

Cortical Encoding of Retinal Output from Natural Scenes with Sparse Representation

Wenxue Wang*
 Department of Electrical and Systems
 Engineering, Washington University
 St. Louis, MO 63130
 ww1@zach.wustl.edu
 ww1@netra.wustl.edu

B. K. Ghosh**
 Department of Electrical and Systems
 Engineering, Washington University
 St. Louis, MO 63130
 ghosh@netra.wustl.edu

Abstract—The visual cortex of a freshwater turtle, when stimulated by a pattern of light, produces waves of activity that have been recorded experimentally and simulated using a model cortex. It is believed that these activity waves encode features of the visual scene, viz. position and velocity of targets. The goal of this paper is to explore how to estimate target velocity using the activity pattern in the model cortex. We consider five natural video scenes and represent them using sparse, over-complete set of basis functions. The associated coefficients are KL-decomposed to provide appropriate cortical signals. The signals are fed as input to a model of the visual cortex and the associated cortical response of a large number of pyramidal cells are generated. Finally, the cortical response has been displayed as a spatiotemporal signal. The paper concludes with a sketch of an outline as to how the motion field of the input visual scene could be reconstructed from the activities of the cortical cells with two steps of processes: estimation of the conductance patterns of pyramidal cells from the activities of the pyramidal cells and estimation of motion field from the estimated conductance patterns of pyramidal cells.

I. INTRODUCTION

The turtle visual cortex responds to visual scenes of the natural world. It is well known that the visual cortex of freshwater turtles, when stimulated by an input pattern of visual activity, produces wave of activity. These activities have been experimentally observed assuming stationary and moving flash as an input. They have been well studied by Prechtl and his colleagues with recording from the external surface of the brain using multielectrode arrays ([10],[11]) or imaging the external surface of the cortex with voltage sensitive dyes ([12]), and by Senseman et al with imaging the ependymal surface of the cortex using voltage sensitive dyes ([13],[14],[15]). A large scale model of the cortex, the UNG model, has also been constructed with software package, GENESIS ([1]), that has the ability to simulate cortical waves with the same qualitative features as the cortical waves seen in experimental preparations. Z. Nenadic, Bijoy K. Ghosh and Philip Ulinski ([6],[7],[8]) and Xiuxia Du and Bijoy K. Ghosh ([2]) studied the dynamics, estimation and detection problems on activity of waves by simulating the UNG cortex model with input of flash pattern. It is believed that the activity waves of a turtle visual cortex encode features of the visual scenes, viz. position and velocity of targets. Wenxue

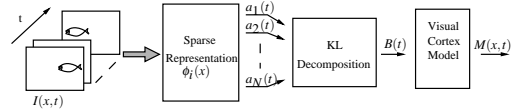


Fig. 1. System Diagram

Wang, Bijoy K. Ghosh and Philip Ulinski ([16],[17]) modified the UNG model by adding inhibitory neurons, subpial cells. In this paper, we feed visual scenes to the visual cortex model as inputs. The purpose of this paper is to estimate the information such as motion field contained in the visual input from the associated neural response by constructing a Time Varying Autoregressive and Moving Average Model. In order to simulate the cortical response, a suitable cortical input has to be constructed from visual input of natural scenes. The input to the cortex has to be of sufficiently low dimension and yet has to maintain the spatiotemporal information of visual inputs. Then cortical inputs were fed to the cortex model to produce cortical activity waves which we use to estimate the motion field in the visual space. The schematic diagram of the visual system is described in Fig 1. With Sparse over-complete representation, the natural scenes can be represented as linear superposition of a set of sparse basis functions by temporal coefficients. The temporal coefficients are treated as the activities of retinal neurons and the cortical inputs. The KL-decomposition is used to lower the dimension of the cortical inputs while maintaining the spatiotemporal information of the visual scenes and the reduced cortical inputs are fed to the WUNG model to produce cortical responses.

