A Unified Model of the Joint Development of Disparity Selectivity and Vergence Control

Yu ZHAO[1], Constantin A. ROTHKOPF[2,3], Jochen TRIESCH[2] and Bertram E. SHI[1]

Abstract — **Reinforcement learning is a prime candidate as a general mechanism to learn how to progressively choose behaviorally better options in animals and humans. An important problem is how the brain finds representations of relevant sensory input to use for such learning. Extensive empirical data have shown that such representations are also adapted throughout development. Thus, learning sensory representations for tasks and learning of task solutions occur simultaneously. Here we propose a novel framework for efficient coding and task learning in the full perception and action cycle and apply it to the learning of disparity representation for vergence eye movements. Our approach integrates learning of a generative model of sensory signals and learning of a behavior policy with the identical objective of making the generative model work as effectively as possible. We show that this naturally leads to a self-calibrating system learning to represent binocular disparity and produce accurate vergence eye movements. Our framework is very general and could be useful in explaining the development of various sensorimotor behaviors and their underlying representations.**

I. INTRODUCTION

Active interaction with the environment is critical for proper neural development [1]. However, the vast majority of past work has treated the two problems of learning perception and learning behavior in isolation. This is in stark contrast to developing organisms, where these processes are tightly coupled. While visual neuronal processing may develop to match the statistics of the visual input [2], these statistics are determined by the agent's actions within the environment [3]. To our knowledge, very little work has been done in developing an integrated framework addressing this "chicken-and-egg" problem of perceptual and behavioral development. Some work has been done in combining unsupervised learning methods with reinforcement learning methods running simultaneously, but still treat the two problems as largely independent, with little explicit coupling between them. One path towards a more integrated approach may be found in the ideas of artificial curiosity [4]. The curiosity signal is based on the idea of compression progress, which measures how efficiently a recurrent neural network can learn to account for the past history of its observations. The compression progress is a formal framework capturing the notion of "interestingness" or the "potential for the discovery of novel patterns". This signal can be used within a reinforcement learning framework to direct the agent to behave so as to obtain the largest gains in compression progress. Recently, this approach has been used in an integrated approach to autonomous perceptual and cognitive development [5].

In this paper, we propose an integrated approach to the joint development of perception and behavior in the context of eye movements. In particular, we model the joint development of stereo disparity perception and vergence eye movements. There is a large discrepancy between the statistics of the absolute disparities in the natural environment, which range over tens of degrees, with the statistics of the preferred retinal disparities in disparity selective primary visual cortical neurons, which range over only about one degree around fixation. Thus, for a passive observer presented with the natural environment, we might expect the distribution of retinal disparity selectivities to be much larger than is actually observed, or that, given the small receptive field sizes in the fovea, binocular cells might not develop. Indeed, kittens with squint that is artificially induced by severing one of the extraocular muscles develop much fewer (20%) binocularly driven cells than normal cats (80%) [6]. One way to resolve this discrepancy is to assume that during development, binocular vergence movements between the two eyes serve to keep surfaces in the environment within a small range of fixation, thus leading to the development of disparity selective neurons tuned to a small range of retinal disparities.

Most past work has considered the development of stereo disparity perception and the development of vergence eye movements in isolation. We are aware of only two studies making steps towards an integrated framework. Franz and Triesch [6] showed that binocular disparity tuned neurons emerge in a reinforcement learning based model of the development of vergence eye movements. However, this work was not fully developmental, as it assumed the existence of a pre-existing set of binocular disparity energy neurons tuned to zero disparities, which was used to determine the reward maximized by the reinforcement learning algorithm. Sun and Shi [8] showed that it is possible to remove this assumption, by using the total activation in the neural population as the reward function. However, this work did not present a unified framework, since the reward being maximized by behavior (total activation) was not the same as the reward being maximized by neural development (sparsity).

