Berthouze, L., Kaplan, F., Kozima, H., Yano, H., Konczak, J., Metta, G., Nadel, J., Sandini, G., Stojanov, G. and Balkenius, C. (Eds.)
Proceedings of the Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems
Lund University Cognitive Studies, 123. ISBN 91-974741-4-2

Event Prediction and Object Motion Estimation in the Development of Visual Attention

Christian Balkenius **Birger Johansson**
Lund University Cognitive Science
Kungshuset, Lundagård
222 22 LUND, Sweden

Abstract

A model of gaze control is describes that includes mechanisms for predictive control using a forward model and event driven expectations of target behavior. The model roughly undergoes stages similar to those of human infants if the influence of the predictive systems is gradually increased.

1. Introduction

At what age do infants become able to anticipate the motion of visual stimuli and use these predictions to control eye-movements? There are two lines of research that have investigated this question. Both have their origin in Piaget's (1937) studies of infants that looked at a toy train moving through a tunnel but differ in how they interpret the situation.

According to Piaget, the child is able to predict that the train that disappears at one end of the tunnel will appear at the other end. This can either be explained by a tracking mechanism that continues to track the motion of the train when it has disappeared, or as form of event learning where the child learns to predict that the disappearing train predicts the subsequent reappearance of the train. There exits evidence for both types of mechanisms. For example, Wentworth amd Haith (1998) found that three-month-old infants could learn spatiotemporal expectations. Other researchers have mainly studied the ability of infants to smoothly track a moving target using predictive models of the motion (von Hofsten & Rosander & 1997, Rosander & von Hofsten, 2004).

Our aim is to design a control system for the visual system of a humanoid robot that can learn and use both continuous models of the behavior of target objects and discrete event based models. The goal of this paper is to lay the foundation for such a system and relate it to the development of visual attention in infants.

Although there are many types of eye movements that are of use to a robot including the vestibulo-ocular reflex and the optokinetic reflex (Shibata et al., 2001), we will only consider smooth pursuit and saccade motions here.

Smooth pursuit occurs when the eyes track a moving target with a continuous motion, which ideally is centered directly on the target and makes the image of the target stationary on the retina. Smooth pursuit movements cannot normally be generated without a moving stimulus although they can appear without a moving stimulus a short moment before the target is expected to appear (Wells & Barnes, 1998).

Smooth pursuit is complicated by the fact that the initial visual processing in the human brain delays the stimulus by approximately 100 ms before it reaches the visual cortex (Wells & Barnes, 1998, Fukushima et al., 2002). If smooth pursuit movements were solely controlled by the position error on the retina, the eye would constantly lag the target.

To overcome this problem, the brain makes use of prediction (Deno et al., 1995, Mehta & Schaal, 2002, Poliakoff, Collins & Barnes, 2004). Because eye control is based on predicted target location rather than the actual target position which is not yet known, it is possible for the gaze to overshoot when the target disappears unexpectedly or changes direction. This does not happen when the disappearance of the target is controlled by the subject, for example by a button (Stork, Neggers & Müsseler, 2002). In this case, the gaze velocity slows down before the target disappears which shows that their expectations control the velocity of the smooth pursuit. Subjects are also able to learn to anticipate the velocity a targets will have when it appears, and in the case of several targets, subjects can produce predictive eye movements of approariate velocity when one of the targets are cued (Poliakoff, Collins & Barnes, 2004).

Infants as young as one month can exhibit smooth pursuit, but only at the speed of 10 degrees/s or less and with low gain (Roucoux et al., 1983). A three month old infant does not follow if a target abruptly changes its direction of movement. Instead it continues in the original direction for a quarter of a second before adjusting its tracking (Aguiar & Bailargeon, 1999). However, at five months of age, the infant

This work was supported by the EU project MindRaces, FP6-511931.

learns the abrupt turn and its lag is reduced. (von Hofsten & Rosander 1997). The ability to smoothly track a target thus develops very rapidly, and at five month of age it approaches that of adults.

Before the infant can use smooth pursuit, it follows moving targets using small saccade movements that rapidly moves the gaze from one position to another (Dayton & Jones, 1964). As the smooth pursuit system develops, these saccades become less frequent to moving targets but are still used to catch up if the lag becomes too large.

