A Developmental Approach to Dynamic Scene Understanding

Christian Balkenius & Birger Johansson
Lund University Cognitive Science
Kungshuset, Lundagård
S-222 22 Lund, Sweden
christian.balkenius@lucs.lu.se

To accurately track a moving target despite delays in the visual processing, it is necessary to predict how the target will move in the future and to direct the gaze toward the anticipated location. We have addressed this problem within a simulated visual scene were a ball is dropped from a random location and subsequently bounces or rolls on different objects before it comes to rest. A developmental system has been designed that attempts to track the location of the ball as closely as possible despite a delay of 250 ms from the the visual input to the control of gaze. The system is an extension of an earlier model of gaze development (Balkenius and Johansson, 2005).

A standard Kalman filter-based tracking framework is used that gradually develops a set of parallel predictive models of the environment. The use of multiple models is similar to the MOSAIC architecture (Wolpert et al., 2003). These models are used to predict the future position of a target object, and by generalizing over many observations, the system can implicitly extract the laws of Newtonian physics. The different models handle such cases as free fall, bounces and rolling, and the visual context is used to select between the different models. The model extends earlier work in this area in that it learns the mapping from visual features in the scene to the selection of the correct dynamic model.

When something unexpected happens to the tracked object, the cause is sought for in the visual scene. Elements in the scene are parsed and recognized to the extent that they contribute to the prediction of target motion. As a consequence, the model will learn to recognize objects that influence the movement of the ball, but will ignore others. The learned relations between objects in the visual scene and their influence on the target results in a basic understanding of the visual scene. This scene understanding is highly task specific.

A critical role is played by deviations from anticipated motion. Such unexpected motion triggers learning of new models that explain how the motion is influenced by different visual features. For example, the visual edges of a surface will make the ball bounce and change direction in a particular way. The models will adapt to include an increasing number of such environmental relations until the dynamics of the scene is fully covered. The processes combines the development of continuous dynamic models with the gradual inclusion of relevant contextual factors. An interesting aspect of the mechanism that creates new models when deviations from expectations are to large is that it shields earlier learning from forgetting. This is a fundamental property of any developmental system.

We believe that this model constitutes a good candidate for the understanding of the development of naive physics. The model also has implications for such phenomena as habituation and object constancy and the notion of objecthood. In the extension, an understanding of the processes involved in the movement of passive objects may lay the ground for the recognition of other classes of causal and noncausal motion.

Acknowledgements

This work was supported in by the EC funded project MindRaces, FP6-511931.

References

Balkenius, C. and Johansson, B. (2005). Event prediction and object motion estimation in the development of visual attention. In *Proceedings of the Fifth International Conference on Epigenetic Robotics*. Lund Univeristy Cognitive Studies, 123.

Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Phil. Trans. R. Soc. Lond. B*, 358(593–602).