A Cognitive Model of Gaze Following and Object Permanence

L. A. Munoz and Markus Vincze
Automation and Control Institute
Gusshausstrase 27-29/376
1040 Vienna, Austria
{munoz, vincze}@acin.tuwien.ac.at

Abstract

The intention of this work is to propose a new cognitive model able to cope with occlusion in Hide & Seek tasks. Object permanence is the awareness that objects continue to exist even if they become occluded or out of sight and such a property is added by introducing a memory model that allows an agent not to be completely disturbed by the presence of an occluder. The key idea is to detect, in advance, the occlusion event. The main challenge for this approach is to achieve the same flexibility as observed in the human cognitive capabilities. While learning OP might be another challenge, we propose to use OP capability to achieve advanced gaze following behaviour.

Children start playing "Hide and Seek" even before they arrive to talk (Trafton et al., 2004). One cognitive capability shown during such a play is known as object permanence (OP). OP is the awareness that objects continue to exist even if they become occluded or out of sight. A further goal is to use computer vision and robotics to better understand the main cognitive properties of OP, e.g. Intelligent Survilliance(Dee and Hogg, 1995).

Spelke (Baillargeon et al., 1985) showed that infants learn the OP capability between months four and six. Gredeback (Gredeback, 1994) demonstrated how four-months old infants can predict the reappearance of a moving object after 660 ms of non visibility. Those studies suggest the existence of an accurate spatio-temporal representation.

A very good example of modelling the gaze following problem in between a Care Giver (CG) and an infant has been presented by Lau and Triesch (Lau and Triesch, 2005). Their model arrives to learn a way to overcome the "Butterworth error" (Butterworth and Jarrett, 1991) (the error related with a possible distraction *before* the infant joined the CGs attention) as behavioral experiments have confirmed in infants up to 18 months old. They proposed an Infants Agent Architecture based on a Bayesian approach composed by three components: a vision system for the perception of the object and saliencies; a system for determining both the care givers head pose and line of sight; and a memory and action selection system for shifting gaze.

The intention of this work is to propose a new cognitive model that is also able to cope with occlusion that could appear *after* the joint attention of CG and infant is established (see Fig. 1).

OP is added by introducing a memory model that allows the infant not to be completely disturbed by the presence of an occluder. The key idea is to detect, in advance, the occlusion event. If the visual area between the infant and the object of interest is occluded - *after* the joint attention is done - the memory level and focus of attention will make the infant gaze towards another salient object.

The main challenge for this approach is to achieve the same flexibility as observed in the human cognitive capabilities. While learning OP might be another challenge, we propose to use OP capability to achieve advanced gaze following behaviour.

We are conducting the gaze learning for a task-specific guidance of visual attention. When the first sub-task (i.e. fix attention on the target object) is achieved, the model increases the objects saliency and allows the infant agent to learn how to follow coherent spatio-temporal reasoning. Due to this kind of task-biasing strategy (also refer to (Navalpakkam and Itti, 2005)), the saliency of the target makes it more likely to show OP capability. If an occluder appears and even if the CG can be attracted to change attention, the infant will maintain the target object in mind.

Once the memory creates the conditions for the infant to show OP, she can either wait until the object reappears or the occluder is removed. In this way the infant agent "knows" the fact that the object of interest is behind the occluder and can retrieve this information to act according to the task, i.e., change position to obtain direct gaze at the object again.

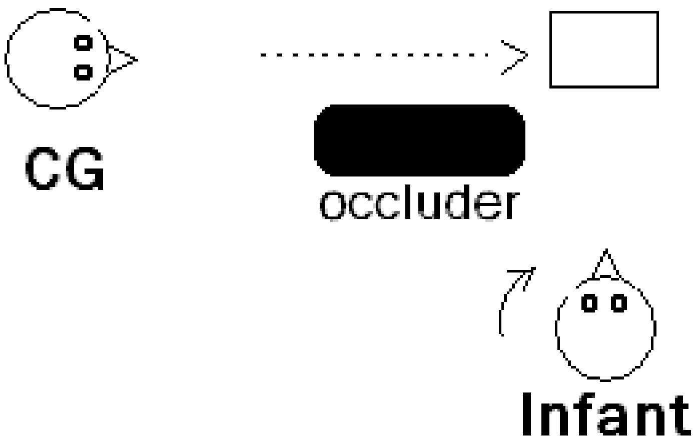

Figure 1: Graphical Architecture

Presently we are extending the methodology to have a 3D Bayesian approach by adding other knowledge representations (associative rules and connectionism) and implement them in our vision and mobile robot system.

For the 2D to 3D extension we are integrating hand, face and object localisation and tracking. We expect such an integration will provide sufficient empirical results to extend this approach towards a more complex theory of Experimental Object Permanence.

Acknowledgements

This work was supported by the Austrian Science Foundation FSP/JRP "Cognitive Vision", grant S9101-N04.

References

Baillargeon, R., Spelke, E. S., and Wasserman (1985). Object permanence in five-month-old infants. *Cognition*, 20:191–208.

Butterworth, G. E. and Jarrett, N. (1991). What minds have in common in space: Spatial mechanisms serving joint visual attention in infancy. *British Journal od Developmental Psychology*, 9:55–72.

Dee, H. M. and Hogg, D. C. (1995). On the feasibility of using a cognitive model to filter surveillance data. In *Intl. Conference on Advanced Video and Signal based Surveillance*, Como, Italia.

Gredeback, G. (1994). *Infants Knowledge of Occluded Objects: evidenc from early spationtemporal Representations*. PhD thesis, Uppsala.

Lau, B. and Triesch, J. (2005). Learning gaze following in space: acomputational model. In VDI-Verlag, (Ed.), *Proc. 3rd Workshop on Self-Organization of AdaptiVE Behavior*, volume 10, Germany, Ilmenau. Fortschritt-Berichte VDI.

Navalpakkam, V. and Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45:205–231.

Trafton, J, G., Schultz, A. C., Perzanowski, D., and Bugajska, M. D. (2004). Children and robots learning to play hide and seek. Submitted to Cognitive Systems Research.