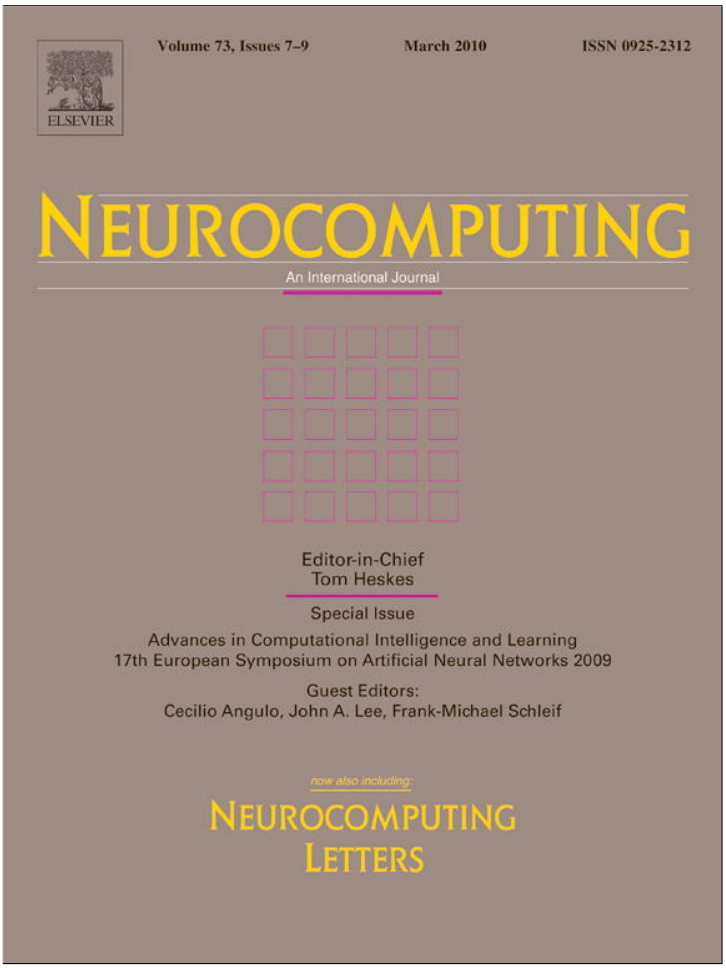

Neurocomputing 73 (2010) 1065–1073

Contents lists available at ScienceDirect

Neurocomputing

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

A cortical 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

ARTICLE INFO

Available online 18 January 2010

Keywords:
Vergence movements
Retinal disparity
Population coding

ABSTRACT

A computational model for the control of horizontal vergence, based on a population of disparity tuned complex cells, is presented. Since the population is able to extract the disparity map only in a limited range, using the map to drive vergence control means to limit its functionality inside this range. 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 to achieve depth perception and to control the movement of the eyes in order to actively get a better perception of the scene, on the basis of the characteristics of the scene itself. Experimental evidences show that, though the complex cells of the primary visual cortex (V1) are the computational substrate for both stereopsis [1] and vergence [2], these two tasks are carried out by two separate cortical mechanisms. The difference between these two mechanisms is that, while the former is capable of producing a single percept from two different retinal images, only when the disparities are within a limited range known as Panum fusional area [3], the latter allows us to extract a control of vergence for wider disparities. In this way, the system is brought back in the fusible range (FR) to ensure again the singleness of vision.

Previous vergence models that are based on a population of disparity detectors require first the computation of the disparity map for the extraction of the control signals [4,5], thus limiting the functionality of the vergence system within the range of disparities in which the system is able to fuse the left and right images. Making a parallel with the biological system, this means that vergence eye movements would be reliable only inside the Panum's area, where they are not useful.

Although each neuron of Medial Superior Temporal (MST) area sensitive to retinal disparity has been found to encode only some limited aspects of the motor response for vergence eye movements, the activity of the whole population directly correlates with the magnitude, direction, and time course of the initial vergence motor response [6,7]. In this paper, mimicking the behavior of the cells of the MST area, we present a model that, by combining the response of a population of complex cells, does not take a decision on the disparity values (disparity map), but extracts disparity-vergence responses that allows us to nullify the disparity in the fovea, even when the stimulus disparities are far beyond the FR. Furthermore, [8,9] suggest that vergence eye movements are not controlled by a simple continuous feedback system [10], but they exhibit dual-mode slow and fast responses. From this perspective, the proposed model provides two distinct vergence control signals: a "fast" signal enabled in the presence of large disparities, and a "slow" signal enabled in the presence of small disparities. The dual-mode theory postulates that a portion of the vergence eye movements response is controlled by a switching control strategy similar to the one that governs the saccadic response [11]. Hence, the model presented in this paper includes a trigger signal to switch the control between the two modes, according to the stimulus disparities.

The paper is organized in sections as follows: Section 2 presents the overall description of the model, including (1) the coding and decoding strategies of binocular disparity through a population of energy complex cells, (2) the proposed approach to construct disparity-vergence responses usable to drive vergence eye movements in a wide range of disparities, and (3) how to combine a set of stereotyped disparity-vergence curves in a dual-mode so as to have fast reaction for large vergence and slower but high precision control for smaller vergence changes. Section 3 reports the results of the model both for synthetic stimuli and a virtual environment, and in Section 4 the behavior of our model is comparatively discussed against psychophysical and neurophysiological data.

* Corresponding author.
E-mail address: agostino.gibaldi@unige.it (A. Gibaldi).

0925-2312/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2009.11.016

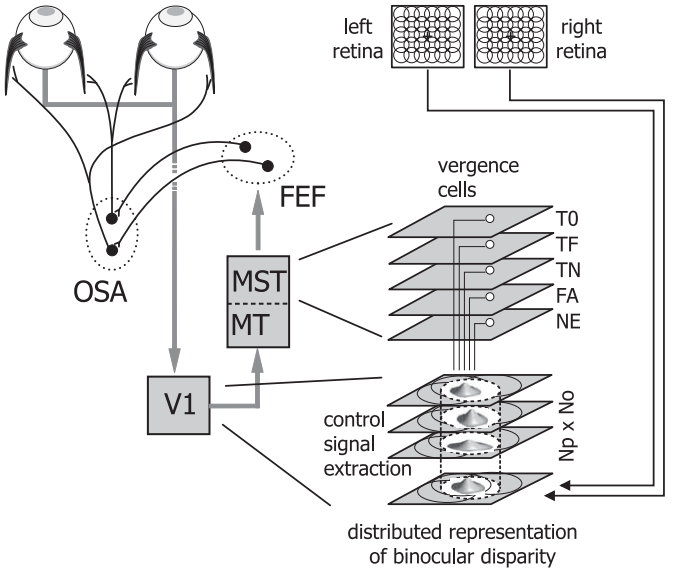

Fig. 1. (left) Simplified scheme of the neural circuitry involved in the control of vergence eye movements. Abbreviations: V1, primary visual cortex; MT, medio temporal area; MST, middle superior temporal area; FEF, frontal eye fields; SOA, supraoculomotor area. The grayed blocks represent the core components of the model present in this paper. (right) Schematic representation of the modeled. The left and right images are processed by a population of disparity detectors, inspired by complex cells of area V1. The population produces a distributed representation of the retinal disparity, and through convolution with weighting kernels it is decoded in order to obtain a family of vergence cells that are able to provide a direct vergence motor response, in accordance with the experimental evidences in area MST. Since the task is to drive vergence eye movements so as to improve the fixation and the estimation of disparity, the information is gathered only from the central (parafoveal) portion of the visual field (white areas).

