Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 13 (2012) 212 – 221

Procedia Computer Science

Proceedings of the International Neural Network Society Winter Conference (INNS-WC 2012)

# How a population-based representation of binocular visual signal can intrinsically mediate autonomous learning of vergence control

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

*The Physical Structure of Perception and Computation Group*
*DIBRIS - University of Genova - Via all'Opera Pia 11/A - 16145 Genova, Italy*

## Abstract

Designing an active visual system, able to autonomously learn its behavior, implies to make the learning controller independent of an external signal (e.g. the error between the actual and the desired vergence angle) or of perceptual decisions about disparity (e.g. from the response of a previously trained network). The proposed approach is based on a direct use of a computational substrate of modeled V1 complex cells that provide a distributed representation of binocular disparity information. The design strategies of the cortical-like architecture, including uniform coverage in feature space and divisive normalization mechanisms, allow the global energy of the population to effectively mediate the learning process towards the proper motor control. Since the learning controller is based on an intrinsic representation of the visual signal, it comes to overlap and coincide with the system that is learning the behaviour, thus closing at an inner cycle the perception-action loop necessary for learning. Experimental tests proved that the control architecture is both able to learn an effective vergence behavior, and to exploit it to fixate static and moving visual targets.



*Keywords:* vergence control, autonomous learning, distributed coding, binocular energy model

## 1. Introduction

From a computational point of view, the complex cells of the primary visual cortex (V1) are the processing substrate for the retinal binocular disparity, and enable both stereopsis and vergence [1, 2]. The former mechanism relies on a proper fixational posture to allow for a complex interpretation of the disparity information provided by V1, that is when the disparity of the retinal images is within Panum's fusional area. Such a posture is obtained when both the eyes look the same object in space, and allows for a precise and effective depth perception. The latter mechanism operates to change the vergence angle, that is the angle between the optical axes, so to bring the fixation point on the object of interest and in this way to restore the posture for singleness of vision and depth perception [3].

*Corresponding author
*Email address:* agostino.gibaldi@unige.it (Agostino Gibaldi)
*URL:* www.pspc.unige.it (Agostino Gibaldi)

1877-0509  doi:10.1016/j.procs.2012.09.131

In humans infant, the learning process of vergence eye movement cannot have any explicit teacher, but it relies on the only "supervision" of a feedback on previous errors defined by an expectation of the movement, *i.e.* a reward given when the eyes fixate an object in the proper manner. Nevertheless, the capability of interpretation of depth relies directly on the proper fixation posture, *i.e.* on a correct vergence posture, thus it is not likely to think that the reward comes from it. Indeed, a human infant starts developing a finely tuned disparity system during the period from 2 to 6 months [4, 5], *i.e.* when the convergence and divergence movements have been already developed. Hence, as the vergence system in an adult might follow a fast reactive stream that directly involves V1 cells without resorting to a high level interpretation of depth, it is plausible to think that even its development relies on an early mechanism of reward. Such a process should ensure that an infant will learn the correct vergence behaviour [4] so to influence the succeeding visual experience for a normal development of the stereopsis.

Regarding the computational substrate, our approach is based on a population network of V1-like complex cells that has been shown to be effective in guiding vergence through a weighted sum of the population response that lead to proper servos [6, 7]. In the context of autonomous learning, it is important to assess whether the network itself can learn the proper behaviour for guiding vergence eye movements, what are the design specification that allows the learning, which signal can be used to measure the system's performance and to drive the learning process.

Instead of an external signal, like the difference between the desired and the actual vergence angle [8], an internal controller based on the visual signal is more suitable to drive an autonomous learning. In this framework, in [9] the authors used a stereo matching index, both to evaluate the state of the system, and to reward the system during the training phase, while in [10] they used as the reward the response of a population of complex cells tuned to zero disparity. Notwithstanding the effectiveness of the method, these approaches imply that the learning controller derives from a correct perception of the disparity that, in a parallel with biological system, is not able to develop without even coarse but effective vergence movements. A more recent model [11] adopts as reward signal an intrinsic parameter of the system, *i.e.* the population normalized peak response. The limit of the approach is in the resulting set of discrete controls, obtained by the combination of the activities of three sub-populations, rather than a continuous control yielded by a single "comprehensive" population. Moreover, the position shifts that characterize the three sub-populations are introduced on the basis of an *a priori* knowledge.

Grounding the control on a single distributed representation of binocular disparity, we propose to learn a proper vergence behaviour from the population response without any external supervision, and without taking any decision (*i.e.* estimation) on the stimulus properties. The information used for the learning phase is the modulation of the input disparity, that is, the more the population activity increases, the more the system goes towards the proper fixation posture and its behaviour is rewarded. To allow for a proper intrinsic learning controller, the sensitivity of the complex cells is chosen to optimally tile the feature space [12], and the architecture is endowed with a normalization mechanism [2] so to increase noise resistance and response stability.

