Robotics and Autonomous Systems 71 (2015) 3–12

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Self-calibrating smooth pursuit through active efficient coding

C. Teulière [a,c,*], S. Forestier [a,d], L. Lonini [a], C. Zhang [b], Y. Zhao [b], B. Shi [b], J. Triesch [a]

[a] Frankfurt Institute For Advanced Studies, Goethe University, Frankfurt am Main, Germany
[b] Department of Electronic and Computer Engineering, HK University of Science and Technology, Hong Kong, China
[c] Clermont Université, Université Blaise Pascal, Institut Pascal, Clermont-Ferrand, France
[d] École normale supérieure de Rennes, Bruz, France

HIGHLIGHTS

- Efficient coding principle is used as a criterion for learning smooth pursuit eye movements.
- A multi-scale approach allows to perceive a large range of motions.
- The model is fully self-calibrating and autonomously recovers from perturbations in the perception/action link.
- Experiments on both simulation and iCub robot demonstrate the approach.

ARTICLE INFO

Article history:
Available online 15 November 2014

Keywords:
Autonomous learning
Active perception
Smooth pursuit
Efficient coding
Robotics

ABSTRACT

This paper presents a model for the autonomous learning of smooth pursuit eye movements based on an efficient coding criterion for active perception. This model accounts for the joint development of visual encoding and eye control. Sparse coding models encode the incoming data at two different spatial resolutions and capture the statistics of the input in spatio-temporal basis functions. A reinforcement learner controls eye velocity so as to maximize a reward signal based on the efficiency of the encoding. We consider the embodiment of the approach in the iCub simulator and real robot. Motion perception and smooth pursuit control are not explicitly expressed as tasks for the robot to achieve but emerge as the result of the system's active attempt to efficiently encode its sensory inputs. Experiments demonstrate that the proposed approach is self-calibrating and robust to strong perturbations of the perception–action link.

1. Introduction

Since the development of information theory, the idea of exploiting redundancy of information in signals to encode them in an efficient manner has been widely applied to different scientific areas. Concepts such as sparse coding techniques are now well known tools in mathematics, signal processing and computer science.

In neuroscience, efficient coding was proposed as a principle for the encoding of sensory information in the brain [1–3]. In particular, one popular expression of the *efficient coding hypothesis* posits that only a few neurons fire at a given time, thus representing sensory inputs with a "sparse code". Several studies proposed to use a sparse coding mechanism on natural stimuli by decomposing them as a linear combination of a small number of basis functions from an over-complete dictionary. Interestingly, it was shown that the basis functions that were learned when optimizing the reconstruction of the input using such a sparse code resemble the receptive fields of sensory neurons in visual, auditory, or olfactory systems [4–6]. In particular, this efficient coding hypothesis implies that the sensory representation in the brain captures the statistics of the sensory inputs. Those statistics obviously depend on the environment of the agent. Importantly, they are also shaped by its behaviour [7,8], which can be directed to control the incoming sensory information. However, there is still little work that accounts for the role of behaviour in the development of an efficient sensory coding.

Recently, [9] used the efficient coding of the causal relation between motor actions and sensory feedback as a drive for the co-development of sensory and motor maps. [10] applied the efficient coding principle to active vision and showed that this principle can lead to the joint learning of an efficient depth representation and

* Corresponding author at: Clermont Université, Université Blaise Pascal, Institut Pascal, Clermont-Ferrand, France.
E-mail address: celine.teuliere@univ-bpclermont.fr (C. Teulière).

http://dx.doi.org/10.1016/j.robot.2014.11.006
0921-8890/

eye vergence movements. This model was embodied into a robot binocular vision system in [11] and showed strong robustness properties [12]. Even more recently, [13] suggested that this model can be extended to a larger range of action–perception loops, by simulating the emergence of smooth pursuit behaviour as the result of an efficient encoding criterion. Note that smooth pursuit control in humans appears to improve in the same time period as motion perception which suggests a co-development [14].

In this paper, we extend the approach of [13] and take it to a next step through its embodiment in a robotic system. We believe that the autonomous learning of active perception loops is of great interest in the robotics context in order to develop self-calibrating systems that can flexibly adapt to their environments.

