2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Intrinsically Motivated Learning of Visual Motion Perception and Smooth Pursuit

Chong ZHANG, Yu ZHAO, Jochen TRIESCH and Bertram E. SHI

***Abstract*—Developmental robots require cognitive structures that can learn perception-action cycles via interactions with the environment. Here, we extend the efficient coding hypothesis, which has been used to model the development of sensory processing in isolation, to model the development of the perception-action cycle. Our extension combines sparse coding and reinforcement learning so that sensory processing and behavior co-develop to optimize a shared intrinsic motivational signal: the fidelity of the neural encoding of the sensory input under resource constraints. Applying this framework to a model of a robot actively observing a time-varying environment leads to the simultaneous development of visual smooth pursuit behavior and model neurons similar to cortical neurons selective to visual motion. We suggest that this general principle may form the basis for a unified and integrated approach to learning many other perception/action loops.**

I. Introduction

Developmental robotics seeks to enable robots to learn to organize their interactions with the environment automatically as they interact with it, much like a child [1]. It draws inspiration from fields such as neuroscience, developmental psychology, and sociology, as well as artificial intelligence and robotics. A key focus is the design a self-developing cognitive structure that enables the robot to represent itself and its environment and to use this representation to guide intelligent interactions with the environment. In contrast, most current approaches to robotics embed the robot designer's understanding of the physics of the robot and how it should interact with its environment explicitly within a fixed control structure. Thus, progress in developmental robotics may enable robots that are considerably more flexible and adaptive than they are currently.

In this paper, we suggest that the efficient encoding hypothesis may be one important component of such a self-developing cognitive structure. The efficient encoding hypothesis posits that neurons respond so that they best represent sensory data while requiring as few neurons to respond as possible [2]. Mathematically, this concept has been captured by sparse coding algorithms, which seek to represent input vectors as sparse linear combinations of basis functions drawn from a possibly overcomplete dictionary. Basis functions from dictionaries learned to represent natural input stimuli closely resemble the receptive fields of neurons found in the primary sensory cortices for vision [3] and audition [4]. The concept of efficient encoding is attractive as a model for biological neural systems, since neural firing is metabolically expensive.

A key prediction of the efficient encoding hypothesis and sparse coding models is that the sensory processing in the brain reflects the statistics of the sensory input. An important question then is what determines these statistics. The composition of the natural environment is certainly a key determinant of the statistics. However, this picture is incomplete without accounting for the equally important effect of behavior, which shapes the statistics by directing the sensory apparatus preferentially towards some regions of the environment. Since the behavior may be driven by sensory neurons, it is highly likely that behavior and sensory processing co-develop in an organism. To date, this problem of co-development has received little attention.

In this paper, we address the problem of co-development of behavior and perception by extending the efficient encoding hypothesis to include the effect of behavior. Our model posits that not only does sensory processing develop to best represent its input under constraints of limited resources, but behavior develops simultaneously to shape the statistics of the input so that it easier to encode. Both perception and behavior develop so as to maximize the fidelity of the sensory representation, as measured by reconstruction error. This developmental model is intrinsically motivated, since the reconstruction error is generated within the agent.

We study this problem specifically in the context of visual motion perception and smooth pursuit behavior. This is an ideal test bed for this model, since evidence indicates that motion perception and smooth pursuit co-develop. Smooth pursuit in infants does not appear until about 2 months of age, consistent with the onset of behavioral motion discrimination [5]. It becomes more prominent and refined with age, concurrently with the development of more advanced motion extrapolation mechanisms [6]. Our experiments reported here indicate that both smooth pursuit and motion perception can develop as emergent properties of our framework.

II. Methods

The embodiment of our framework for smooth pursuit eye movements is illustrated in Figure 1. An eye which can rotate in both pan and tilt senses a time varying pattern of light projected onto its retina from the environment. The environment contains targets moving in the environment at random speeds.

This work was supported in part by the Hong Kong RGC (619111), the HKRGC/ German DAAD (G HK25/10), the European Community (FP7-ICT-IP-231722) and the German BMBF (FKZ 01GQ0840).

Chong ZHANG, Yu ZHAO and Bertram E. SHI are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon Kong Hong (email: eebert@ust.hk). BES is also with the Division of Biomedical Engineering.

Jochen Triesch is with the Frankfurt Institute for Advanced Studies in Frankfurt am Main, Germany.