2. Model description

An overview of the proposed model's architecture is sketched in Fig. 1. The left side diagram schematically frames the core components of our model (grayed blocks) in the cerebro-midbrain circuitry responsible for the generation of the vergence eye movements: the visual information from the retina reaches the primary visual cortex (V1) and from there it projects to the extrastriate areas, middle temporal (MT) and medial superior temporal (MST). MST projects to the frontal eye field (FEF), which, in turn, projects to the supraoculomotor area (SOA) and adjacent reticular formation around the oculomotor nucleus that contain premotor neurons related to vergence eye movements [12,13]. The right side diagram shows the three components of the model, from the distributed representation of the binocular disparity to the generation of the vergence signal, as detailed in the following subsections.

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 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 (see Fig. 2a):

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

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 ($\Delta\psi = \psi_L$-ψ_R), along the rotated axis x_θ, which confers to the cell its specific tuning to a disparity $\delta_\theta^{pref} = \Delta\psi/2\pi k_0$, 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

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

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 [14] (see Fig. 2 b)

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

Accordingly, r_c has its maximum when the product of the magnitude of the stimulus disparity $\boldsymbol{\delta}$ and the spatial peak frequency equals the phase difference in the binocular receptive field [15]. The population is, in this way, capable of providing reliable disparity estimates in the range between -Δ and Δ, where $\Delta = \Delta\psi_{\max}/2\pi k_0$ can be defined as the maximum detectable disparity of the population.

It is worth noting that the population of complex cells are, by construction, tuned to oriented disparities, i.e. jointly tuned to horizontal (δ_H) and vertical disparities (δ_V). In general, indeed, the retinal disparity is a two-dimensional (2D) feature and the full decoding of the population response would require the proper solution of the aperture problem [17]. This can be achieved, by example, through the intersection of the constraints provided by the different orientation channels (cf. [4]). If one proceeds in such a way that is by recovering the full disparity vector, the disparity detectability range would still be limited to $\pm\Delta$, and the horizontal (vertical) component of the full disparity vector will then used for the control of horizontal (vertical) vergence. Unless one uses computationally expensive multiscale techniques for widening the disparity detectability range, this approach would considerably limit the working range of the vergence control. 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. To this end, we can map the 2D disparity feature space into the 1D space of the projected horizontal disparities, where the orientation θ plays the role of a parameter. More precisely, by assuming $\delta_V = 0$, the dimensionality of the problem of disparity estimation reduces to one, and the orientation of the receptive field is used as a degree of freedom to extend the sensitivity range of the cells' population to horizontal disparity stimuli. In this way, each orientation channel has a sensitivity for the horizontal disparity that can be obtained by the projection of the oriented phase difference on the (horizontal) epipolar line in the following way:

$$\delta_H^\theta = \frac{\Delta\psi}{2\pi k_0 \cos\theta} \tag{4}$$

Fig. 3a shows the horizontal disparity tuning curves obtained from the population for different orientations of the receptive fields. 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 δ_H^{est} is

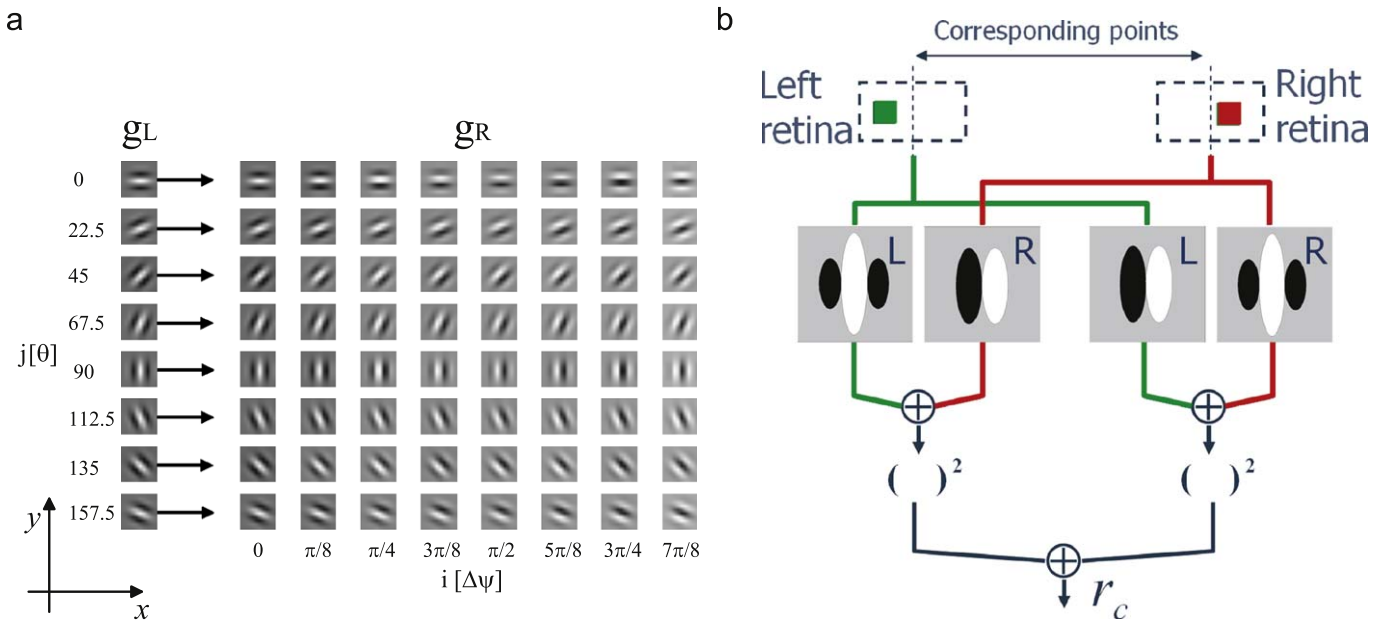

Fig. 2. Architectural resources for the distributed representation of the disparity. (a) The population of the model binocular simple cell receptive fields for each retinal location. For each spatial orientation θ, the binocular simple cell response is obtained by summing the response of a left receptive field characterized by a phase $\psi_L = 0$, with a right receptive field characterized by one of the N_p phases in the range $(-\pi, \pi)$. (b) The complex cells binocular energy unit for a single orientation channel is constructed as the squared sum of a quadrature pair of simple cells. The green and the red pathways relate to the monocular "quadrature pair" of left and right simple cell receptive fields, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

obtained by

$$\delta_H^{est} = \frac{\sum_{i=1}^{N_p}\sum_{j=1}^{N_o}\frac{\Delta\psi_i}{2\pi k_0 \cos\theta_j} r_c^{ij}}{\sum_{i=1}^{N_p}\sum_{j=1}^{N_o} r_c^{ij}} \tag{5}$$

where r_c^{ij} denotes the response of the complex cell characterized by the i-th phase difference and by the j-th orientation. The dashed line plots in Fig. 3 b–c show the resulting disparity curves obtained by population decoding. The estimate of the disparity can be considered correct when the stimulus disparity is within $\pm\Delta$. By analyzing the tuning curves of the population (see Fig. 3 a) we observe that the peak sensitivity of cells that belong to a single orientation channel is uniformly distributed in a range that increases with the orientation angle θ of the receptive field, as the horizontal projection of the frequency of the Gabor function declines to zero. Thus, applying the center of mass decoding strategy, separately for each orientation, we can obtain j different estimates of the disparity:

$$\delta_{H,\theta_j}^{est} = \frac{\sum_{i=1}^{N_p}\frac{\Delta\psi_i}{2\pi k_0 \cos\theta_j} r_c^{ij}}{\sum_{i=1}^{N_p} r_c^{ij}} \tag{6}$$

