VISUAL ATTENTION IN 3D SPACE
Using a Virtual Reality Framework as Spatial Memory

M. Zaheer Aziz and Bärbel Mertsching
GET Lab, Universität Paderborn, 33098 Paderborn, Germany

Keywords: Visual attention, Virtual reality simulator, 3D space, Spatial memory, Robotic vision.

Abstract: This paper presents a conceptual framework to integrate a spatial memory, derived from a 3D simulator, with a visual attention model. The proposed system is inspired from brain research that explicitly accounts for the use of spatial memory structures in intelligent object recognition and navigation by humans in the three dimensional space. The experiments presented here extend the capability of visual attention modeling to work in 3D space by connecting it to simulated maneuvers in virtual reality. The introduction of this concept opens new directions for work to reach the goal of intelligent machine vision, especially by mobile vision systems.

1 INTRODUCTION

Spatial memory is an important part of the human brain that is responsible to store three dimensional structures of environments, landmarks, and objects (Moscovitch et al., 2005). The stored environments in this memory help during navigation through known routes and maps like walking through corridors or driving through streets of everyday routine. The object data in this memory is also useful for collision avoidance in an automatic way, for example during car driving a decision to overtake a long vehicle is made quite involuntarily after estimating the vehicle's length using its memorized 3D model recalled by looking at its rear only.

Construction of scenes and objects in the spatial memory has a close relation with visual attention. Research on vision systems of primates reveals that the natural vision views and recognizes objects (especially large ones) by fixating on their constituent parts rather than perceiving them as a whole. This is managed by the visual attention mechanism that selects salient portions of objects (or scenes) and focuses upon them one after the other. In artificial vision systems, such selective viewing can help to filter out redundant and non-relevant data.

This paper presents design of a memory driven vision system that integrates artificial visual attention with a 3D spatial memory. A robotic vision system able to perform overt visual attention will focus on salient objects or their visible parts and use this visual information to activate the complete 3D model of the object from the spatial memory. Utilization of a virtual reality simulation framework is proposed as storage mechanism for the learned environments and objects. Such a proposal not only leads to knowledge driven machine vision but introduces a very useful utility of 3D simulation engines as well.

Literature in psychophysics has described the role of spatial memory and its role in navigation, object recognition, self localization, and intelligent maneuvers. The work presented in (Oman et al., 2000) shows capability of learning three dimensional structure not only by looking at the target object or environment from different directions but by imagining to view it from these orientations as well. The ability of the brain to visualize a scene from an orientation in space that was not actually experienced by it is shown in (Shelton and Mcnamara, 2004). A relation between visual attention and spatial memory is established in (Aivar et al., 2005) with a conclusion that a detailed representation of the spatial structure of the environment is typically retained across fixations and it is also used to guide eye movements while looking at already learnt objects. Formation of object representation by human vision through snapshots taken from different view angles is discussed in (Hoshino et al., 2008) and it is suggested that such procedure is followed only for the objects under visual attention while the unattended scene may be processed as a 2-D representation bound to the background scene as a texture.

The natural visual attention mechanism rapidly analyzes the visual features in the viewed scene to determine salient locations or objects (Treisman and

Zaheer Aziz M. and Mertsching B. (2009).
VISUAL ATTENTION IN 3D SPACE - Using a Virtual Reality Framework as Spatial Memory.
In *Proceedings of the 6th International Conference on Informatics in Control, Automation and Robotics - Robotics and Automation*, pages 471-474
DOI: 10.5220/0002247204710474

