Learning Optimal Gaze Decomposition

Eric Wiewiora*, David Berg†, Jochen Triesch†, and Tomonori Hashiyama‡
*Computer Science Department, UC San Diego, USA
Email: ewiewior@cs.ucsd.edu
†Cognitive Science Department, UC San Diego, USA
‡Institute of Natural Sciences Nagoya City University, Japan

I. Introduction

When the head is free to move, subjects frequently engage in coordinated head and eye movements to bring a target object to the fovea. Freedman and Sparks [2] found that the relative contributions of head and eye movements to the total gaze shift are non-linear functions of the initial eye position and the total gaze displacement. Freedman [1] and Wang and colleagues [8], [9] have recently proposed descriptive mathematical models for the decomposition of total gaze shift into head and eye movements. It is however an open question a) if and how this decomposition can be seen as resulting from an optimality principle, b) if this decomposition strategy is learned and c) if so, what learning mechanisms are responsible for its acquisition. We show that the rather complex behaviorally observed gaze decomposition can be understood as the result of optimizing a simple cost function. We propose a simple model for the simultaneous learning of the calibration of goal directed head/eye movements and the optimal gaze shift decomposition based on a reinforcement learning mechanism [7]. In our model, the cerebellum plays a key role in learning a gaze shift decomposition that accurately brings the desired target to the fovea while at the same time minimizing this cost function. Our model is roughly consistent with the known anatomy of oculomotor control systems. The model learns gaze shift decompositions observed experimentally and makes a number of testable predictions. The model is also implemented and tested in an anthropomorphic robot head that autonomously learns to calibrate its gaze shifts.

II. Learning Gaze Decomposition as Optimization

Although the redundancy in neck and eyes allows for a wide range of movements to achieve a desired gaze shift, the decomposition of the task shows little variance in highly trained monkeys [2] (but also see [5]). Generally, eye movements are solely responsible for small shifts, while the head becomes increasingly involved in larger movements. Based on the data in [2], Wang and Jin [9] derived a system of equations that describe the decomposition of horizontal gaze shifts based on the initial angle of the eye relative to the head and the location of the target. We show that this gaze decomposition results from optimizating the following simple cost function:

$$C(H, E, E_i) = aH + \frac{1}{b + E_i} H^2 + c(E_i + E)^2,$$

where H and E are the head and eye movement amplitudes, and E_i is the initial eye position with respect to the head.

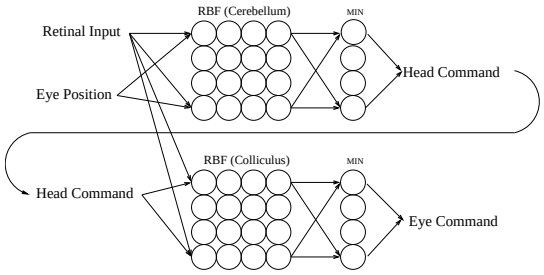

Fig. 1. Schematic of the Learning Architecture

This cost function has terms that are linear and quadratic in the head movement and a term that is quadratic in the final eye position $E_i + E$. Note that cost terms that are linear and quadratic in the motor signal are common in the motor control literature, e.g. [3]. The definition of a cost function allows us to view the decomposition of gaze shifts as an optimization problem. In the following, we propose a simple reinforcement based learning architecture for learning gaze shifts with the optimal decomposition.

III. The Learning Architecture

Our model is based on the following architecture proposed in [8]. The target location is encoded in the buildup cells of the superior colliculus (SC) and transmitted to the cerebellum (CB) and the burst cells of the SC. The CB determines the head contribution to the gaze shift, and sends this information to the burst cells of the SC. The burst cells compute an appropriate eye movement given the target location and the planned head movement. While this model is certainly overly simplistic in many respects it serves well as a basis for an initial exploration of putative learning mechanisms involved in gaze decomposition.

In our model, the CB and the burst cells of the SC are modeled with coarse population coding networks employing basis function units [6]. An overview of the architecture is given in Fig. 1. Both networks have two sources of input. The two basis function networks project to two other networks, that predict the costs for specific head and eye movements. The movements with the smallest estimated costs are executed. The system learns by updating the weights between the basis function networks and the cost prediction networks according to the total cost incurred for the movement. This total cost is the sum of $C(H, E, E_i)$ and a term that punishes inaccurate gaze shifts. The overall architecture is modular, in the sense that

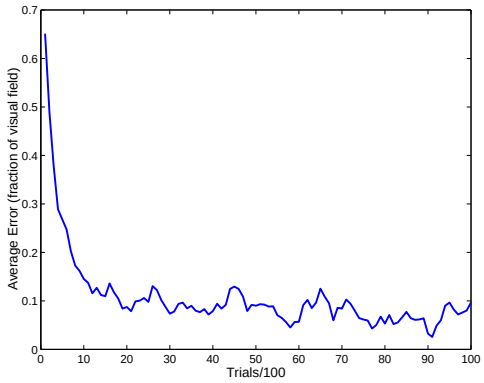

Fig. 2. Simulation results.