It is worthy to note that 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 and accuracy of the measure (as an extreme case, horizontal receptive fields are unable to detect horizontal disparities, i.e., $\delta_H^{\theta=0} \to \infty$). In any case, the estimate of the disparity can be considered correct in a range around $[-\Delta, \Delta]$, only.

Moreover, since the 1D tuning curves of the population were obtained under the assumption of horizontal disparity only, when the vertical disparity in the images differs from zero, the correctness of estimate of the actual component of the horizontal disparity has to be verified. We observe that (see Fig. 3 b and c, top row), the disparity estimated by the whole population is unaffected by nonnull vertical disparities, as well as the estimate obtained by the orientation $\theta = 0$ (vertically oriented cells are indeed, by definition, sensitive to horizontal disparity only). On the contrary, the estimated disparity obtained for $\theta \neq 0$ shows a dependence on vertical disparity that increases with θ (see Fig. 3 c, middle and bottom row), and leads to a systematic error response.

2.2. Control signal extraction

A desired feature of disparity-vergence curves is an odd symmetry with a linear segment passing smoothly through zero disparity, which defines critical servo ranges 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. Starting from the estimated disparity curves shown in Fig. 3 b, we can exploit the responses at different orientations to design linear servos that work outside the reliability range of disparity estimation. Yet, we have to cope with the attendant sensitivity to vertical disparity, which is an undesirable effect that alters the control action. Hence, given a stimulus with horizontal and vertical disparity δ_H and δ_V, we want to combine the population responses in order to extract a vergence control proportional to the δ_H to be reduced, regardless of any possible δ_V. We demonstrate that such disparity vergence response can be approximated by proper weighting of the population cell responses where disparity tuning curves act as basis functions. 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 disparity vergence response curves r_v^k are obtained by a weighted sum of the complex cell responses, for the different orientations and disparity tuning, over a spatial region related to the parafoveal area (see Fig. 4) of approximately 5° [3]:

$$r_v^k = \sum_{i=1}^{N_p}\sum_{j=1}^{N_o} w_{ij}^k r_c^{ij} \tag{7}$$

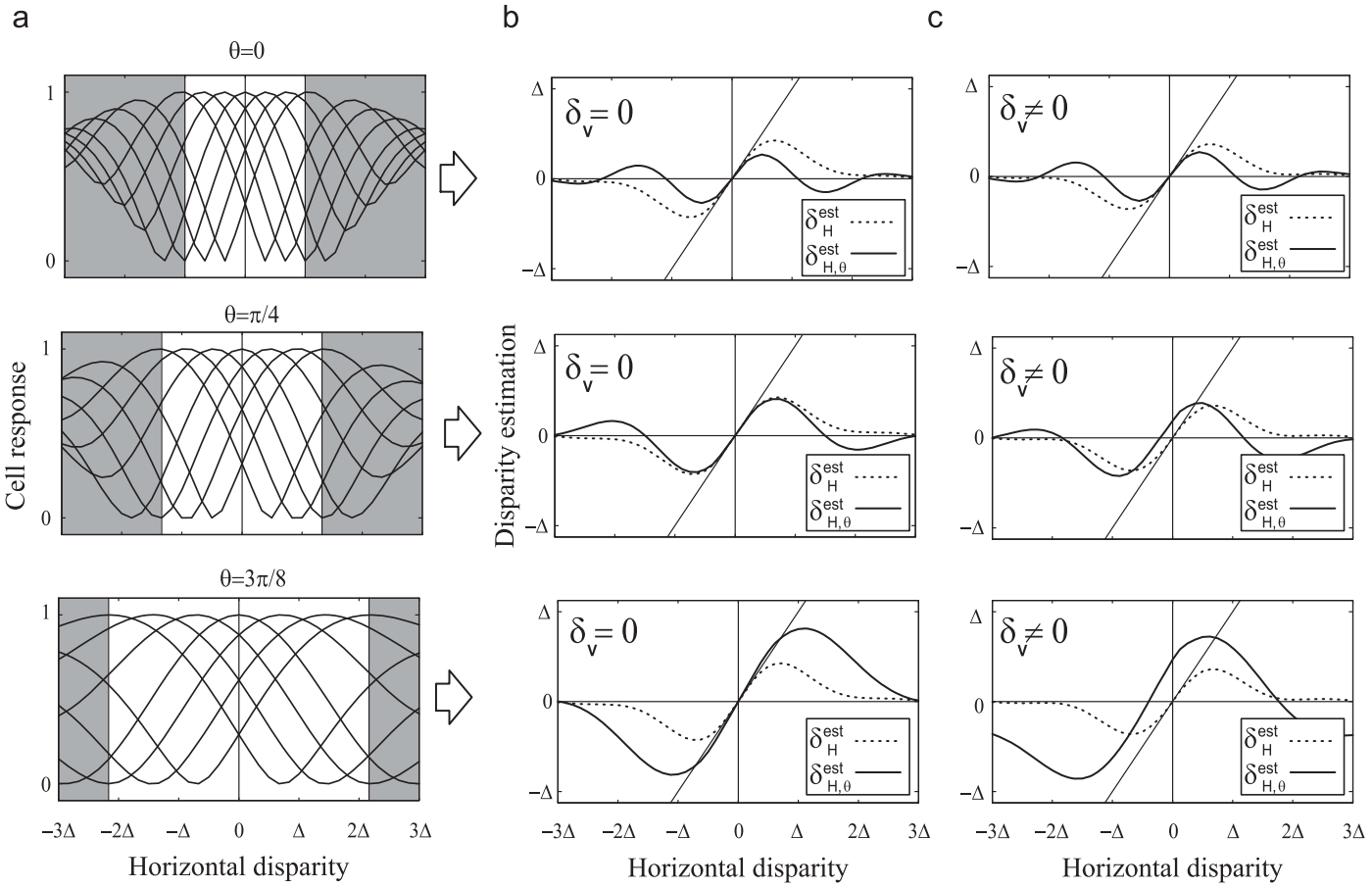

Fig. 3. (a) Disparity tuning curves of complex cells at different orientations: the increase of the slant of the receptive field produces an apparent increase of the range of reliability of the estimated disparity. (b) Estimated horizontal disparity using single orientation channels (solid lines) in the presence of horizontal disparity only ($\delta_V = 0$). (c) Estimated horizontal disparity using single orientation channels (solid lines) in the presence of a fixed amount of vertical disparity ($\delta_V \neq 0$). Dashed line plots refer to the horizontal disparity estimates obtained by combining all the orientation channels.