Our unified model of the development of binocular disparity tuning and vergence control seeks both a neural representation and a behavior driven by that representation that results in the "best" encoding of the input, subject to a constraint on the complexity of the representation. The

[1] Dept. of Electronic and Computer Engineering, HK University of Science and Technology, Clear Water Bay, Hong Kong, eebert@ee.ust.hk

[2] Frankfurt Institute for Advanced Studies, Frankfurt, Germany

[3] Institute of Cognitive Science, University Osnabrück, Osnabrück, Germany

measure of the quality of the encoding that we use is the squared magnitude of the difference between the image reconstructed from the neural representation and the original binocular input image patches. The complexity is constrained by limiting the number of neurons active in representing the input.

II. METHODS

The inputs to the model are stereo image sequences. For the experiments we describe here, stereo image pairs are generated artificially from 20 images taken from the van Hateren database [9]. The left eye sees the original image, while the right eye sees a horizontally shifted version of this image. We refer to this shift as the input disparity, d_i. Binocular images at different input disparities correspond to viewing planar objects at different depths, assuming perspective projection and that the optical axes of the two cameras are parallel.

Stereo image sequences are created by choosing an image and an input disparity randomly and keeping these constant for 10 frames. After each set of 10 frames, a new image is chosen randomly from the set of 20, and a new input disparity is chosen randomly either independently according to a fixed distribution or according to a random walk process depending upon the experiment.

From each image pair in the sequence, we extract two 55 by 55 pixel windows, which are taken to model the left and right foveae. Assuming that 10 pixels correspond to about 1 degree of visual angle, the fovea region is about 5.5 degrees. We model eye movements by changing the location of these windows. The location of the left eye window is fixed at a random location in the left eye image during each set of 10 frames. The right eye window has the same vertical location as the left eye window. Its horizontal location is offset from that of the left eye window by an amount we refer to as the vergence, v. The actual retinal disparity, d_r, between the images contained in the two windows is the difference between the input disparity and the vergence: $d_r = d_i - v$. Because we choose the left eye window locations and the input disparities randomly, the number of distinct windows seen is much larger than 20, the number of images used to generate the stereo image pairs.

The vergence changes from frame to frame depending upon the action chosen. We consider A = 11 possible actions, corresponding to changing the value of the vergence from its value in the previous frame by integer pixel shifts ranging from -5 to +5 pixels. If the system develops a vergence control policy to maintain binocular fixation, the retinal disparity should generally be zero: $d_r \approx 0$. Equivalently, the vergence should generally be equal to the input disparity, $v \approx d_i$.

At each frame, a vergence command is generated from the information contained in the two foveal windows, as illustrated in Figure 1. The generation of the vergence command can be divided into two stages. First the information in the two foveal windows is encoded

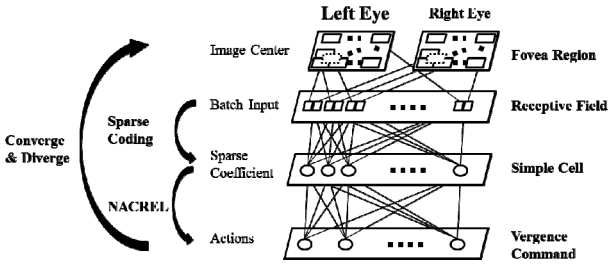


Figure 1: The mapping from stereo images to vergence command and the feedback control of the stereo images by vergence command.

binocularly using a sparse coding model with matching pursuit [10]. Second, this encoding is mapped to a vergence command by a neural network whose weights are learned using the natural actor critic reinforcement learning (NACREL) algorithm [11]. These two stages are described in more detail below.

A. Perceptual Encoding and Learning

Each 55-by-55 pixel window is divided into P = 100 10-by-10 pixel patches, whose locations are generated by 5 pixel shifts horizontally and vertically so that the patches cover the window with 5 pixel overlap between neighboring patches. The intensities within each 10-by-10 monocular patch are shifted and normalized so that the 100 dimensional vector of intensities has zero mean and unit norm.

At each time t, corresponding patches from the left and right images are then combined into a single 200 dimensional binocular image patch, $x_i(t)$, where $i \in \{1,\ldots,P\}$ indexes the patch. In the following, we will generally drop the subscript when referring to the entire collection of patches, i.e. $x(t) = \{x_i(t)\}_{i=1}^{P}$. The first 100 entries of the vector are the intensities of the left eye patch and the second 100 entries are from the right eye.