## 2. Efficent coding of binocular disparity

We consider a distributed approach in which the retinal binocular disparity $\boldsymbol{\delta}(\mathbf{x})$, in its horizontal $\delta_H$ and vertical $\delta_V$ components, is never measured, but is implicitly coded by the population activity of cells that act as "disparity detectors". On the basis of neurophysiological evidences [2, 13], the V1-like binocular energy complex cells are obtained with two quadrature pairs of simple cells with Gabor-like receptive fields (RF). Defining the complex-valued Gabor filter $h(\mathbf{x};\theta,\psi) = \eta e^{\left(-\frac{1}{2\sigma^2}\mathbf{x}_\theta^T\mathbf{x}_\theta\right)} e^{j(k_0 x_\theta+\psi)}$ where ${\mathbf{x}_\theta}^T = [x_\theta, y_\theta]$ is the coordinate system rotated at an angle $\theta$ about its center, $k_0$ is the radial peak frequency of the filter, $\eta$ is a normalization constant, and $\psi$ is the phase value. The response of the left and right RFs centered in $\mathbf{x}$ to the images $I^{L/R}$, are: $r_{L/R}(\mathbf{x};\theta,\psi^{L/R}) = I^{L/R} * h^{L/R}(\mathbf{x};\theta,\psi^{L/R})$ and the response of a modeled binocular complex cell is:

$$r_c(\mathbf{x};\theta,\Delta\psi) = \left|r_L(\mathbf{x};\theta,\psi^L) + r_R(\mathbf{x};\theta,\psi^R)\right|^2 . \tag{1}$$

Extending the classical formulation showed by [14], and taking into account the preferred orientation $\theta$ of the filter, it is possible to represent the tuning curve of a complex cell:

$$r_c(\mathbf{x};\mathbf{k}_\theta,\Delta\psi) = |r_L(\mathbf{x})|^2 + 2\left|r_L(\mathbf{x})r_R^*(\mathbf{x})\right|\cos(\delta^\theta k_0 - \Delta\psi) + |r_R(\mathbf{x})|^2 \tag{2}$$

where $\delta^\theta$ is the projection of the full disparity along the orthogonal direction with respect to $\theta$. Accordingly, along its orientation the cell is tuned to a specific stimulus disparity $\delta^\theta = \lfloor \Delta\psi \rfloor_{2\pi}/k_0$, defined by the phase shift difference between the left and right RFs, $\Delta\psi = \psi^L - \psi^R$. From Eq. 2 turns out that, unlike simple cells, the response of complex cells depends only on the phase differences $\Delta\psi$ between the left and right filters, and not on the monocular Fourier phases of the input stimuli. Since phase shifts are unique in $(-\pi, \pi]$, the maximum disparity to which the cells could be selective is $\pm\Delta = \delta^\theta_{pref}\big|_{\Delta\psi=\pm\pi} = \pm\pi/k_0$.

Considering the two-dimensional (2D) nature of the binocular disparity, we extended the classical formulation showed by [14] so to represent the tuning curve of a complex cell. In the vector disparity domain $\{\delta_H, \delta_V\}$, the tuning curves present a sinusoidal modulation given by the vector disparity $\boldsymbol{\delta}(\mathbf{x})$ projected along $\theta$, but also a Gaussian modulation given by the magnitude of the retinal disparity $|\boldsymbol{\delta}(\mathbf{x})|$ (see Fig. 1). In order to characterize the complex cell's response, we used a synthetic ideal stimulus, that is a white Gaussian noise defined by constant Fourier power spectrum. With this stimulus it is possible to approximate the analytical expression of the tuning curve:

$$r_c(\mathbf{x}; \theta, \Delta\psi) \approx \frac{16\pi^4 |\tilde{I}|^2}{\sigma^4} \left[ 1 + e^{-\frac{\left|\boldsymbol{\delta}(\mathbf{x})\right|^2}{\sigma^2}} + 2e^{-\frac{\left|\boldsymbol{\delta}(\mathbf{x})\right|^2}{2\sigma^2}} \cos[\delta_H(\mathbf{x}) k_0 \sin\theta + \delta_V(\mathbf{x}) k_0 \cos\theta - \Delta\psi] \right] \tag{3}$$

where $|\tilde{I}|^2$ is the constant power spectrum of the input noise images, assuming for the sake of simplicity that locally $\tilde{I}^L \approx \tilde{I}^R = \tilde{I}$.

Grounding on the phase-shift model, we constructed a population of $N_p \times N_o$ disparity detectors with $N_p = 9$ phases and $N_o = 8$ orientations, equally spaced between $-\pi$ and $\pi$ and between 0 and $\pi$, respectively.