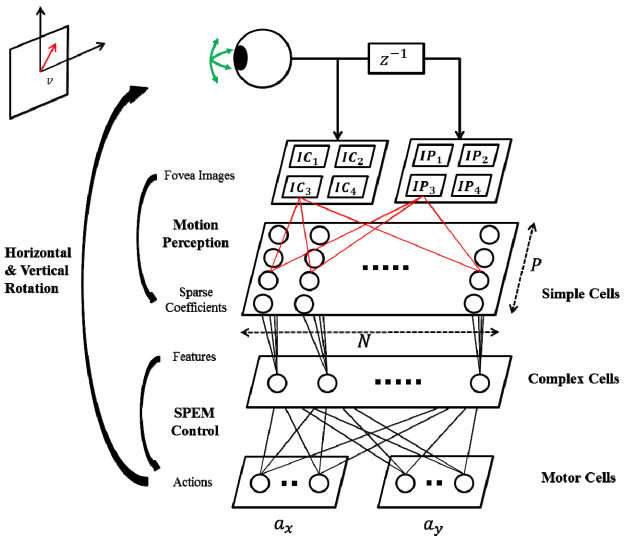


Figure 1 The developmental model for smooth pursuit. P is the number of patches we extract in the fovea region, N is the size of the dictionary, a_x and a_y are actions controlling the pan and tilt of the eye, IP and IC are fovea images obtained at previous and current time. For clarity we show four patches, but our systems uses 100.

The time varying retinal images from a foveal region are divided into overlapping spatial patches. Each patch is passed to a sparse coding stage, which represents information from the patch in the current and previous frame by a sparse linear combination of spatio-temporal basis functions. In a biological neural system, a possible substrate for this computation are the simple cells in the primary visual cortex. Each basis function corresponds to one simple cell, whose response at time t encodes the coefficient of the basis function in the linear combination representing the input at time t [3].

The simple cells for different patches share the same basis functions. Coefficients corresponding to the same basis functions are squared and pooled across all patches to yield a response analogous to that of a complex cell in the visual cortex. Complex cells are thought to achieve position invariance by combining the responses from many simple cells whose receptive fields have similar properties, but at slightly different retinal positions [7]. The complex cell responses are then fed to neural networks which determine the probability of different actions which may be taken to alter the rotational velocities of the eye.

The basis functions of the sparse coding stage and the weights of the action neural networks are initialized randomly. Thus, sensory processing and behavior at the start of development are independent of the environment. The basis functions and weights co-evolve to maximize the same criterion: fidelity of the linear combination in the sparse coding stage. As described in more detail below, basis functions are updated similarly to past work in sparse coding. The weights of the actor neural network are updated using the natural actor-critic reinforcement learning algorithm, where the reward to be maximized is the discounted sum of the negative of the current and future reconstruction errors.

We emphasize that the criterion optimized in the model is generic, and is not specific to smooth pursuit. Nonetheless, our experimental results in Section 3 will demonstrate (1) that the system develops smooth pursuit tracking behavior and (2) that the basis functions that develop have properties similar to those associated with the linear spatio-temporal receptive field profiles of motion tuned neurons in the primary visual cortex. In addition, because the smooth pursuit behavior shapes the statistics of the retinal slip (the difference between the projected target velocity and the eye rotational velocity) towards low retinal image velocities, the distribution of velocity tunings in the basis functions is biased towards lower retinal image velocities than would be presented to a stationary eye in the same environment. In the discussion, we suggest that this finding may provide a developmental explanation for the presence of a perceptual bias towards lower image velocities that has been found in psychophysical experiments.

The following describes the individual components of this model in more detail. We discretize time so that foveal images are presented at a sampling rate of 30 frames per second. We assume that the foveal region covers 11 degrees of visual angle, and is sampled spatially at 5 pixels per degree. Thus, the dimension of each foveal image is 55 by 55 pixels. Each fovea image is further divided into $P = 100$ 10 by 10 pixel patches, each covering two degrees of visual angle (10 by 10 pixels). The patches cover the whole fovea image with 1 degree (5 pixel) overlap between neighboring patches both horizontally and vertically.

A. *Sparse coding*

Sparse coding is applied to each patch independently, but all patches share the same dictionary, in the same way that neighboring image patches in a convolutional neural network share the same weights. For each patch, the sparse coding seeks to jointly encode information from both the current and previous time step. Thus, the input vector is 200 dimensional (10 by 10 pixels by 2 frames). We combine information from two frames because direction selective neurons combine input from two distinct sub populations of cells whose temporal peak responses (68 and 93ms) differ by 25ms [8], which corresponds to about one frame in our simulations.

