Robotics and Autonomous Systems 71 (2015) 23–34

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Autonomous learning of disparity–vergence behavior through distributed coding and population reward: Basic mechanisms and real-world conditioning on a robot stereo head

Agostino Gibaldi *, Andrea Canessa, Fabio Solari, Silvio P. Sabatini

Physical Structure of Perception and Computation Group, Department of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Italy

HIGHLIGHTS

- We implemented a cortical model for the vergence control based on a population of disparity detectors.
- The model is able to autonomously learn its behavior by means of an internal parameter.
- The speed of convergence and the precision of the control precision were evaluated on different disparity ranges and learning signals.
- The informative content of the different orientation channels was assessed.
- The learning capabilities on real robot stereo pairs demonstrate an adaptation to the stimulus characteristics.

ARTICLE INFO

Article history:
Available online 17 January 2015

Keywords:
Vergence control
Stereo vision
Neural models
Binocular disparity
Energy model
Hebbian learning
Active vision

ABSTRACT

A robotic system implementation that exhibits autonomous learning capabilities of effective control for vergence eye movements is presented. The system, directly relying on a distributed (*i.e.* neural) representation of binocular disparity, shows a large tolerance to the inaccuracies of real stereo heads and to the changeable environment. The proposed approach combines early binocular vision mechanisms with basic learning processes, such as synaptic plasticity and reward modulation. The computational substrate consists of a network of modeled V1 complex cells that act as oriented binocular disparity detectors. The resulting population response, besides implicit binocular depth cues about the environment, also provides a global signal (*i.e.* the overall activity of the population itself) to describe the state of the system and thus its deviation from the desired vergence position. The proposed network, by taking into account the modification of its internal state as a consequence of the action performed, evolves following a differential Hebbian rule. The overall activity of the population is exploited to derive an intrinsic signal that drives the weights update. Exploiting this signal implies a maximization of the population activity itself, thus providing an highly effective reward for the developing of a stable and accurate vergence behavior. The role of the different orientations in the learning process is evaluated separately against the whole population, evidencing that the interplay among the differently oriented channels allows a faster learning capability and a more accurate control. The efficacy of the proposed intrinsic reward signal is thus comparatively assessed against the ground-truth signal (the actual disparity) providing equivalent results, and thus validating the approach. Trained in a simulated environment, the proposed network, is able to cope with vergent geometry and thus to learn effective vergence movements for static and moving visual targets. Experimental tests with real robot stereo pairs demonstrate the capability of the architecture not just to directly learn from the environment, but to adapt the control to the stimulus characteristics.

1. Introduction

The primary visual cortex (V1) is considered the processing substrate for the retinal binocular disparity, and enables both stereopsis and vergence eye movements [1,2]. While stereopsis is the process that allows for the perception of depth by disparity information, vergence movements are responsible for ensuring the

* Corresponding author.
E-mail addresses: agostino.gibaldi@unige.it (A. Gibaldi), andrea.canessa@unige.it (A. Canessa), fabio.solari@unige.it (F. Solari), silvio.sabatini@unige.it (S.P. Sabatini).

http://dx.doi.org/10.1016/j.robot.2015.01.002
0921-8890/

singleness of vision and for maintaining stable fixations. These two mechanisms mutually influence and develop together in a process in which a finer vergence movement helps improving the stereoacuity, and a finer stereoacuity provides effective information for finer vergence movements.

When an infant starts looking around, both the mechanisms are not yet developed, and have to be learned. Since the infant has no explicit teacher in learning how to control his/her own eyes [3], the only "supervision" is gathered from interaction, *i.e.* from a direct sensorimotor connection with the environment [4]. A plausible learning process should count on a reward given when the eyes fixate an object in the proper manner, thus when the singleness of vision is guaranteed. This process ensures that an infant will learn the correct vergence behavior [5,6], and that an adult can adapt the control to the stimulus characteristics [7]. In fact, adaptation in the vergence system is *"essential for an organism to maintain optimal visuomotor function"* [8].

Consequently, the retinal binocular disparity is used by the brain as a source of information, both to gain depth perception and to control the eye movements, in order to actively get a better perception of the scene, on the basis of the characteristics of the scene itself. Indeed, a visual stimulus containing a disparity, like a random dot stereogram (RDS), is equally effective in providing depth perception and in triggering the correct vergence eye movements [1,2,9].

From a computational point of view, although the complex cells of V1 are the processing substrate for both stereopsis [10] and vergence [1,2], these two tasks are carried out by two separate cortical mechanisms. While the former is capable of producing a single perception from two different retinal images, only within a *small* range of disparities (Panum's fusional area), the latter allows us to extract a vergence control for *large* disparities as well [11]. Accordingly, disparity–vergence responses might follow a fast reactive stream that directly involves V1 cells without resorting to a high level interpretation of depth. In this way, the system is brought back to the fusible range in order to ensure again the singleness of vision.

From a modellistic point of view, even though the stereo and vergence mechanisms are supposed to develop in parallel and to refine and calibrate each other, the approaches that jointly address this issues are very seldom [12–14]. Indeed, early works showed, on the one hand, how it is in principle possible to learn the receptive fields of simple and complex cells from the statistical properties of the natural images [15–18] and, on the other hand, how a network of disparity detectors is an effective substrate to guide vergence eye movements [19,20]. At an intermediate level, we can start from a large population of already developed V1-like cells and make the network to develop effective vergence control.

In [21–23], the authors proposed a read-out mechanism of the response of a population network of V1 complex cells so to specialize it for an effective vergence behavior. The vergence control is computed through a weighted linear summation of the population response. The weights are obtained by a least mean square (LS) algorithm. The population tuning curves are used as basis function and to approximate a desired behavior. On the one hand, imposing a behavior allows the architecture to take full advantage of the resources, so to cope with a wide range of disparities. On the other hand, a behavior that is externally imposed, requires a complete knowledge of the available resources.

In [24], a convolutional neural network is trained by exploiting the vergence error (difference between the actual and the desired vergence) as the learning signal. Notwithstanding the effectiveness of the approach, it requires a ground-truth knowledge of the robotic system in relation with the environment, that is unlikely to be available in the real world.

From the perspective of an active vision system, a key point is that the control should be learned *actively* through a direct interaction with the environment, not by imposing an external predefined behavior or by providing an external error. In [25] the authors focus on how a learned sensory representation is able to guide vergence movements in a behaving organism, using a biological and unsupervised reward. To this end, the response of a population of complex cells tuned to zero disparity is used to obtain the emergence of disparity tuning in a three-layer neural network, obtaining in such a way, a plausible substrate for guiding vergence movements. The network is able to specialize different servos with high sensitivity for a small disparity and broad curves for large disparities, thereby validating the approach.

The more recent models of [26,27] propose to use the output of different populations of complex cells, both for vergence control and for the reward. In particular, the reward is directly computed at each time step as the average activity of the population response, computed after a normalization and a half-wave rectification stage. As a consequence, the learning process relies on an internal parameter and requires no previous knowledge about the environment. However, since they use receptive fields with a vertical orientation ($\theta = 0$), their populations are tuned to the horizontal disparity component only, thus limiting the complexity of the problem. Their approach was extended in [28], where the role of neurons tuned to different orientations has been explored, demonstrating a meaningful role of the non-vertical receptive fields.

It is worth noting that these models rely on a selected set of resources, characterized by specific binocular properties, or specific classes of disparity tuned cells. Much more appealing, in the context of autonomous learning, is to assess (1) whether a network of disparity detectors is able to learn the proper behavior without any external supervision, and (2) what can be an ideal signal to evaluate the performance of the system and drive the learning phase. In [29] the authors demonstrated how the activity of a population of disparity detectors, designed according to biologically inspired specifications, can be exploited not only to compute the control for the vergence movement, but also as a metric of the relative position of the cameras with respect to the observed scene, and consequently to evaluate the effect (positive or negative) of the movement on the status of the system. Nevertheless, the proposed architecture was trained by using a *particle swarm optimization*, which is not likely to be a possible neural strategy.

In this paper, we demonstrate how a single scale neural architecture of disparity detectors, that can be generally used for different early vision tasks (*e.g.* disparity and optic flow estimation [30], 3D object recognition and scene understanding [31]), implicitly allows for a learning strategy of the control of vergence eye movements, grounding on basic cortical mechanisms. Mimicking the encoding of retinal disparity of primary visual cortex [32–34], the population of disparity detectors is tuned to different two dimensional disparity magnitudes along different orientations, and includes a normalization stage that ensures that the population activity is stable and sensitive to retinal disparity only. The vergence control is provided by a linear network whose weights are updated following a Hebbian rule. Since the mechanism has to evaluate the effect of an action, the basic rule has been modified to take into account a temporal asymmetry, so to drive the architecture by means of an internal reward signal that is provided by the overall population activity. From this perspective, the learning capability of each oriented channel has been evaluated, evidencing that the whole population response is critical to gather a robust and precise control in a lower convergence time, than what can be achieved by considering resources tuned to horizontal disparity, only. Indeed, a correct reward for the actions taken by the system is internally obtained by the overall increase of the population activity, which can be used to autonomously train the system towards an effective solution. Moreover, while single-scale approach allows for a reduced computational time with respect to a multi-scale one, the