where the index k denotes the different kind of the desired vergence response curves. Referring to a common classification [16] we divide the V1 cells into five categories: near (*NE*) and far (*FA*) dedicated to coarse stereopsis, and tuned near (*TN*), tuned far (*TF*) and tuned zero (*T* 0) for fine stereopsis. The weights w_{ij}^k are obtained through a recursive LMS algorithm. Actually, to decrease the computational cost, the complex cell responses can be first averaged in a parafoveal region and then decoded to compute the vergence cells r_v^k, producing the same results. From the control point of view, we assume that small values of vertical disparities do not affect the disparity-vergence curves. Moreover, to mildly constraint the solution of the problem and, in the meantime to ensure a good control stability, we pose the VD independence constraint for HD $\simeq 0$, only. Under this assumption, we can design the disparity-vergence curves that define the visual servos by considering the tuning curves obtained separately for VD=0 and HD=0 (i.e. the orthogonal cross-section of the oriented 2D disparity tuning curves of the binocular energy model). More precisely, the profile of the desired vergence curve v_H^k 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 formula. This term ensures that the sum of the vertical disparity tuning curves $r^c(\delta_V;\theta,\Delta\psi)$, weighted with the same $\mathbf{w}^k$, approximates v_V^k. To overcome the difficulties of approximating a constant with a combination of a limited number of periodic basis functions, we impose v_V^k to have a profile that is mildly constant as the one that can be obtained by summing the tuning curves all together ($v_V^k = \sum_{i=1}^{N_p}\sum_{j=1}^{N_o} r_c^{ij}(\delta_V)$). Hence, the weights $\mathbf{w}^k$ are obtained by minimizing the following functional:

$$E(\mathbf{w}^k) = \left\| \sum_{i=1}^{N_p}\sum_{j=1}^{N_o} r_c^{ij}(\delta_H) w_{ij}^k - v_H^k \right\|^2 + \cdots + \lambda \left\| \sum_{i=1}^{N_p}\sum_{j=1}^{N_o} r_c^{ij}(\delta_V)(w_{ij}^k - 1) \right\|^2 \tag{8}$$

where $\lambda > 0$ balances the relevance of the second term over the first. In our simulations we fixed $\lambda = 1$ in order to give the same relevance to both δ_H and δ_V. To test the functionality of the model, at this stage, we used the same kind of stimuli adopted to compute the disparity tuning curves of the cells, so that we expect the disparity vergence tuning curve to be the same we drew from the minimization. The stimuli have a disparity varying in the same range used for the tuning curves, and the control computed has the same shape of the desired curves (Fig. 4 b). A drawback that arises is that if the image contrast is lowered, disparity vergence tuning curves hold the same shape, but their gain is consequently lowered, with the effect that the speed of the vergence movements is modulated by the image contrast. The estimated disparity does not show this effect because the center of mass decoding strategy means to take a decision on the disparity value, regardless to the contrast of the stimulus (cf. [18]). By analogy with the formula used to decode the disparity, we can introduce the same normalization term to let the system work in the proper way independently

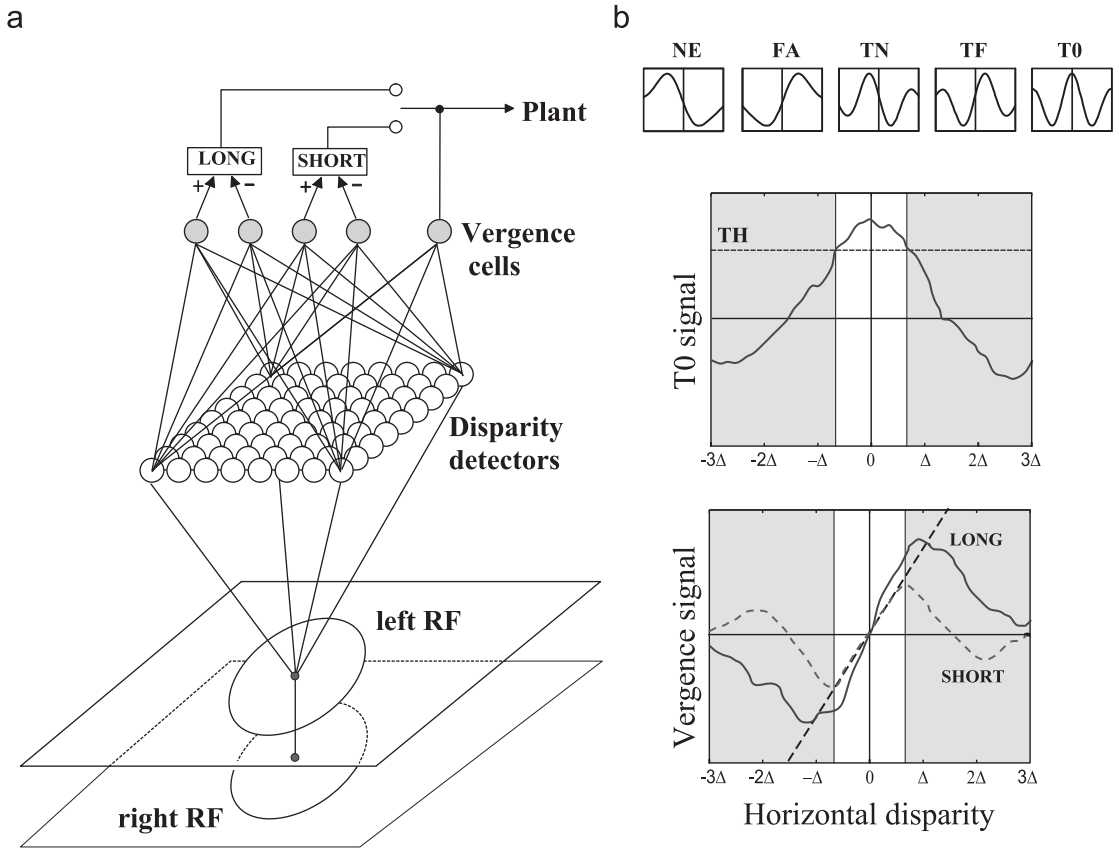

Fig. 4. (a) Extraction of the vergence control signals: in each image location, the disparity is encoded by the population of disparity detectors, i.e. the complex cells r_c^{ij}. The population response is decoded through different weighted summations in order to extract five signals r^{NE}, r^{FA}, r^{TN}, r^{TF}, and r^{T0}. These signals are combined in a differential way to obtain the LONG and SHORT controls, whose interplay is regulated by the $T0$ signal. (b) The v_H^k target curves to be approximated by the LMS minimization (top row), and the effective LONG (solid line), SHORT (dashed lines), and $T0$ signals, computed by the model stimulated with a random dot stereograms (RDS). When the disparities in the scene are small $T0$ is above a proper threshold TH and it enables the SHORT control (white region), otherwise it enables the LONG control (gray region).

