Neurocomputing 112 (2013) 179–188

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Kinesthetic teaching of visuomotor coordination for pointing by the humanoid robot iCub

Andre Lemme [a,*], Ananda Freire [a,b], Guilherme Barreto [b], Jochen Steil [a]

[a] Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld University, Universitätsstr. 25, 33615 Bielefeld, Germany
[b] Federal University of Ceará, Department of Teleinformatics Engineering, Av. Mister Hull, S/N - Center of Technology, Campus of Pici, Fortaleza, Ceará, Brazil

ARTICLE INFO

Available online 27 February 2013

Keywords:
Pointing gesture
Extreme learning machine
Static reservoir network
Visuomotor coordination
iCub

ABSTRACT

Pointing at something refers to orienting the hand, the arm, the head or the body in the direction of an object or an event. This skill constitutes a basic communicative ability for cognitive agents like, e.g. humanoid robots. The goal of this study is to show that approximate and, in particular, precise pointing can be learned as a direct mapping from the object's pixel coordinates in the visual field to hand positions or to joint angles. This highly nonlinear mapping defines the pose and orientation of a robot's arm. The study underlines that this is possible without calculating the object's depth and 3D position explicitly since only the direction is required. To this aim, three state-of-the-art neural network paradigms (multilayer perceptron, extreme learning machine and reservoir computing) are evaluated on real world data gathered from the humanoid robot iCub. Training data are interactively generated and recorded from kinesthetic teaching for the case of precise pointing. Successful generalization is verified on the iCub using a laser pointer attached to its hand.

1. Introduction

Learning of visually guided sensorimotor behaviors is an active field of research in cognitive robotics. In this regard, common approaches to sensorimotor learning relate sensory signals and motor commands, and the direction of the mapping plays an important role for defining the task of interest. A forward mapping transforms motor commands into sensory states, while an inverse mapping provides for a desired sensory state the motor commands to achieve this very state. The latter mapping typically is redundant and therefore not easy to learn globally. In visual servoing, however, the end-effector position of a robot (e.g. the gripper) is continuously updated based on visual information provided by cameras (see [1,2] for a survey), a procedure that demands a feedback loop and is feasible because locally an inverse model can be well applied.

Approaches for dealing with sensorimotor coordination tasks can coarsely be divided into three major groups: model-based, model-free and hybrid approaches. In model-based approaches [3,4] mathematical models of the robot and the cameras are derived. This approach typically involves separate stereo-matching algorithms for depth calculation, which in turn require precise calibration of the camera system and computationally expensive search for the best stereo match. Alternatively, model-free approaches involve learning of forward or inverse input–output mappings by means of supervised [5–7] or self-organizing neural networks [8–10]. Finally, hybrid approaches combine neural learning of the robot model with conventional mathematical modeling of the task of interest [11] or visual servoing through local corrections [12].

Reaching a particular location in 3D space is a widely studied sensorimotor learning problem [13–16]. *Pointing*, which corresponds to orienting one of the hands or arms, the head or the whole body in the direction of the object or event of interest, is much less studied than reaching. Nevertheless, pointing has been an important topic of research in cognitive science, especially in what concerns the development of infants' preverbal (i.e. prior to speech) communicative skills [17,18]. It is a skill that is fully acquired by children around the age of 12 months and that is crucial for the development of further social and language abilities [19].

Classically, infants are thought to point for two main reasons [19,20]. Firstly, they point when they want a nearby adult to do something for them (e.g. give them something). This is called *imperative pointing* and consists in extending the arm as if reaching for an object. It has been proposed that children use imperative pointing as a result of not being able to reach objects that are too far away to grasp. Secondly, infants point when they want an adult to share attention with them to some interesting event or object.

It is worth mentioning, however, that orienting the hand to an object too distant to each is a fundamental component of a variety

* Corresponding author.
E-mail addresses: alemme@CoR-Lab.Uni-Bielefeld.de (A. Lemme), anandalf@gmail.com (A. Freire), guilherme@deti.ufc.br (G. Barreto), jsteil@CoR-Lab.Uni-Bielefeld.de (J. Steil).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.12.040

of behavioral responses. It is relevant as protective reflex (e.g. when intercepting an object that comes into the agent's direction [21]) or to block something from the sight (e.g. a bright light, children's game *peek-a-boo*) [7,22]. Besides, any cognitive and interacting agent needs to be able to perform pointing either by moving arms or hands in active gesturing or by orienting the body or the head towards a relevant direction in the environment, e.g. during handshaking or when speaking with a partner.

To the best of our knowledge, few studies have addressed the acquisition of pointing skills by robots [22,23,7]. Shademan et al. [22] and Marjanović et al. [7] use visual information to compute error signals in a closed loop control scheme. Marjanović et al. [7] implement two mappings (from images coordinates to the eye motors, and then to the coordinates of arm motors) using radial basis functions networks. Shademan et al. [22] do not rely on neural networks to learn the desired mappings, but rather on the locally least squares-based Jacobian estimation method [24]. These authors have demonstrated how a robot can learn primitive skills and how to augment them. The proposed approach was validated with real uncalibrated camera mounted on the end-effector (eye-in-hand scenario) of a 4-DOF robot.

Of particular interest to this paper is the work of Doniec et al. [23]. These authors address the task of pointing in order to accelerate the learning of joint attention by an upper torso humanoid robot, called Nico. Joint attention is achieved when one person alerts another to an object or event by means of eye-gazing, pointing or other verbal or non-verbal indications. Their approach to joint attention learning required three phases.