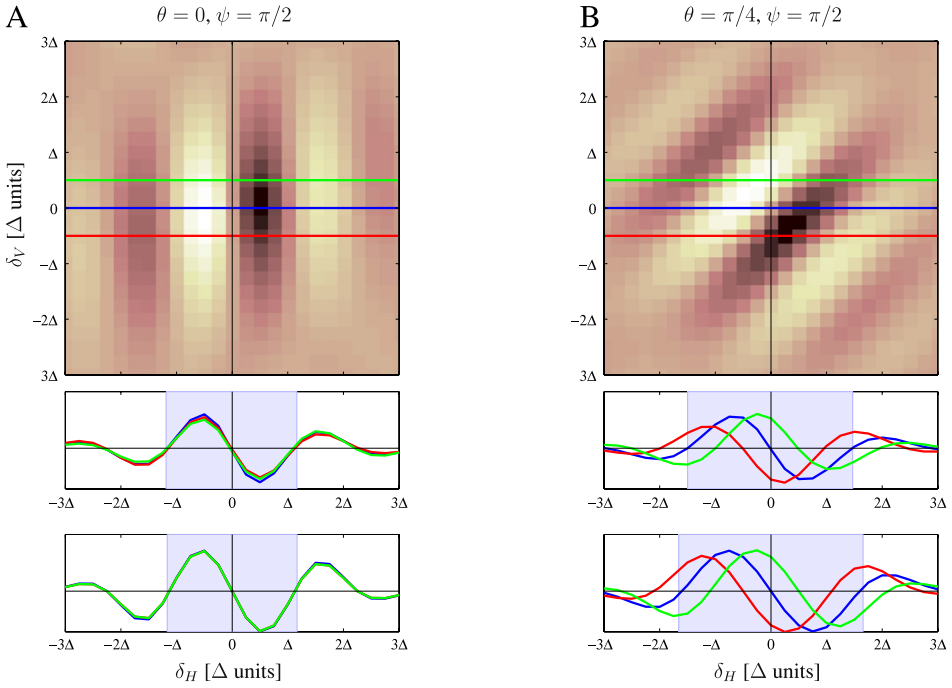


Fig. 1. The 2D response profile of two V1 modeled complex cells to a RDS stimuli. The stimulus disparity varies in the range $[-3\Delta, 3\Delta]$ for both δ_H and δ_V. The parameters used for the two represented cells are: (A) $\theta = 0$ and $\psi = \pi/2$, and (B) $\theta = \pi/4$, $\psi = \pi/2$. The top insets show the horizontal cross section of the actual (*i.e.* simulated) responses for a fixed amount of vertical disparity equal to $\Delta/2$ (green), 0 (blue) and $\Delta/2$ (red), whereas the bottom ones represent the same cross sections for the analytical model response. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

proper use of the cells tuned to non-vertical orientations allows for a wide working range of the resulting control. The robustness and adaptability of the architecture that arise from the population approach, are demonstrated on a real stereo pairs acquired by a robot head in vergent geometry, in which the learned behaviors show a conditioning from the characteristics of the actual stimulation.

The paper is organized as follows: Section 2 describes the cortical architecture of V1 complex cells and the role of the divisive normalization, Section 3 describes the implementation of the learning algorithm, based on the population response, Section 4 analyzes the results obtained both with synthetic stimuli and with the real robot stereo head, and finally in Section 5 we present our conclusions.

2. Efficient coding of binocular disparity

Taking inspiration from neurophysiological evidences [33–36], our architecture of disparity detector consists of a population of V1-like binocular complex cells energy units obtained by summing the squared responses of quadrature pairs of simple cells. These complex cells, each with its own sensitivity to a specific vector disparity $\boldsymbol{\delta}(\mathbf{x}) = (\delta_H, \delta_V)$, allow the population to implicitly code the full disparity information. The population is sensitive to disparities bounded in a circular sub-region of the 2D disparity domain centered around zero. The region radius is directly related to the size of the simple cell's receptive field (RF), modeled by a complex-valued Gabor filter:

$$h(\mathbf{x};\theta,\psi) = \frac{1}{2\pi\sigma^2} e^{\left(-\frac{1}{2\sigma^2}\mathbf{x}_\theta^T\mathbf{x}_\theta\right)} e^{\jmath(k_0 x_\theta + \psi)} \tag{1}$$

where $\mathbf{x}$ are the spatial coordinates on the image plane, ${\mathbf{x}_\theta}^T = [x_\theta, y_\theta]$ are the coordinates rotated by an angle θ about its center, $k_0 = 2\pi f_0$ with f_0 radial peak frequency of the filter, σ is the spatial extent of the Gaussian envelop, and ψ is the phase value. In particular, $\theta = 0^\circ$ corresponds to a vertically oriented RF. The responses of the left and right RFs centered in $\mathbf{x}$ in the images $I_{L/R}$ is:

$$r_{L/R}(\mathbf{x};\theta,\psi_{L/R}) = I_{L/R} * h_{L/R}(\mathbf{x};\theta,\psi_{L/R}) \tag{2}$$

where $*$ denotes the spatial convolution operator; the response of a modeled binocular complex cell is:

$$r_c(\mathbf{x};\theta,\Delta\psi) = |r_L(\mathbf{x};\theta,\psi_L) + r_R(\mathbf{x};\theta,\psi_R)|^2 . \tag{3}$$

Scaling up to the 2D case the 1D approach proposed in [9], it is possible to analytically derive the complex cell's response by using as a synthetic ideal stimulus, a white Gaussian noise characterized by a constant Fourier power spectrum. Expanding Eq. (3), the resulting tuning curve of a complex energy cell approximates to:

$$r_c(\mathbf{x};\mathbf{k}_\theta,\Delta\psi) \approx \frac{4\pi^2|\tilde{I}|^2}{\sigma^4}\left[1 + \exp\left(-\frac{|\boldsymbol{\delta}(\mathbf{x})|^2}{\sigma^2}\right) + 2\exp\left(-\frac{|\boldsymbol{\delta}(\mathbf{x})|^2}{2\sigma^2}\right)\cos\left(\mathbf{k}_\theta^T\boldsymbol{\delta}(\mathbf{x}) - \Delta\psi\right)\right] \tag{4}$$

where $|\tilde{I}|^2$ is the constant power spectrum of the input noise images, assuming for the sake of simplicity that locally $\tilde{I}_L \approx \tilde{I}_R = \tilde{I}$, and $\mathbf{k}_\theta = [k_0\cos\theta, k_0\sin\theta]$ is the frequency vector. From Eq. (4), considering $\mathbf{k}_\theta^T\boldsymbol{\delta}(\mathbf{x}) = k_0\delta^\theta$, where δ^θ is the projection of the full disparity along the direction of θ, we can observe that the complex cell is tuned to a specific stimulus disparity $\delta^\theta = \lfloor\Delta\psi\rfloor_{2\pi}/k_0$, depending only on the phase difference between the left and right RFs, $\Delta\psi = \psi_L - \psi_R$ (see Fig. 1). Since the phase is constrained to its principal value in the interval $(-\pi, \pi]$, the maximum disparity to which a cell can be selective is $\pm\Delta = \delta^\theta_{pref}\big|_{\Delta\psi=\pm\pi} = \pm\pi/k_0$. In order to obtain an efficient and complete representation of the visual signal with a reduced number of channels, we implemented a population of complex cells, tuned to different oriented disparities, following the specification of [37]. The Gabor receptive fields were designed in order to obtain an optimal coverage of the spatial frequency space, without loss of information [38]. In particular, the extent of the Gaussian envelope σ and the radial peak frequency $\mathbf{k}_\theta$ were defined in order to have filters with both a cut-off frequency corresponding to half of the amplitude spectrum $\Delta\mathbf{k}_\theta = 2\sqrt{2\log 2}/\sigma$, and constant relative bandwidth

$\sigma = \frac{1}{\mathbf{k}_\theta}\frac{2^\beta+1}{2^\beta-1}\sqrt{2\log 2}$, with the constant relative bandwidth β set at one octave. Under these specifications, the coverage of the angular frequency is obtained by $\Delta\theta = 2\arctan\frac{2\sqrt{2\log 2}}{\mathbf{k}_\theta\sigma}$, leading to a minimum number of oriented channels of $N_o = 2\pi/\Delta\theta$. The resulting population is composed by of $N_p \times N_o$ disparity detectors, where N_p are the phase shifts $\Delta\psi$ along and N_o orientations, equally spaced between $-\pi$ and π and between 0 and π, respectively.

Although the complex cell tuning curve presents a peak of response at the preferred disparity, from Eq. (4) we note that it is proportional to the energy of the image $|\tilde{I}|^2$, too. Using a divisive normalization stage (cf. [39,40]), we can rescale the activity of any single cell by the energy E_{bin} of the entire population, pooled over all the phases and the orientations:

$$E_{bin}(\mathbf{x}) = \frac{1}{\pi}\int_0^\pi \frac{1}{2\pi}\int_{-\pi}^{\pi} r_c(\mathbf{x};\mathbf{k}_\theta,\Delta\psi)d\Delta\psi d\theta = \frac{4\pi^2}{\sigma^4}\left(1+\exp\left(-\frac{|\boldsymbol{\delta}(\mathbf{x})|^2}{\sigma^2}\right)\right)|\tilde{I}|^2. \quad (5)$$

This normalization stage allows us to obtain the normalized complex cell response:

$$\hat{r}_c(\mathbf{x}) = \frac{r_c(\mathbf{x})}{E_{bin}} = 1+\cos(\mathbf{k}_\theta^T\boldsymbol{\delta}(\mathbf{x}-\Delta\psi))\,\mathrm{sech}(\boldsymbol{\delta}^2/\sigma^2) \quad (6)$$

independent of the stimulus luminance, but preserving the tuning on the stimulus disparity.