The neural representation is based on the idea that the neurons encode image patches based upon a sparse linear combination of basis functions drawn from an over-complete dictionary [2]. More specifically, given an over-complete dictionary of unit norm basis vectors $\phi(t) = \{\phi_n(t)\}_{n=1}^{N}$ where N = 300 is the dictionary size, matching pursuit tries to approximate the binocular image patch as the weighted sum

$$x_i(t) \approx \sum_{n=1}^{N} a_{i,n}(t)\phi_n(t) \qquad (1)$$

where at most C = 10 of the scalar coefficients $a_{i,n}(t)$ are nonzero. All patches are approximated using elements from the same dictionary. We use the matching pursuit algorithm proposed by Mallat & Zhang [9] to choose the coefficients $a_{i,n}(t)$. Although we could use a more formal generative model and inference method, we use matching pursuit because it has two advantages. First, it captures the main ideas of sparse coding in a computationally simple way. Second, it enables us to constrain the complexity of the representation easily by choosing the value of C.

We make a loose analogy between the coefficients $a_{i,n}(t)$ and the responses of disparity selective simple cells in the

primary visual cortex. For fixed patch index *i*, the coefficients $a_{i,n}(t)$ are the responses of *N* simple cells in one hypercolumn, a set of cells responding to the same visual spatial location but with different selectivity along other dimensions, such as orientation, spatial frequency or disparity. Since the $a_{i,n}(t)$ can be both positive and negative, they might more precisely be thought of as encoding the responses of pairs of opponent simple cells. The basis function $\phi_n(x,y)$ determines the selectivity, and is roughly analogous to the receptive field of the neuron. The selection process for the *C* nonzero coefficients can be considered an approximation to the effect of lateral feedback interconnections between neurons, which can be used to ensure sparseness [2]. Following this analogy, we implement a model of the pooled activity of complex cells serving the visual locations covered by the foveal windows. Complex cells display the same selectivity as the simple cells, but more position invariance. We model the pooled activity of complex cells with the same selectivity as basis functions ϕ_n, by summing their squared coefficients over the patches in the window

$$f_n(t) = \sum_{i=1}^{P} a_{i,n}(t)^2 \tag{2}$$

We adapt the dictionary of basis functions to minimize reconstruction error using an on-line two step procedure similar to that used by Olshausen & Field (1996). In the first step, we find the coefficients $a_{i.n}(t)$ using matching pursuit. In the second step, we assume the coefficients $a_{i.n}(t)$ are fixed, and adapt the basis functions to minimize the average squared reconstruction error over patches

$$r(t) = \frac{1}{P}\sum_{i=1}^{P} \frac{\left\| x_i(t) - \sum_{n=1}^{N} a_{i,n}(t)\phi_n(t) \right\|^2}{\left\| x_i(t) \right\|^2} \tag{3}$$

After each update, the basis functions are re-normalized so that they are unit norm.

B. Behavior

Behavior is defined by a policy, which maps the state of the actor in its environment to an action. In our system, the state is represented by $f(t)$, the *N*-dimensional vector of complex cell outputs. Actions are chosen from the set of *A* vergence commands. The mapping from state to an *A*-dimensional vector of action probabilities is performed by a two layer neural network, with *N* neurons in the input layer and *A* neurons in the output layer. Output activity is computed via linear combination of the inputs followed by a softmax activation function. Mathematically, if we denote the probability of choosing the *a*-th action by π_a, then

$$\pi_a\left(f(t)\right) = \frac{\exp(T^{-1}z_a)}{\sum_{b=1}^{A}\exp(T^{-1}z_b)} \tag{4}$$

where *T* is a positive scalar called the temperature and z_a is the activation of the *a*-th output neuron, which is given by

$$z_a = \sum_{n=1}^{N} \theta_{a,n}(t) f_n(t) \tag{5}$$

where each $\theta_{a,n}(t)$ is a scalar valued weight coefficient. The soft-max activation ensures that the outputs of the neural network are positive and sum to one. We obtain a stochastic policy by sampling the action according to the probability distribution $\pi\left(f(t)\right)$. For a fixed set of coefficients $\theta(t)$, the temperature parameter determines the entropy of is probability distribution. As the temperature decreases, the entropy decreases, and the probability of the action with the largest value of z_a approaches 1 while the others approach zero.