In Eq. 3 the energy of the image $|\tilde{I}|^2$ acts as a multiplicative gain on the cell response. In order to remove such a dependence, it is possible to include a divisive normalization stage [2, 15], in which the activity of the single cell is modulated by the pooled activities of the cell in a surrounding. The resulting cell's selectivity is still attributed to the energy stage obtained by the linear summation and the squaring operations, as in Eq. 1, and its energy invariance is attributed to division, *i.e.* the normalization stage. The response of the normalized complex cell $\hat{r}_c(\mathbf{x})$, is obtained by dividing $r_c(\mathbf{x})$ by the activity of the population ($E_{bin}$), pooled over all the phases and the orientations:

$$E_{bin}(\mathbf{x}) = \frac{1}{\pi} \int_0^{\pi} \frac{1}{2\pi} \int_{-\pi}^{\pi} r_c(\mathbf{x}; \theta, \Delta\psi)\, d\Delta\psi\, d\theta = \frac{1}{\sigma^8} \left( 1 + e^{-\frac{\left|\boldsymbol{\delta}(\mathbf{x})\right|^2}{\sigma^2}} \right) |I|^2 \tag{4}$$

Being $E_{bin}(\mathbf{x})$ proportional to the local Fourier energy of the stimulus $|\tilde{I}|^2$, the normalization rescales the cell responses with respect to the stimulus luminance, preserving the dependence on the stimulus disparity $\boldsymbol{\delta}$.

## 3. Learning the vergence behaviour

### *3.1. Population coding for learning*

The proposed neural architecture, exploited in a proper way, is an effective substrate for driving vergence eye movements both in simulated environments [6] and with real systems [7]. The vergence control can be obtained by proper weighting of the normalized population responses within a perifoveal region ($\Omega$):

$$r_v = \sum_{\mathbf{x}\in\Omega} \sum_{i=1}^{N_p} \sum_{j=1}^{N_o} G(\mathbf{x}) w_{ij} \hat{r}_c^{ij}(\mathbf{x}) \tag{5}$$

where $G(\mathbf{x})$ is a Gaussian profile centered in the fovea, and $w_{ij}$ are the connection weights. To derive the weights with an LS algorithm allows the system to cope with disparity in a wide range of horizontal disparities ($[-3\Delta, 3\Delta]$), and to be insensitive to the vertical component of disparity, in a range of about $[-\Delta, \Delta]$ (see Fig. 3A). Taking inspiration from psychophysical experiments [16], it is possible to implement two different vergence controls: a LONG control that should work in a fast and coarse manner with large disparities, and a SHORT control, that allows for smooth movements with small disparities, and precise and stable fixations. Such a strategy means to