3. The learning algorithm

Grounding on a population of complex cells tuned to N_p phases along N_o orientations, as introduced in Section 2, the goal of the learning algorithm is to derive a proper disparity–vergence control signal v_S, by a weighted linear summation of the population response in a foveal neighborhood Ω, so that:

$$v_S = \sum_{\mathbf{x}\in\Omega}\sum_{i=1}^{N_p\times N_o} G(\mathbf{x})w_i\hat{r}_c^i(\mathbf{x}) \equiv \sum_{i=1}^{N_p\times N_o} w_i\sum_{\mathbf{x}\in\Omega}G(\mathbf{x})\hat{r}_c^i(\mathbf{x}) = \sum_{i=1}^{N_p\times N_o} w_i\hat{R}_c^i \quad (7)$$

where $\hat{R}_c^i$ is the summation of the response of the i_{th} cell $\hat{r}_c^i$ over the Ω, weighted by Gaussian profile $G(\mathbf{x})$ centered in the fovea. In [29] we demonstrated that the network, for its implicit characteristics, can be exploited to drive the progress of learning and to move the system towards an effective behavior. Indeed, by analyzing the normalized complex cell response Eq. (6), we can see that it oscillates about an unitary value, and is composed of a sinusoid and a term in sech that defines the region on the disparity plane where the cell is responsive. The latter term, presenting a maximum for null disparity and decreasing as a function of the disparity modulus, can be informative about the distance of the system from a correct vergence position. Nevertheless, its computation requires the reliable knowledge of the ground truth stimulus disparity, and, since our aim is to endow an active visual system with autonomous learning capabilities, we have to consider that the these data are not available and thus we cannot rely upon them.

It is worth noticing from Eq. (6), that we can derive the standard deviation (STD) of the population response as a function of disparity for a white noise stimulus:

$$\mathrm{STD}(\boldsymbol{\delta}) = \sqrt{\frac{1}{\pi}\int_0^\pi\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\hat{R}_c(\mathbf{k}_\theta,\Delta\psi)\right)^2 d\Delta\psi d\theta - 1} = \sqrt{\frac{1}{2}\,\mathrm{sech}\left(-|\boldsymbol{\delta}|^2/\sigma^2\right)} \quad (8)$$

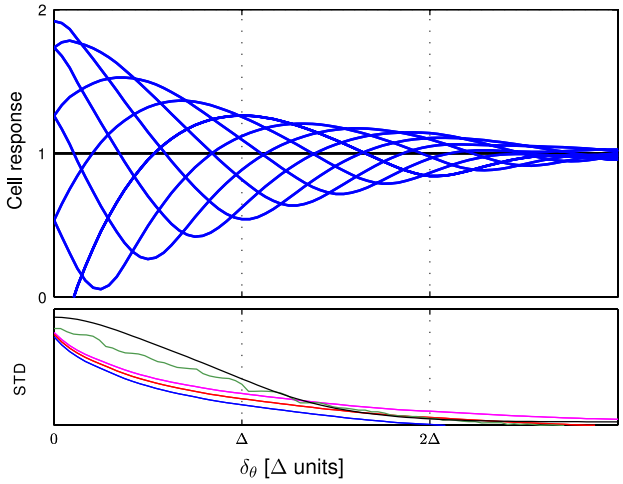


Fig. 2. (Top) Response of a single orientation channel of the population (*i.e.* for a fixed θ and with phase shifts $\Delta\psi \in [-\pi,\pi]$) to a disparity stimulus δ_θ that varies in the range $[0, 3\Delta]$, along that orientation. (Bottom) The associated STD of the whole population response. Each color plot corresponds to a different image set: RDS (blue), pink noise (magenta), a natural image (red), stereo pairs acquired by a robot head (green), and the theoretical STD as in Eq. (8) (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that, as desired, reflects the variability of the population activity (see Fig. 2, black line), but can be computed directly by the *internal* coding of the visual information, *i.e.* the population response, without requiring any external knowledge from the environment. Accordingly, this signal can be used as an effective reward for the system, since it provides a proper measure of the distance from the goal of the vergence movements. Indeed, while the system moves towards zero disparity, the population activity moves towards its maximum activity, both for indicating how much the system is close to the desired vergence (*i.e.* zero disparity) and evaluating how much effectively and rapidly it guides the action towards the fixation. In order to verify and validate the stability of this measure, we computed the response profiles, and thus the STD, not only for a white noise stimulus (see Fig. 2, blue line), but also to more natural and real stimuli, *i.e.* to pink noise (magenta line), to a natural image (blue line), and to real stereo pairs acquired by a robot head (green line). The slightly different shapes of these signals with respect to the theoretical STD (black line) computed accordingly to Eq. (8) is given by the fact that they are computed on real tuning curves, and that they results from a summation instead of an integral over the parameter space.

Differently from [29] where we rely on a *particle swarm optimization* algorithm for the development of the weights, which is far from being a plausible cortical mechanism, here we adopt a more biologically plausible Hebbian rule to learn the vergence control. Specifically, since the effect of the performed vergence movements can be evaluated at an instant of time subsequent to the one that triggered it, we introduced a temporal asymmetry in the algorithm, obtaining a differential Hebbian rule [41]. Analyzing the temporal trend of the population, it turns out that the vergence control directly correlates with the variation of the activity of the complex cells, *i.e.* the v_S computed at $t-1$, correlates with the variation of the activity of the complex cells between the time $t-1$ and t. This observation leads to the following rule for the update of the weights $\boldsymbol{w}$:

$$w_i|_t = (1-\eta)w_i|_{t-1} + \eta v_S(\mathbf{r}_c|_{t-1})\Delta r_c^i \quad (9)$$

where $\Delta r_c^i = r_c^i|_t - r_c^i|_{t-1}$. Accordingly, the single synaptic weight w_i at the instant t is modified by its value at the previous instant $w_i|_{t-1}$ plus the update term. Such a term is composed of the activity

of the postsynaptic neuron (*i.e.* the vergence control) computed at the previous instant $v_S(\mathbf{r}_c|_{t-1})$, multiplied by the variation of the activity of the presynaptic neuron (complex cell) Δr_c^i elicited by the vergence control. The parameter η models the learning rate. Since the variation of the cell's activity can be positive or negative, as well as the vergence control, such a rule positively reinforces the synapses of presynaptic neurons whose variation correlated to output of the postsynaptic neuron, and negatively those with anti-correlated variations.

In order to avoid the weights to diverge to infinity, Eq. (7) has been modified adopting, at each time step, a normalization, as in Oja's rule [42], $\hat{w}_i = w_i/\bar{w}$, where:

$$\bar{w}|_t = \sum_{j=1}^{N_p \times N_o} (w_j^p|_{t-1})^{1/p} \tag{10}$$

with the exponent p set to 2.

At the beginning of the learning phase, the weights are randomly initialized with uniform distribution between -1 and 1 with zero mean. Consequently, the initial vergence control is far from yielding the correct movement. Nevertheless, the learning rule enhances the influence of the cells with correlated and anticorrelated variation, no matter what is the result of the action taken. Instead of a fixed learning rate, our purpose is to include a dynamic rate that is derived from the activity of the population itself, the variation of STD, *i.e.* $\eta = \eta|_t = \Delta\text{STD}$, that modulates the learning algorithm [43], leading to:

$$\hat{w}_i|_t = \left[(1-\eta)w_i|_{t-1} + \eta v_S(\mathbf{r}_c|_{t-1})\Delta r_c^i\right]/\bar{w}. \tag{11}$$

If in two consecutive time steps the STD grows, it implies that the vergence movement has been correct, and the learning rule increases/decreases the weights corresponding to the cells with correlated/anticorrelated variation. Otherwise, when the vergence movement is incorrect, the sign of the learning rate is opposite, as well as the update of the weights.

Considering the reward term directly in the learning rule, provides an update of the weights that depends on the resulting action in two ways. First qualitatively, because the sign of the variation of the reward ΔSTD inverts the effect on the synapses. Second, quantitatively because the weights are strengthened proportionally to the effectiveness of the action, *i.e.* on the modulus of ΔSTD. The resulting process rapidly converges to a vergence control able to move the system towards null disparity, and to provide stable and correct fixations.

An interesting feature of the computational substrate that must be taken into account is that the standard deviation is effective in driving the learning because of the characteristic of the tuning curves, that yield a preference for zero disparity in the overall population response. Indeed, if the network were composed of cells with ideal bell-shaped response (*e.g.* Gaussian) tuning curves, the standard deviation of the population activity would be flat in the region of interest (see Fig. 3, magenta line), providing no information about the state of the system, and thus being useless for the learning process.

4. Results

The learning algorithm is tested in three different conditions. The first experiment is conducted with synthetic stimuli characterized by constant disparity, and aims to assess: (1) the informative content of each orientation channel separately, (2) the capability to converge to a proper solution, and (3) the convergence velocity with different learning signals. In the second experiment we test the robustness of the algorithm by using a reduced subset of cells, randomly chosen from the original population. The third

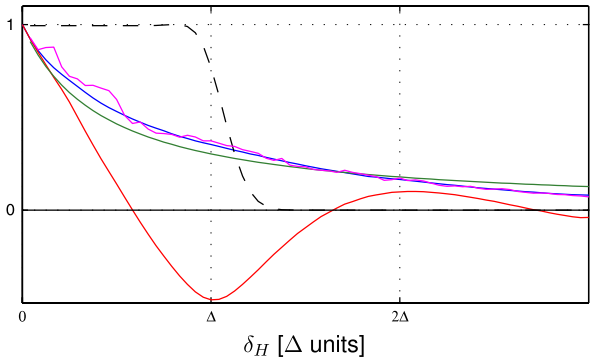


Fig. 3. The learning signals used to train the network: STD is the standard deviation of the population response (blue line), HW is the sum of the half-wave rectified activity of the same population (magenta line), PH0 is the response of the cell defined by $\Delta\psi = 0$ and $\theta = 0$ (red line), and δ_{GT} is a function of the ground truth disparity (green line). Those signals are compared to the standard deviation of a population of cells defined by Gaussian tuning curves (black dashed line). For the sake of representation, the signals are normalized in order to have all a unitary maximum. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

experiment analyzes the behavior of the vergence control in a realistic situation, *i.e.* while the eyes are looking at a plane in a simulated environment resembling the optical characteristics of the iCub stereo head. Finally, the effectiveness of the learning algorithm is validated on real stereo image pairs acquired directly by the iCub stereo head, and the learned control is tested for real-time behavior on the robot.