We learn behavior using a version of the natural actor critic reinforcement learning algorithm (NACREL) [1] to find a policy that minimizes the long term discounted sum of the reconstruction errors

$$\sum_{\tau=-0}^{\infty} \gamma^{\tau} r(t+\tau) \tag{6}$$

where γ is the discount factor. In particular, we use a modification of Algorithm 3 as described in [1] that adds the discount factor and regularizes the policy network weights.

III. Results

We present the results of three sets of experiments. The first set of experiments examines the properties of the dictionary of basis functions that develop under sparse coding. The second set of experiments examines the development of vergence control policies when the basis functions are fixed. Here, we assume that the basis functions developed separately under the disparity statistics considered in the first set of experiments. The final set of experiments examines joint learning, where both the basis functions and the policy are initialized randomly and are updated every frame to minimize the reconstruction error.

A. Binocular Receptive Field Development

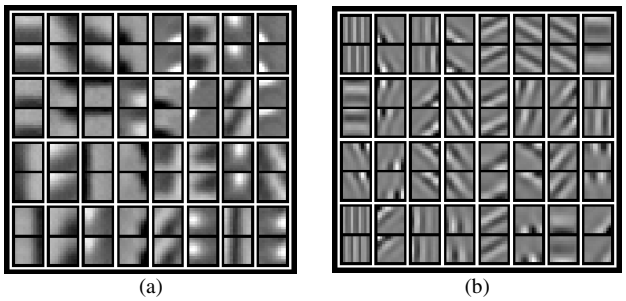
In this subsection, we examine the characteristics of the binocular basis functions that develop using matching pursuit as the statistics of the input disparity vary.

For these experiments, input disparities, *d*, are chosen randomly between -18 and +18 pixels according to a discrete truncated Laplacian distribution

$$P(d) = Me^{-|d|/D} \tag{7}$$

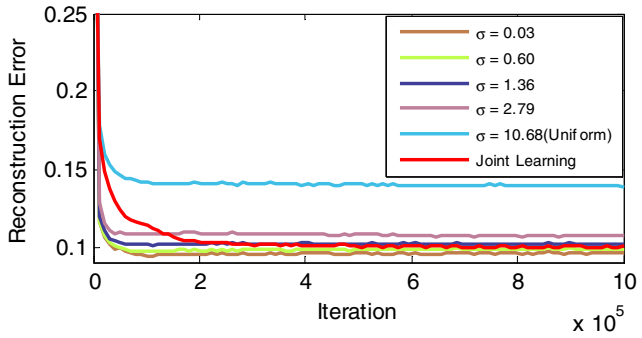
where *M* is a normalization factor that ensures that $P(d)$ sums to unity. The vergence is fixed at zero so that the input disparity and retinal disparities are identical. We consider ten different input disparity distributions where D varies among 0.125, 0.25, 0.5, 1, 2, 4, 8, 16, 64, and ∞. The corresponding standard deviations σ are 0.03, 0.2, 0.6, 1.4, 2.8, 5.2, 7.6, 9.1, 10.3 and 10.7. When $D = \infty$, the input disparity distribution is uniform over ±18 pixels.