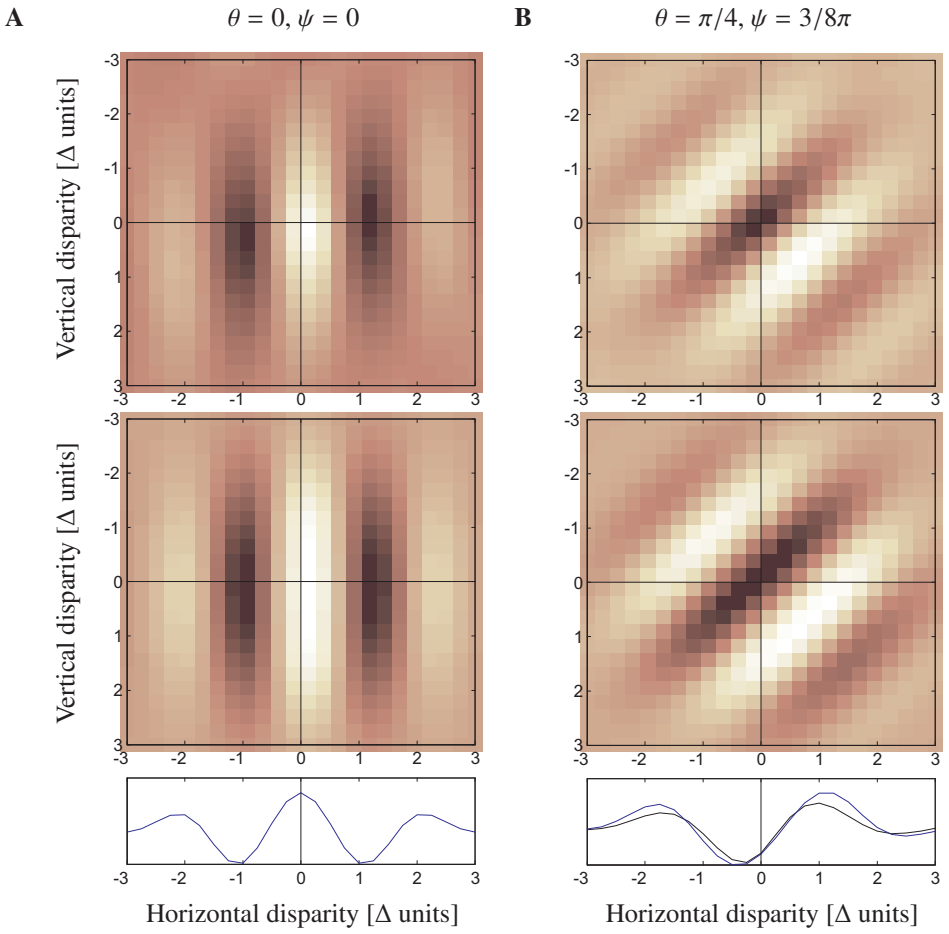


Fig. 1. The actual 2D response profile of the V1 modeled complex cells to a RDS stimuli (top row), and the response obtained by the analytical model (bottom row). The stimulus disparity varies in the range $[-3\Delta,\ 3\Delta]$ for both $\delta_H$ and $\delta_V$, so to gather an extensive characterization of the resources. The two cells represented are defined by: (A) $\theta = 0$ and $\psi = 0$, and (B) $\theta = \pi/4$, $\psi = 3/8\pi$. The bottom row shows the horizontal cross section of the actual responses (black line) and of the analytical models (blue line).

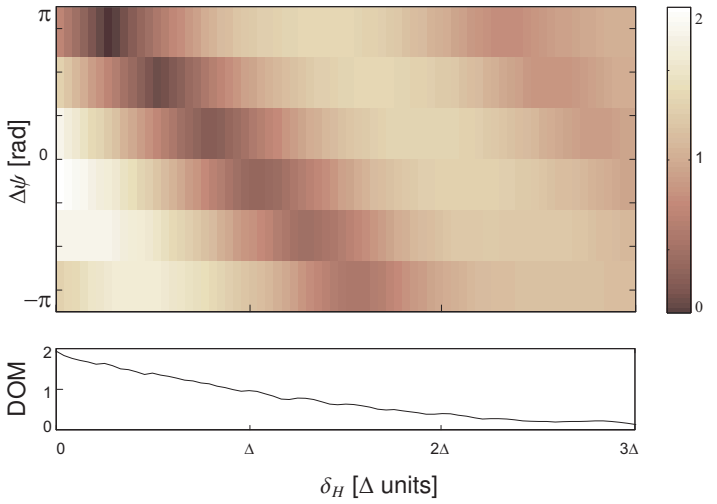


Fig. 2. Response of a single orientation channel of the population with $\Delta\psi \in [-\pi, \pi]$(top) and its depth of modulation (bottom).

*impose* a behaviour of the system in response to $\delta_H$ and a tolerance to $\delta_V$. However, an active vision system should be able to *learn* a proper motor control instead of following a predefined behaviour. From the perspective of reinforcement learning, the choice of the reward that drives the evolution of the system plays a key role. Since the goal to be reached is a viewing posture with zero mean disparity in a foveal area, the ground truth disparity as well as the desired vergence angle [8] provide an ideal evaluation of the distance from the desired position. Besides, aiming to allow the network to learn its behaviour autonomously, it is not possible to use these signals, because they require an extern and perfect knowledge of the environment respect to the system. The learning controller should not be based on an external task, but rather on an intrinsically rewarding mechanism, that acts as a driving force for the learning process. For instance, the response of a complex cell tuned to zero disparity and with vertical orientation has a maximum for zero disparity, and decreases while the modulus of disparity increases (see Fig. 1A). From this perspective, it is well suited for the purpose [10]. Nevertheless, this signal is effective in a limited range of disparity only because for larger values it starts oscillating, thus voiding its capability in driving the learning. Moreover, a single cell is affected by noise and by false matches, thus providing a reward that is not always stable and reliable.

Our aim here is to design and exploit the characteristics of the population network so to obtain a learning signal that (1) monotonically increases as the goal is approaches, and (2) is modulated solely by the the driving stimulus (*i.e.* disparity). A salient feature of the population response is the depth of modulation (DOM), that can be defined as defined as $\mathrm{DOM} = \left(\max r_c^{ij} - \min r_c^{ij}\right)/E\left\{r_c^{ij}\right\}$, that including the normalization stage, becomes $\mathrm{DOM} = \left(\max r_c^{ij} - \min r_c^{ij}\right)$. Concerning to the first point, the uniform distribution of the population selectivity over the feature space (*cf.* [13]) allows the DOM to provide an optimal measure of the distance to null disparity (see Fig. 2), that is to the goal of the vergence movements. Regarding the second point, a direct use of the population response, implies that a stimulus with a high energy would evoke a high population response, and consequently would provide a high gain control that may lead to instability. On the other hand, a low energy stimulus would provide a slow and inefficient control. Similarly, the DOM would be drastically changed by the stimulus energy, affecting the learning rate. Having included an energy normalization stage in the architecture allows the vergence control to be modulated by disparity only, and the learning signal is therefore exclusively informative about it.