4.1. Learning on synthetic stimuli

The algorithm is first tested with synthetic stimuli randomly selected among those used to derive the population tuning curves in Section 2 (RDS, pink noise images and natural images). Such a configuration can be considered a facilitated one, because the disparity is directly controllable and constant over the whole image, and it is used to demonstrate the learning capabilities of the network. From an implementation point of view, the filters have been specified according to Section 2, in order to provide the system with a sufficient working range in peripersonal space, for the test in the simulated environment and with the iCub stereo head. Indeed, the range of disparities encoded by the population is defined by the radial peak frequency, set to $f_0 = 1/16$ cycles per pixel, resulting in $\Delta = 8$ pixels, performed by $N_p = 9$ phase shifts. The standard deviation of the Gaussian envelope is computed accordingly ($\sigma \approx 9$ pixels), and the angular coverage the disparity is granted by using $N_o = 8$ orientations. The filter size is set to 43×43 pixels, in order to cover the full spatial extent of the Gaussian envelope ($\approx 5\sigma$). The binocular stimulus is created from two images, of size 256×256 pixels, where the left is an identical copy of the right one, shifted by an equal amount of binocular disparity. For not-integer values of disparity, the right image is obtained by a bilinear interpolation of the left one. Each trial starts from a randomly initialized disparity value in a continuum within the range $[-\Delta, \Delta]$ for the horizontal component. The system evolves for 8 time steps, where the disparity content of the stimulus is modified at each t according to the vergence control computed by Eq. (7). In order to qualitatively take into account the vertical disparity that arises from the vergent geometry, the range for the vertical component is set to 1/3 of the horizontal one, following the disparity statistics that occurs in natural images [44].

Fig. 4 shows the evolution of a set of weights (column A) using the STD as the reward signal. The effectiveness of the weights along their development can be assessed from the tuning curves of the resources. Indeed, the resulting vergence control can be

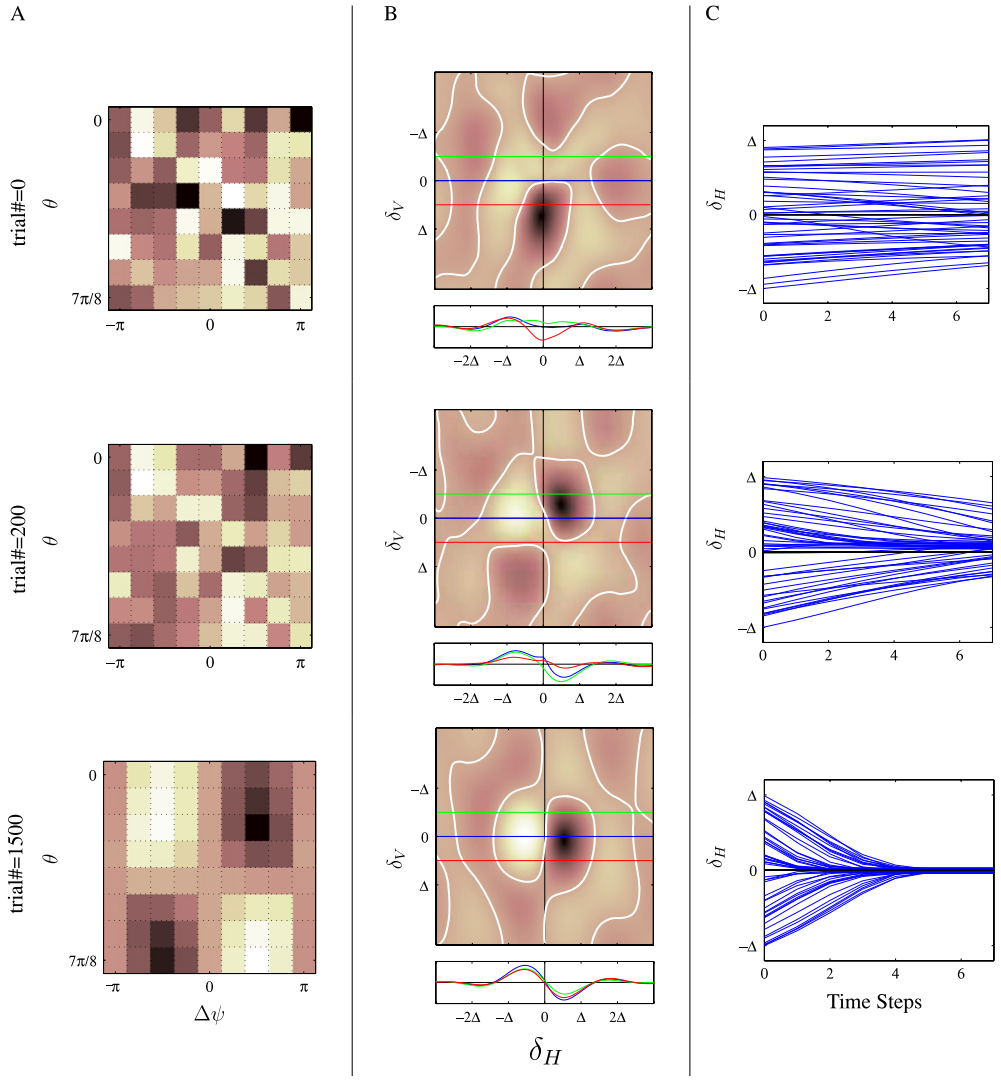


Fig. 4. Evolution of the learning process for a single set of weights at trial 0 (top panel), 200 (middle panel) and 1500 (bottom panel), using STD as reward signal. A. Representation of the learned set of weights with respect to the tuning characteristics of the cells' RFs, *i.e.* the orientation θ and the binocular phase difference $\Delta\psi$. Bright and dark colors correspond to positive and negative weights, respectively. B. The resulting vergence control (*i.e.* disparity–vergence response) over the 2D disparity domain. Bright and dark colors correspond to positive (convergence) and negative (divergence) controls, while white lines represent the zero crossing of the profile. The bottom inset represents the corresponding horizontal cross sections of the control for a fixed vertical disparity pedestal equal to $-\Delta/2$ (green), 0 (blue) and $\Delta/2$ (red). C. Temporal evolution of the vergence trajectories. Note that each panel shows a distinct stage of the learning process: trials 0–50 (top panel), 150–200 (middle panel), and 1450–1500 (bottom panel). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

derived by a weighted summation of the tuning curves (column B). Since at the beginning the weights are initialized randomly, the vergence control derived has not a specific shape, both in term of the 2D profile and for what concerns its horizontal cross section (column B, top). As a consequence, the resulting control, instead of moving the system towards zero disparity, produces wrong vergence trajectories (column C, top). After a limited number of trials (*e.g.* 200 for the figure), the profile of the disparity–vergence control evolves to more a symmetric shapes that approximates linear servos, and produces vergence trajectories that drive the system towards zero disparity (column B–C, middle). At the end of the learning process (*i.e.* at trial 1550, for the figure), the weights converge to a stable configuration, and the system is able to complete effectively and precisely the vergence movement within the given number of time steps (bottom row).

A qualitative analysis of the learned set of weights and their capability, pointed out that the resulting profiles are characterized by two salient features typical of an effective vergence control: they develop an odd symmetric control with a zero crossing for zero disparity (see Fig. 4(B), bottom). At a first glance, the learned set of weights have a characteristic structure that emerges after a sufficient learning time. Indeed, so as the control, the weights exhibit an odd symmetry with respect to zero phase shift and vertical orientation (see Fig. 4(A), bottom). Cells characterized by the same orientation provide opposite contribution, depending in the sign of the zero phase shift, so as cells with the same phase shift provide opposite contribution with respect to the sign of the deviation from the vertical orientation. As expected, cells with an odd tuning curve ($\Delta\psi \neq 0, \pm\pi$) contribute and compete in realizing the control, whereas cells with even tuning curve ($\Delta\psi = 0, \pm\pi$) provide null contribution, as well as cells tuned to pure vertical disparity. Moreover, the robustness of the learned control to vertical disparity is assessed by the fact that a meaningful vertical disparity pedestal (*e.g.* $\pm\Delta/4$) does not affect the effectiveness of the control, which maintains its odd symmetry and the zero crossing (see Fig. 4(B), blue, green and red cross section plots).

4.2. Role of the orientation channels

While a multi-scale approach allows implicitly for a larger working range of the control, the proposed algorithm relies on a

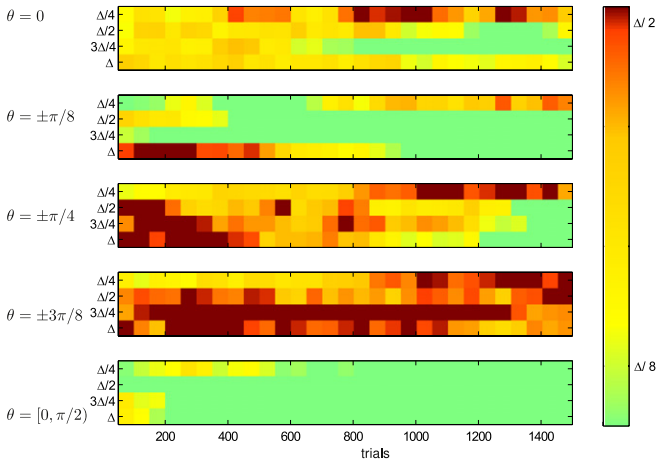


Fig. 5. Residual absolute error (in Δ units) of the vergence control during the training phase, over 1500 trials, where the different orientations are used separately. The color represents the error according to the colorbar on the right, while on each plot the x-axis represents the different consecutive trials, and the y-axis indicates the four different ranges of disparities used in the learning process, from $[-\Delta/4, \Delta/4]$, to $[-\Delta, \Delta]$. From top to bottom: the error for $\theta = 0, \theta = \pm\Delta/4, \theta = \pm\Delta/2, \theta = \pm 3\Delta/4$, and, for comparison, the error obtained when considering the whole orientation channels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