We denote the input vector of image intensities at time t from patch $i \in \{1,...,P\}$ by $x_i(t)$. We reconstruct each input as a weighted sum of unit norm basis functions taken from an over-complete dictionary $\phi_n(t)$, where $n \in \{1,...,N\}$ indexes the basis functions. For convenience, we fix $N = 300$, but it would be interesting to investigate making the size adaptive. The approximation is given by

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

We use the matching pursuit algorithm proposed by Mallat and Zhang [9] to choose the coefficients $a_{i,n}(t)$ such that at most 10 of them are nonzero. We update the dictionary of bases using an on-line two step procedure similar to that used by Olshausen [3]. 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

bases to minimize the average normalized squared reconstruction error over patches

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

Since all patches share the same dictionary, basis function updates are first pooled over patches before updating the dictionary. After each update, the bases are re-normalized so that they are unit norm.

Each basis function yields one complex cell response, resulting in a feature vector

$$\mathbf{f}(t) = \left[f_1(t) \quad f_2(t) \quad \ldots \quad f_N(t)\right]^T, \quad (3)$$

where

$$f_n(t) = P^{-1}\sum_{i=1}^{P} a_{i,n}(t)^2. \quad (4)$$

B. Reinforcement learning

The weights in the action network evolve according to the natural actor critic reinforcement learning algorithm [10]. Reinforcement learning algorithms seek a policy mapping the current state of the agent to a probability distribution over potential actions by the agent, such that the discounted sum of future rewards is maximized. In our framework, the state of the agent is represented by the complex cell feature vector **f**(*t*). The actions are accelerations applied to modify the rotational velocities of the eye in the pan and tilt directions. The instantaneous reward at time *t* is the negative of the squared reconstruction error (2). The discount factor is 0.3.

We use neural networks to represent the policy (the actor) and the value function (the critic). The input to the neural networks are the complex cell feature vector **f**(*t*). We compared the use of two different policy parameterizations: a softmax policy and a Gaussian policy.

For the softmax policy, actions are chosen from a discrete set of *K* possible actions whose probabilities are given by the outputs of a linear network with a softmax output nonlinearity, where the number of outputs is *K*. Since separate actions update the rotational velocities of the pan and tilt axes of the eye, we use two separate neural networks for each axis. Denoting the probability of choosing the i^{th} action at time *t* by $\pi_i(t)$,

$$\pi_i(t) = \frac{\exp\left(z_i(t)/T\right)}{\sum_{j=1}^{K}\exp\left(z_j(t)/T\right)}, \quad (5)$$

where the temperature *T* is a positive scalar controlling the entropy of the policy (see section 2.3 of [11]), and $z_i(t)$ is the activation of the i^{th} output neuron, which is computed by

$$z_i(t) = \boldsymbol{\theta}_i^T(t)\mathbf{f}(t) \quad (6)$$

where $\boldsymbol{\theta}_i(t)$ denotes the weights connecting to the i^{th} neuron.

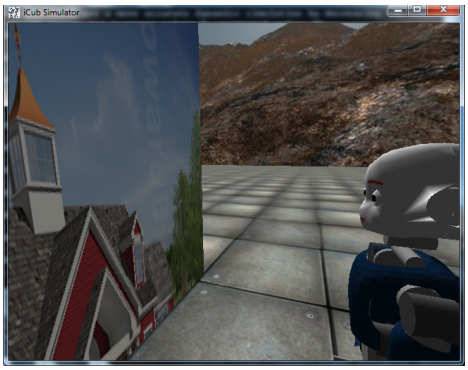


Figure 2. A screenshot of the iCub simulator environment

For the Gaussian policy, continuous-valued actions are chosen by sampling from a Gaussian distribution whose mean varies with **f**(*t*) but whose variance is fixed. The mapping from **f**(*t*) to a mean vector is performed by a three layer neural network, with *N* neurons in the input layer, *H* = 5 neurons in the hidden layer whose activation functions are hyperbolic tangents, and two neurons in the output layer, which is linear. The two neurons encode the means of 1D Gaussian distributions describing the distributions of the pan and tilt acceleration actions.

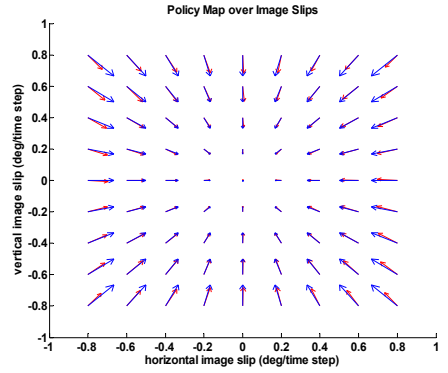
C. Environment model

We use two different models for the robot and environment. Most experiments reported in the next section are with using the first model. The second model was used in the final set of experiments reported in Figure 7.