As discussed earlier, the basis functions are models of the receptive fields of binocular disparity selective neurons. Figure 2 shows examples of the bases which develop in joint learning, but the results are similar for learning with a fixed distribution of disparity values where $D \approx 2$. The monocular RFs are Gabor-like and show a diversity of spatial frequencies. Because the distribution of input disparities is clustered around zero, the left and right eye components of

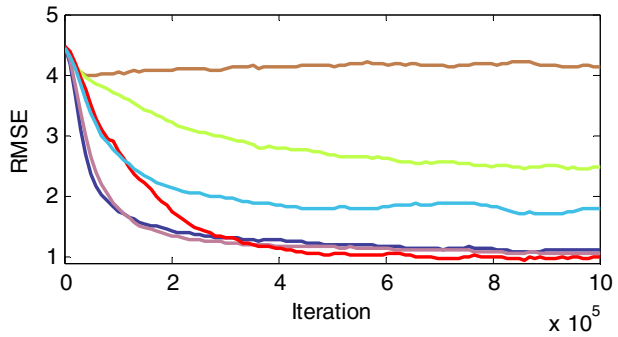


(a) (b)

Figure 2: Example bases trained in joint learning. The left and right parts of each basis are shown as 10-by-10 pixel images, which are aligned vertically. Bright regions correspond to positive values. Dark regions correspond to negative values. Neutral grey corresponds to zero. (a): The most commonly chosen 32 bases. (b) The least commonly chosen 32 bases.



(a)



(b)

Figure 3: (a) The evolution of the average squared reconstruction error for inputs with zero retinal disparity. (b) The RMSE of policies learned under separate and joint development of disparity perception and vergence.The legend in (b) is the same as that in (a).

the most commonly chosen basis functions are nearly identical, suggesting that the corresponding simple cells are tuned excitatory cells for zero disparity. On the other hand, the left and right eye components of the least frequently chosen basis functions display horizontal shifts or are additive inverses, corresponding to cells tuned to non-zero disparities or tuned inhibitory cells.

Figure 3(a) shows the decrease in reconstruction error during training of the sparse coding. The steady state value increases with the standard deviation of the input disparity distribution. Inputs whose distributions are closely clustered around zero are easier to encode as the left and right eye images are highly redundant. Figure 4 shows the steady state average squared reconstruction error for inputs with different disparities. Basis functions trained with input disparity

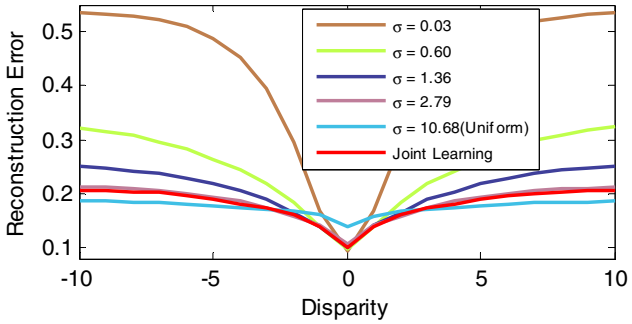


Figure 4: The reconstruction error as a function of the input disparity for basis functions learned under different input disparity distributions.

distributions with small standard deviations are very effective at encoding inputs with small disparities, but less effective at inputs with large disparities. In comparison, basis functions trained with input disparity distributions with large standard deviations are more effective at larger input disparities, but less effective at small input disparities.

B. Learning Vergence Control with Fixed Basis Functions

In this section, we demonstrate that given bases learned via sparse coding as described previously, the natural actor critic algorithm can learn policies which control vergence to maintain binocular fixation on the input as it changes in disparity. We refer to this as separate learning, since the two developmental processes are decoupled in time. For these experiments, the image sequences are created similarly as described previously, except that instead of selecting the disparity according to the truncated Laplacian distribution, the disparity changes every 10 frames according to a random walk, where the change in disparity is uniformly distributed between +5 and -5 pixels. The absolute input disparity is limited to lie between +12 and -12 pixels.

Figure 5(a)-(c) shows estimates of the policies that emerge during reinforcement learning using basis functions developed by sparse coding of inputs with different input disparity standard deviations. Each image pixel corresponds to a particular input disparity (column) and action (row). The intensity of the pixel indicates the probability that the action is chosen when the input is at a particular retinal disparity. This probability is estimated by averaging the action probabilities at the output of the policy neural network over 50 binocular image windows

$$\tilde{\pi}_a(d) = \frac{1}{50}\sum_{w=1}^{50}\pi_a(f_w(d)) \qquad (8)$$

where $f_w(d)$ is the feature vector of complex cell activations calculated from a binocular window pair w which has input The largest 95% confidence interval across all disparity-action pairs is 0.07. As a reference standard, we consider a policy that would zero out the retinal disparity after one step. Such a policy would have probability one for all of the green boxes, and zero otherwise.