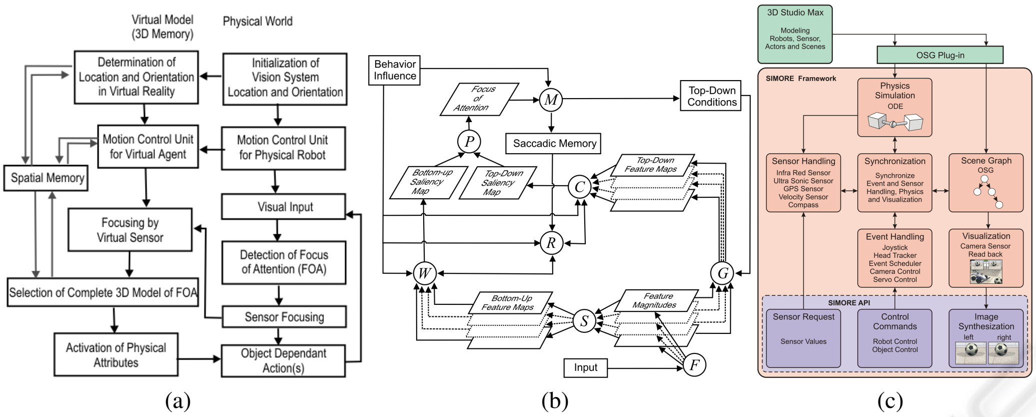

Figure 1: (a) Proposed model for integrating visual attention with 3D spatial memory. (b) Architecture of the region-based attention model used as the selection mechanism for regions of interest in real world (Aziz and Mertsching, 2007). (c) Architecture of the simulation framework SIMORE that is used as spatial memory of the mobile robot (Kutter et al., 2008).

Gelade, 1980)(Wolfe and Horowitz, 2004). After attending the current focus of attention a process of inhibition of return (IOR) (Cutzu and Tsotsos, 2003) suppresses the attended location so that other less salient objects may also get a chance to be attended. Existing attention models have shown success in selection and inhibition of return in two dimensional view frames. The natural visual attention, on the other hand, works in three dimensional world despite its perception of a two dimensional projection on the retina. In order to make advancement in the state-of-the-art, the model proposed in this paper attempts to integrate a spatial memory with the visual attention process in order to extend the scope of attention and IOR towards 3D.

2 PROPOSED MODEL

The objective of the current status of the proposed model is to associate a spatial memory to the visual attention process and activate the three dimensional structure of the attended object for use in decision making procedures. Figure 1(a) shows the architecture of the proposed model. As this model involves visual attention and a spatial memory to perform its task, the design of the two involved components is also discussed here.

The architecture of the attention model is presented in figure 1(b). The primary feature extraction function F produces a set of regions (Aziz and Mertsching, 2006) and associates feature magnitudes of color, orientation, eccentricity, symmetry, and size with each region. Computation of the bottom-up saliency using rarity criteria and bottom-up contrast of region features with respect to its neighborhood is performed by the group of processes S (see (Aziz and Mertsching, 2008a) for details) whose output is combined by the procedure W. The function G considers the given top-down conditions to produce fine grain saliency maps that are combined by the function C. The function P combines the saliency maps into a master conspicuity map and applies a peak selection mechanism. Inhibition of return (IOR), denoted by R, suppresses the already attended location(s) using a saccadic memory. Explanation of the internal steps and functions of this attention model can be seen in (Aziz and Mertsching, 2007) and (Aziz and Mertsching, 2008b).

The simulation system used as spatial memory is a 3D robot simulation framework SIMORE (SImulation of MObile Robots and Environments) developed in our group (Mertsching et al., 2005) (Kutter et al., 2008). Figure 1(c) shows its architecture with its major components exposed. The core of simulator is based upon the open source library Open Scene Graph (OSG) (Burns and Osfield, 2004) which is a hierarchical graph consisting of drawable meshes in a forward kinematic order. The physics simulation component represents the dynamic engine for collision detection and force based physics using Open Dynamics Engine (ODE) which is an open source library for simulating rigid body dynamics (Smith, 2009). Extensions for sensor and meta information have been done using specialized nodes for these purposes. These enhancements of the existing scene graph allows to rely on an existing library and enables the system to import and export from available 3D modeling software such as 3D Studio Max.

The simulation framework SIMORE has the ability to maneuver a simulated robot in the virtual environment by driving, taking turns, rotating its camera head, and turning other movable sensors by control commands from an external computer program.

The master control program, for example a computational model of visual attention, manipulates the physical robot in the real world and maneuvers the virtual agent in the simulation framework. Therefore the simulated robot can act as an agent of the real platform. The readings from the simulated sensors are obtained according to their position and direction in the virtual environment while they are aligned with the physical ones. Such a mechanism allows the vision system to recall a complete 3D model of an attended object even by looking at only a part visible from the current view angle.

