Chapter 7
Early Perception–Action Cycles in Binocular Vision: Visuomotor Paradigms and Cortical–Like Architectures

Silvio P. Sabatini
DIBRIS – University of Genoa, Italy

Fabio Solari
DIBRIS – University of Genoa, Italy

Andrea Canessa
DIBRIS – University of Genoa, Italy

Manuela Chessa
DIBRIS – University of Genoa, Italy

Agostino Gibaldi
DIBRIS – University of Genoa, Italy

ABSTRACT

Pushed by ample neurophysiological evidence of modulatory effects of motor and premotor signals on the visual receptive fields across several cortical areas, there is growing attention for moving the active vision paradigm from systems in which just the effects of action influence the perception, to systems where the acting itself, and even its planning, operate in parallel with perception. Such systems could close the loops and take full advantage of concurrent/anticipatory perception-action processing. In this context, cortical-like architectures for both vergence control and depth perception (in the 3D personal space) that incorporate adaptive tuning mechanisms of the disparity detectors are presented. The proposed approach points out the advantages and the flexibility of distributed and hierarchical cortical-like architectures against solutions based on a conventional systemic coupling of sensing and motor components, which in general poses integration problems since processes must be coupled that are too heterogeneous and complex.

INTRODUCTION

There is increasing interest in moving the active vision field from systems in which just the effects of action influence the perception, to systems where the acting itself, and even its planning, operate in parallel with perception, thus really closing the loops and taking full advantage of a concurrent/anticipatory perception-action processing. From this perspective, the motor system of a humanoid should be an integral part of its perceptual machinery (e.g., see Int. Journal of Humanoid Research Special issue on the “Active
Traditionally, however, in robot vision systems, perception-action loops close at a “system level” (by decoupling *de facto* the vision modules from those dedicated to motor control and motor planning), and the computational effects of the eye movements on the visual processes are rarely implemented in artificial systems.

The limitation of this approach is that solving specific high-level tasks usually requires sensory-motor shortcuts at the system level, and specific knowledge-based rules or heuristic algorithms have to be included to establish behaviorally consistent relationships among the extracted perceptual features and the desired actions. The risk is to abandon distributed representations of multiple solutions to prematurely construct integrated descriptions of cognitive entities and commit the system to a particular behavior. Conversely, our claim is that early/complex interactions between vision and motor control are crucial in determining the effective performance of an active binocular vision system with a minimal amount of resources and coping with uncertainties and inaccuracies of real systems.

There is ample evidence for the pivotal role of programmed eye movements in the computations that are performed in the process of seeing (as opposed to “looking at”). Yet, how to profitably integrate this accumulating evidence with the computational theories of stereo vision has not been fully exploited, as rectified images are still calculated in current humanoid active-disparity vision modules, relying solely on the encoder’s data. The complexity of integrating efficiently and with flexibility the different aspects of binocular active vision indeed prevented until now a full validation of the visuomotor approaches to 3D perception in real world situations. We believe that the advantages of binocular visuomotor strategies could be fully understood only if one jointly analyzes and models the problem of neural computation of stereo information, and if one takes into account the limited accuracy of the motor system. Unfortunately, models in this joint field are few (Theimer & Mallot, 1994; Hansard & Horaud, 2008; Read & Cumming, 2006) and rarely address all the computational issues.

In this work we defend a visuomotor approach to 3D perception by proposing the instantiation of visuomotor optimization principles concurrently with the design of distributed neural models/architectures that can efficiently embody them. Specifically, we present two case studies that show how large-scale networks of V1-like binocular cells can provide a flexible medium on which to base coding/decoding adaptation mechanisms related to sensorimotor schema. At the coding level, the position of the eyes in the orbits can adapt the disparity tuning to minimize the necessary resources, while preserving reliable estimates (i.e., adjustable tuning mechanisms based on the posture of the eyes to improve depth vision; see Chessa, Sabatini & Solari, 2009). At the decoding level, read-out mechanisms of the disparity population code can specialize to gain vergence control servos over a wider range of disparities than what would be possible through an explicit calculation of the disparity map (Gibaldi, Canessa, Chessa, Solari & Sabatini, 2010).

**BACKGROUND**

Purposive (active) vision is an important source of information and provides a number of cues about the 3D layout of objects in a scene that could be used for planning and controlling goal-directed behaviors. Although computer and robot vision are today making technological progress, the current active vision solutions are increasingly driven by applications like robotic manipulation or surveillance and are still far from reaching a real purposive behavior. More precisely, research in robot vision is more geared towards active “looking at” scenes and objects rather than “seeing”, gathering the visual data and having to act in a limited time and space. According to the senso-