Information Sciences 352–353 (2016) 79–97

Contents lists available at ScienceDirect

Information Sciences

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

# Grounding humanoid visually guided walking: From action-independent to action-oriented knowledge

Hendry Ferreira Chame*, Christine Chevallereau

*IRCCyN, Ecole Centrale de Nantes, CNRS, Nantes, France*

ARTICLE INFO

*Article history:*
Received 6 May 2015
Revised 8 December 2015
Accepted 25 February 2016
Available online 4 March 2016

*MSC:*
68T40
68T45
68T30

*PACS:*
Computing methodologies
Cognitive robotics & Computer vision
Computer vision problems
Vision for robotics

*Keywords:*
Cognitive robotics
Embodied cognition
Humanoid robotics
Ego-localization
Top-down visual attention
Robot vision

ABSTRACT

In the context of humanoid and service robotics, it is essential that the agent can be positioned with respect to objects of interest in the environment. By relying mostly on the cognitivist conception in artificial intelligence, the research on visually guided walking has tended to overlook the characteristics of the context in which behavior occurs. Consequently, considerable efforts have been directed to define action-independent explicit models of the solution, often resulting in high computational requirements. In this study, inspired by the embodied cognition research, our interest has focused on the analysis of the sensory-motor coupling. Notably, on the relation between embodiment, information, and action-oriented representation. Hence, by mimicking human walking, a behavior scheme is proposed and endowed the agent with the skill of approaching stimuli. A significant contribution to object discrimination was obtained by proposing an efficient visual attention mechanism, that exploits the redundancies and the statistical regularities induced in the sensory-motor coordination, thus the information flow is anticipated from the fusion of visual and proprioceptive features in a Bayesian network. The solution was implemented on the humanoid platform Nao, where the task was accomplished in an unstructured scenario.



## 1. Introduction

The automation of visually guided walking has arguably adopted in its infancy the so called cognitivist approach to artificial intelligence (AI), which, under the Cartesian dualist influence, has tended to look at physical and mental processes as belonging to different realms. Significant progress has been obtained from this view, though the gains are still distant from the sophistication observed in natural behavior.

Among the several challenges reported in the literature, one is undoubtedly to achieve reliable perception from noisy sensory data. Since the sensory input goes through a process of symbolization, and cognition involves computations over symbols, the physical context at which the latter emerged is no longer available to the cognitive process. In other words, in abstracting cognition from the context, information is inevitably lost.

* Corresponding author. Tel.: +33240376932.
*E-mail addresses:* hendry.ferreira-chame@irccyn.ec-nantes.fr, hendryf@gmail.com (H. Ferreira Chame), christine.chevallereau@irccyn.ec-nantes.fr (C. Chevallereau).

http://dx.doi.org/10.1016/j.ins.2016.02.053

To cope with the difficulties of perceiving the object while moving, several methodologies that flourished in the machine learning research have been employed (e.g. Markovian models, support vector machines, among others). These attempts have produced impressive results, although, by keeping intact the fundamental premise of decoupling between bodily and mental processes, they have relied on expensive resources in the form of context-free explicit models, knowledge databases, and intensive computation. Thus, the processing bottleneck has impacted the autonomy and the reactivity of the agent, and extraneous variables (i.e. unmodeled phenomena) have been controlled by adapting the scene to the task, which has compromised the generality of the solution.

In the last decades, a different perspective has been adopted to study natural behavior from the multi-disciplinary research on embodied cognition (EC), where knowledge representation is thought to be grounded in the physical interaction with the environment. Unlike the cognitivist approach, emergent behavior would not be explicitly represented nor planed in advance. The analysis of the sensory-motor coupling in natural tasks, from a dynamic system perspective, appears as a promising research direction that can provide more efficient, robust, and autonomous solutions.