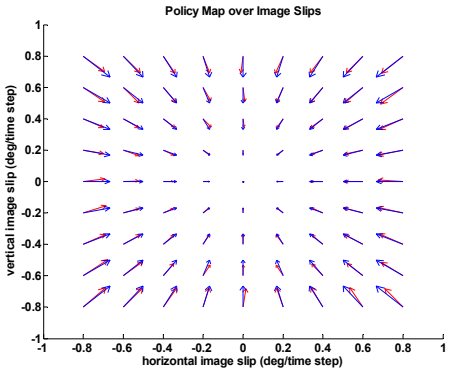
In the first model, we assume that the targets are textures applied to a sphere centered at the center of the eye and extend to fill the fovea. In this case, object translations can be modelled simply by shifting the image. For textures, we use 20 images taken from van Hateren database [12]. Every 1/3 of a second (10 frames), we change the target texture and choose a new rotational velocity for the target. The rotational velocity is kept constant until the next change. We constrained the target rotational velocity to be at most 24 deg/s in the pan and tilt directions. The rotational velocity of the eye was constrained to the same range, which is consistent with the maximum speed of around 30 deg/s for smooth pursuit in humans. The maximum retinal slip is thus 48 deg/s in each dimension, if the eye and target move in opposite directions. Using our discretization, this corresponds to a range of horizontal and vertical velocities ranging between +8 and -8 pixels per frame.

The eye's rotational velocity at time t is obtained by taking the rotational velocity at time *t*-1 and adding the acceleration action obtained by sampling from the policy distribution. For the softmax policies, the network encodes *K* = 11 commands, which are equally spaced accelerations ranging from -900 to +900 deg/s^2 (-5 to +5 pixels/frame^2).

For the second model, we used the iCub simulator [13]. This is a more realistic model, as it includes the effects of perspective projection in image formation. We move a planar object in the frontal parallel plane one meter away from the robot at a random velocity for 2/3 second before changing

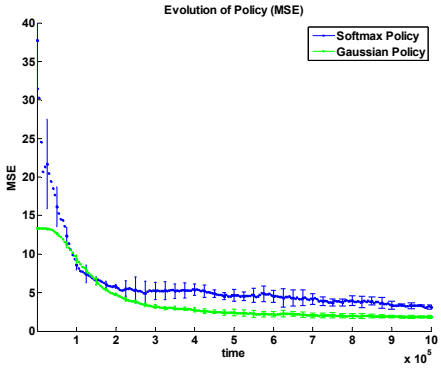


(a)

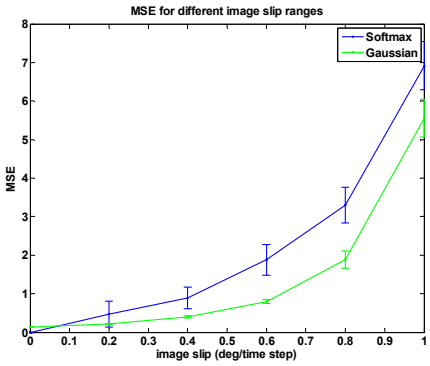


(b)

Figure 3 The policy obtained at the end of training for the Gaussian (a) and softmax (b) policy networks. The base of each arrow indicates the retinal slip. The direction and magnitude of each arrow shows the change in image slip after changing the eye's rotational velocity according to the policy assuming the target velocity remains constant. Blue arrows indicate an ideal policy, which would zero out retinal slip in one time step. Arrows point directly towards the origin with length proportional to the magnitude of image slip. Red arrows indicate the effect of the average command taken by the final policy.



(a)



(b)

Figure 4(a) Evolution of the mean squared error (MSE) between the learned policy and the "ideal" policy during training. (b) MSE for inputs with different magnitudes of image slip. Error bars indicate the standard deviation in the MSE over trials.

textures. The horizontal and vertical velocities are chosen independently from a uniform distribution between ±0.5m/s. The rotational eye velocities were constrained between ±60 deg/s. The environment is shown in Figure 2.

III. Results

We simulated the model using both the softmax and Gaussian policies starting from random initial conditions. For each case, we conducted three trials. The total processing time required for sparse coding and reinforcement learning is around 10ms per frame running in MATLAB on a 3.3GHz i5 PC with 8GB RAM.