The policy that develops for basis functions that have been exposed only to inputs very close to zero disparity (Figure 5(a)) is far from the reference policy. It favors either zero or

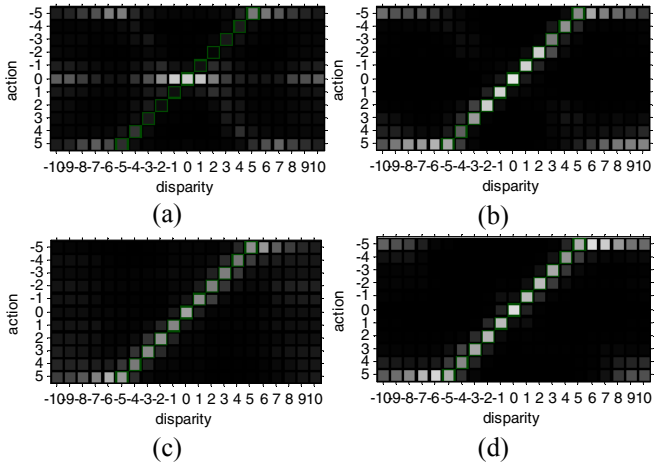


Figure 5: Examples of policies learned by reinforcement learning. The intensity of white color represents probability of corresponding disparity-action pair, while full intensity means 1. (a) The policy when basis functions are fixed after learning sparse coding for inputs with disparity standard deviation 0.03 pixels. (b) Similar to (a) except standard deviation is 0.60 pixels. (c) Similar to (a) except standard deviation is 10.7 pixels. (d) The policy from joint learning.

directionally symmetric vergence change actions. Because these basis functions have been exposed only rarely to stimuli with non-zero disparities, the encoding that develops cannot distinguish between positive and negative disparities. At a slightly larger input disparity standard deviation (Figure 5(b)), the policy that develops is closer to the reference policy for disparities between +5 and -5, but does not choose the correct action at larger input disparities as often as a policy that develops for the largest input disparity standard deviation (Figure 5(c)).

We evaluate the quality of the policy according to the root mean squared error (RMSE) between the actions taken by the policy and the actions taken by the reference policy that brings the input inside the binocular windows to zero disparity in one step. Mathematically, we define

$$\mathrm{RMSE} = \sqrt{\frac{1}{11}\sum_{d=-5}^{5}\sum_{a=-5}^{5}\tilde{\pi}_a(d)(d+a)^2} \qquad (9)$$

Figure 3(b) shows the evolution of the RMSE during training. For policies using basis functions trained over inputs with a large disparity standard deviation, the RMSE decreases over time. However, as described above, for policies using basis functions trained only on zero disparities, the RMSE remains nearly constant. The RMSE of the final policy after convergence depends upon the input standard deviation, with the policy that uses basis functions trained with input disparity standard deviation of 2.8 pixels achieving the lowest RMSE.

Figure 6 shows the average reconstruction error (ARE) of input images as well as the reconstruction error of sparse coding tested with input only at disparity zero and with input only at disparity ten, where there is no overlap between left eye and right eye images. ARE is calculated in such a way that after initialized to an input disparity, the system controls itself by actions generated by reinforcement learning, while the reconstruction errors at each time step during the testing

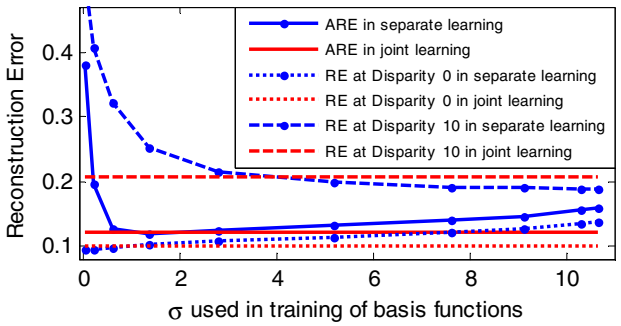


Figure 6: The reconstruction error as a function the input standard deviation of the disparities used to train the sparse coding bases during separate learning.