Of particular interest is how infants behave when the target disappears, for example behind an occluder. Infants that are 7-9 weeks old continue to look at the edge of the occluder where the object disappears for 1 second before finding it it again (Rosander & von Hofsten, 2004). Infants that are 12 weeks old move their eyes as soon as the target becomes visible again. This delay also decreases with each trial which indicates that the infant starts to anticipate where the objects will reappear. Some of these effects have been seen in younger infant as well but they have not been reliable. It is possible that the younger infant would have performed better if the object was made invisible instead of occluded since the occluder distracts attention from the target (Jonsson & von Hofsten 2003).

Contrary to earlier speculations, infant do not continue track a object over occlusion with smooth pursuit. Instead the tracking stops and one or two saccades are made to the other side (Rosander & von Hofsten, 2004). These saccades are made to anticipate when the object reappear – not when it disappears.

2. A Model of Visual Development

We have developed a model of gaze control that is able to build complicated models of target motions and to use these to control the motion of the eyes. An overview of the system is shown in Fig. 1. There are three main ways in which the visual stimulus gains access to the eye-control system. The saccade pathway generates saccades based on the location of salient features in the image. The second pathway consists of a target detections system together with a predictive model and a control system that is used to generate smooth pursuit movements. Finally, the third pathway consists of an event detection system together with an event predictor that can learn arbitrary relations between visual events.

The position of the eye is updated according to

$$p(t+1) = p(t) + v(t) + s(t) + a(t) + n(t),$$

where $v(t)$ is the velocity from the pursuit pathway, $s(t)$ is the bursts from the saccade generator, $a(t)$ is a saccade generated by the event predictor pathway, and $n(t)$ is a noise term.

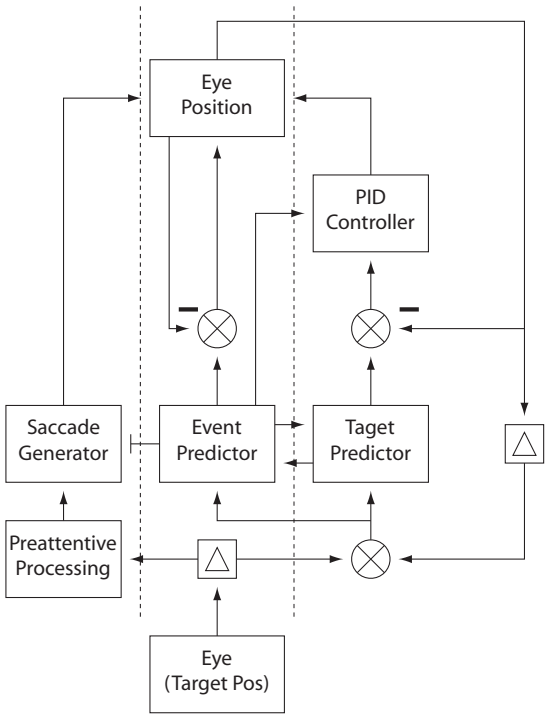

FIGURE 1: *The model consists of a number of interacting modules for target detection, target prediction, event prediction, as well as eye control.*

2.1 The Saccade Pathway

The first pathway is controlled by a pre-attentive system that selects salient features or stimuli in the image and directs attention to them.

Preattentive Processing This part of the model consists of a number of simple filtering operations including the detection of oriented contrast, curvature, foreground elements, and motion (Balkenius, Eriksson & Åström, 2004). The resulting pre-attentive maps are added together to form a salience map from with the next target location is selected. The probability of selecting a region in the image is proportional to the salience of that region which is given by the salience map.

Saccade Generation When the selected target location is sufficiently far away from the center of the eye, a saccade toward the target is generated by the saccade generator. In the *intermediate region* around the center, catch-up saccades are generated to tracked objects (cf. Smeets & Bekkering, 2000) and in the *peripheral region*, an orienting system is used to roughly direct attention in the direction of any transient event (Balkenius & Kopp, 1996). When this is the only system available, it will produce small saccades that will track any salient stimulus in view.