By the divisive normalization, the mean activity of the population gets to be almost constant around a unit value, while the DOM is strictly related to the absolute value of the stimulus disparity (see Fig. 2). In such a way, the population response is stable and insensitive to contrast, luminance and texture frequency content, and it can consequently provide both a control signal with constant gain and a constant learning rate. To overcome the problem of noise sensitivity in the DOM, we used the standard deviation of the population response that depends on the whole population activity, even though from now onwards we will refer to it as DOM. Such a dependence

can be derived from the complex cell's response given by Eq. 3. The average standard deviation

$$\rho = \sum_{\mathbf{x}\in\Omega} \sqrt{\frac{1}{\pi}\int_0^{\pi}\frac{1}{2\pi}\int_{-\pi}^{\pi}(\hat{r}_c(\mathbf{x};\theta,\Delta\psi))^2\; d\Delta\psi\, d\theta - 1} = \sum_{\mathbf{x}\in\Omega}\sqrt{\frac{1}{1+\cosh\left(-|\boldsymbol{\delta}|^2/\sigma^2\right)}} \tag{6}$$

can be used as an ideal reward for the system, both to indicate how much the system is close to the correct state (zero disparity) and to evaluate how much effectively and rapidly an action drives the system toward the correct state.

### *3.2. The Learning Algorithm*

The proposed approach relies on a particle swarm optimization algorithm (PSO) [17] to evolve a *swarm* of candidate particles (*i.e.* sets of weights $\boldsymbol{w}$). The evolution toward the solution is driven by a Monte Carlo reinforcement learning technique [18]. From a generation to the next, the optimality of each set of weights is evaluated through the DOM. Even though the PSO is not likely to be the learning mechanism of the cortex, it is a natural candidate for our algorithm because it can cope with high dimensional space of possible solutions like the ones we have in population-based algorithms. Moreover, not relying upon gradient descent like techniques, it does not require the optimization criterion to be differentiable, as for classic optimization methods. Following the PSO specifications, we implemented a population $\boldsymbol{\Pi}$ (swarm) of 20 candidate solutions (particles). Each particle $\pi$ is a different set of weights $\mathbf{w}$, that are randomly initialized. At each generation, the particles are tested with 100 different trials, consisting of a synthetic stimulus (a random dot stereogram) with horizontal disparity distributed randomly within the range $[-3\Delta, 3\Delta]$. On the single trial, the stimulus disparity is modified according to the control produced by the particle $\pi$ considered $\delta_H(t) = \delta_H(t-1) + r_v$, along 10 time steps. The swarm evolves for 500 generations in which, step by step, each particle is moved in the search-space taking into account four different factors: (1) an inheritance from the same particle at the previous generation (progressive exploration and exploitation of acquired knowledge); (2) a random component to escape from the local minima (exploration of the search space); (3) a tuning towards the particle that achieved the best result on the whole set of trials ($Q_1$); (4) a tuning towards the particle that achieved the best action on the single time step of the trial ($Q_2$). Each of these factors is multiplied by a coefficient that defines its weight on the new generation of particles.

Following the Monte Carlo exploration, after each generation, the observed returns are used to evaluate the particles, that are subsequently improved at all the states visited in the trial. In such a way, the single learned control is capable of providing the correct movements, but in a smaller range respect to [6]. To improve the performance of the system, we exploited the DOM to include the dual-mode behaviour [16]. If $\rho$ is below a proper threshold, the system is in a state where a LONG-like control is needed, otherwise it is in a state where a SHORT-like control is needed. This policy raises two specialized controls for the two different disparity ranges. The particles are evaluated considering: (1) which particle yields the best action on each single trial, *i.e.* approached the system closer to the optimal state (zero disparity) through $Q_1^{\pi}(s) = E\left\{\sum_{t=0}^{T}\gamma^t\rho_t \mid s_t = s\right\}$ where $\gamma$ is the discount rate that determines the value of the past rewards; (2) which particle at a single time step provides the best movement toward the optimal state, through $Q_2^{\pi}(s,t) = \rho_t - \rho_{t-1} \mid s_t = s$, where $t$ is the time, and $s_t$ determines the membership to a state. In this way, the swarm evolves not just toward the best position, but rather toward a mean position of all the particles that provided good actions during the generation.

The overall rule for the improvement of the particles at the generation $g$ is:

$$\boldsymbol{\Pi}^g = e\boldsymbol{\Pi}^{g-1} + \boldsymbol{\zeta} + \mu\boldsymbol{\Pi}^{g-1} * \boldsymbol{Q}_1^{\mathrm{OPT}} + \xi\boldsymbol{\Pi}^{g-1} * \boldsymbol{Q}_2^{\mathrm{OPT}} \tag{7}$$

where $e$, $\mu$ and $\xi$ are the weights that determine the predominance of inheritance, the $Q_1^{\pi}(s)$ and the $Q_2^{\pi}(s,t)$ terms, and $\boldsymbol{\zeta}$ is the random component. The operator $*$ computes the center of gravity of the particles of the swarm weighted by the vectors $\boldsymbol{Q}_1^{\mathrm{OPT}}$ and $\boldsymbol{Q}_2^{\mathrm{OPT}}$, that is the best solution with respect to $Q_1^{\pi}(s)$ and $Q_2^{\pi}(s,t)$.