According to the proposed model, the vision system selects an object to attend from its viewed scene and performs overt attention using its pan-tilt camera. Whenever the vision system finds an object of interest (or its part) it consults the spatial memory by looking at the virtual scene through its simulated camera. The visible features of the attended object in the real camera view, information about location of robot (in real and virtual environment), and angles of camera direction can guide to pick the right object from the virtual scene matching with the object under the focus of attention. Knowing the identity of the modeled object its complete set of attributes will be loaded into the working memory. Using the current status of our experimental platforms we demonstrate a spatial inhibition of return on the previously focused object(s) so that they remain inhibited even after the robot motion in 3D space.

3 CURRENT SYSTEM STATUS

In the current status, the interface between the visual attention module and the simulation engine is successfully established and work is underway to enable the synchronized selection of the attended objects from the spatial memory. We are able to present here the expected results from the proposed model with manual configurations in the synchronization part.

Figure 2 shows the arrangement in which the vision system appearing at the right side of the subfigure (a) drives forward while searching for red objects (the ball and the robot in the left-bottom corner). Figure 2(b) shows the global camera view of the arrangement in the simulation framework.

Results of the first attempt of attentional search are provided in figure 3 in which subfigure (a) shows camera view of real robot and (b) is the view from the aligned simulated camera. Figure 3(c) shows the region selected by the attention mechanism and (d) shows the camera view after bringing the ball into center of view frame. Figure 3(e) shows the virtual camera view after aligning it with the current status of the real camera. Figure 3(f) demonstrates the selected 3D model from the spatial memory whose activation not only exposes its hidden portion to the vision system but its volume information as well.

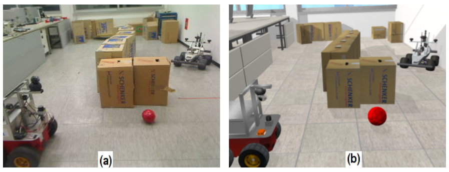

Figure 2: Initialization of the robotic platform and alignment of the agent in terms of location and orientation. (a) Real robot at initialization (b) Global view in virtual reality.

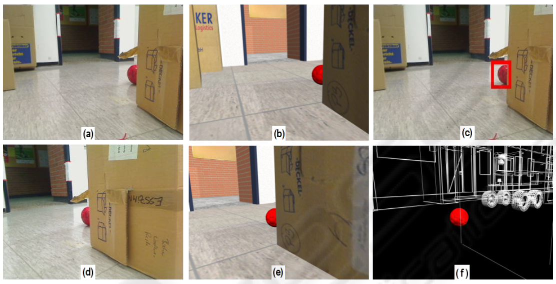

Figure 3: Results of first attempt of attention. (a) Camera view of real robot (b) Camera view of virtual robot (c) Focus of attention detected by real robot (d) Overt attention to object of interest (e) View of the simulated environment (spatial memory) after synchronized rotation of the simulated camera (f) 3D model of first FOA activated (inactive objects are shown in wireframe).

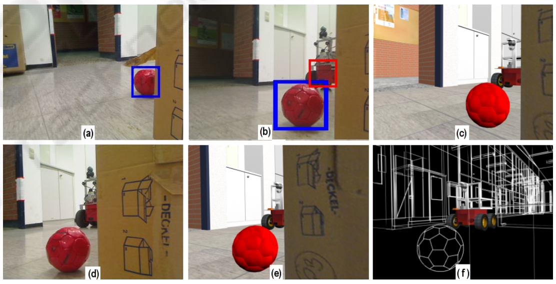

Figure 4: Results of second attempt of attention. (a) Camera view of real robot after moving ahead (previous FOA inhibited). (b) After moving further ahead the second target found (previous FOA still remains inhibited). (c) Aligned camera view of virtual robot. (d) Overt attention to second object of interest. (e) Orientation synchronization of simulated camera. (f) 3D model of second FOA activated.

Figure 4 shows results of the second attempt of attention after moving ahead subsequent to attending the first target (the ball). The first focus of attention remains under inhibition of return during this attempt (shown by dark (blue) rectangle). In subfigure (a) the vision system moves ahead but no new object of interest comes into view, whereas the already attended ball remains under inhibition of return due to the use of spatial memory even when its 2D location in the view frame and size has changed with respect to its last attended instance. Figures 4(b) shows attention on the second target whereas the ball is still under inhibition. The subfigure (c) shows the view in the spatial memory after aligning its sensors to the real world. Figures 4 (d) and (e) demonstrate views in the real world and the simulation framework after overt attention on the second target while the activated model of the attended object (the robot) can be seen in figure 4(f).