of the image contrast:

$$r_v^k = \frac{\sum_{i=1}^{N_p}\sum_{j=1}^{N_o} w_{ij}^k r_c^{ij}}{\sum_{i=1}^{N_p}\sum_{j=1}^{N_o} r_c^{ij}} \tag{9}$$

2.3. Signal choice

With reference to the five categories of the disparity-vergence curves, it is plausible to think that the first two generate the fast and coarse component and the others the slow and fine component of the vergence movements. In practice the fast-coarse control is given by long $= r^{NE}$-r^{FA}, while the slow-fine is given by short $= r^{TN} - r^{TF}$ (see Fig. 4). 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. 4 b). Out of its range of linearity, the SHORT signal decreases and loses efficiency to the point where it changes sign, thus generating a vergence movement opposite to the desired one. On the contrary for small disparities the LONG control signal yields overactive vergence signal that make 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. 4 b). On the contrary, for large stimulus disparities, r^{T0} is below the threshold and it enables the LONG control (see gray regions in Fig. 4 b).

An straightforward but meaningful effect that arises from calculating the SHORT and the LONG controls in a differential way is a strong robustness to noise. If we add a Gaussian white noise to the population response, both the decoding of the disparity and the computation of the r_v^k signals, would be affected. Since the weights $\mathbf{w}^k$ are normalized, it is easy to demonstrate that the noise terms on r^{NE} and r^{FA} cancel each other while differentiating to compute the LONG control, and so it happens for the SHORT one. Simulation results evidenced that, when one adopts the differential SHORT and LONG control signals, the S/N ratio is ~ 6 dB higher than the input S/N ratio for the complex cell responses.

3. Results

3.1. Test with RDS

We tested the proposed model with synthetic stimuli consisting of random dot stereograms (RDS) in which the stereo image pairs are shifted horizontally. Specifically, we applied horizontal disparity steps varying from -3Δ to 3Δ. The model works in a perception-action loop in which the vergence movements are simulated, reducing step by step the disparity between the left and right images by an amount proportional to the vergence control. The vergence movement is computed both through the estimation of the disparity δ_H^{est} and through the vergence signals r_v^k. For a direct comparison of the vergence control performance, we showed the percentage of vergence movement accomplished by the two mechanisms for different

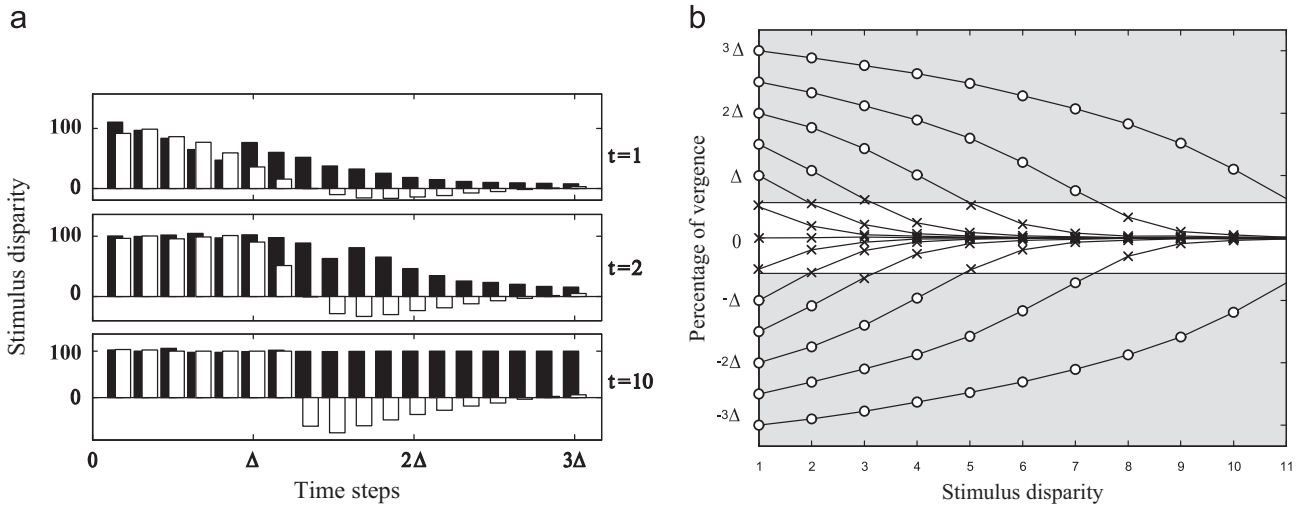

Fig. 5. Behavior of the vergence control tested with a RDS characterized by a disparity step in a range varying from -3Δ to 3Δ. (a) Comparison of the percentage of vergence achieved by the model using the estimated disparity δ_H^{est} (white bars), and using the r_v^k signals (black bars). Only the positive axis is represented because the response is symmetric around zero disparity. The graphs represent the status of the system after 1, 2 and 10 time steps. At each time step the r_v^k signals are able to reach the target in the whole tested range, while δ_H^{est} yields a wrong control for disparities larger than Δ. (b) Time course of the stimulus disparity step, after step-by-step the vergence corrections. On each trace the interplay between the LONG and SHORT control is evidenced. The LONG control (open circles) is active when disparities are above a defined threshold value (gray area, cf. Fig. 4 b), and the SHORT control (crosses) is active when the disparities are below that threshold (white area).

time steps (see Fig. 5a). Since the behavior of the two mechanisms is symmetric with respect to zero disparity, we show the positive semiaxis, only. A percentage value higher than 100 indicates an overshoot of the movement, whereas a value lower than zero indicates a movement in the opposite (i.e. wrong) direction. After the first time step (Fig. 5 a top row), if the stimulus disparity is within Δ, the behavior is slightly better for δ_H^{est} (white bars), whereas outside this range it produces a vergence movement that is the opposite of the one requested. The r_v^k signals (black bars) produce almost the same movement of δ_H^{est} for small disparity steps, but they are able to achieve slow but effective vergence movements up to the limit of the tested range. At the second time step (Fig. 5 a middle row), for disparity steps smaller than Δ, both the mechanisms reach the target, and for higher disparities the behavior is similar to the previous time step. After 10 time steps (Fig. 5 a bottom row), we observed that δ_H^{est} was able to work in the proper way only for disparities within Δ, whereas r_v^k was able to reach the target in all the tested range. The switching behavior from the LONG to the SHORT mode in a simulated vergence response is shown in Fig. 5 b. The SHORT mode plays the role when the stimulus disparity falls into the linear servo range of its disparity-vergence response (see Fig. 4 b), otherwise the control relies on the LONG mode.

3.2. Test with a frontoparallel plane

Considering 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. 6a). 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 $b = 70$ mm, the nodal length is $f_0 = 17$ mm, and the stimulus is projected onto the retinal plane, with a size of 6 mm, thus considering a field of view of almost 20°. 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: $\Delta\alpha_R = -\Delta\alpha_L = -\arctan(r/2f_0)$, where r is computed by considering the weighted average of the vergence responses r_v^k or of the estimated disparities δ_H^{est}. The area where the average is computed, is a neighborhood of the fovea of 5°, and its size is based on physiological experiments [3] that show that it is the maximum extent of the retina where the disparity stimulus is integrated to drive vergence eye movements in humans.