## 4. Results

In a comparison with the SHORT and LONG controls (see Fig. 3A), the result at the end of the evolution is that almost each particle evolves toward a SHORT-like or a LONG-like control (see Fig. 3, B-D), validating the

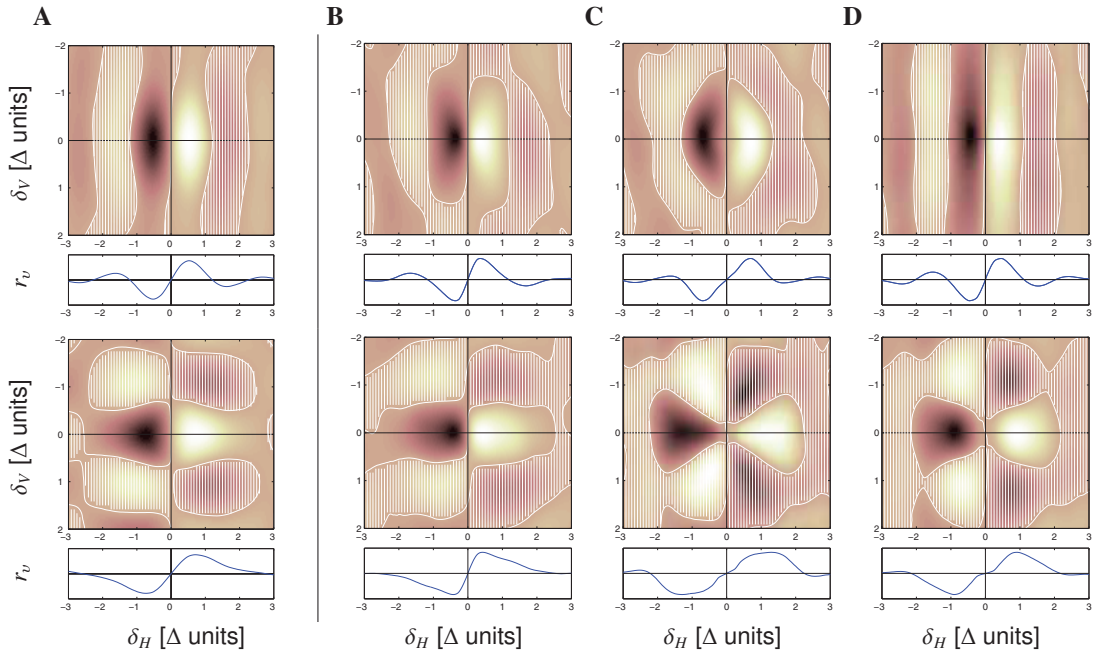


Fig. 3. The 2D response curve and horizontal cross sections for the SHORT (top) and LONG (bottom) imposed (A) and learned (B-D) controls. The white hatched areas mark where the control produces the wrong vergence movement.

approach. From a behavioural point of view, the emerging of two different controls, that are close to the imposed ones, can be considered as the capability of the system to meet the contingent situation of the task at hand. In fact, each particle evolves towards a solution that is defined by its affinity with one of the two states. From this perspective, a SHORT-like control can be considered the results of an evolution in an situation characterized by small disparities only, whereas a LONG-like one turns out from an situation with large disparities. As pointed out by [19], each cell of the population cooperates to generate the motor control, not only those characterized by zero phase difference [10] or by a vertical orientation selectivity [11]. Similarly, in our network the weights that can be associated to SHORT/LONG control signals show a continuous distribution spread over the whole population. Increasing the dimension of the state space to more than two controls, would lead to an intermediate distributed representation of the vergence control channels, each specialized for a restrained range of disparities.

To assess the efficacy of the learned weights, we tested the vergence control performance in a simple virtual environment, as in [6]. The geometrical system of the binocular vision system is characterized by common tilt for the left and right cameras ($\alpha_L$, $\alpha_R \equiv \alpha_L$), and independent pan angles ($\beta_L$, $\beta_R$), as in the Helmholtz reference frame [20]. This configuration yields a simplified parameterizations of visual direction in term of version $\upsilon$ and vergence $\chi$ angles [7]:

$$\begin{cases} \upsilon = 1/2(\beta_L + \beta_R) \\ \chi = \beta_L - \beta_R \end{cases} \quad \text{or} \quad \begin{cases} \beta_L = \upsilon - \chi/2 \\ \beta_R = \upsilon + \chi/2. \end{cases} \tag{8}$$

In fact, the vergence control needed to move the fixation point, while keeping constant gaze direction, is a quantity $\Delta\chi$ to be applied symmetrically on both the eyes: $\Delta\chi = \Delta\beta_R = -\Delta\beta_L = -\arctan(\frac{r_V}{2f_0})$. The test considers a frontoparallel plane whose position in depth varies in time as a ramp (with different slopes) and as a sinusoid (with different frequencies) or as a step (see Fig. 4). The learned sets of weights are equally able to produce both fast changes of the fixation point (LONG-like control) and smooth movements as well as stable fixations (SHORT-like control). In the same way, the frequency of the sinusoid that controls the depth of the plane was varied between 7 and 38 time steps (not shown), and again the simulated results are qualitatively similar to the experimental data [16]. The learned vergence control (red line), similarly to the imposed one (blue line), ensures