However, adopting the EC methodology also poses important challenges to robotists, in particular, when fulfilling the requirements underlying the physical grounding hypothesis. Firstly, the autonomous evolution of the system, as it happens in natural beings, would ideally occur under a phylogenetic architecture that can modify itself. Secondly, the ontology of the system must ensure knowledge acquisition for diverse purposes, by fusing information from different sensory modalities. Lastly, enactment conditions the development of cognitive skills to the sensory-motor coupling and the interaction with the environment, thus knowledge acquisition is a slow process analogous to the natural one.

In view of the advantages and the challenges encountered in the aforementioned research paradigms, we have opted for an intermediate perspective, aiming at obtaining both generality and autonomy for applications in service robotics. Thereby, in this work we study the task of approaching and positioning in relation to visual stimuli. For this, we adopt the cognitivist assumption that human employs action-independent knowledge for localizing the object in the scene. But, as an embodied being, human can resort to contextual information to discriminate and perceive the object. Therefore, we explore action-oriented representations in the form of embodied features that capture bodily sensations emerging in the task. Furthermore, we examine the redundancies and the statistical regularities induced in the sensory-motor coordination for obtaining a more efficient perceptive processing. We describe a behavior scheme that includes a proposal for resources organization, and, through a case study, we consider the information fusion within a Bayesian network in charge of discriminating the object.

This document is organized as follows. In Section 2 some related contributions and challenges encountered from the cognitivist approach are discussed. Section 3 starts by the conceptualization of our research within the literature of embodied cognition, it then proceeds with the methodological analysis of the task where the behavior scheme is proposed and described. A case-study has been designed including simulations and experiments with the robot Nao, which is presented in Section 4. The results obtained are discussed in Section 5 along with the research perspectives. Finally, the conclusions of the study are given in Section 6.

## 2. Related work

Vision-based locomotion is a challenging task for walking robots. Unlike natural beings, which are in possession of extremely sophisticated sensory organs, the vast majority of the research in robotic vision has been carried out with quite inferior equipment, usually employing general purpose cameras. Moreover, the body structure and the actuation system utilized is much less stable, fine, and accurate, when compared to the natural musculo-skeletal system. In view of such limitations, several difficulties have been reported when processing images captured on-board, such that illumination noise and motion blur have affected the localization. Thereby, some studies have resorted to capture information from external cameras (e.g., Lewis & Simo [19], and Michel et al. [21]). Unfortunately, the robot may occlude the sensors, thus compromising the solution of the task. Other sensor modalities have been investigated as well. In a work by Allen et al. [1] a support vector machine was trained to recognize the body intrinsic noise, as captured from external microphones. Though, occlusions also affected the task, and more importantly, unmodeled noise disturbed the localization. Despite the advantages in the use of extra-corporeal sensors, in our opinion the strategy has two major drawbacks: (a) it neglects the humanoid body metaphor, and (b) it affects the autonomy of the agent, since having sensors in the scene is a form of rigorous control over extraneous variables.

On-board solutions have been proposed under the visual servoing (VS) framework (see Chaumette & Hutchinson [6] and Corke [9]), which relates variations on image features to the spatial instantaneous velocity of a flying (dismembered) camera. A study by Dune et al. [11] has considered a monocular vision task with the robot HRP2. Given the walk style of the robot, the solution involved the cancellation of the oscillatory contribution to the control signal (also called the *sway motion*). In order to handle the image noise, the tracking technique by Comport et al. [8] was employed. The algorithm requires of the 3D model of the object, which is projected as virtual contours on the actual image. The estimation for the object pose in relation to the camera is obtained from local (heuristic) search. However, the results of the technique are based on premises that may be a bit restrictive. Firstly, a realistic model of the object is required, which prevents the reusability of the solution to similar objects. Secondly, the initial observation of the localization has to be very precise, so assisted by a manual procedure. Lastly, the projection of the model on the image requires of accurately estimating the spatial evolution of the camera. Since the later is considered to be dismembered, visual odometry was used. Thus, a relatively