2.2 The Pursuit Pathway

We used a control system for target tracking similar to that of Shibata et al. (2001), where a forward-model is used to predict target velocity based on image-slip and gaze, as the human tracking system (Fukushima et al., 2002). Unlike the model of Shibata et al. (2001), all calculations are made with positions rather than velocities and velocities are only used implicitly as differences between locations.

In this system, like in the human gaze-control system, the prediction of target motion is separate from the motor control of the eyes, which is very important for the success of our system since learning of target motions should not interfere with the learning of eye control.

The pursuit pathway consist of two main modules. One is used for target prediction and the second is used to control the smooth movements of the eye. There are also two modules that performs transformations between eye coordinates and world coordinates. In the present model, these transformations are made by simply adding or subtracting the position for the eye from the target location in the eye. There is also a delay in the pathway from the eye position to the target predictor that matches that of the delay in the visual processing.

Target Prediction The target predictor consists of a linear predictor which attempts to predict the current location of the target stimulus based on a number of target locations that are delayed by Δ time steps.

$$\hat{T}(t+1) = \sum_{i=0}^{\tau} c_i(t) T(t-i)$$

The coefficients c_i are learned on line based on the error of the prediction. Thus, at time t, we know the correct target location at time $t-\Delta$ and can update the constants based on the prediction error at that time:

$$E_\Delta(t) = T(t) - \hat{T}(t-\Delta)$$

The weights of the predictor are updated according to the following equation,

$$c_i(t+1) = c_i(t) + \alpha E_\Delta(t) T(t-i).$$

Note that the predictor is simultaneously operating in two different time frames, one for learning and one for prediction.

Finally, the uncertainty $C(t)$ of the prediction is calculated as an exponentially decaying average of the prediction error

$$U(t+1) = (1-\beta)U(t) + \beta E(t).$$

which is subsequently used to set the confidence of the prediction

$$C(t+1) = e^{-U(t+1)}.$$

PID Controller The eye control system uses an proportional-integral-derivative (PID) controller. In such a controller, the control signal is the sum of three different terms,

$$v(t+1) = v(t) + c_P E(t) + c_I I(t) + c_D D(t)$$

where $I(t)$ is the integral of the error over time

$$I(t) = \sum_{i=0}^{1} E(i)$$

and $D(t)$ is the derivative of the error

$$D(t) = E(t) - E(t-1).$$

The values c_P, c_I, and c_D are parameters that define how much each type of error term contributes to the control signal. Although it would be possible to include learning of the parameters in the controller, this was not used in the current implementations. Instead, the three parameters were set to a constant multiplied with the confidence from the target predictor to model the development of smooth pursuit,

$$c_P(t) = c'_P C(t),$$

$$c_I(t) = c'_I C(t),$$

$$c_D(t) = c'_D C(t).$$

2.3 Event Prediction Pathway

The event prediction pathway contains one main module that detects events and forms associations between them. When this module predict that the target will appear it will produce a saccade $a(t)$ to the expected location and simultaneously inhibit the saccade generator.

Event Detection An event is defined as any abrupt change in any variable within the system (cf. Prem et al., 2002). In the present model we have included two signals that are used to detect events: the tracking error and the recognition of the target objects. Fast changes of the tracking error will thus be considered as an event as will the appearance or disappearance of the target. When an event occurs the type of event and corresponding location is saved to potentially be correlated with other later events.

Learning to Anticipate The anticipated changes in target motion and location are learned as associations between two events: $E_1 \rightarrow E_2$, where E_1 may be the disappearance of the target or the fact that the target reaches a certain location, and E_2 is the reappearance of the target or the expected new position of the target.

The learned associations does not only code that a target disappearing at a location $\langle x, y \rangle$ will appear at another $\langle x', y' \rangle$, but also the time between the two events Δt and the expected velocity when the target reappears v (Poliakoff, Collins & Barnes, 2004):

$$\langle x, y \rangle \rightarrow \langle x', y', \Delta t, v \rangle$$

This learning is driven by the rewarding property of the target, i. e. when the target appears it will generate a reward that will drive the learning of the event associations. This is consistent with the observation that all brain systems involved in the linking between visual stimulation and oculomotor behavior encode the expected value of the target (McCoy & Platt, 2005).

