses. Given a stimulus with an horizontal disparity δ_H, we want to combine the population responses in order to extract a vergence control proportional to the disparity to be reduced, regardless of a possible vertical disparity δ_V. To this end, we can use the disparity population response curve (see solid thick curve in Fig.1c). Thought, the control capability, as it can be gathered from the width of the estimated disparity curve, is again limited inside a small range. By exploiting the other orientations, we can design linear servos working over wider disparities, so that we can increase

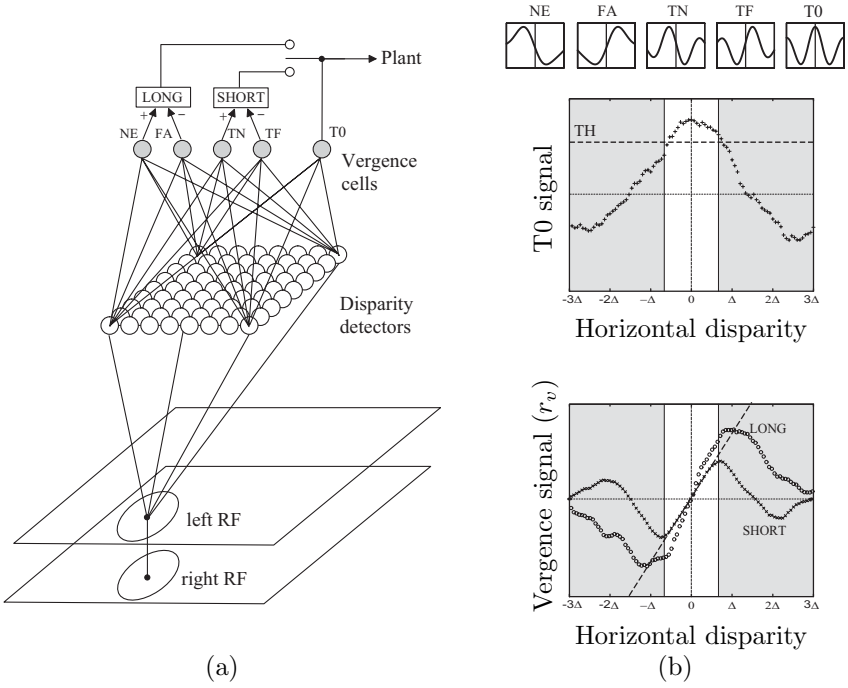


Fig. 2: Extraction of the vergence control signals.

the control capability to one extent larger than the FR. Yet, the problem arises that the resulting curves on which we base our vergence control have a sensitivity to vertical disparity, which is an undesirable effect that alters the control action (see Fig.1d). Due to these considerations, the population responses are combined with two very specific goals: (1) to obtain signals proportional to horizontal disparities, (2) to make these signals be insensitive to the presence of vertical disparities. The desired disparity vergence response curves r_v^k are obtained by a weighted sum of the complex cell responses (see Fig.2):

$$r_v^k = \sum_{i=1}^{N_p} \sum_{j=1}^{N_o} w_{ij}^k r_c^{ij} \tag{5}$$

and the weights w_{ij}^k are obtained through a recursive LMS algorithm. More precisely, given the profile of the desired vergence curve $r_v^k(\delta_H)$, such curve is approximated by a weighted sum of the tuning curves for horizontal disparity $r_c(\delta_H; \theta, \Delta\psi)$. To gain the insensitivity to vertical disparity we add a constraint term in the minimization to ensure that the sum of the vertical disparity tuning curves $r^c(\delta_V; \theta, \Delta\psi)$, weighted with the same w^k, is approximately constant. Hence, the weights w^k are obtained by minimizing the following functional:

$$E(w^k) = \| \sum_{i=1}^{N_p} \sum_{j=1}^{N_o} r_c^{ij}(\delta_H) w_{ij}^k - r_v^k(\delta_H) \|^2 + \lambda \| \sum_{i=1}^{N_p} \sum_{j=1}^{N_o} r_c^{ij}(\delta_V)(w_{ij}^k - 1) \|^2 \quad (6)$$

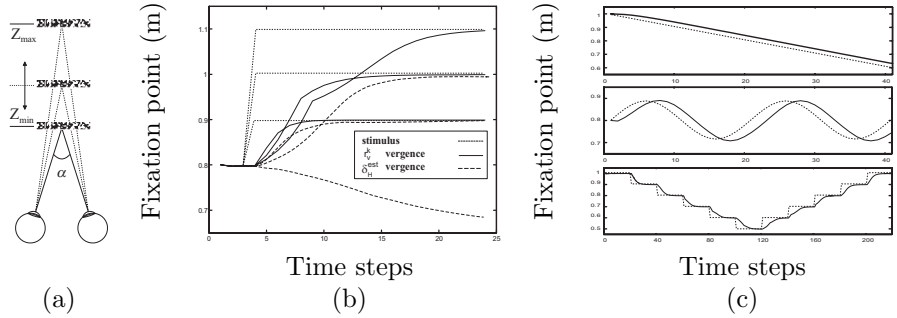
where $\lambda > 0$ balances the relevance of the second term over the first. To ensure a good resistance to noise, a positivity constraint on the weights has been introduced. In our simulations we fixed $\lambda = 1$.