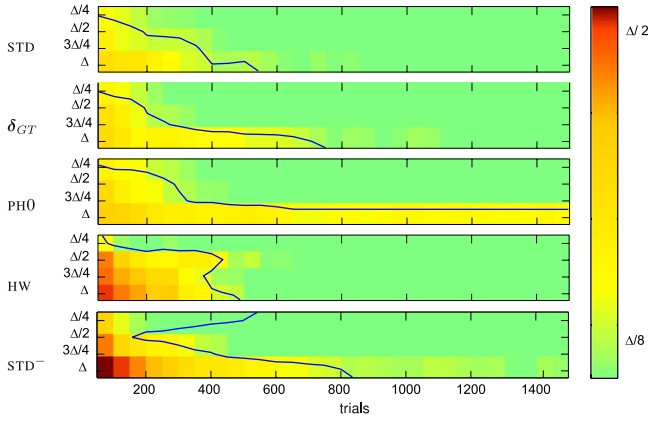


Fig. 6. Residual absolute error (in Δ units) of the vergence control during the training phase (as in Fig. 5), where different learning signals are used. From top to bottom: the STD; a function of the ground truth disparity δ_{GT}; the HW summed activity of the half-wave rectified complex cells' response; the response PH0 of the cell defined by $\Delta\psi = 0$ and $\theta = 0$; and the STD$^-$ of the same population as STD, but when a reduced set of resources (50%) is considered. The blue contour line represents a threshold of $\Delta/8$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

single scale network in order to keep the computational cost minimal. It is the use of resources tuned to different orientations that allows us to obtain a broader working range by the non-vertical orientated channels [22,28]. Indeed, the more the 2D Gabor RF deviates from the vertical, the most its tuning curve for horizontal disparity (*i.e.* $\delta_V \approx 0$) can be described by a 1D Gabor profile with a lower frequency with respect to the nominal peak frequency f_0 of the RF (see Fig. 1, blue cross-section plot). Considering that in vergent viewing conditions the vertical disparity is not zero, the tuning curve of the non vertically-oriented complex cells is considerably modified (see Fig. 1, red and green cross section plot), and a vergence control that relies on such resources have to cope with this effect to maintain its effectiveness.

In order to assess the role of the different orientation channels with respect to the stimulus disparity, we have extensively and separately tested their contribution for different disparity ranges of the input stimulus: $[-\Delta/4, \Delta/4]$, $[-\Delta/2, \Delta/2]$, $[-3\Delta/4, 3\Delta/4]$ and $[-\Delta, \Delta]$. The average behavior of the learning algorithm is evaluated repeating the procedure presented in Section 4.1 50 times for each of the disparity range, and each set of weighs, randomly initialized, is let evolve over 1500 different trials. The performance of the resulting control with respect to the achieved vergence is evaluated by the residual vergence error at the end of each trial, *i.e.* the residual horizontal disparity at fixation. Fig. 5 shows, for each orientation channel as well for the whole population, the evolution of the residual vergence error along the trials (x-axis), averaged over the 50 different sets of weights, for the four different disparity ranges considered (y-axis).

Fig. 5 shows a trade-off in the learning capability of the control between the extent of the working range and the orientation of the resources. The more the orientation deviates from the vertical, the most its working range can increase, but its sensitivity to vertical disparity increases together, weakening the effectiveness of the learning capability. Indeed, the vertical oriented units, *i.e.* the ones tuned to horizontal disparities only, even if not sensitive to vertical disparity, are characterized by tuning curves to horizontal disparity with a high frequency (see Fig. 1), and consequently with a small working range. Accordingly, such units fall short in learning a proper control when the disparity range approaches the theoretical limit Δ (*i.e.* $\delta > 3\Delta/4$). Conversely, the cells oriented to $\theta = 3\pi/8$, which are the ones that are able to provide the widest working range, are completely unable to drive the learning of a proper control, due to their strong sensitivity to the vertical disparity. The two intermediate orientations show a good learning capability for disparities $> \Delta/4$, where the sensitivity to the vertical disparity simply results in a longer time for convergence. Finally, the results show how each single orientation channel may not learn a stable vergence behavior for small disparities ($< \Delta/4$).

Therefore, so as the signals from different orientations facilitate the disparity decoding [45], they play a key role on vergence control and its learning process. The comparison between the performance of the whole population and those achieved by the different orientations separately, confirms the advantage of the population coding. In agreement to [28], including other orientations than the vertical one, decreases the time required to learn an effective vergence behavior and increases the working range of control. The main difference is that in [28] the added orientations have no effect on the resulting precision, whereas with our approach the use of the whole population makes it possible a higher precision, particularly for small disparities (see Fig. 5, bottom). Such an effect is likely to arise from the different training stimuli used. Indeed, [28] employ images containing horizontal disparity, only, whereas the synthetic stimuli we used is characterized by a vertical disparity pedestal. It is the interplay among the different orientation channels that allows us to gain an insensitivity to the vertical disparity, which is a mandatory feature to obtain an effective and precise horizontal vergence control, able to cope with real stereo image pairs acquired in a vergent viewing geometry.

4.3. Learning through different rewards

In order to have a direct comparison with the other methods proposed in literature, the algorithm is tested using four different reward signals (see Fig. 3), *i.e.* the ground truth disparity (δ_{GT}), the response of a zero disparity complex cell with $\theta = 0$ and $\Delta\psi = 0$ (PH0) [25], the sum of half-wave rectified population activity (HW) [27], and finally the proposed standard deviation of the population response (STD) [29].

As in Figs. 5, 6 shows for the whole population, the performance of the different rewards for the considered disparity ranges. At a first glance, it turns out that the larger is the range of disparities the system has to cope with, the longer it takes to converge to an effective control. The proposed algorithm, exploiting the STD, is able to

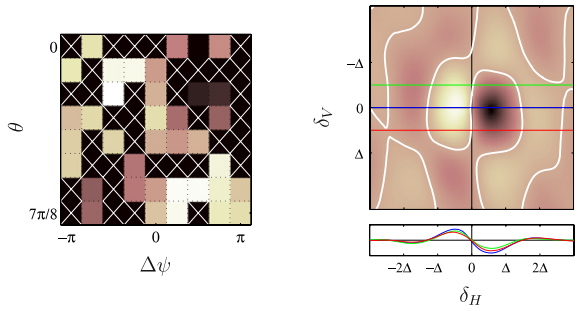


Fig. 7. An example of a learned set of weights (left) and the resulting two-dimensional profile of the vergence–disparity response (right), when considering a population subset of 50% of the resources, presented with the same notation of Fig. 4(A)–(B). The white crosses represent the cells that are removed from the population. The bottom inset represents the corresponding horizontal cross sections of the control for a fixed vertical disparity pedestal equal to $\Delta/2$ (green), 0 (blue) and $\Delta/2$ (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

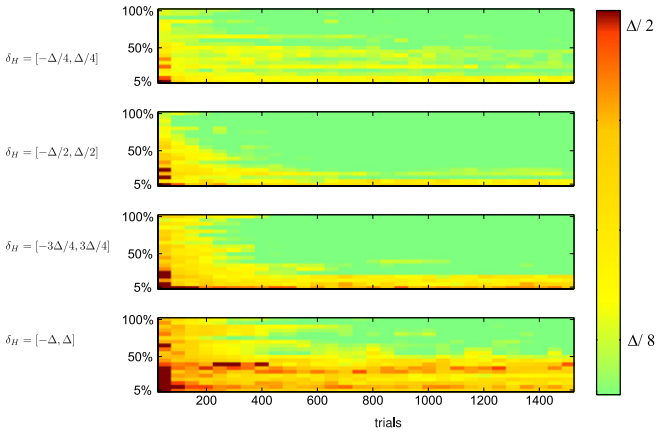


Fig. 8. Residual absolute error (in Δ units) of the vergence control during the training phase (as in Fig. 5), for the four different ranges of disparities, but using different random percentages of the resources (from 100% to 5% of the whole population of complex cells, at steps of 5%).

develop an effective vergence control (residual error smaller than $\Delta/8$), within the first 20 trials, 250 trials, 500 trial and 550 trials, for the four ranges of disparity tested, respectively. As expected, whereas the STD, the HW and the δ_{GT} are able to develop an effective control for all the tested ranges, the PH0 signal takes a slightly longer time and it is not able to make the system to converge for a large disparity range ($\pm\Delta$). This happens because the PH0 signal is not monotone and it changes sign at about $\delta = \Delta/2$ (see Fig. 3), providing a correct reward for small disparity ranges, only. The STD provides a performance comparable to HW, but with a lower error at the beginning of the learning phase and a slightly shorter time to convergence, particularly for small disparities. Indeed, we have to consider that, being the population tuned to different phase shifts (*i.e.* disparities) the cells that respond below the average are informative just like those that respond above average. As a consequence, even if both STD and HW carries similar information content, the first signal, not using the half-wave rectification, always exploits the whole population response, thus yielding a more robust signal (see Fig. 3) that provides a slightly improved performance.

Finally, comparing the STD and δ_{GT}, the velocity of convergence can be considered equal for both signals, thus validating the proposed unsupervised approach. The system, by exploiting an internal parameter, the STD, is able to learn a correct and effective vergence behavior, with similar capabilities as if it were exploiting a supervised ground truth signal.

4.4. Learning with a limited number of resources

A proper distribution of the resources over the feature space may result in a oversimplification of the algorithm. Indeed, since the system is pushed to develop a symmetric control through a symmetric exploitation of the resources, it is unlikely to suppose that in a real neural architecture the resources, although covering the whole feature space, are so regularly distributed over it (see for example [32]). In order to further validate the proposed learning algorithm, we repeated the tests removing a subset of cells from the population (see Fig. 7, right). The removed cells are randomly chosen among the whole set, so to prevent the system to rely on symmetric resources. Nevertheless, the developed control provides the two required salient features (see Fig. 7, left), and shows a capability that is qualitatively equivalent to that obtained by the whole population.