are averaged. ARE should be a weighted average of the reconstruction errors at different disparities as are shown in Figure 4, where the weights are the statistics of disparity distribution which is controlled by action policy as is shown in Figure 3(b). Neither good reconstruction, nor good policy guarantees low ARE. This can be seen in Figure 6 and ARE achieves the lowest value with basis functions trained with inputs at a standard deviation of 1.4 pixels.

C. *Joint Learning of Perception and Behavior*

Figure 5(d) shows the policy that develops under joint learning, where both the sparse coding and the NACREL algorithm run simultaneously. The policy that emerges is close to the reference policy for disparities close to zero. However, it is not as close to the reference policy The RMSE of this policy, shown as the red line in Figure 3(b), is comparable to the lowest RMSE obtained for the NACREL running on previously developed basis functions. The average reconstruction error of joint learning, shown by the red line in Figure 6, is also comparable to the lowest value found for separate learning.

IV. Discussion

This paper has introduced a unified model for the joint development of visual perception and eye movement control. We have shown that model binocular disparity selective cells can develop simultaneously with a vergence control strategy that maintains binocular fixation. These two developmental processes are coupled. The vergence control policy interacts with the environmental statistics to determine the statistics of the retinal disparities seen by the disparity selective neurons. These statistics determine the development of the basis functions modeling the receptive fields. At the same time, vergence control commands are generated solely based on the output of the disparity selective neurons. The two developmental processes are unified in that they both seek to optimize the same objective: minimal reconstruction error under a fixed complexity constraint.

We emphasize that the development of vergence control is not an explicit objective of this system, but rather emerges through the system's efforts to learn how to behave so that its inputs become easily encoded. The reason a vergence control

policy emerges can be explained through Figure 4. At the beginning of training, the vergence control policy is essentially random and we expect retinal disparities to have a distribution with a large standard deviation. Nonetheless, for inputs with this large spread, the reconstruction error achieves a minimum value, albeit shallow, for inputs with zero disparity. By seeking to minimize reconstruction error, the NACREL learns a vergence control policy that tends to move the retinal disparity towards zero. This leads to input disparity statistics with a smaller standard deviation, which results in basis functions for which the minimum at zero disparity is deeper, further favoring policies which drive the retinal disparity towards zero.

Our experimental results have demonstrated that this joint process results in basis functions that are comparable to those produced when the sparse coding process operates alone (Figure 4), as well as a vergence control policy that achieves performance comparable to the best policy learned using fixed basis functions (Figure 3(b)). This jointly learned policy also results in the lowest average reconstruction error over time, since vergence control policy can adapt to changes in the neural responses due to changes in the basis functions, while the basis functions can also adapt to best encode inputs with the disparity distribution determined by the vergence control policy.

To our knowledge, this is one of the first models of unified joint development. Past work has either studied the development of perception in isolation[2][12][13], behavior in isolation [14][15][16], or the development of perception and behavior through processes decoupled both in time and objective, i.e. where the development of behavior started only after the development of perception had concluded, and the two developmental processes sought to optimize different objective functions [17][18]. A notable exception can be found in [20]. The present work has shown that by coupling both perceptual and behavioral development through the same objective function and allowing the two to evolve simultaneously, we create a positive feedback loop that ensures the process proceeds robustly.

Evidence from development in human infants lends support for a coupled model as we suggest here. First, it appears that visuomotor experience is essential to calibrate the eye movement systems for saccades and smooth pursuit, since the optics of the eye and the retinal receptive field locations undergo large postnatal changes. Second, accurate eye movements seem to co-develop with visual perception. The eye movement control of infants at birth is quite undeveloped, and does not match that of adults until about four months. For example, the accuracy of vergence eye movements in young infants seems to be limited by the deficiencies in disparity perception [19].