4 DISCUSSION

A conceptual framework of integrating a spatial memory with the vision procedures has been presented here and the feasibility of using a 3D simulator as a spatial memory is introduced. The area of integration of vision and spatial memory, their interaction, and cooperation needs to be explored further as there are many issues to be resolved. For example the physical system can gain error of localization and orientation over time due to inaccuracy in its sensors and wheel slippages that lead to synchronization problem between the real robot and its agent.

Using the spatial memory can increase the potentials of vision in 3D world and intelligence in autonomous decision making. Work needs to be done for handling further complexities in the scenario. For example, activation of the 3D models of objects will be more useful when positions of movable objects are not known in advance. Using the visual information from the camera, the robot could recognize an object and activate its whole model there. This can be helpful in navigation planning while roaming in known environments in which a bunch of known objects are moving around or located at arbitrary locations, for example 3D models of different types of vehicles could be used for intelligent autonomous drive on a known road map.

ACKNOWLEDGEMENTS

We gratefully acknowledge the funding of this work by the German Research Foundation (DFG) under grant Me 1289/12-1(AVRAM).

REFERENCES

Aivar, M. P., Hayhoe, M. M., Chizk, C. L., and Mruczek, R. E. B. (2005). Spatial memory and saccadic targeting in a natural task. *Journal of Vision*, 5:177–193.

Aziz, M. Z. and Mertsching, B. (2006). Color segmentation for a region-based attention model. In *Workshop Farbbildverarbeitung (FWS06)*, pages 74–83, Ilmenau - Germany.

Aziz, M. Z. and Mertsching, B. (2007). Color saliency and inhibition using static and dynamic scenes in region based visual attention. *Attention in Cognitive Systems, LNAI 4840*, pages 234–250.

Aziz, M. Z. and Mertsching, B. (2008a). Fast and robust generation of feature maps for region-based visual attention. *Transactions on Image Processing*, 17:633–644.

Aziz, M. Z. and Mertsching, B. (2008b). Visual search in static and dynamic scenes using fine-grain top-down visual attention. In *ICVS 08, LNCS 5008*, pages 3–12, Santorini - Greece. Springer.

Burns, D. and Osfield, R. (2004). Tutorial: Open scene graph. In *Proceedings Virtual Reality*, pages 265–265.

Cutzu, F. and Tsotsos, J. K. (2003). The selective tuning model of attention: Psychophysical evidence for a suppressive annulus around an attended item. *Vision Research*, pages 205–219.

Hoshino, E., Taya, F., and Mogi, K. (2008). Memory formation of object representation: Natural scenes. *R. Wang et al. (eds.), Advances in Cognitive Neurodynamics*, pages 457–462.

Kutter, O., Hilker, C., Simon, A., and Mertsching, B. (2008). Modeling and simulating mobile robots environments. In *3rd International Conference on Computer Graphics Theory and Applications (GRAPP 2008)*, Funchal - Portugal.

Mertsching, B., Aziz, M. Z., and Stemmer, R. (2005). Design of a simulation framework for evaluation of robot vision and manipulation algorithms. In *International Conference on System Simulation and Scientific Computing*, Beijing-China.

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., McAndrews, M. P., Levine, B., Black, S., Winocur1, G., and Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *Journal of Anatomy*, pages 35–66.

Oman, C. M., Shebilske, W. L., Richards, J. T., Tubré, T. C., Bealli, A. C., and Natapoffi, A. (2000). Three dimensional spatial memory and learning in real and virtual environments. *Spatial Cognition and Computation*, 2:355–372.

Shelton, A. L. and Mcnamara, T. P. (2004). Spatial memory and perspective taking. *Memory & Cognition*, 32:416–426.

Smith, R. (last accessed March 2009). Open Dynamics Engine, Version 0.8. http://www.ode.org.

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

Wolfe, J. M. and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews, Neuroscience*, 5:1–7.