The first test considers a fixed frontoparallel plane at a given distance, while the eyes are fixating on the surface of the plane. The plane steps back and forth by an amount that varies from trial to trial. Fig. 6 b shows that a control based on the disparity computation produces the correct change of the vergence angle (dotted lines), when the size of the step is restrained. On the other hand, the implemented model is able to produce a faster change of the fixation point (solid lines), and, even for larger depth steps, the model is able to ensure a reliable vergence control. Moreover, once the fixation point has reached the plane in depth, the disparity in the fovea is approximately zero and the system is able to ensure a stable fixation.

The second test considers a frontoparallel plane whose position in depth varies continuously in time as a ramp and a sinusoid. The slope of the ramp is varied from 0.5 cm per time step to a pure step (Fig. 7a). While for small values only the SHORT control is enabled, in the other cases the initial part of the vergence is produced by the LONG one, and the interplay between the two controls is very similar to the one observed in the transient and sustained components of the physiological responses [9]. In support of this hypothesis, in case of both a divergent and a convergent ramp, the simulated vergence movements are qualitatively very similar to the results obtained in physiological experiments. In the same way, the frequency of the sinusoid that controls the depth of the plane was varied between 7 and 38 time steps, and again the simulated results (Fig. 7 b) are qualitatively similar to the experimental data [9]. Increasing the frequency, it is evident a transition from a slow and smooth tracking of the plane, due to the SHORT control, to a combination of the LONG and SHORT controls. When the frequency becomes too high, the system is no more able to follow the stimulus in depth.

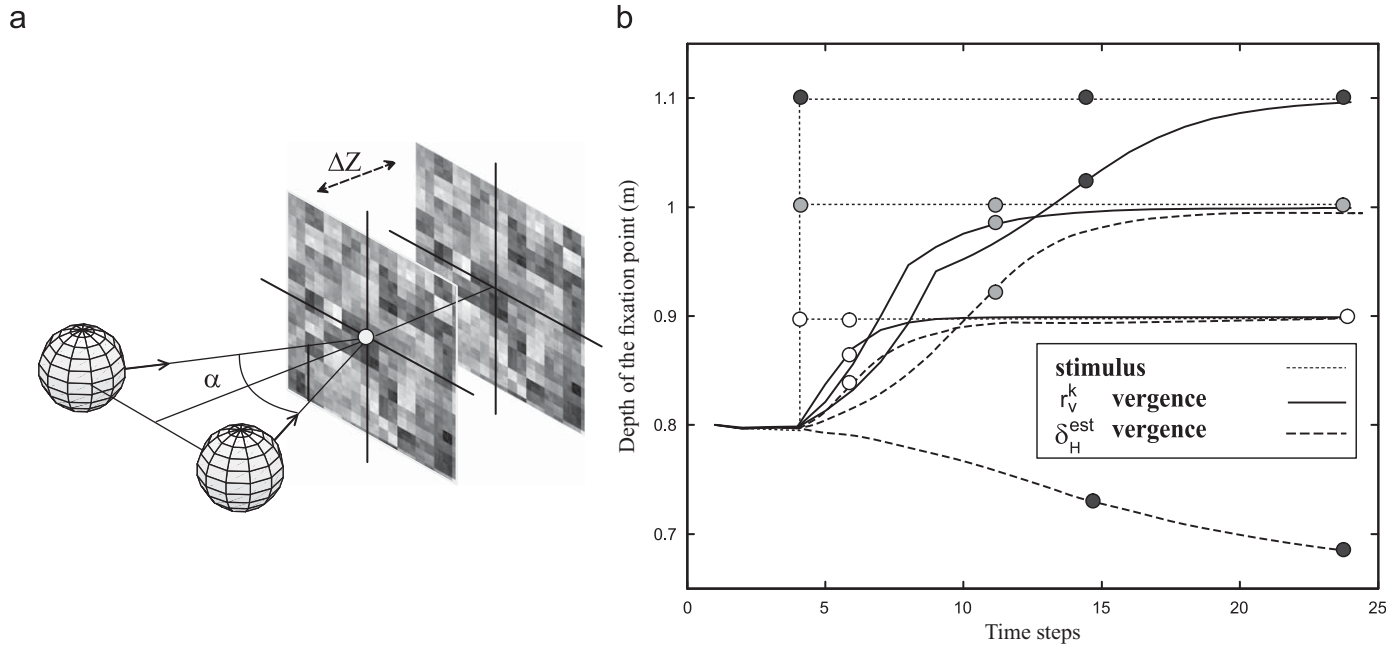

Fig. 6. (a) The simulated experimental setup, consisting of the eyes looking at a plane characterized by an RDS pattern, and perpendicular to the binocular line of sight. (b) Behavior of the vergence control using r_v^k *vs.* δ_H^{est} in case of a diverging step. The eyes are forced to keep the straightforward gaze direction while changing the vergence angle. The short-dashed line plots represent the three depths of the stimulus plane. The disparity-based and model vergence controls are denoted by long-dashed and solid line plots, respectively. The symbols on the stimulus/control plots distinguish the amount of the disparity step: white circles refer to a step of $\Delta Z = 0.1$ m, gray circles to a step of $\Delta Z = 0.2$ m, and black circles to a step of $\Delta Z = 0.3$ m. The r_v^k controls yield correct fixations on the target planes in all the presented cases, whereas the δ_H^{est} controls produce a wrong movement for a depth step above a certain threshold (long-dashed lines, black circles).