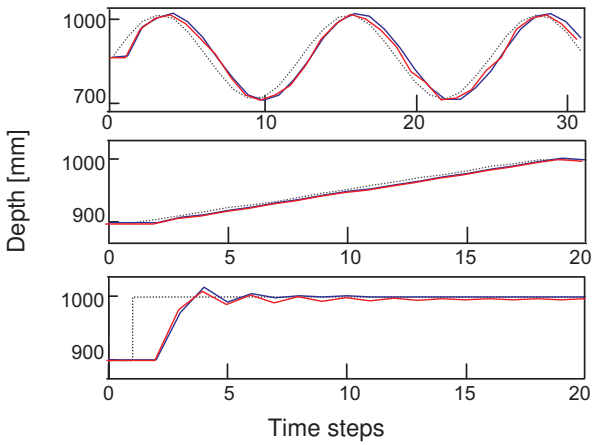


Fig. 4. Depth of the fixation point moved by the imposed (blue) and learned (red) vergence control to sinusoid, ramp and step movement of a frontoparallel plane (dotted).

both the ability to track a moving stimulus in depth (dotted line), and to provide robust and stable fixations on steady stimuli.

In order to validate the use of the standard deviation of the population response respect to other kind of rewards, we made a comparison with other learned controls. The performance of the learned signals is compared with the effectiveness of the control computed relying on the extensive knowledge of the resources [6]. As expected, the ground truth disparity is an highly effective signal to measure the distance from the goal (*i.e.* zero disparity in fovea), thus driving the learning. The controls that specialize at the end of the evolution (not shown), can be clustered in two different types, LONG-like and SHORT-like, and do not differ from those learned exploiting the DOM. Using the response of the cell defined by $\theta = 0$ and $\Delta\psi = 0$, on the other side, allows to learn a control that is effective only if the stimulus disparity is small (about $[-\Delta, \Delta]$), that is considerably smaller respect to the limits supported by the LONG-like controls. Moreover, the learned signals, that are all close each other as shape, run out of the zero crossing, that is a mandatory feature for the correct vergence posture (see Fig.5).

The setup considered to compare the effectiveness of the controls consists of a frontoparallel plane that oscillates around the depth of 600mm with constant frequency, but with increasing amplitude (see Fig. 6).

The effect of such stimulus on the different vergence controls is that, while the amplitude increases, the disparity that the system experiences increases with it, up to the point the the control is no more able to follow the plane in depth. Such a motion in depth is effective in comparing the effectiveness of the different controls, because the earlier a single control loses the plane, the lesser it is effective in providing the correct control for vergence. The control computed with the LS minimization (blue line) is able to follow properly the stimulus in depth (black dotted line) up to the last oscillation, with a small delay. The control learned with the ground truth disparity (green line) presents a bigger delay, but it is able to cope with larger disparities, in fact it follows properly the stimulus to the end of the trajectory. The control learned through the standard deviation (magenta line) is able to follow the stimulus almost to the end of the trajectory, but with a slightly bigger delay. Finally, as it can be deduced by Fig. 5, the control learned with the response of the cell with $\theta = 0$ and $\Delta\psi = 0$ (red line) is able to work in a very limited range, and consequently show a limited capability of following the stimulus in depth. Moreover, as shown at the beginning of the trajectory, lacking of zero crossing, it is not able to keep the fixation point on a steady stimulus.

In conclusion, by exploiting a precise and extensive knowledge either of the environment (ground truth disparity) or of the resources (tuning curves) allows to obtain an effective control. Nevertheless, such techniques are grounded on a kind of knowledge that prevent the system to learn autonomously its behaviour. On the other side, the whole activity of the network, designed mimicking the neural parameters, is an internal parameter of the

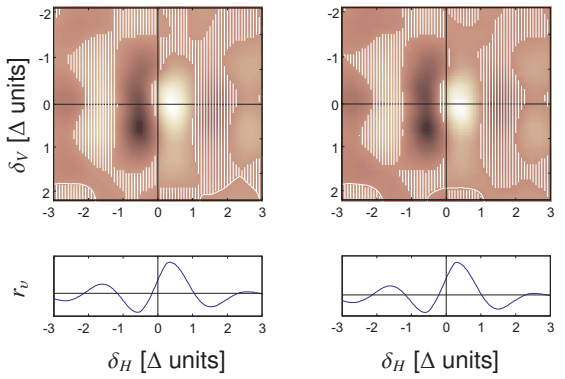


Fig. 5. A choice of two particles of the swarm learned by the system, using the response of the complex cell defined by $\theta = 0$ and $\Delta\psi = 0$ as reward signal. The behaviours can be qualitatively compared to the SHORT one. The areas where the control is ineffective are marked by white hatches (top row). The vergence control is measured in [Δunits/time steps], *i.e.* the movement that the control produces in a single step of time. The bottom row show the horizontal cross sections of the controls.