Figure 3 shows the effect on retinal slip in the next time step for the average command taken by the final policy for different initial values of retinal slip. This figure illustrates clearly the emergence of a smooth pursuit behavior, since the average actions taken by both the softmax and Gaussian policies are quite close to those of an "ideal" smooth pursuit policy that would zero out retinal slip in one time step. Both models led to the emergence of smooth pursuit in all trials, indicating that the exact structure of the model is not critical for the emergence of this behavior. The range of slips tested ranged between -0.8 to 0.8 deg/frame (-24 to 24 deg/s) in each direction in steps of 0.2 deg/frame (6 deg/s). For each value of retinal slip, the output of the policy network may vary with the foveal image content. To reduce this variability, we averaged the greedy action over 50 pairs of current and past foveal images with the same retinal slip. For the Gaussian policy, the greedy actions were the means of the Gaussians for pan and tilt. For the softmax policy, the greedy actions corresponded to the maximum output of the softmax network. The foveal images were obtained by taking pairs of 55 by 55 pixel subwindows from images in the van Hateren database not used during training, where the locations of the subwindows differ by a translation corresponding to the image slip. We also averaged over the three trials.

We can obtain a single quantitative measure of smooth pursuit behavior using the mean squared error (MSE) between greedy action taken by learned policy and the action of the ideal policy that would zero out retinal slip in one time step. The MSE is obtained by averaging over three dimensions. First, we obtain an overall measure by averaging over the 81 different retinal slip conditions shown in Figure 3. Second, we reduce the effect of input variability by averaging over the squared error for 50 actions taken for different pairs of current and past foveal images. Finally, we reduce the effect of stochastic variability introduced by differences in the random

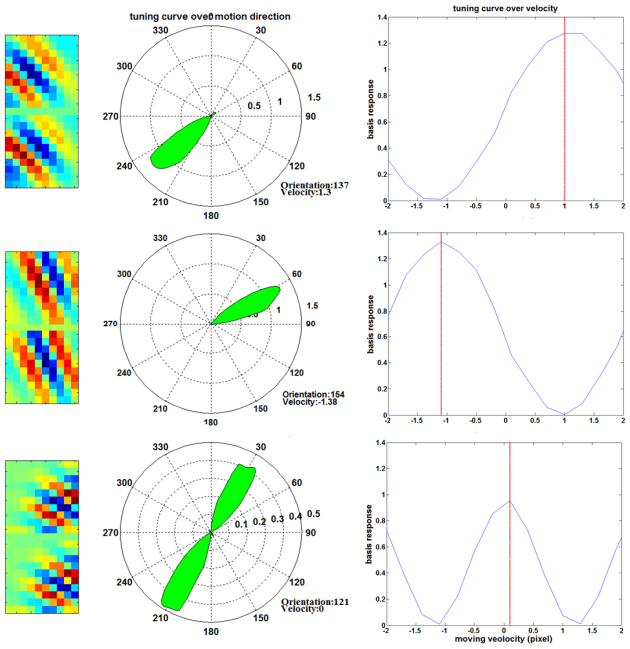


Figure 5 Three representative learned basis functions. The first column shows the spatiotemporal bases as 20 by 10 images, where the top half corresponds to the patch input from time *t*-1 and the bottom to time *t*. The second column shows the directional tuning curve for each basis in polar coordinates where radius indicates response magnitude and angle indicates input velocity direction. The third column shows the velocity tuning curve (blue) for each basis. The red line indicates the peak tuning predicted by the fitted Gabor parameters.

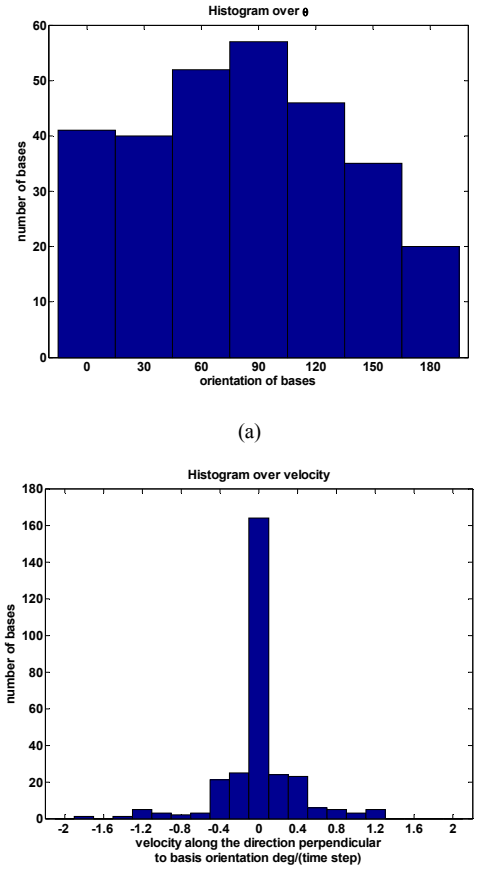


Figure 6 Histograms of the preferred orientations and velocities of the learned bases.

initial conditions and choices of target velocities by averaging over the three trials.