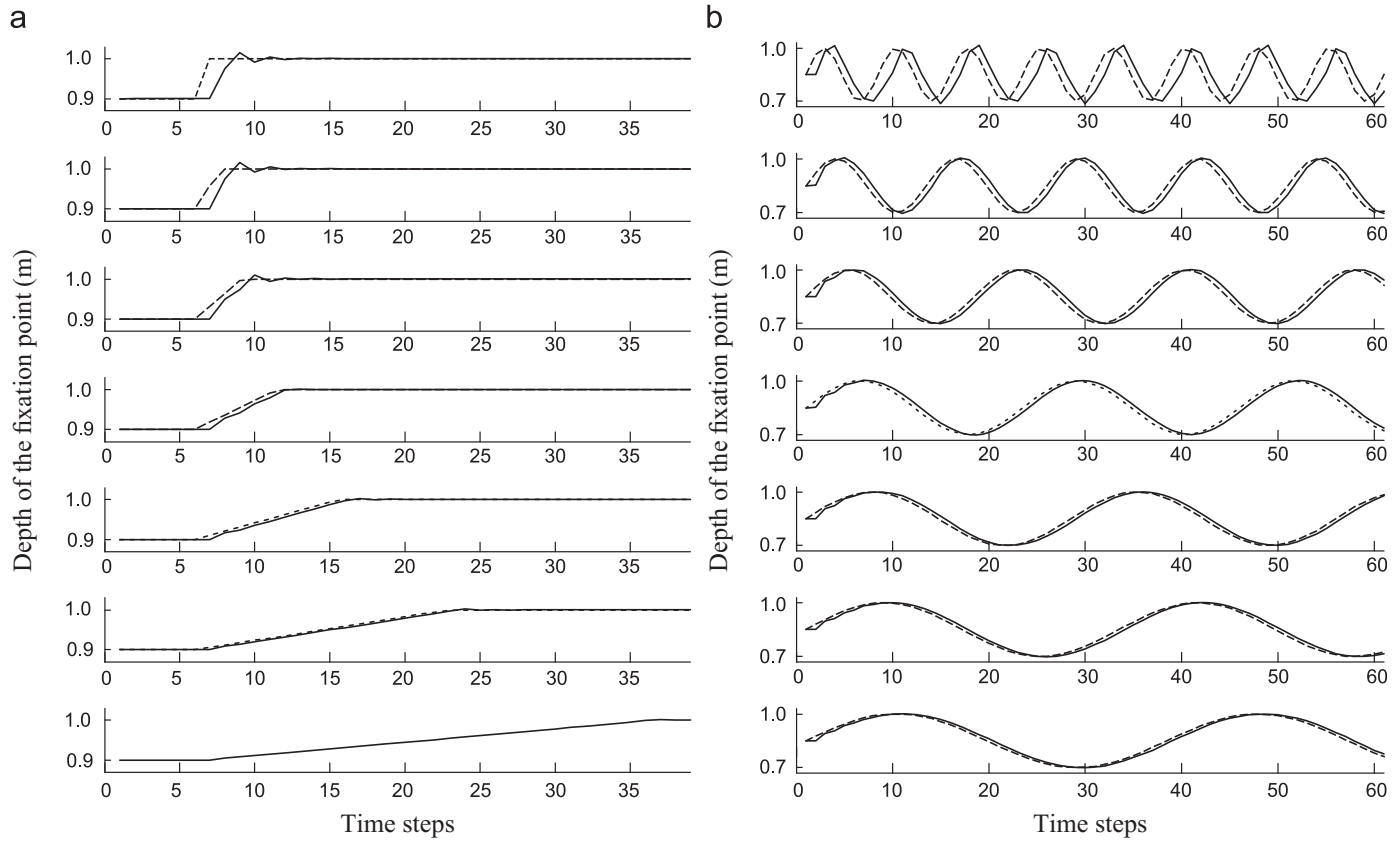

Fig. 7. (a) Time course of the fixation point to diverging ramps with different slopes, and (b) to sinusoids characterized by different periods. The dashed line represents the depth of the stimulus and the solid one is the depth of the fixation point.

4. Discussion and conclusions

In conclusion, instead of a continuous and linear feedback system, where a single vergence control proportional to the estimated disparity is used [4,5], the model here proposed relies on two controls with different and specific tasks. The LONG signal produces a coarse control to yield inaccurate but fast vergence movements for disparities that are far outside the FR. The SHORT signal, instead, yields a precise ad smooth movements for small vergence request.

The proposed model can be compared with neurophysiological evidences in visual cortical areas.

Disparity selective cells are ubiquitous in the dorsal stream of the visual system (area V1 [1,16,19–22], MT [23,24], MST [25–27]). It has been often pointed out that these cells provide a direct measure of vergence error and so have the potential to provide the primary drive for vergence eye movements like those described in our paper. More specifically, Takemura reported that the earliest vergence eye movements were attenuated after chemical lesions in MST [27]. A further quantitative analysis [7] revealed that 20% of the neurons recorded in MST were sensitive to disparity steps applied to large textured patterns, and the associated vergence velocity response were well correlated with their summed activity. Although there is no direct neurophysiological evidence supporting the weighted averaging of complex cell responses, the hierarchical convergence principle can be advocated to motivate as a first approximation, the biological plausibility of our model. The receptive field at a given stage can be obtained by aggregating overlapped receptive fields of cells of the previous stage. In this way, the disparity tuning curves observed in MT and MST cells can be inherited by the previous area (V1). From this perspective, the archetypal disparity tuning curves obtained by our model (r^{NE}, r^{FA}, r^{TN}, r^{TF}, r^{NE}T0, see Fig. 4 b) are referable to the different groups used by Takemura [6] to cluster MST disparity sensitivities. From direct combination of these archetypes we derive the vergence motor commands, in a similar but simplified way with respect to the one proposed by Takemura, because neurophysiological tuning curves have a larger variability than our archetypes.

From the analysis of the response amplitude of vergence movements to step disparity stimuli in humans, Alvarez et al. [11] pointed out that at least two control modes may mediate a single vergence response, and the second component is generated if the primary one does not reach at least the 80% of the stimulus amplitude. Since the saccadic and the vergence systems share several features, this behavior predicts that the vergence eye movements should be controlled by a switching component similar to the one that characterizes the saccades, even if there are no experimental evidences of neural correlates of such a mechanism. Moreover, the amplitude and the frequency of the vergence response in humans increase with stimulus velocity, and this suggests that the switching mechanism is triggered by an error between the stimulus and the response, and that this error could be obtained from visual feedback. From this perspective, the T 0 signal, indeed produces the desired error, because when the stimulus amplitude is large, it is under the defined threshold and activates the LONG signal, which can be related to the primary component. The resulting vergence movement reduces the stimulus amplitude. The visual feedback produces an increase of T 0, switching the vergence control to the SHORT signal, related to the secondary component.

Even though the model does not take into account the eye balls and the eye muscles, so that the movements take place with no inertia, the vergence angle shows a behavior similar to the one that has been observed in psychophysical data. More precisely, considering the experimental responses of different subjects to a divergent ramp, reported in [9], the model produces a similar overshoot for ramps with high speed, due to the LONG control, after which the SHORT control activates and produces a fine refinement of the vergence to the extent to which it yields a stable fixation on the surface of the target plane. On the other side, with both ramps with lower speed, and with sinusoids with a large period, the SHORT control is able to achieve a smooth and continuous tracking of the plane moving in depth. Moreover, for both ramps and sinusoids, it is possible to observe an interplay between the two controls similar to the one between the transient and the sustained components of vergence eye movements.