2.3 Signal Choice

In analogy to a common classification [9], we distinguish five categories of r_v cells: near (r^{NE}) and far (r^{FA}) dedicated to coarse vergence, tuned near (r^{TN}), tuned far (r^{TF}) and tuned zero (r^{T0}) for fine vergence. More precisely, r^{NE} and r^{TN} drive convergence movements, while r^{FA} and r^{TF} divergence movements, in a push-pull system (see Fig.2). In practice the fast-coarse control is given by LONG= $r^{NE} - r^{FA}$, while the slow-fine is given by SHORT= $r^{TN} - r^{TF}$. The SHORT control signal is designed to proportionally generate, in a small range of disparities, the vergence to be achieved, and allows a precise and stable fixation (Fig.2b). Out of its range of linearity, the SHORT signal decreases and it loses efficiency to the point where it changes sign, thus generating a vergence movement opposite to the desired one. Instead for small disparities the LONG control signal yields overactive vergence signal that makes the system to oscillate, whereas for larger disparities it provides a rapid and effective signal. The role of the r^{T0} signal, is to act as a switch between the SHORT and the LONG controls. When the binocular disparities are small, r^{T0} is above a proper threshold TH, and it enables the SHORT control (see white regions in Fig.2b). On the contrary, for large stimulus disparities, r^{T0} is below the threshold and it enables the LONG control (see grey regions in Fig.2b).

3 Results and Discussion

We tested the proposed model in a virtual environment in which the eyes, characterized by null version and elevation angle, and by a vergence angle α, look at a plane with a random dot texture (Fig.3a). The plane is at a depth Z with respect to the cyclopic position, and perpendicular to the binocular line of sight. The interocular distance is $b = 70mm$, the nodal length is $f_0 = 17mm$, and the stimulus is projected onto the retinal plane, with a size of $6mm$, thus considering a field of view of almost 20 degree. At the first time step the plane and the fixation point are at the same Z, then the plane is moved to a new depth, and the vergence angle starts to change step by step, until the fixation point reaches the depth of the plane. Considering the position of the eyes, the vergence variation is applied symmetrically: $\Delta\alpha_R = -\Delta\alpha_L = -\arctan(\frac{r}{2f_0})$, where r is computed by considering the weighted average of the vergence responses r_v^k or of the estimated disparities δ_H^{est}, in a neighborhood of the fovea. Fig.3b shows that the estimate of the disparity produces the correct change of the vergence



(a) (b) (c)

Fig. 3: (a) Simulated experimental setup. (b) Behaviour of the vergence control using r_v^k vs δ_H^{est}, and (c) for different moving stimuli: ramp, sinusoid and staircase

angle (dotted lines) when the depth of the plane is not too large. On the other hand, the r_v^k signals are able to produce a fast and reliable change of the fixation point, even for larger depths (solid lines). Once that the fixation point reaches the plane, the disparity in fovea is zero and the system is able to ensure a stable fixation. Finally we tested the model by changing the Z of the plane in time along a ramp, a sinusoid and a staircase (Fig.3c). In the first two tests the Z of the plane varies continuously, and the fixation point follows its depth in the proper way [6].

Acknowledgements
This work has been supported by the EU Project FP7-ICT-217077 "EYESHOTS"

References

[1] I. Ohzawa, Freeman R. D., and G. C. DeAngelis. Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science*, 249(4972):1037–1041, 1990.

[2] G. S. Masson, C. Busettini, and F. A. Miles. Vergence eye movements in response to binocular disparity without depth perception. *Nature*, 389:283–286, 1997.

[3] W. M. Theimer and H. A. Mallot. Phase-based vergence control and depth reconstruction using active vision. *CVGIP, Image understanding*, 60(3):343–358, 1994.

[4] S. S. Patel, H. Ogmen, and B. C. Jiang. Neural network model of short-term horizontal disparity vergence dynamics. *Vision Research*, 37(10):1383–1399, 1996.

[5] A. Takemura, Y. Inoue, C. Quaia, and F. A. Miles. Single-unit activity in cortical area mst associated with disparity-vergence eye movements: Evidence for population coding. *J. Physiol.*, 85(5):2245–2266, 2001.

[6] G. K. Hung, J. L. Semmlow, and K. J. Ciuffreda. A dual-mode dynamic model of the vergence eye movement system. *Trans. on Biomedical Engineering*, 36(11):1021–1028, 1986.

[7] N. Qian. Computing stereo disparity and motion with known binocular cell properties. *Neural Computation*, 6(3):390–404, 1994.

[8] D. J. Fleet, H. Wagner, and D. J. Heeger. Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. *Vision Research*, 36(12):1839–1857, 1996.

[9] G. F. Poggio. Mechanism of stereopsis in monkey visual cortex. *Cerebral Cortex*, 5:193–204, 1995.