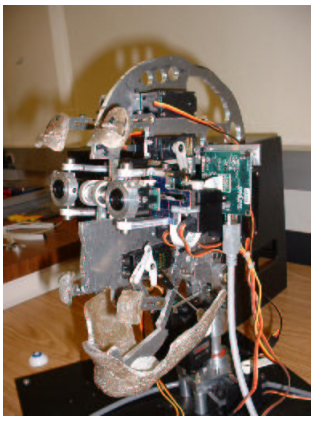

Fig. 3. Robot Head

eye and head movement are computed by different networks, but they are not computed independently since the output of one network serves as input to the other. Figure 2 shows the results of a simulation of the model. It depicts the average gaze error as a function of learning time. The trade-off between the number of units in the networks and the resulting gaze shift accuracy, and its dependence on the model's parameters are currently being investigated.

IV. Implementation on Robotic Head

We are also testing the model on a 9 degree of freedom humanoid robot head (Fig. 3). The robot head has miniature cameras fitted as eyes and is capable of making fast saccade-like eye movements. Each eye has two degrees of freedom (pan and tilt) and the neck has two degrees of freedom (also pan and tilt). The remaining three degrees of freedom allow facial gestures (opening and closing the mouth, raising/lowering the eyebrows, smiling/frowning).

The robotic implementation forces us to solve the visual correspondence problem, an important subproblem in learning gaze control strategies which most other models do not address. Initially a target for a saccade is selected based on a interest operator. After a saccade has been made the intended target of the saccade needs to be found in the new retinal image in order to compute how successful the saccade was in bringing the object to the center of gaze. To this end, we are using a Graph Matching technique [4] based on a Gabor wavelet representation. The distance between the matched target and the center of gaze enters the computation of the cost of the gaze shift as an additional term. The model is fully implemented and experiments are about to start.

V. Conclusion

While previous work has provided descriptive models of how gaze shifts are decomposed into head and eye movements, we propose to view the task as an optimization problem and present a simple cost function, whose optimization is consistent with experimentally observed gaze decomposition patterns. We propose a modular reinforcement learning architecture utilizing population coding networks to learn optimal gaze shifts. Our model makes a number of predictions. First, our model predicts gaze decomposition patterns for a range of conditions that have not been investigated experimentally. Second, the model predicts that physical changes to the plant (e.g. increased head inertia) will change the observed gaze decomposition patterns in systematic ways.

References

[1] E.G. Freedman. Interactions between eye and head control signals can account for movement kinematics. *Biological Cybernetics*, 84:453–462, 2001.

[2] E.G. Freedman and D.L. Sparks. Eye-head coordination during head-unrestrained gaze shifts in rhesus monekys. *J. Neurophysiol.*, 77:2328–2348, 1997.

[3] C.M. Harris and D.M. Wolpert. Signal-dependent noise determines motor planning. *Nature*, 394:780–784, 1998.

[4] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Computers*, 42:300–311, 1993.

[5] J. Pelz, M. Hayhoe, and R. Loeber. The coordination of eye, head, and hand movements in a natural task. *Exp. Brain Res.*, 139:266–277, 2001.

[6] A. Pouget and L. Snyder. Computational approaches to sensorimotor transformations. *Nature Neuroscience*, 3:1192–1198, 2000.

[7] W. Schultz, P. Dayan, and P. R. Montague. A neural substrate of prediction and reward. *Science*, 275:1593–1599, 1997.

[8] X. Wang, J. Jin, and M. Jabri. Neural network models for the gaze shift system in the superior colliculus and cerebellum. *Neural Networks*, 15:811–832, 2002.

[9] X. Wang and J.S. Jin. A quantitative analysis for decomposing visual signal of the gaze displacement. In D.D. Feng, J. Jin, P. Eades, and H. Yan, editors, *Selected papers from 2001 Pan-Sydney Workshop on Visual Information Processing*, volume 11, pages 153–159. ACS, 2002.