Robotics and Autonomous Systems 71 (2015) 49–59

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Learning agent's spatial configuration from sensorimotor invariants☆

Alban Laflaquière [a,b], J. Kevin O'Regan [c], Sylvain Argentieri [a,b], Bruno Gas [a,b], Alexander V. Terekhov [c,*]

[a] Sorbonne Universités, UPMC Univ Paris 06, UMR 7222, ISIR, F-75005, Paris, France

[b] CNRS, UMR 7222, ISIR, F-75005, Paris, France

[c] Laboratoire Psychologie de la Perception, CNRS UMR 8242, Université Paris Descartes, F-75006, Paris, France

HIGHLIGHTS

- Autonomous robots should develop perceptual notions from raw sensorimotor data.
- Environment-dependency of visual inputs complicates acquisition of spatial notions.
- Agent can learn its spatial configuration through invariants in sensorimotor laws.
- Approach is illustrated on a simulated planar multijoint agent with a mobile retina.

ARTICLE INFO

Article history:
Available online 19 January 2015

Keywords:
Space
Perception
Robotics
Learning
Developmental robotics
Sensorimotor theory

ABSTRACT

The design of robotic systems is largely dictated by our purely human intuition about how we perceive the world. This intuition has been proven incorrect with regard to a number of critical issues, such as visual change blindness. In order to develop truly autonomous robots, we must step away from this intuition and let robotic agents develop their own way of perceiving. The robot should start from scratch and gradually develop perceptual notions, under no prior assumptions, exclusively by looking into its sensorimotor experience and identifying repetitive patterns and invariants. One of the most fundamental perceptual notions, space, cannot be an exception to this requirement. In this paper we look into the prerequisites for the emergence of simplified spatial notions on the basis of a robot's sensorimotor flow. We show that the notion of space as environment-independent cannot be deduced solely from exteroceptive information, which is highly variable and is mainly determined by the contents of the environment. The environment-independent definition of space can be approached by looking into the functions that link the motor commands to changes in exteroceptive inputs. In a sufficiently rich environment, the kernels of these functions correspond uniquely to the spatial configuration of the agent's exteroceptors. We simulate a redundant robotic arm with a retina installed at its end-point and show how this agent can learn the configuration space of its retina. The resulting manifold has the topology of the Cartesian product of a plane and a circle, and corresponds to the planar position and orientation of the retina.

1. Introduction

The classical approach to the control of robotic systems consists in developing an electro-mechanical model of the robot, defining the range of operating conditions, and building algorithms for robot/environment state estimation and robot control. This approach, while efficient in highly controlled environments (such as robotized factories), can lead to severe failures when operating under unforeseen circumstances. Making robots more autonomous and capable of operating in unstructured a priori unknown environments is the main objective of modern robotic research.

The application of techniques from machine learning and artificial intelligence has enabled significant progress in this direction over recent decades. Numerous solutions have been proposed for online robot self-modeling [1–5] (see also [6] for a comprehensive review) and recovery from unknown damages [7–9]. These solutions usually define a set of building blocks (such as rigid bodies) and the rules of their connection (e.g. joints) and try to find the

☆ This work was conducted within the French/Japan BINAAHR project under Contract No. ANR-09-BLAN-0370-02 funded by the French National Research Agency. J.K. O'Regan & A.V. Terekhov were financed by ERC Advanced Grant 323674 "FEEL".

* Corresponding author.
E-mail addresses: alban.laflaquiere@gmail.com (A. Laflaquière), jkevin.oregan@gmail.com (J.K. O'Regan), sylvain.argentieri@upmc.fr (S. Argentieri), bruno.gas@upmc.fr (B. Gas), avterekhov@gmail.com (A.V. Terekhov).

http://dx.doi.org/10.1016/j.robot.2015.01.003
0921-8890/

combination of these blocks that best accounts for the incoming sensory information given the motor commands. Examples using such an approach can be found for instance in [10,11] and more recently in [12,13]. Another approach avoids defining building blocks but instead adds some pre-processing of the sensory flow to avoid facing its raw complexity, generating inputs that suit the task. Such pre-processing can for instance be used to define the coordinates of the robot's hand in the visual field to learn a kinematic model of the arm [14,15] or a target object for reaching [16,17].

Although this type of approach often produces spectacular results, its robustness and efficiency strongly depend on the choice of the building blocks and pre-processing algorithms. This makes the entire approach heavily biased by the designer's intuition, which is rooted in human perception and which is not necessarily best suited for robots, whose sensors and effectors differ significantly from those of human beings. To understand the implications of this difference, consider the fact that the entire field of computer vision has been biased by the false perceptual intuition that seeing is similar to having a photo of the visual scene [18,19] and that stepping aside from this paradigm can yield unexpectedly fruitful results [20–22]. Thus, to make robotic systems truly autonomous and robust, we need to shed the biases imposed on us by our own perceptual system and let robots develop their own ways of perceiving the world. Few studies adopting such a radical approach to control robots have been proposed [23–25]. Although these studies are in line with our approach, they do not directly address the problem of space perception (or only implicitly through the robot's ability to move in its environment). This will be the main focus of the present paper.