Figure 4(a) shows the evolution of the MSE during training for the softmax and Gaussian policies evaluated at 200 equally spaced points along the training. In the beginning of the simulations, the rotational velocity of the eye is essentially random. The initial MSE of the softmax policy is much larger than the initial MSE of the Gaussian policy. For the Gaussian policy, the small random initial weights make the mean action close to zero for almost all images. On the other hand, for the softmax policy the actions are chosen randomly over the entire discrete action range. In both cases, we observe a gradual reduction in MSE as the smooth pursuit behavior emerges.

The final MSE of the soft max policy is larger than that of the Gaussian policy. Figure 4(b) shows the MSE of the final policy evaluated for image slips with different magnitudes. The Gaussian policy is both more accurate (lower MSE) and more precise (lower standard deviation). The advantage becomes more significant as the retinal slip magnitude increases.

Thus, while both policies result in the emergence of smooth pursuit, the Gaussian policy has several advantages over the softmax policy. In addition to the performance advantages noted above, the Gaussian policy also (1) requires fewer parameters: $(N+2)H$ versus $2NK$ where N=300 is the number of basis functions, H=5 is the number of hidden units and K=11 is the number of discrete actions, (2) does not require us to predefine the possible set and range of actions, which may bias the results of learning, and (3) can give continuous rather than discrete actions.

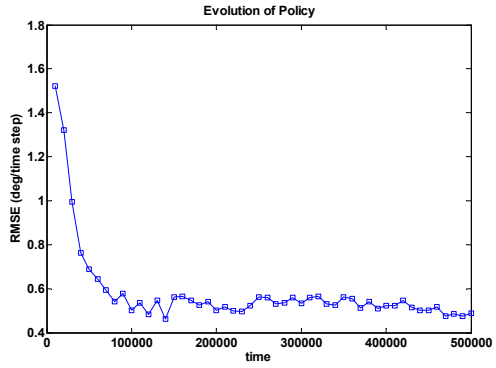
Figure 5 shows three representative bases and their tuning curves obtained using the Gaussian policy. We observe that the bases exhibit spatio-temporally profiles very similar to the spatio-temporal Gabor functions used to model the spatio-temporal receptive fields of motion selective visual cortical neurons. This is true in general over all bases (not shown due to space constraints.) We fitted the learned bases using 2D spatial Gabor functions for current and previous frames, assuming the same parameters except for a phase shift, which determines the velocity tuning. The MSE of the residual was around 0.06, indicating excellent fits (recall that bases have unit norm).

We calculated direction and velocity tuning curves for the bases by computing their squared correlation with moving cosine grating with different orientations, directions and velocities. For the direction tuning curves, gratings at the optimal spatio-temporal frequencies were used. For the velocity tuning curves, gratings at the optimal orientation and spatial frequency were used. The first two bases are tuned to similar velocities, but different directions. The third basis is tuned to velocities near zero, and shows little direction selectivity.

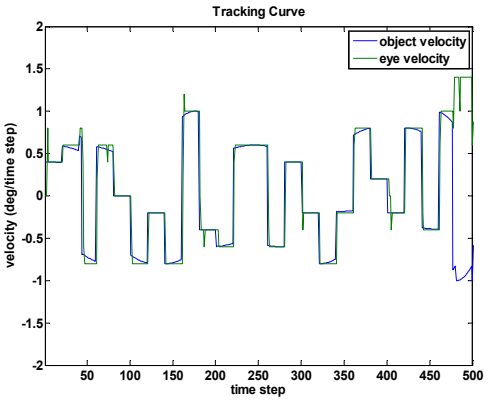
Figure 6 shows the histograms of preferred orientations and velocities computed over all bases whose Gabor fit error was smaller than a threshold (0.3). Preferred orientations and velocities were computed based on the fitted Gabor parameters. The peak tuning velocity is given by $2\pi\Delta\varphi/\lambda$, where $\Delta\varphi$ is the phase difference between Gabor fits at time t and t-1 and λ is the spatial wavelength. The distribution of orientations is close to uniform, and most of the bases are tuned to velocities close to zero. This velocity tuning distribution is a consequence of the joint development of perception and behavior. If the eye were stationary, the statistics of retinal slip velocities would be the same as the statistics of the target velocities (evenly distributed between -0.8 and 0.8 deg/s). The efficient encoding hypothesis without our extension to include behavior would predict a more uniform distribution of velocity tunings.

