ICSC 2006
Dynamics in Spatial Interactions

ORAL PAPER

A biologically plausible robot attention model, based on space and time

Anna Belardinelli • Fiora Pirri

Keywords Human–robot interaction · Vision-based attention · Spatial cognition

In this paper we describe a biological inspired approach to robot attention, developed on the basis of several experiments mapping human gaze search onto robot behaviours, as described in previous works (see Belardinelli et al., 2005, 2006). In particular in this paper we show how a gaze search process can be defined as a Markov random process weighetd with utility functions. The role of utilities is to account for the optimization of the visual endeavours.

Visual attention has been intensively investigated in last decades in order to understand how humans orient their gaze. Strategies employed in visual attention account for the ability in human vision to detect interesting spots in the visual field. When observing a scene the ability to focus immediately on salient regions is crucial in many search or surveillance tasks and furthermore it clearly speeds up object recognition.

The human skill of selective attention deployment has been deeply researched in order to determine which mechanisms allow us to capture meaningful details at a glance, without needing to process whole regions of the visible scene. The model that gained most credit in the past years is due to Treisman (Treisman and Gelade, 1980), according to whose theory focused attention is driven by perception of several separable features, such as intensity, color, shape, edge orientations, and conjunctions of features. Therefore mechanisms for feature extraction and recombination have been implemented in artificial systems as well, according to computational models for the control of bottom-up attention such as the construction of saliency maps (Itti and Koch, 2001; Niebur et al., 2001).

Gaze orienting is fundamental in guiding spatial attention in visual search tasks. During observation of a scene, attention moves through a sequence of gaze shifts and fixations. While shifting the fovea, visual information is not processed, being rather filtered out until a fixation occurs, allowing to focus on a salient location selected in the previous fixation. This phenomenon is known as change blindness and it accounts for the lack of change detection during eye movements while observing images or real world scenes (Simons and Levin, 1997; Rensink et al., 2000). In these moments focused attention is not applied and therefore our mind cannot build a coherent and detailed representation of the world, tending instead to assemble a kind of sampled representation of perceptually relevant locations and features. Similarly, fast translational or rotational movements of cameras mounted on robot heads can be left out in the processing, since corresponding frames result to be blurred and therefore they cannot deliver meaningful information. We focused thus on fixations and researched underlying mechanisms in order to infer a model of attention that could be implemented on a robotic platform. Only in recent years studies in the field of Cognitive Sciences began to investigate whether depth could play a role as a further feature in projecting gaze shifts and fixations in a three-dimensional space. Theeuwes et al. (1998) reported that grouping mechanisms help segregating contiguous objects, focusing attention on a subset of the visible elements. In particular they showed that attention can select a particular depth plane determined by binocular disparity, even if other relevant features on different planes can act as distractors if identical to the target ones. Furthermore, several authors (among them, Maringelli et al., 2001; Couyoumdjian et al., 2003)

A. Belardinelli (✉) • F. Pirri
Dipartimento di Informatica e Sistemistica,
Università di Roma "La Sapienza", Italy
E-mail: belardinelli@dis.uniroma1.it

suggested that different spatial representations are used for near and far space and relative information is probably processed separately, since near space is more related to motor tasks while far space is mainly perceptual. Evidence of such a different representation is a greater reaction time obtained when shifting the fovea from peripersonal space (about 1 m) to extrapersonal space. This performance decrease could be justified by both a cognitive effort and an ocular effort. More specifically, Previc (1998) proposes four behavioural realms: peripersonal, focal extrapersonal, action extra-personal and ambient extra-personal.

The third one is the only one that can shift between near and distal space along with foveation and its function is mainly devoted to searching and recognizing objects and visual targets. The latter task, particularly, requires attentional resources and visual memory.

Depth as source of visual attention has been computationally modelled by Ouerhani et Hügli (2000), who integrated it as ulterior feature in the construction of the saliency map. Frintrop et al. (2004) applied visual attention to range images in order to perform efficient object recognition.

In this paper we present an experiment based on the above mentioned considerations. Our goal was to find out scanning strategies in task-driven attention. This achievement shall lead to automatic generation of likely scanning paths in terms of a sequence of rotation angles efficiently performable by a pan-tilt unit, on which a stereo camera is mounted.

Allegedly visual search progresses aggregating contiguous objects in terms of each of the three dimensions and passes from a clique to another minimizing the overall effort and maximizing utility according to the task. Once recorded visual data from the subject point of view and the pan and tilt angles of his head, performed during fixations, we computed velocity of gaze shifts and paths between fixations on a mosaic scene. Further we have mapped these shifts onto a

Markov random process which is weighted by utility functions. Thus transitions denote optimized gaze shifts.

References

Belardinelli A, Pirri F, Carbone A (2005) Bottom-up gaze shifts and fixations learning, by imitation. University of Rome ”La Sapienza”, Tech. Rep. 191205, December 2005.

Belardinelli A, Pirri F, Carbone A (2006) Spatial discrimination in task-driven attention. University of Rome ”La Sapienza”, Tech. Rep. 210306, March 2006

Brown M, Lowe D (2003) Recognising panoramas. Proceedings of the 9th International Conference on Computer Vision, vol. 2, pp. 1218–1225

Couyoumdjian A, Nocera FD, Ferlazzo F (2003) Functional representation of 3d space in endogenous attention shifts. The Quarterly Journal of Experimental Psychology 56A(1):155–183

Fintrop S, Nüchter A, Surmann H (2004) Visual attention for object recognition in spatial 3d data. Proceedings of the 2nd International Workshop on Attention and Performance in Computational Vision (WAPCV), vol. 3368, pp. 168–182

Itti L, Koch C (2001) Computational modeling of visual attention. Nature Reviews Neuroscience 2(3):194–203

Koch C (2004) Models of motion perception: Introduction and the reichardt correlation model. Notes of the Course ”Vision: from computational theory to neuronal mechanisms CNS/Bi/EE 186 (4-4-4), 2004, caltech.

Maringelli F, McCarthy J, Steed A, Slater M, Umiltà C (2001) Shifting visuo-spatial attention in a virtual three-dimensional space. Brain Research Cognitive Brain Research 10:317–322

Mozer M, Vecera S (2005) Object- and space-based attention. In: Itti L, Rees G, Tsotsos J (eds) The encyclopedia of the neurobiology of attention. Elsevier Press, pp. 130–134.

Niebur E, Itti L, Koch C (2001) Controlling the focus of visual selective attention. In: Hemmen LV, Domany E, Cowan J (eds) Models of Neural Networks IV, Springer Verlag.

Ouerhani N, Hügli H (2000) Computing visual attention from scene depth. Proceedings of the 15th International Conference on Pattern Recognition, vol. 1, pp. 375–378.

Previc F (1998) The neuropsychology of 3-d space. Psychological Bulletin 124:123–164.

Rensink R, O'Regan J, Clark J (2000) On the failure to detect changes in scenes across brief interruptions. Visual Cognition 7:127–145.

Simons D, Levin D (1997) Change blindness. Trends in Cognitive Sciences 1:261–267.

Theeuwes J, Atchley P, Kramer F (1998) Attentional control within 3-D space. Journal of Experimental Psychology: Human Perception and Performance 24(5):1476–1475.

Treisman A, Gelade G (1980) A feature-integration theory of attention. Cognitive Psychology 12:97–136.