Moving forward, this model may lay the groundwork for algorithms for self-calibrating robotic systems that have no need for manual calibration, which is costly and must be performed every time the robot's physical configuration changes. The general framework we have proposed seems widely applicable to other visual-motor tasks, e.g. the joint development of motion perception and smooth pursuit.

ACKNOWLEDGEMENTS

This work was supported in part by the Hong Kong RGC and the German DAAD through the Germany/Hong Kong Joint Research Scheme (project number G_HK25/10).

REFERENCES

[1] R. Held and A. Hein, "Movement-produced stimulation in the development of visually guided behavior," *Journal of Comparative and Physiological Psychology,* vol. 56, pp. 872-876, 1963.

[2] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?" *Vision Res.,* vol. 37, pp. 3311-3325, 1997.

[3] C. A. Rothkopf, T. H. Weisswange and J. Triesch, "Learning independent causes in natural images explains the spacevariant oblique effect," in *IEEE International Conference on Development and Learning*, 2009.

[4] J. Schmidhuber, "Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990–2010)," *IEEE Transactions on Autonomous Mental Development,* vol. 2, pp. 230-247, 2010.

[5] M. Luciw, V. Graziano, M. Ring and J. Schmidhuber, "Artificial curiosity with planning for autonomous perceptual and cognitive development," in *IEEE International Conference on Development and Learning,* 2011.

[6] D. H. Hubel and T. N. Wiesel, "Binocular interaction in striate cortex of kittens reared with artificial squint," *J. Neurophysiol.,* vol. 28, pp. 1041-1059, 1965.

[7] A. Franz and J. Triesch, "Emergence of disparity tuning during the development of vergence eye movements," in *IEEE 6th International Conference on Development and Learning,* 2007, pp. 31-36.

[8] W. Sun and B. E. Shi, "Joint development of disparity tuning and vergence control," in *IEEE International Conference on Development and Learning,* 2011.

[9] J. H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proceedings of the Royal Society of London.Series B: Biological Sciences,* vol. 265, pp. 359-366, 1998.

[10] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing, ,* vol. 41, pp. 3397-3415, 1993.

[11] S. Bhatnagar, R. S. Sutton, M. Ghavamzadeh and M. Lee, "Natural actor-critic algorithms," *Automatica,* vol. 45, pp. 2471-2482, 2009.

[12] H. B. Barlow, "Possible principles underlying the transformation of sensory messages," in *Sensory Communication*, W. A. Rosenblith, Ed. Cambridge, MA: MIT Press, 1961.

[13] J. J. Atick and A. N. Redlich, "Towards a theory of early visual processing," *Neural Comput.,* vol. 2, pp. 308-320, 1990.

[14] W. Schultz, P. Dayan and P. R. Montague, "A neural substrate of prediction and reward," Science, vol. 275, pp. 1593-1599, 1997.

[15] N. D. Daw and K. Doya, "The computational neurobiology of learning and reward," Curr. Opin. Neurobiol., vol. 16, pp. 199-204, 2006.

[16] E. S. Bromberg-Martin, M. Matsumoto and O. Hikosaka, "Dopamine in motivational control: rewarding, aversive, and alerting," *Neuron,* vol. 68, pp. 815-834, 2010.

[17] R. Legenstein, N. Wilbert and L. Wiskott, "Reinforcement learning on slow features of high-dimensional input streams," PLoS Computational Biology, vol. 6, pp. e1000894, 2010.

[18] N. J. Gustafson and N. D. Daw, "Grid Cells, Place Cells, and Geodesic Generalization for Spatial Reinforcement Learning," PLoS Computational Biology, vol. 7, pp. e1002235, 2011.

[19] R. N. Aslin, "Anatomical constraints on ocular motor development: Implications for infant perception," in *Perceptual Development in Infancy*, A. Yonas, Ed. Hillsdale, N.J.: L. Erlbaum Associates, 1988, pp. 67-99.

[20] S. Saeb, C. Weber, and J. Triesch, "Goal-directed learning of features and forward models," *Neural Networks* vol. 22, pp.586-592, 2009.