In the supplemental material, we have provided a movie tracing the simultaneous evolution of both the basis and policy. At the beginning, basis functions begin with random initial conditions, and the eye movements are random. Some of the basis functions evolve quite quickly to exhibit spatio-temporal Gabor like profiles, and the system learns to steer eye movements so that the fovea follows the general direction of the target albeit with some errors. As more basis functions develop Gabor like profiles, the smooth pursuit behavior improves. The video also illustrates how the distribution of basis functions shifts to become more concentrated around zero velocity as smooth pursuit behavior improves.

Figure 7(a) shows the development of smooth pursuit during training using the iCub simulator platform with the soft max policy. Each data point shows the performance of the learned basis and policy weights at the corresponding iteration. Performance was measured by applying the fixed basis and policy weights to the same set of test images and object trajectories. We computed the RMSE between the eye's angular velocity and the object's angular velocity over 10,000 time steps. As training goes on, the tracking performance becomes better and finally stabilizes at around 0.5 degrees per time step. Figure 7 (b) shows the tracking performance at the end of 500,000 iterations of training. The eye velocity follows the object velocity, although small changes in the object velocity cannot be tracked due to the discrete nature of the actions in the soft max network.



(a)



(b)

Figure 7 (a) The evolution of the smooth pursuit behavior for the soft max policy trained on the iCub simulator. (b) An example trajectory for the pan angle of the eye for the basis set and policy found at the end of training. To facilitate comparison, the object velocity is converted into the equivalent angular velocity of the eye required to maintain perfect tracking at the fovea center.

IV. DISCUSSION

We have described a framework for the intrinsically motivated co-development of sensory processing and behavior. To our knowledge, this framework is the first extension of the efficient encoding hypothesis to include the development and effects of behavior. We have demonstrated that this framework results in the emergence of neurons selective to visual motion via mechanisms similar to those hypothesized to be used by visual cortical neurons, as well as smooth pursuit behavior, in a model of an active robot vision system. We emphasize that this development is an emergent phenomenon, governed by the properties of the environment and the agent's interactions with it. Nothing in the model itself is specific to motion perception or smooth pursuit. Both sensory processing and behavior co-develop according to a generic criterion: maximizing the fidelity of the neural representation of the sensory input under limited resource constraints.

The smooth pursuit model we have presented is consistent with current hypotheses about the neural mechanisms of smooth pursuit. Smooth pursuit eye movements can be divided into two stages, which are divided by a catch-up saccade. In presaccadic pursuit, the eye accelerates to match the velocity of the target. The catch up saccade centers the target in the fovea. In postsaccadic pursuit, velocity matching by smooth pursuit is more accurate. Recent work by Wilmer and Nakayama suggests that pre-saccadic pursuit may be driven by low-level motion-energy-based signals, whereas post-saccadic pursuit may be driven by high-level feature or position based signals [14]. Our model is consistent with the behavior in presaccadic pursuit. The sensory neurons driving the pursuit signal have basis functions and tuning curves similar to the receptive fields and tuning curves of neurons in the primary visual cortex, which are often modeled using motion energy mechanisms (Figure 5).

By stabilizing the image on the retina, smooth pursuit eye movements make the time varying image sequences easier to encode by increasing the correlation between current and

delayed visual inputs. This stabilization reduces retina slip, resulting in a population of neurons with motion selectivities biased near zero velocity (Figure 6). This result is consistent with Bayesian models of visual motion perception which posit a velocity prior that favors low speeds [15], which have been used to account for psychophysical experiments showing that low contrast stimuli appear to move slower than high contrast stimuli [16][17]. Thus, this model provides a possible explanation for the developmental origins of this velocity prior.

One of the most promising avenues for future development of this framework is that it may eventually provide a unified account for the development of a wide range of eye movement behaviors. For example, it has already been shown to account for the joint development of stereo disparity tuned neurons and vergence behavior [18]. The results in this paper provide evidence for the generality of this framework. Simply by changing the input images and output actions, we have demonstrated the emergence of a different behavior. In addition, we have also demonstrated that the framework applies under more complex action spaces (2D smooth pursuit vs 1D vergence), and that the results are robust to the specific choice of the action network.

While different types of eye movement, e.g. pursuit, saccades and vergence are often examined separately, recent studies suggest that these systems share extensive amounts of neural circuitry in their implementation. For example, the majority of pursuit neurons in the frontal eye field (FEF) discharge not only for frontal pursuit but also for vergence eye movements [19]. It has also been demonstrated that pursuit and saccades are not controlled entirely by different neural pathways but rather by similar networks of cortical and subcortical regions, and even by the same neurons in some cases [20]. By providing an integrated account of different behaviors, the proposed intrinsically motivated learning framework for efficient coding in active perception may also shed light on the development of this shared processing.