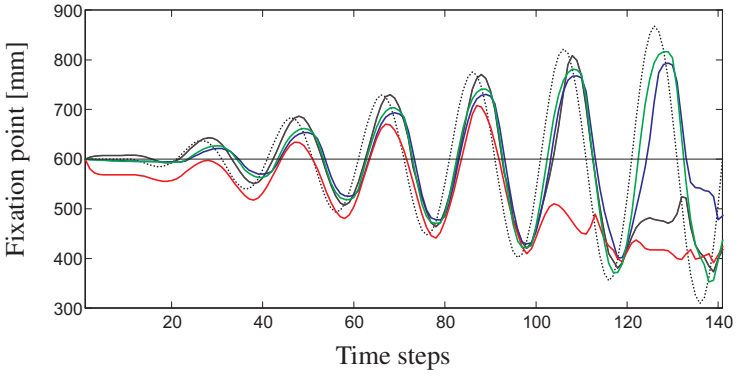


Fig. 6. Vergence trajectories achieved by the system relying on different controls, while looking to a plane moving in depth: ground truth depth of the stimulus (black dotted line), control obtained with the LS algorithm (blue line), control learned using the ground truth disparity as reward (green line), control learned by the algorithm using the standard deviation of the population response as reward (magenta line), control learned using the response of the cell with $\theta = 0$ and $\psi = 0$ as reward (red line).

architecture, that is able to drive the learning of the behaviour and to obtain equivalent performances.

## 5. Conclusion

We presented an architecture that is able to learn a proper vergence behaviour without external supervision. The proposed approach exploits the advantages and the flexibility of distributed cortical-like architectures against solutions based on a conventional systemic coupling of sensing and motor components. The choice of the (uniform) distribution of the resources in the parameter space, together with the normalization circuit, allow the computational substrate to provide effective and stable signals both for vergence movements and for driving the learning of the control itself. Indeed, the activity of the whole population defines the state of the vision system with respect to the environment, which is exploited both to compute the control for the vergence movement, and to evaluate the effect (positive or negative) of the movement on the system.

By integrating directly early vision modules and motor control, the architecture not only closes the perception-action loop in a continuous (*i.e.* not sequential) way, with a reciprocal benefit for perception and action, but also at an inner cycle, exploiting the interaction with the environment to evaluate the effect of the movement, and to

eventually learn a proper behaviour. In such a way, an artificial system earns the capability of perceiving the 3D structure of the environment so to coordinate the camera movements and to better exploit its resources.

## Acknowledgements

This work has been partially supported by the Italian MIUR (PRIN 2008) project “Bio-inspired models for the control of robot ocular movements during active vision and 3D exploration ”.

## 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] D. Fleet, H. Wagner, D. Heeger, Modelling binocular neurons in the primary visual cortex, Jenkin and Harris, Cambridge, 1996.
[3] L. Wilcox, R. Allison, Coarse-fine dichotomies in human stereopsis, Vis. Res. 49 (2009) 2653–2665.
[4] R. Held, E. Birch, J. Gwiazda, Stereoacuity of human infants, PNAS 77 (1980) 5572–5574.
[5] R. N. Aslin, Development of binocular fixation in human infants, J. of Experimental Child Psychology 23 (1) (1977) 133 – 150.
[6] 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.
[7] 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.
[8] 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.
[9] J. Piater, R. Grupen, K. Ramamritham, Learning real-time stereo vergence control, in: Intelligent Control/Intelligent Systems and Semiotics, 1999, Cambridge, MA , USA, 1999, pp. 272–277.
[10] 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.
[11] Y. Wang, B. Shi, Improved binocular vergence control via a neural network that maximizes an internally defined reward, Autonomous Mental Development, IEEE Transactions on 3 (3) (2011) 247 –256.
[12] 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, CVIU 114 (6) (2010) 681 – 699.
[13] 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.
[14] D. Fleet, H. Wagner, D. Heeger, Neural encoding of binocular disparity: Energy models, position shifts and phase shifts., Vision Research 36(12) (1996) 1839–1857.
[15] M. Kouh, T. Poggio, A canonical neural circuit for cortical nonlinear operations, Neural Computation 20 (2008) 1427–1451.
[16] 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.
[17] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Neural Networks, 1995. Proceedings., IEEE International Conference on, Vol. 4, 1995, pp. 1942–1948.
[18] N. Metropolis, S. Ulam, The monte carlo method, J. Am. Stat. Ass. 44 (247) (1949) 335–341.
[19] A. Takemura, Y. Inoue, C. Quaia, F. 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.
[20] M. Hansard, R. Horaud, Cyclopean geometry of binocular vision, J. Opt. Soc. Am. 25 (2008) 2357–2369.