Acknowledgement

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

References

[1] I. Ohzawa, R.D. Freeman, G.C. DeAngelis, Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors, Science 249 (4972) (1990) 1037–1041.
[2] G.S. Masson, C. Busettini, F.A. Miles, Vergence eye movements in response to binocular disparity without depth perception, Nature 389 (1997) 283–286.
[3] R.S. Allison, I.P. Howard, X. Fang, The stimulus integration area for horizontal vergence, Exp. Brain Res. 156 (2004) 305–313.
[4] W.M. Theimer, H.A. Mallot, Phase-based vergence control and depth reconstruction using active vision, CVGIP Image understanding 60 (3) (1994) 343–358.
[5] S.S. Patel, H. Ogmen, B.C. Jiang, Neural network model of short-term horizontal disparity vergence dynamics, Vis. Res. 37 (10) (1996) 1383–1399.
[6] A. Takemura, Y. Inoue, C. Quaia, F.A. Miles, Single-unit activity in cortical area MST associated with disparity-vergence eye movements: evidence for population coding, J. Physiol. 85 (5) (2001) 2245–2266.
[7] A. Takemura, K. Kawano, C. Quaia, F.A. Miles, Population coding of vergence eye movements in cortical area MST, in: L. Harris, M. Jenkin (Eds.), Levels of Perception, 2006.
[8] J.L. Semmlow, G.K. Hung, K.J. Ciuffreda, Quantitative assessment of disparity vergence components, Invest. Ophthalmol. Vis. Sci. 27 (1986) 558–564.
[9] G.K. Hung, J.L. Semmlow, K.J. Ciuffreda, A dual-mode dynamic model of the vergence eye movement system, Trans. Biomed. Eng. 36 (11) (1986) 1021–1028.
[10] V.V. Krishnan, L. Stark, A heuristic model for the human vergence eye movement system, IEEE Trans. Biomed. Eng. 24 (1977) 44–49.
[11] T.L. Alvarez, J.L. Semmlow, W. Yuan, Closely spaced, fast dynamic movements in disparity vergence, J. Neurophysiol. 79 (1998) 37–44.
[12] P.D.R. Gamlin, Neural mechanisms for the control of vergence eye movements, Ann. N.Y. Acad. Sci. 956 (2002) 264–272.
[13] S. Ferraina, M. Par, R.H. Wurtz, Disparity sensitivity of frontal eye field neurons, J. Neurophysiol. 83 (2000) 625–629.
[14] N. Qian, Computing stereo disparity and motion with known binocular cell properties, Neural Comput. 6 (3) (1994) 390–404.
[15] D.J. Fleet, H. Wagner, D.J. Heeger, Neural encoding of binocular disparity: energy models, position shifts and phase shifts, Vis. Res. 36 (12) (1996) 1839–1857.
[16] G.F. Poggio, Mechanism of stereopsis in monkey visual cortex, Cerebral Cortex 5 (1995) 193–204.
[17] M.J. Morgan, E. Castet, The aperture problem in stereopsis, Vis. Res. 37 (1997) 2737–2744.
[18] D.J. Fleet, H. Wagner, D.J. Heeger, Modelling binocular neurons in the primary visual cortex, in: M. Jenkin, L. Harris (Eds.), Computational and Biological Mechanisms of Visual Coding, Cambridge University Press, Cambridge, 1996.
[19] Y. Trotter, S. Celebrini, B. Stricanne, S. Thorpe, M. Imbert, Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1, J. Neurophysiol. 76 (1996) 2872–2885.
[20] E.L. Smith III, Y. Chino, H. Cheng, Binocular combination of contrast signals by striate cortical neurons in the monkey, J. Neurophysiol. 78 (1997) 366–382.
[21] S.J. Prince, A.D. Pointon, B.G. Cumming, A.J. Parker, The precision of single neuron responses in cortical area V1 during stereoscopic depth judgments, J. Neurosci. 20 (2000) 3387–3400.
[22] B.G. Cumming, A.J. Parker, Binocular neurons in V1 of awake monkeys are selective for absolute, no relative disparity, J. Neurosci. 19 (1999) 5602–5618.
[23] G.C. DeAngelis, W.T. Newsome, Organization of disparity-selective neurons in macaque area MT, J. Neurosci. 19 (1999) 1398–1415.

[24] J.H.R. Maunsell, D.C. Van Essen, Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity, J. Neurophysiol. 49 (1983) 1148–1167.
[25] S. Eifuku, R.H. Wurtz, Response to motion in extrastriate area MSTl: disparity sensitivity, J. Neurophysiol. 82 (1999) 2462–2475.
[26] J.P. Roy, H. Komatsu, R.H. Wurtz, Disparity sensitivity of neurons in monkey extrastriate area MST, J. Neurosci. 12 (1992) 2478–2492.
[27] A. Takemura, Y. Inoue, K. Kawano, C. Quaia, F.A. Miles, Evidence that disparity-sensitive cells in medial superior temporal area contribute to short-latency vergence eye movements, Soc. Neurosci. Abstr. 25 (1999) 1400.

Agostino Gibaldi received his degree in Biomedical Engineering from the University of Genoa, Genoa, Italy, in 2007. Since then, he is with the Physical Structure of Perception and Computation (PSPC) Group in the Department of Biophysical and Electronic Engineering (DIBE) of the University of Genoa, where he is working on his Ph.D. His research interests are related to cortical models of V1, MT and MST areas, in relation with the estimation of disparity, the perceptio-action loop for the control of vergence eye movements, and the optic flow analysis for navigation.

Manuela Chessa received the Ph.D. in Bioengineering from the University of Genoa, Genoa, Italy, in 2009. Since 2005, she is with the Physical Structure of Perception and Computation (PSPC) Group in the Department of Biophysical and Electronic Engineering (DIBE) of the University of Genoa. Her primary research interests are the development of cortical models for the estimation and the analysis of optic flow and disparity, inspired by the processing of cortical areas, the study of active binocular vision systems, the simulation of robotic systems, and the developing of neuromorphic algorithms.

Andrea Canessa was born in Genoa, Italy, in 1983. He received his Bachelor Degree in Biomedical Engineering at the University of Genoa in 2004 and his Master Degree in Bioengineering (curriculum "Neuroengineering") at the University of Genoa in 2007. He is currently carrying out the Ph.D. in Bioengineering at the University of Genoa. His research interests include neuromorphic models for disparity and motion and eye movements influences on depth perception.

Silvio P. Sabatini graduated in Electronic Engineering at the University of Genoa, Italy (1992); Ph.D. in Electronic Engineering and Computer Science at the Department of Biophysical and Electronic Engineering (DIBE) of Genoa (1996). Since 1999 Assistant Professor in Computer Science at the University of Genoa. In 1995 he promoted the creation of the "Physical Structure of Perception and Computation" (PSPC) Research Group at DIBE (http://pspc.dibe.unige.it/) to develop models that capture the "physicalist" nature of the information processing that takes place in the visual cortex, to understand the signal processing strategies adopted by the brain and build novel algorithms and hardware devices for artificial perception machines. His current research interests include biocybernetics of vision, theoretical neuroscience, neuromorphic engineering, and artificial vision. He is author of more than 60 international papers in peer-reviewed journals and conferences.

Fabio Solari received the Laurea degree in Electronic Engineering from the University of Genoa, Italy, in 1995. In 1999 he obtained his Ph.D. in Electronic Engineering and Computer Science from the same University. Since 2005, he has been appointed as Assistant Professor in Computer Science at the Department of Biophysical and Electronic Engineering (DIBE), University of Genoa. His research activity concerns the study of the physical processes of biological vision to inspire novel algorithms and artificial perceptual machines based on neuromorphic computational paradigms. In particular, he is interested in: (i) developing computational models of neural architectures, mainly in the dorsal stream of visual cortex, and the study of visuomotor transformations (eyes, head, arm coordination); (ii) algorithms for motion and depth computation, exploiting processing techniques based on adaptive and multichannel filtering; (iii) robotic systems for active vision and for autonomous navigation based on visual information.