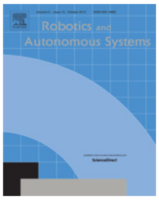
Robotics and Autonomous Systems 71 (2015) 23–34

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