Analyzing the convergence velocity of the system when exploiting only the 50% of the resources (see Fig. 6, bottom), it is evident how it takes a longer time to develop an effective control, but the final behavior is equal to that obtained when using the whole set.

In order to assess the robustness of the learning algorithm, as long as the effectiveness of the reward signal, the learning phase where repeated for different random percentages of resources used on the four disparity ranges used. Fig. 8 shows how for small disparities ($[-\Delta/4, \Delta/4]$ and $[-\Delta/2, \Delta/2]$), the 15% of the population is sufficient to learn a proper control, whereas for $[-3\Delta/4, 3\Delta/4]$ the percentage raises to 25%, and finally for large disparities ($[-\Delta, \Delta]$) a 50% is needed.

4.5. Test on a frontoparallel plane

In order to assess the effectiveness of the learned control in realistic environment, characterized by vergent geometry and varying azimuth and elevation of the gaze direction, we implemented a virtual environment (Fig. 9(A)) that resembles the geometrical and optical characteristics of the iCub stereo head used in the next section. The interocular distance is $b = 70$ mm, the focal length is $f_c = 6$ mm, and the stimulus is projected onto the retinal plane, with a size of $1/3''$ mm, thus with a field of view of about $\approx 80°$, horizontally, and $\approx 60°$, vertically. The images are rendered with a resolution of 120×160 pixels, thus leading to a square pixel of 0.5°. The binocular head look at a planar surface characterized by a pink noise texture. According to Eq. (7), the vergence control is computed by weighting the cells' response with a Gaussian profile centered in the fovea and with a standard deviation of 1.5°. From a functional point of view, the filter size is selected starting from the assumption that the architecture should be at least able to learn the proper vergence control in range of disparities encoded by the population, *i.e.* Δ. Accordingly, it has been set to $f_0 = 0.0.125°$ cycles/degree, leading to $\Delta = 4°$, in order to provide the system with a sufficient vergence working range in peripersonal space.

The geometry of the modeled binocular vision system is characterized by a common elevation for the left and right cameras, and independent azimuth angles, as in the Helmholtz reference frame [46]. This configuration yields a simplified parametrization of the visual direction in terms of version and vergence (α) angles, where, the vergence control necessary to move the fixation point, while keeping constant the gaze direction, is applied symmetrically on both the eyes [23]: $\Delta\alpha = -\arctan(v_S/2f_c)$.

Unlike the case with parallel optical axes, where the proper calibration allows a cancellation if the vertical disparity component across the whole image, in a realistic system with vergent geometry the arising pattern is more complex [29], and would require a continuous recalibration. Hence a vertical disparity tolerant system is highly desirable. Adopting the Helmholtz geometry, the optical axes are always intersecting, and the vertical disparity is zero

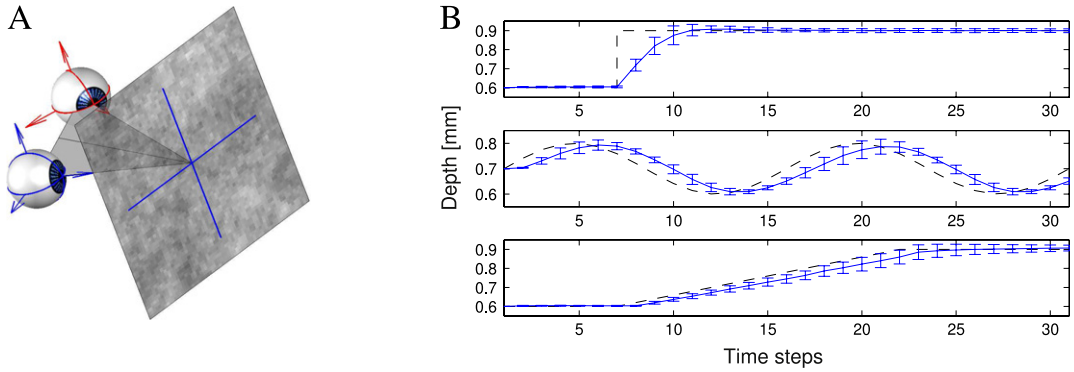


Fig. 9. A. Simulated experimental setup for the test of the learned vergence control on a textured plane. B. Mean and standard deviation of the fixation point trajectory in depth, computed over different gaze directions and plane orientations, according to the learned vergence control (blue solid line), when following a step, a ramp and a sinusoid movement of a textured plane stimulus (dashed black line).

along the horizontal meridian, only. More precisely, the vertical disparity grows with the retinal eccentricity, and it is influenced by three factors: (1) a gaze direction that deviates from the straight-ahead, (2) the non-perpendicularity of the observed surface with respect to the binocular line of sight, (3) large vergence angles. Thus, aiming to learn a vergence control that would be able to cope with a changeable and unknown environment, the vertical disparity is a major concern that is mandatory to take it into account. In order to test the effectiveness of the learning algorithm in a more realistic situation, we moved the gaze direction from -30° to 30° at steps of 5° both for the azimuth and the elevation. Moreover, the orientation of the plane in space has been chosen not perpendicular to the binocular line of sight, but its orientation is changed between -20° and 20° at steps of 2° with respect to both the degrees of freedom of orientation.

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 moving. Working in a visually-closed loop, the control is updated at each time step and, particularly for small disparities, it has a value proportional to the binocular disparity in the fovea. As a consequence, when the fixation point reaches the depth of the stimulus, the vergence control intrinsically turns to zero and the fixation point remains steady at the depth of the plane.

Taking inspiration from psychophysiological experiments [47], the effectiveness of the learned control is evaluated with three different tests, *i.e.* with a frontoparallel plane whose position in depth varies in time as a step, as a ramp and as a sinusoid. Fig. 9(B) shows the mean trajectory of the fixation point and its standard deviation (blue), computed over all the gaze directions and plane orientations considered, with respect to the ground truth depth of the plane (black). These tests show how the system developed the two salient features of the control. In fact, it is capable of producing fast changes of the fixation point for abrupt changes of the stimulus depth (Fig. 9(B), top panel) and smooth movements to follow the depth of the stimulus (Fig. 9(B), center and bottom panel). Likewise, it is able to produce stable fixations on a steady stimulus (Fig. 9(B), top panel) with the fixation point laying correctly at the depth of the stimulus. Moreover, the small standard deviation of the trajectories validates the robustness of the control to the varying vertical disparity present in the different trials.

4.6. Learning on real stereo image pairs

The effectiveness of the learning algorithm is tested on the iCub stereo head [48]. Being designed to resemble the human head, it has the interesting feature of a baseline of 70 mm, *i.e.* similar to the baseline of a human being. This allows the system, working in the peripersonal space, to experience binocular images with disparities close to those that would fall on the human retinas in similar conditions. The original images from the Dragonfly cameras (768×1024), are sub-sampled to 120×160 pixels, in order to allow the system to achieve real-time performances [23]. The filter specifications are the same of that used in the previous experiments, thus leading to $\Delta = 4^\circ$.

The learning algorithm is trained offline, following similar principles to those described in Section 4.1. The training set of images consists of stereo pairs acquired from the iCub head, in a setup similar to that used in the simulated environment. A textured plane is kept fixed at a depth of approximately 600 mm, and perpendicular to the binocular line of sight. The images are acquired with the fixation point encompassing the depth range necessary to obtain stereo images characterized by a horizontal disparity between $[-\Delta, \Delta]$, *i.e.* $\approx$[350, 1500] mm, with steps of $\approx$10 mm.

The flexibility of the population of disparity detectors provides an effective substrate to gather the reward signal for the learning phase also with real robot stereo image pairs (see Fig. 2, green line). The fixation point, starting over a set of 1500 different depths randomly chosen within a predefined range, is updated by means of v_S, applied as in Section 4.5, and should converge towards the actual depth of the stimulus, *i.e.* 600 mm.

The obtained control is tested with a textured plane, moved smoothly by hand following a trajectory that oscillates about a mean depth of $\approx$600 mm with an almost constant frequency of 0.15 Hz, and with increasing amplitude from 0 to $\approx$600 mm (see Fig. 11). The stereo image pairs are acquired at a frame rate of $\approx$14 fps, and the vergence command is applied to the azimuth motors at the same frequency. The depth of the stimulus with respect to the stereo head is measured with a Microsoft *Kinect* sensor device, precisely calibrated for the task at hand [49]. The actual depth of the fixation point (red solid line) with respect to the ground truth of the stimulus (black dotted line), is estimated by the position of the azimuth axes provided by the magnetic encoders of the motors, returned by the robot head. The stimulus motion is initiated after a few seconds, that are given to the system so that the fixation point precisely reaches the actual depth of the plane. Notwithstanding the high motor backlash ($\leq \pm 1^\circ$) that produces both biased evaluation of the depth of the fixation point, and a delay between the stimulus and the fixation point trajectories, the control shows the capability of correctly following a real stimulus in depth, for both small and large movements. The robustness of the computed control is demonstrated by the smooth the motion of the fixation point, which resembles the stimulus trajectory.

Notwithstanding the different sources of error, the algorithm is able to learn an effective control from real stereo image pairs. Indeed, representing the learned control in feature space (see Fig. 10), shows how the behavior is qualitatively similar to those

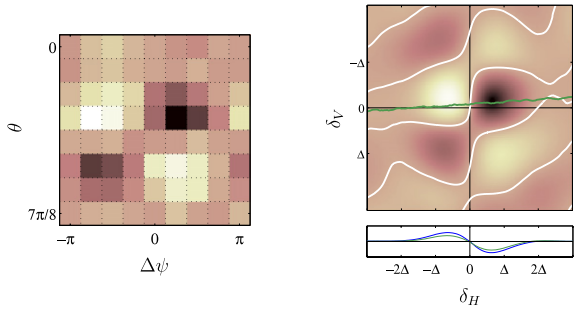


Fig. 10. An example of a learned set of weights (left) and the resulting two-dimensional profile of the vergence–disparity response (right), when considering real robot stereo image pairs, presented with the same notation of Fig. 4(A)–(B). The green line represents the actual full disparity experienced by the robot for different vergence angles.