The anticipatory saccades constitute a form of adaptive switching control strategy, where the anticipatory saccade controller quickly sets the parameters of the smooth pursuit controller to immediately obtain good tracking performance (Huang & Lin, 2004). Evidence that arbitrary stimuli can be used to predict the appearance, time and velocity of a stimulus in adults comes from experiment by Barnes and Donelan (1999).

2.4 Eye Control

The cooperating modules result in an eye controller which is similar to an earlier design that has been implemented in a stereo vision head (Balkenius & Kopp, 1996). The visual field is divided into three regions and different control strategies are used depending on in which region the target is located.

In the focal region, a controller for smooth pursuit and fixation is used which is in line with evidence that the same mechanism is used for these two behaviors (Smeets & Bekkering, 2000). In the intermediate and peripheral regions, different types of saccades are generated instead.

3. Results

3.1 Taget Prediction

The target prediction module was tested with a sinusoidally moving target and a 100 ms delay of the visual input. A new frame was processed every 20 ms to parallel the human visual bandwidth of approximately 50Hz. One cycle lasted for 5 second and the target moved over 60 degrees of the visual field from end to end.

The constants were set as $\alpha = 0.1$, $\Delta = 5$, and $\tau = 3$. The theoretical prediction coefficients c_i were calculated for three types of predictors: (A) a simple linear extrapolation of position based on the calculated velocity between the last two target positions and the known delay Δ, (B) another linear predictor that averaged the last two estimated velocities, and (C) a prediction that also included the estimated acceleration of the target. As can be seen in Table 1, the theoretical model that takes acceleration into account gives the smallest error. However, when noise is added to each target location, the performance of this predictor deteriorates considerably. In this case, the predictor that averages over two velocity estimates gives the best result.

When the predictor is allowed to learn the prediction coefficients, they resembles predictor B, except that c_1 is not zero indicating that the current position is a weighed average between the last two target locations. The performance of this learned predictor is close to that of predictor B.

The predictor was also allowed to learn the coefficients with noise. In this case, it is able to learn coefficients that results in a lower average error than any of the other predictors since it learns to essentially use the average of the last three target positions as the estimate. Although the error is reduced, the target estimation now lags the real target location and is no longer anticipatory.

These results show that a linear predictor can learn the appropriate coefficients to make anticipatory estimates of the target location. These coefficients favor an estimation that averages over several target locations and will thus only become anticipatory when the target locations are sufficiently reliable. The learned model is a compromise between limiting the sensitivity to noise making an accurate prediction.

3.2 Smooth Pursuit

The development of the pursuit system was simulated for the model when it continuously viewed a a sinusoidal movement. As the confidence of the prediction increased, so did the gain of the smooth pursuit system. The parameters were set as in the previous simulations. The target moved either according to a sinusoidal path or in a triangular way (Fig. 2).

Fig. 2 shows the development of smooth pursuit from initial saccadic tracking to the final model based tracking. Even when a predictive model is used to control the gaze, there is still a small overshoot that is casued by the abrupt change of direction in the triangular motion. This overshoot almost completely disappears when the event detection system is added (data not shown). In this case, the predicted target location at the end of the envelope triggers an event that will associate to a new velocity and direction of

TABLE 1: *Theoretical and learned prediction coefficients and the corresponding prediction errors with and without noise with a sinusoidally moving target. The noise was additive with a range from -10 to 10 degrees.*

Type	c_0	c_1	c_2	LMS Error	With Noise
A. Linear	6	0	-5	$2.0 \cdot 10^{-3}$	$3.3 \cdot 10^{-2}$
B. Averaged	3.5	0	-2.5	$2.2 \cdot 10^{-4}$	$1.8 \cdot 10^{-2}$
C. Acceleration	18.5	-30	12.5	$3.3 \cdot 10^{-5}$	$3.5 \cdot 10^{-1}$
Learned	3.2	0.33	-2.53	$2.4 \cdot 10^{-4}$	$1.9 \cdot 10^{-2}$
Lrn w noise	0.36	0.32	0.24	$5.3 \cdot 10^{-3}$	$5.3 \cdot 10^{-3}$