First, a neural network is trained by a fast method for learning iterative reaching through motor babbling [6] in order to learn the forward model of the arm, $\dot{\mathbf{x}} = \mathbf{J}(\boldsymbol{\theta})\dot{\boldsymbol{\theta}}$, where $\boldsymbol{\theta} \in \mathbb{R}^6$ is the vector of joint angles, $\mathbf{x} \in \mathbb{R}^3$ is the vector of Cartesian coordinates of the end-effector, and $\mathbf{J}(\boldsymbol{\theta})$ is the corresponding 3×6 Jacobian matrix. Once trained, the desired arm displacement in joint angles can be computed by locally inverting the forward mapping

$$\Delta\boldsymbol{\theta} = (\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T(\mathbf{x}_{obj} - \mathbf{x}_{eff}), \tag{1}$$

where $\mathbf{x}_{obj}$ and $\mathbf{x}_{eff}$ are, respectively, the current positions of the object and of the end-effector.

Second, the forward model is used to produce imperative pointing, that is, the robot is commanded to approach an object that is out of reach. For this purpose, it has to stretch its elbow. However, by trying to do so, the robot arm movement will eventually stop as the elbow joint reaches singularity. The alternative found by Doniec et al. was to remove the elbow joints from the Jacobian matrix calculation at this point and continue iterative movement by simply using the shoulder joints. As a consequence, one should replace the original Jacobian matrix $\mathbf{J}_{3\times6}$ in Eq. (1) with $\mathbf{J}'_{3\times4}$, and the corresponding joint vector $\boldsymbol{\theta}_{6\times1}$ with $\boldsymbol{\theta}'_{4\times1}$. Finally, a second neural network is trained to associate head pose with the position of the object of interest.

It is worth emphasizing that, according to Doniec et al.'s approach, the neural forward model (and, by extension, the robot) is not explicitly trained to point to an object. The pointing skill, in this case, is a consequence of a clever adaptation of the forward model to cope with singularities arising when the robot stretches the elbow. Notwithstanding this fact, the resulting neural forward model is accurate enough to allow pointing without visual feedback.

In this paper we also deal with learning of pointing skills without resorting to visual feedback. However, we follow a different approach. We show that the first and second phases of Doniec et al.'s approach can be merged into a single phase in order to provide the robot with imperative pointing skills. For this purpose, we tackle this problem from the viewpoint of learning a sensorimotor mapping without visual feedback *and* without computing object's depth and 3D position, since this task does not require a completely accurate positioning of the robot's hand in order to give a correct indication of the direction of the object. Additionally, we take one step forward and show how to use kinesthetic teaching to improve robot's pointing accuracy considerably.

It is important to make the distinction clear between positioning the hand in the direction of an object without a precise orientation (a.k.a. imperative pointing) and precise pointing to an object. Imperative pointing has been previously approached in [25,23] and the present work can be understood as an attempt to improve the accuracy of imperative pointing even further through kinesthetic teaching. This distinction is crucial because it leads to different task representations and different requirements for training data. Notwithstanding this distinction, an important common feature of both tasks is that computation of 3D information (e.g. object's depth and 3D position) reveals unnecessary, since only the orientation along some direction needs to be adjusted, as illustrated in Fig. 1(c).

We aim at showing that both tasks, imperative and precise pointing, can be learned as a mapping from the object's pixel coordinates in the visual field directly to hand positions (or to joint angles) in an eye-to-hand configuration. This highly nonlinear mapping defines the pose and orientation of the robot's arm. To this aim, three state-of-the-art neural network paradigms (multilayer perceptron, extreme learning machine and static reservoir computing) are evaluated on data gathered from the humanoid robot iCub [26]. For the sake of data collection and

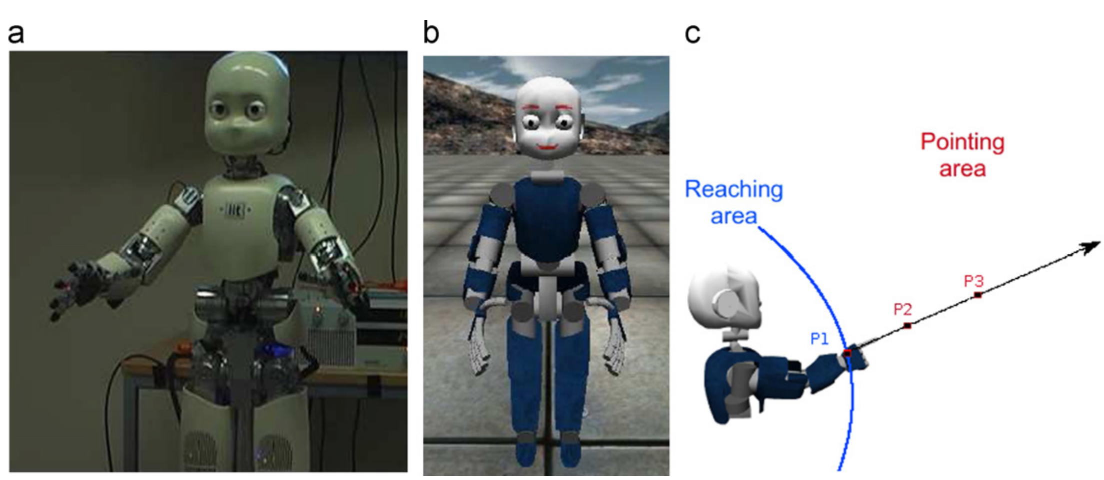

Fig. 1. (a) Humanoid robot iCub in CoR-Lab's installation and (b) its simulated version. (c) Illustration of pointing task characteristics.