In addition to the hope for a better understanding of the joint development of visual perception and behavior in humans, we believe that our model is also of interest in the current search for artificial systems such as robots that autonomously learn and adapt to their environment. In that regard, one essential property of our model is that it is fully self-calibrating. In an analysis of the vergence model in [18], Lonini et al. showed that the model can adapt to different kind of perturbations such as blur, or eye misalignments [21][22]. This property is an important step towards autonomous robots capable of open ended learning.

ACKNOWLEDGMENT

We would like to give special thanks to Luca Lonini, Constantin Rothkopf and Céline Teulière for their helpful discussions.

REFERENCES

[1] M. Asada, K. F. MacDorman, H. Ishiguro and Y. Kuniyoshi, "Cognitive developmental robotics as a new paradigm for the design of humanoid robots," Robotics and Autonomous Systems, vol. 37, pp. 185-193, 2001.

[2] H. Barlow, "Possible principles underlying the transformation of sensory messages," *Sensory Communication*, MIT Press, pp. 217-234, 1961

[3] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?" *Vision Research*, vol. 37, no. 23, pp. 3311-3325, 1997.

[4] M. S. Lewicki, "Efficient coding of natural sounds," *Nature: Neuroscience*, vol. 5, no. 4, pp. 356–363, 2002

[5] B. Oliver, A. Janette, and W. B., "Normal and anomalous development of visual motion processing: motion coherence and dorsal-stream vulnerability," *Neuropsychologia*, vol. 41, no. 13, pp. 1769-1784, 2003.

[6] V. H. Claes and R. Kerstin, "Development of smooth pursuit tracking in young infants," *Vision Research*, vol. 37, no. 13, pp. 1799-1810, 1997.

[7] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology*, vol. 160, no. 1, pp. 106-154, 1962.

[8] R. L. De Valois and N. P. Cottaris, "Inputs to directionally selective simple cells in macaque striate cortex," *Proceedings of the National Academy of Science*, vol. 95, no. 24, pp. 14488-93, 1998.

[9] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing* , vol. 41, no. 12, pp. 3397-3415, 1993.

[10] S. Bhatnagar, R. Sutton, M. Ghavamzadeh, and M. Lee, " Natural actor critic algorithms," *Automatica*, vol. 45, no. 11, pp. 2471–2482, 2009.

[11] R.S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction,", MIT Press, 1998.

[12] 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, no. 1394, pp. 359-366, 1998.

[13] V. Tikhanoff, A. Cangelosi, P. Fitzpatrick, G. Metta, L. Natale, F. Nori, "An open-source simulator for cognitive robotics research: The prototype of the icub humanoid robot simulator," in: *Proceedings of IEEE Workshop on Performance Metrics for Intelligent Systems Workshop*, ACM, 2008.

[14] J. B. Wilmer and K. Nakayama, "Two distinct visual motion mechanisms for smooth pursuit: Evidence from individual differences," *Neuron*, vol. 54, no. 6, pp. 987-1000, 2007.

[15] Y. Weiss, E. Simoncelli, and E. Adelson, "Motion illusions as optimal percept," *Nature Neuroscience*, vol. 5, no. 6, pp. 598–604, 2002.

[16] P. Thompson. "Perceived rate of movement depends on contrast," *Vision Research*, vol. 22, no. 3, pp. 377–380, 1982.

[17] L.S. Stone and P. Thompson, "Human speed perception is contrast dependent," *Vision Research*, vol. 32, no. 8, pp. 1535–1549, 1992.

[18] Y. Zhao, C. Rothkopf, J. Triesch and B. E. Shi ," A Unified Model of the Joint Development of Disparity Selectivity and Vergence Control," *IEEE Joint International Conference on Development and Learning – Epigenetics and Robotics*, San Diego, CA, USA, Nov. 2012.

[19] K. Fukushima K, T. Yamanobe, Y. Shinmei, J. Fukushima, S. Kurkin and B. W. Peterson, "Coding of smooth eye movements in three-dimensional space by frontal cortex," *Nature* vol. 419, no. 6903, pp. 157-162, 2002

[20] R. J. Krauzlis, "The control of voluntary eye movement: new perspectives," *Neuroscientist*, vol. 11, no. 2, pp. 124-137, 2005.

[21] L. Lonini, S. Forestier, C. Teulière, Y. Zhao, B. E. Shi and J. Triesch, "Robust Active Binocular Vision through Intrinsically Motivated Learning," *Frontiers in Neurorobotics*, 2013.

[22] L. Lonini, Y. Zhao, P. Chandrashekhariah, B. E. Shi and J. Triesch, "Autonomous learning of active multi-scale binocular vision", *IEEE Joint International Conference on Development and Learning – Epigenetics and Robotics*, 2013.