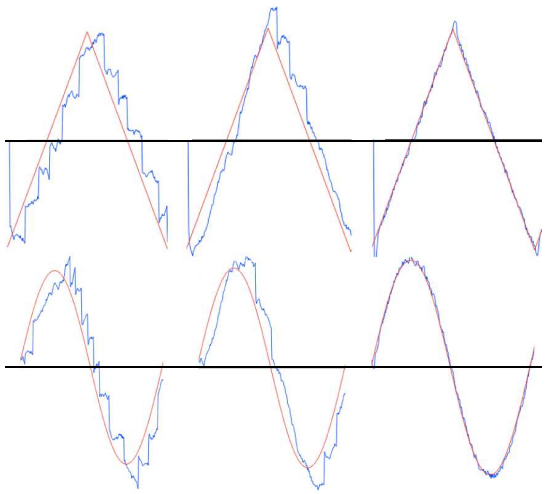

FIGURE 2: *The development of tracking behavior of a sinusoidal and triangular target motion. Left: No smooth pursuit. Middle. An intermediate stage. Right: Fully developed predictive smooth pursuit. The difference between the tracking motions is mainly a result of an increased gain of smooth pursuit. Note that there is still a small overshoot in the triangular case.*

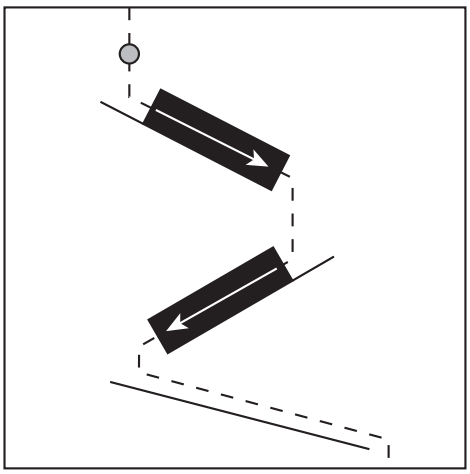

FIGURE 3: *En example scene with a ball falling down and rolling through two pipes and a slide. The white arrows indicate locations where the system will learn associations between the disappearance and appearance of the ball.*

the target.

3.3 Motion and Events

The complete system was tested with a more complicated version of the train in the tunnel scene. An animation was shown to the system where a ball falls down and rolls through two pipes before falling again and rolling on a slide before falling out of the scene (Fig. 3). The system learns both to track the target ball when it is visible and to anticipate the reappearance of the ball when it is hidden in one of the pipes.

4. Discussion

The model shows how it is possible to combine three different types of gaze control: reactive, continuous model based, and event driven control and how a system can learn to autonomously shift between the different types of control depending on the visual scene. In addition, we have shown that the development of the system roughly follows that of human infants.

In the future, we want to further investigate how the system can learn and use several different models concurrently (cf. Wolpert et al., 2003) and how the task context can influence what model is used (Doya, et al., 2003, Balkenius & Winberg, 2004). This type of model may also form a basis for the study of the development of synchronization and imitation (Barnes & Donelan, 1999).

We also want to investigate the relation between this type of switching control and reinforcement learning and how the system can learn to generalize from previously learned scenes to new ones, which may eventually make it able to track moving objects perfectly on the first trial.

A limitation of the current model is that the association mechanism is very simplistic since it only associates two subsequent events with each other. It can not learn regularities over longer times if they are interrupted by other events. This limitation will be addressed in the future when a more complete associative mechanism will be included in the model.

References

Aguiar, A. & Baillargeon, R. (1999). *Cognitive Psychology, 39*, 2, 116–157.

Balkenius, C., Eriksson, A. P. & Åström, K. (2004). Learning in Visual Attention. In *Proceedings of LAVS '04*. St Catharine's College, Cambridge, UK.