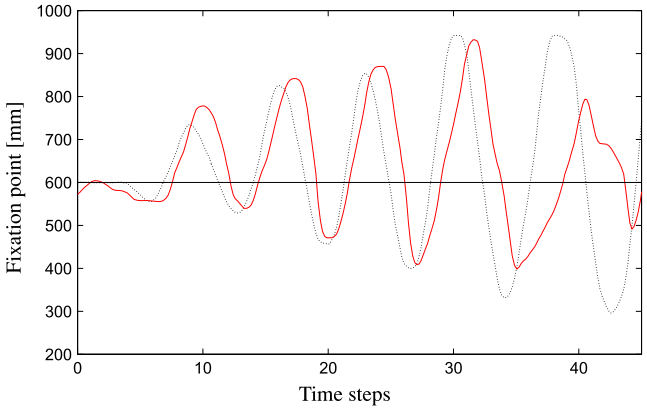


Fig. 11. Vergence trajectories achieved by the real robot system (red solid line), while it is binocularly tracking a frontoparallel plane (black dotted line) moving with constant a frequency and an increasing amplitude.

controls obtained on synthetic data. First of all, our population network, encoding explicitly the vertical disparity, allows us to cope: (1) with real images not compensated for lens distortion, (2) with the not-constant vertical disparity pattern that arises in vergence geometry [50], and (3) with the vertical disparity offset due to the misaligned optical axes. Secondly, the zero crossing solves the error given by the motors' backlash. In fact, working in a closed visual loop, when the disparity in the fovea is reduced to a value close to zero, the control consequently becomes be zero, regardless of the real depth of the object and of the actual position of the motors. Thirdly, it is possible to assess a direct relationship between the shape of the obtained 2D response profile (see Fig. 10, left), and the real disparities experienced by the robot in the foveal area. Fig. 10 shows how the learned controls preserve the mandatory feature of zero crossing for zero disparity, both for the cross section defined by $\delta_V = 0$ (blue line), and, more essentially, for the actual vector disparity characterizing the stereo pairs (green line). Indeed, the learned profile is slightly tilted and has a small vertical offset, according to the characteristics of the real stimulus disparities. Such a feature evidences the capability of the algorithm to learn directly from the environment and to be conditioned by it, in order to drive the system towards a control that is not just able to cope with different disparity patterns, but also to adapt to the contingent situation.

Nevertheless, while at a functional level the large backlash is compensated by the closed visual loop, its drawback on the learning capability is that it sever the temporal relation between the computed vergence command and the action achieved (*i.e.* with the reinforce provided according to the action), seriously jeopardizing the online learning on the robot. The problem can be in principle solved by a temporal integration of the learning signal, and this is currently an ongoing work.

A further advice given by this result is that, a Helmholtz geometry with three degrees of freedom like in the iCub stereo head, has many advantages for simplifying the control and the properties of the disparity pattern, but it has a drawback on the learning capability of the presented algorithm. In fact, since the designed architecture is characterized by oriented binocular energy detectors that span the whole 2D disparity plane, such resources are not fully exploited by a Helmholtz system. More precisely, the Helmholtz kinematics bounds the mean disparity on a single stripe, limiting in this way the learning capabilities of the system to cope with large disparities. It is evident how the working range of the learned vergence control comes to be larger when the vertical disparities are constrained (see Fig. 10), respect to when they varies randomly (see Fig. 4, bottom). Accordingly, while an online implementation of the proposed algorithm on a robot head characterized is straightforward for an Helmholtz geometry, more interesting results are expected to rise using different mechanical geometries, as the Fick-like gimbals implemented on the binocular turret of the K-Team Koala [51] or the non-conventional mechatronic binocular eye systems with two degrees of freedom for each camera (*e.g.*, see [52]). Indeed, the extension of the mechanical system to four degrees of freedom would, from the one hand complicate the algorithm is expected to slow down the learning process, but from the other hand it would allow for a more natural study of the problem of binocular coordination in three-dimensional space, so as of the relation existing among depth perception, vergence and oriented disparity.

In conclusion, by exploiting a precise and complete knowledge either of the resources (tuning curves) or of the environment (ground truth disparity), it is possible to obtain an effective vergence control. Nevertheless, such techniques are grounded on a kind of knowledge that prevents the system to autonomously learn its behavior from data. On the other side, the proposed architecture, designed by mimicking the neural parameters, gains the intrinsic characteristics of providing in internal parameter, related to the whole population activity, which is able to drive the learning of the vergence behavior and to obtain equivalent performances, both on synthetic stimuli and real stereo image pairs.

5. Conclusion

In this paper, we proposed a neural architecture that is able to exploit its internal state to drive the development a proper vergence behavior, thus without any external supervision. The proposed approach, grounds on basic cortical mechanisms both for the representation of the visual signal and for the learning strategy.

From an algorithmic point of view, the distributed coding of the visual signal, allows an effective estimation of the state of the system, and, consequently, of the proper action to reach another state. Relying on the efficacy of the divisive normalization mechanism, the state estimation is robust and stable, and invariant with image features other than the disparity. The obtained population response is modulated by the retinal disparity only, both with white noise stimuli and with more natural ones like pink noise, natural images and, above all, with stereo pairs acquired by a real robot head.

More precisely, the estimation of the relationship between the system and the environment (*i.e.* the system's state) is obtained on a visual basis. Indeed, it is directly the internal representation of the visual signal that drives the learning of the motor control, causing a behavior that naturally emerges from the interaction between the system and the environment. The visual exploration performed by the system, also corresponds to an exploration of the capabilities of the system itself, and to an autonomous tuning exercise for a better interaction with the environment.

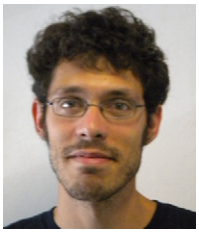
From a methodological point of view, the proposed architecture points out the advantages and the flexibility of distributed cortical-like architectures against solutions based on a conventional systemic coupling of sensing and motor components. Indeed, the linear network, despite its simplicity, does not require any previous knowledge about the resources in use, but rather it is able to exploit the interaction with the environment to *learn* the relevance of each cell of the population, that comes to cooperate for the generation of the motor control. Each orientation channel, indeed, is not sufficient, by itself, to provide the necessary learning capabilities, whereas it is the interplay among them that concurs to develop an effective and precise control in a short time, attesting how a complete representation of the visual signal is instrumental to gather an effective motor control. The cortical-like architecture is thus an effective substrate, able to derive correctly the vergence movements, and to gather the reward signal for the learning phase. The proposed learning rule, extending the standard Hebbian learning, takes into account (1) a temporal asymmetry (and thus the dynamic of the system), and (2) a reward signal that directly affects the learning rate, implementing a direct and mutual influence between the control and the population activity that generated it, *i.e.* with the stimulus that evoked the activity. Indeed, grounding on the flexibility of the distributed approach, the developed vergence behavior is not just able to cope with the mechanical and optical inaccuracies that affect a real vision system, but it is also implicitly suited to the stimulus.

More generally, the proposed architecture, by exploiting an early synergy between the sensing modules and motor control, allows us not just to close the loop between action and perception at system level, but shorten it at an inner one. This allows not just the emergence of spatial competences but provides an instrumental conditioning by which the behavior arises directly from the interaction between the environment and the perceiving system, which could modify its interactions with the world.