Smooth pursuit abilities and gaze control in general are fundamental not only to humans but also humanoids as they condition lots of basic behaviours such as e.g. reaching objects [15]. Motion perception and tracking have been largely studied in the computer vision and robotics communities. Most approaches for motion perception involve either optic flow computation [16,17], or some form of object representation along with matching or tracking techniques [18] which sometimes require calibrated sensors [19]. Visual servoing approaches [20] can then be used to close the loop between perception and action. Such methods provide good performances in terms of accuracy. Some specific controllers have also been designed in the context of smooth pursuit [21,22]. However, the above methods require knowledge of the kinematic link between the camera velocity and the visual change. This requirement implies both the need for a calibration phase and the inability to handle modifications due to, e.g., a mechanical shock to the system. Some attempts have been made to learn this link [23–25]. However, such techniques still require prior knowledge on a specific goal for the robot to reach. In this work, we consider a different paradigm: motion perception and smooth pursuit control are not explicitly expressed as tasks for the robot to achieve but emerge as the result of the system's active attempt to efficiently encode its sensory inputs. A sparse coding model (perception component) encodes sensory information from pairs of successive frames using basis functions at different spatial resolutions, while a reinforcement learner (action component) generates the camera movement based on the output of the sparse coding model. Importantly, perception and behaviour develop in parallel, by minimizing the same cost function: the error between the original stimulus and its reconstruction by the sparse coding model. We call this approach *active efficient coding*.

We extend the work of [13] in the following ways: first, we increase the range of motion that the system can perceive by using a multi-scale approach, and demonstrate its benefits. Secondly, we demonstrate that our system can adapt to drastic perturbations in its perception–action link. Finally, we present experiments both in simulation and on a real robot and show that the model performs well in realistic conditions with the presence of noise and distortions.

2. Model architecture

This section presents the architecture of our model. An overview of the main components is given in 2.1 and their description is detailed in Sections 2.2 and 2.3.

2.1. Overview

Our embodiment of the efficient coding principle to the autonomous learning of smooth pursuit relies on the following idea: when one eye is smoothly pursuing an object of interest, the successive images it senses in its foveal region are very similar and can therefore be encoded efficiently when exploiting this redundancy of information. Our model makes use of this basic idea by considering the encoding quality of pairs of successive image patches as the criterion that drives the movements of the eye.

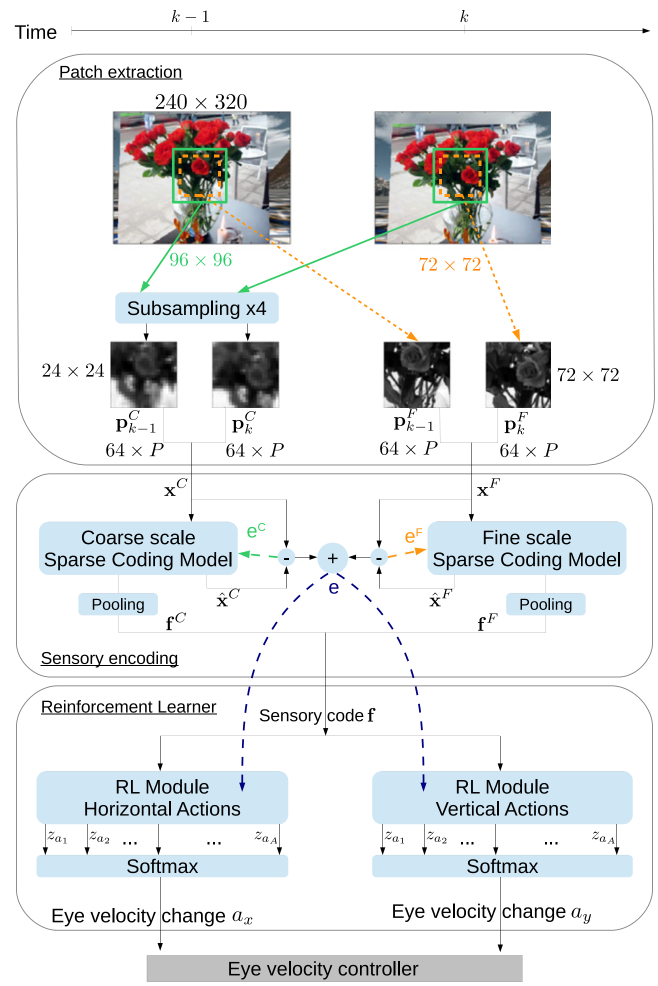

Fig. 1. Overview of the model.

In the following, we will model one eye by a camera, which can rotate in both pan and tilt degrees of freedom. Note that we only consider one eye for convenience but [26] shows that the model can be extended to two eyes. Rotations of the camera around the line of sight as they occur in the primate visual system are left for future work.

Our model consists of two main components (see Fig. 1):

- The sensory encoding component receives image patches of two consecutive frames from the camera, and encodes them as a sparse linear combination of basis functions. It is composed of two sparse coding modules dealing with different spatial resolutions.
- The motor control component is based on a reinforcement learning agent that generates velocity commands for the robot camera according to the encoding of the images. The agent receives a reward for the selected action depending on the efficiency of the encoding of the subsequent image patches by the sensory coding component.

The following subsections describe the two components in more detail.

2.2. Sensory encoding

To allow our system to deal with a large range of motion, we consider two different scales or image resolutions and train one sparse coding model for each scale. The structure of the two models is identical, the only difference is the input they receive: fine or coarse scale image patches.