Balkenius, C., & Kopp, L. (1996). Visual tracking and target selection for mobile robots. In Jörg, K-W. (Ed.), *Proceedings of the First Euromicro Workshop on Advanced Mobile Robots (EUROBOT '96)* (pp. 166-171). Los Alamitos, CA: IEEE Computer Society Press.

Balkenius, C., & Winberg, S. (2004). Cognitive modeling with context sensitive reinforcement learning. In *Proceedings of AILS '04*. Lund: Dept. of Computer Science

Barnes, G. R. & Donelan, S. F. (1999). The remembered pursuit task: evidence for segregation of timing and velocity stoarge in predictive oculomotor saccades. *Exp Brain Res, 129*, 57–67.

Dayton, G. O. & Jones, M. H. (1964). Analysis of characteristics of the fixation reflex in infants by use of direct current electro-oculograpy. *Neurology,14*,1152-1156.

Deno, D. C., Crandall, W. F., Sherman, K. & Keller, E. L. (1995). Characterization of prediction in the primate visual smooth pursuit system. *BioSystems, 34*, 107–128.

Doya, K., Sugimoto, N., Wolpert, D. & Kawato, M. (2003). Selecting optimal behaviors based on context. In *International symposium on emergent mechanisms of communication*.

Fukushima, K., Yamanobe, T., Shinmei, Y. & Fukushima, J. (2002). Predictive responses of periarcuate pursuit neurons to visual target motion. *Exp Brain Res, 145*, 104120.

Guan, Y., Eggert, T., Bayer, O. & Büttner, U. (2005). Saccades to stationary and moving targets differ in the monkey. *Exp Brain Res, 161*, 220–232

Huang, H-P. & Lin, F-Y. (2004). On-line adaptive switching control for open-loop stable and integrating processes. In *Proceedings of the 2004 IEEE International Conference on Control Applications*, Taipei.

Jonsson, B. & von Hofsten, C. (2003). Infants' ability to track and reach for temporarly occluded objects. *Development Science, 6*, 88–101.

McCoy, A. N. & Platt, M. L. (2005). Expectations and outcomes: decision-making in the primate brain. *J Comp Physiol A, 191*, 201–211

Mehta, B. & Schaal, S. (2002). Forward models in visuomotor control. *J Neurophysiol, 88*, 942–953.

Piaget, J. (1937). *La construction du rèel chez l'enfant*. Neuchatel: Delachaux et Niestle.

Prem, E., Hrtnagl, E. & Dorffner, G. (2002). Growing Event Memories for Autonomous Robots. In *Proceedings of the Workshop On Growing Artifacts That Live*, Seventh Int. Conf. on Simulation of Adaptive Behavior, Edinburgh, Scotland.

Roucoux, A., Culee, C. & Roucoux, M. (1983). Develop of fixation and pursuit eye movments in human infants. *Behavioural brain reseach, 10* ,133–139.

Rosander, K. & von Hofsten, C. (2004). Infants' emerging ability to represent occluded object motion. *Cognition, 91*, 1, 1–22.

Shibata, T., Vijayakumar, S., Conradt, J. & Schaal, S. (2001). In *Humanoid Oculomotor Control Based on Concepts of Computational Neuroscience. Humanoids2001, Second IEEE-RAS Intl. Conf. on Humanoid Robots*, Waseda Univ., Japan. 278–285.

Smeets, J. B. J. & Bekkering, H. (2000). Prediction of saccadic amplitude during smooth pursuit eye movements. *Human Movement Science, 19*, 275–295.

Stork, S., Sebastian, F. W., Neggers, S. F. W, & Müsseler, J. (2002). Intentionally-evoked modulations of smooth pursuit eye movements. *Human Movement Science, 21*, 335–348

von Hofsten, C. & Rosander, K. (1997). Development of smooth pursuit tracking in young infants. *Vision Research, 37*, 13, 1799–1810.

Wells, S. G. & Barnes, G. R. (1998). Fast, anticipatory smooth-pursuit eye mevements appear to depend on a short-term store. *Exp Brain Res, 120*, 129–133.

Wentworth, N. & Haith, M. M. (1998). Infants' acquisition of spatiotemporal expectations. *Developmental Psychology, 34, 2*, 247–257.

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