References

[1] G. Masson, C. Busettini, F. Miles, Vergence eye movements in response to binocular disparity without depth perception, Nat. 389 (1997) 283–286.
[2] B. Cumming, A. Parker, Responses of primary visual cortical neurons to binocular disparity without depth perception, Nature 389 (1997) 280–283.
[3] J. Law, P. Shaw, M. Lee, A biologically constrained architecture for developmental learning of eye–head gaze control on a humanoid robot, Auton. Robots 35 (1) (2013) 77–92.
[4] R. Sutton, A. Barto, Reinforcement Learning: An Introduction, MIT Press, Cambridge, MA, 1998.
[5] E. Birch, J. Gwiazda, R. Held, Stereoacuity development for crossed and uncrossed disparities in human infants, Vis. Res. 22 (1982) 507–513.
[6] F. Thorn, J. Gwiazda, A.A. Cruz, J.A. Bauer, R. Held, The development of eye alignment, convergence, and sensory binocularity in young infants, Invest. Ophthalmol. Vis. Sci. 35 (1994) 544–553.
[7] W. Yuan, J. Semmlow, P. Muller-Munoz, Model-based analysis of dynamics in vergence adaptation, IEEE Trans. Biomed. Eng. 48 (2001) 1402–1411.
[8] R. Leigh, D. Zee, The Neurology of Eye Movements, Oxford University Press Inc., 2006.
[9] D. Fleet, H. Wagner, D. Heeger, Neural encoding of binocular disparity: energy models, position shifts and phase shifts, Vis. Res. 36 (12) (1996) 1839–1857.
[10] I. Ohzawa, R.D. Freeman, G.C. DeAngelis, Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors, Science 249 (1990) 1037–1041.
[11] L. Wilcox, R. Allison, Coarse-fine dichotomies in human stereopsis, Vis. Res. 49 (2009) 2653–2665.
[12] W. Sun, B. Shi, Joint development of disparity tuning and vergence control, in: 2011 IEEE International Conference on Development and Learning (ICDL), 2011, pp. 1–6.
[13] Y. Zhao, C. Rothkopf, J. Triesch, B. Shi, A unified model of the joint development of disparity selectivity and vergence control, in: IEEE 8th International Conference on Development and Learning, 2012.
[14] L. Lonini, Y. Zhao, P. Chandrashekhariah, B.E. Shi, J. Triesch, Autonomous learning of active multi-scale binocular vision, in: 2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL), IEEE, 2013, pp. 1–6.
[15] Z. Li, J.J. Atick, Efficient stereo coding in the multiscale representation, Network 5 (1994) 157–174.
[16] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (1996) 607–609.
[17] A. Hyvarinen, P. Hoyer, A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images, Vis. Res. 41 (2001) 2413–2423.
[18] M. Solgi, J. Weng, Developmental stereo: emergence of disparity preference in models of the visual cortex, IEEE Trans. Auton. Ment. Dev. 1 (4) (2009) 238–252.
[19] W. Theimer, H. Mallot, Phase-based vergence control and depth reconstruction using active vision, CVGIP, Image Underst. 60 (3) (1994) 343–358.
[20] 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.
[21] A. Gibaldi, M. Chessa, A. Canessa, S. Sabatini, F. Solari, A neural model for binocular vergence control without explicit calculation of disparity, in: Proc. European Symposium on Artificial Neural Networks (ESANN'09), (Bruges, Belgium), April 2009.
[22] A. Gibaldi, M. Chessa, A. Canessa, S. Sabatini, F. Solari, A cortical model for binocular vergence control without explicit calculation of disparity, Neurocomp. 73 (2010) 1065–1073.
[23] A. Gibaldi, A. Canessa, M. Chessa, S. Sabatini, F. Solari, A neuromorphic control module for real-time vergence eye movements on the iCub robot head, in: 11th IEEE-RAS International Conference on Humanoid Robots, 2011, 2011, pp. 1065–1073.
[24] N. Chumerin, A. Gibaldi, S. Sabatini, M. Van Hulle, Learning eye vergence control from a distributed disparity representation, Int. J. Neural Syst. 20 (2010) 267–278.
[25] A. Franz, J. Triesch, Emergence of disparity tuning during the development of vergence eye movements, in: International Conference on Development and Learning 2007, (London, 11–13 July 2007), 2007, pp. 31–36.
[26] Y. Wang, B. Shi, Autonomous development of vergence control driven by disparity energy neuron populations, Neural Comput. 22 (2010) 730–751.
[27] Y. Wang, B. Shi, Improved binocular vergence control via a neural network that maximizes an internally defined reward, IEEE Trans. Auton. Ment. Dev. 3 (3) (2011) 247–256.
[28] C. Qu, B. Shi, The role of orientation diversity in binocular vergence control, in: The 2011 International Joint Conference on Neural Networks, 2011. Proceedings, 2011, pp. 2266–2272.
[29] A. Gibaldi, A. Canessa, M. Chessa, F. Solari, S. Sabatini, How a population-based representation of binocular visual signal can intrinsically mediate autonomous learning of vergence control, Procedia Comput. Sci. 13 (2012) 212–221.
[30] M. Chessa, S.P. Sabatini, F. Solari, A fast joint bioinspired algorithm for optic flow and two-dimensional disparity estimation, in: Proceedings of the ICVS '09, Berlin, Heidelberg, Springer-Verlag, 2009, pp. 184–193.
[31] M. Antonelli, A. Gibaldi, F. Beuth, A.J. Duran, A. Canessa, M. Chessa, F. Hamker, E. Chinellato, S. Sabatini, A hierarchical system for a distributed representation of the peripersonal space of a humanoid robot, IEEE Trans. Auton. Ment. Dev. (2014) 1–15.
[32] S. Prince, B. Cumming, A. Parker, Range and mechanism of encoding of horizontal disparity in macaque v1, J. Neurophysiol. 87 (2002) 209–221.
[33] J. Prince, A. Pointon, B. Cumming, A. Parker, Quantitative analysis of the responses of v1 neurons to horizontal disparity in dynamic random-dot stereograms, J. Neurophysiol. 87 (2002) 191–208.
[34] J. Durand, S. Celebrini, Y. Trotter, Neural bases of stereopsis across visual field of the alert macaque monkey, Cerebral Cortex 17 (2007) 1260–1273.
[35] G. DeAngelis, I. Ohzawa, R. Freeman, Depth is encoded in the visual cortex by a specialized receptive field structure, Nature 352 (6331) (1991) 156–159.
[36] I. Ohzawa, G.C. Deangelis, R.D. Freeman, Encoding of binocular disparity by complex cells in the cat's visual cortex, J. Neurophysiol. 77 (6) (1997) 2879–2909.
[37] J.G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, J. Opt. Soc. Amer. A 2 (7) (1985) 1160–1169.
[38] S.P. Sabatini, G. Gastaldi, F. Solari, K. Pauwels, M.M.V. Hulle, J. Diaz, E. Ros, N. Pugeault, N. Krueger, A compact harmonic code for early vision based on anisotropic frequency channels, Comput. Vis. Image Underst. 114 (6) (2010) 681–699.
[39] D. Fleet, H. Wagner, D. Heeger, Modelling binocular neurons in the primary visual cortex, 1996.
[40] M. Kouh, T. Poggio, A canonical neural circuit for cortical nonlinear operations, Neural Comput. 20 (2008) 1427–1451.
[41] P. Roberts, Computational consequences of temporally asymmetric learning rules: I. differential Hebbian learning, J. Comput. Neurosci 7 (3) (1999) 235–246.
[42] E. Oja, Simplified neuron model as a principal component analyzer, J. Math. Biol. 15 (3) (1982) 267–273.
[43] R. Legenstein, S. Chase, A. Schwartz, W. Maass, A reward-modulated Hebbian learning rule can explain experimentally observed network reorganization in a brain control task, J. Neurosci. 30 (2010) 8400–8410.
[44] Y. Liu, A. Bovik, L. Cormack, Disparity statistics in natural scenes, J. Vis. 8 (11) (2008) 1–14.
[45] S. Patel, H. Bedell, P. Sampat, Pooling signals from vertically and non-vertically orientation-tuned disparity mechanisms in human stereopsis, Vis. Res. 46 (1) (2006) 1–13.
[46] M. Hansard, R. Horaud, Cyclopean geometry of binocular vision, J. Opt. Soc. Amer. 25 (2008) 2357–2369.
[47] G. Hung, J. Semmlow, K. Ciuffreda, A dual-mode dynamic model of the vergence eye movement system, IEEE Trans. Biomed. Eng. 36 (11) (1986) 1021–1028.

[48] R. Beira, M. Lopes, M. Praga, J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi, R. Saltaren, Design of the robot-cub (iCub) head, in: Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006, 2006, pp. 94–100.
[49] A. Canessa, M. Chessa, A. Gibaldi, S. Sabatini, F. Solari, Calibrated depth and color cameras for accurate 3d interaction in a stereoscopic augmented reality environment, J. Vis. Commun. Image Represent. 25 (1) (2014) 227–237.
[50] M. Hansard, R. Horaud, Patterns of binocular disaprity for a fixating observer, in: BVAI, 2007, pp. 308–317.
[51] K-Team-Corp., K-team mobile robotics, 2010. http://www.k-team.com.
[52] G. Cannata, M. Maggiali, Implementation of listing's law for a tendon driven robot eye, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, October 9 - 15, 2006, Beijing, China, vol. 32, 2006, pp. 3940–3945.



Agostino Gibaldi received his degree in Biomedical Engineering from the University of Genoa, Genoa, Italy, in 2007, and his Ph.D. in 2011. Since the master thesis he has been working with the Physical Structure of Perception and Computation Group (PSPC) where he is actually a post doc. His research interests are related to cortical models of V1, MT and MST areas, in relation with the estimation of disparity, the control of vergence eye movements, and the optic flow analysis for navigation, for their real-time implementation on robot platforms so to obtain active behaviors and adaptation to the environment. Aside, he also worked on neural networks and learning, eye tracking algorithms, camera calibration, 3D data modeling and virtual and augmented reality. Nowadays, he is working with the Computer Vision Group of the University of Bologna, on CT perfusion and image registration.

Andrea Canessa received his Bachelor Degree in Biomedical Engineering in 2004 and his Master Degree in Bioengineering (curriculum "Neuroengineering") in 2007, both at the University of Genoa. In 2011, he obtained the Ph.D. in Bioengineering from the same University. From 2011 to 2013, he was with the PSPC Lab of University of Genoa as a research fellow. In 2013, he joined the ITD of the CNR as a research fellow. His research interests include Serious Game for subjects with cognitive disability, RGBD devices and Virtual Reality for rehabilitation, neuromorphic architectural models of visual cortex, eye movements influences on depth perception.

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 visual perception with the aim to develop computational models of cortical vision processing, to devise novel bio-inspired computer vision algorithms, and to design virtual and mixed reality environments for ecological visual stimulations. In particular, his research interests are related to neuromorphic architectures of visual cortex, space-variant visual processing, motion and depth estimation and interpretation, active vision systems, and augmented reality systems for the visuo-motor coordination in the peripersonal space.

Silvio P. Sabatini received the Laurea Degree in Electronics Engineering and the Ph.D. in Computer Science from the University of Genoa in 1992 and 1996. He is currently Associate Professor of Bioengineering at the Department of Informatics, Bioengineering, Robotics and System Engineering of the University of Genoa, Coordinator of the B.Sc. and M.Sc. programs in Biomedical engineering and Bioengineering and Member of the Board of the Doctoral Course in Bioengineering and Robotics. In 1995, he promoted the creation of the "Physical Structure of Perception and Computation" (www.pspc.unige.it) lab to develop models that capture the "physicalist" nature of the information processing occurring in the visual cortex, to understand the signal processing strategies adopted by the brain, and to build novel algorithms and architectures for artificial perception machines. His research interests relate to visual coding and multidimensional signal representation, neuromorphic computing, early-cognitive models for visually-guided behavior, and robot vision. He recently coordinated the EU FP7 project EYESHOTS on the structural mechanisms of visuospatial cognition, responsible for interacting in the 3D peripersonal space, and he was active in promoting and participating, as a partner, to several other EU FP5, FP6 and FP7 projects: ECOVISION, DRIVSCO, MCCOOP, SEARISE. He is the author of more than 100 papers in peer-reviewed journals, book chapters and international conference proceedings.