In order to minimize a priories about perceptual systems we consider a robotic agent designed as a *tabula rasa* receiving undifferentiated sensory inputs (e.g. not knowing whether a given one of them comes from a video camera or from an encoder in a joint) and sending out undifferentiated motor outputs. Perceptual structures can emerge as stable patterns in the agent's sensorimotor flow. This approach, when applied to visual information, can lead to the discovery of stable features, such as edges, similar to those present in the human visual cortex [26–30]. A similar approach, formalized in the language of information theory [31], can make it possible to describe the topological structure of the agent's surface [32] and certain properties of its interaction with the environment [33–35]. The cited studies clearly show that *tabula rasa* agents can learn basic properties of the available sensorimotor information. However, only the simplest perceptual notions are straightforwardly dictated by the sensory inputs themselves. More complicated notions represent laws linking sensations and actions, rather than particular instances of sensory information [36]. Space, which does not correspond to any particular sensory inputs, is one such a notion.

Can a *tabula rasa* approach be used to model the acquisition of the notion of space? In order to answer this question we first need to decide on what we mean by the term *space*. In mathematics, various spaces are described: topological, metric, linear, etc. Each of these notions captures certain features of what we usually mean by space. For example, topological spaces only feature the notion of proximity, and can be thought of as reflecting an agent with a highly impaired ability to make distance judgments (which is true for humans performing certain tasks). Although rather primitive, topological space nevertheless includes some fundamental aspects of space in general, such as dimensionality, and its notion can be useful in such tasks as the mapping of large spaces [37]. Metric spaces are more complicated objects, which imply precise information of distances. They provide the tool required to work with such notion as the length of a path, and can underlie navigation abilities. In particular, knowledge of a metric space enables odometry and SLAM [38–40]. Linear spaces introduce the notion of the vector, which is an efficient tool for describing motion. The link between motion and linear spaces is used in many studies that address the problem of space acquisition. Thus, Poincaré [41] suggested that spatial knowledge emerges from the agent's capacity to move, with spatial relations such as the distance to an object being internally encoded as potential motor commands. The agent's ability to move has also played an essential role in more recent works on space [37,42,43]. Philipona and co-authors showed in [44] that under certain conditions the dimensionality of space can be estimated by analyzing only sensorimotor information that is available to the agent. This result launched a series of publications by the present authors, extending the conditions of dimension estimation [45] and applying similar ideas to different agents and robotic systems [46–48].

Knowing the number of spatial dimensions is not, however, the same as having the notion of space. It has recently been shown that the notion of space can be learned as a proprio-tactile mapping [43] or as a group of rigid transformations of the environment [49]. Here we focus on a different aspect of spatial knowledge, probably the simplest that can be extracted by a naive agent. In order to introduce it, let us first note that mathematical spaces (topological, metric, etc.) do not emphasize what is special about our subjective experience of space. Mathematical spaces can be applied, for example, to describe the full set of an agent's body postures, or motor commands, or even the outputs of every pixel in the agent's visual sensor (e.g. camera). However, these examples clearly do not correspond to what we usually mean by space. We believe that what characterizes space is the particular structure that it imposes on possible sensorimotor experiences. It can be identified in the laws that govern the way sensory inputs change as the agent moves around.

The first and most basic property of those laws is an invariance: space does not depend on the particular environment, nor on the particular posture of the agent. The agent must somehow know that its sensor is at the same spatial position independently of what objects are around and what are the positions of the other sensors. In other words, the first aspect of the notion of space is the "point of view" from which the agent "looks" at the world (here we adopt visual terminology for simplicity, but the notion must not depend on the particular type of sensors in question: camera, microphone, or taxel array). From now on, when we speak about the notion of space we will be referring to the set of the agent's "points of view". These "points of view" are the precursors to the more convenient notion of "point", which is the basic element of what we call space, and which can be used to build more complex notions of space. Note that in our approach, we are looking at the problem of space from the agent's point of view. Instead of taking the existence of external space for granted, we are trying to identify signs of it in the agent's sensorimotor flow, and to see what makes the notion of space useful to the agent. Our hope is that by learning how the notion of space can be constructed from the sensorimotor flow we will acquire a better understanding of how other perceptual notions can be learned, such as body and object. In this respect, our work is notably different from the field of research on body schema acquisition, which is already the subject of a large literature. The question of space is usually eluded in these studies, as it is supposed to be either an unnecessary prerequisite for action or to already have been acquired.

This study extends our previous work [48], where a neural network was used to learn a mapping between the motor space and an internal representation of the agent's external configuration. This internal representation was generated online during the exploration of multiple environments. The present paper introduces two main improvements. First, it offers a clear definition of the structure of the constraints captured by the agent. In doing so, it makes explicit the mapping that was implicitly captured by the neural network in our previous study. Second, the metrics of the