II. SPARSE, OVER-COMPLETE REPRESENTATION AND KL-DECOMPOSITION

Sparse representation with an over-complete basis set was proposed to explain and examine the receptive field properties in terms of a strategy for producing a sparse distribution of output response to natural images or scenes ([9]). The spatial receptive fields of simple cells in mammalian striate cortex have been reasonably well described physiologically and can be characterized as being localized, oriented, and bandpass, comparable with the basis functions of wavelet transforms. Also it is reasonable to believe that the cortex has discovered efficient coding strategies for representing natural

*Research was partially supported from NSF grant ECS-9976174

**Research was partially supported from NSF grants ECS-9976174, and ECS-0323693

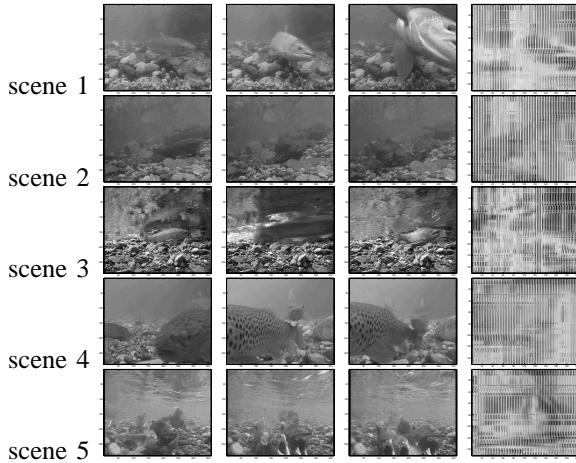


Fig. 2. Frames selected from 5 natural scenes and the corresponding dimension-reduced cortical signals generated with sparse and over-complete representation and KL decomposition. From top to bottom are scene 1 to scene 5. The columns 1 through 3 are scene frame at 1, 65 and 91 ms respectively. The right-most column are the cortical signals fed to the model of the visual cortex.

images. The sparse representation with an over-complete basis set of functions provides such an efficient representation strategy of natural images since this strategy can reduce the redundancy due to complex statistical dependencies. ‘Over-completeness’ means the number of code elements (basis functions) is greater than the effective dimensionality of the input space and ‘sparseness’ means that the representation of a given image may usually utilizes a small number of basis functions([3]). The strategy is based on a probabilistic model to capture the image structure. With this approach, an image patch, $I(x)$, is described as a linear superposition of a set of basis functions, $\phi_i(x)$, with amplitudes a_i :

$$I(x) = \sum_i a_i \phi_i(x) + v(x)$$

where x denotes spatial position within the patch and the variable v represents Gaussian noise(i.i.d.) which is included to the probabilistic model structure in the images that is not well captured by the basis functions. The basis functions, $\phi_i(x)$, may be thought of as a set of spatial features of images. The coefficients a_i represent how much of each feature is contained in the image. The basis functions are trained from the set of images with adapting the probabilistic model to statistics of images so that the distribution of images arising from the linear image model, $P(I|\Phi)$, can match as closely as possible to the actual distribution of images observed in nature, $P^*(I)$.

In this study, we used five natural scenes as visual input to the visual system. Each of natural scenes is 160 ms long and contains 160 sequential images of 288×360 pixels. The targets in the natural scenes are moving fishes. Some

example frames of the natural scenes are shown in Fig 2. Total 800 images were used to learn the set of 64 basis functions which are shown in Fig 3, $\phi_i(x)$, of 8×8 pixels(This benefits from Bruno A. Olshausen’s program). With this set of basis functions, any temporal patch of natural scenes can be described as a linear superposition of the basis functions with temporal amplitudes $a_i(t)$:

$$W(x,t) = \sum_i a_i(t) \phi_i(x) + v(x,t)$$

We think of the basis functions $\phi_i(x)$ as retinal neurons with certain spatial features and the temporal coefficients $a_i(t)$ as activities of these retinal neurons which are inputs to the cortex model. To construct the cortical inputs from the natural scenes, $I_k(x,t)$ where k indexes natural scenes, with the learnt basis functions, a temporal stripe of 152×360 pixels was chosen from each natural scenes as the view range of fresh turtle and the stripe was split into 45 blocks of 152×8 pixels, indexed by p from left to right. Each block within the stripes contains 19 temporal patches of 8×8 pixels, indexed by q from top to bottom and every temporal patch was denoted as $W_k^{p,q}(x,t)$ and can be represented as:

$$W_k^{p,q}(x,t) = \sum_{i=1}^{64} a_{k,i}^{p,q}(t) \phi_i(x) + v(x,t)$$

For every block of the stripe of each natural scene, the all temporal coefficients within the block were arranged together into a matrix, $A_k^p(t)$, of which the 1216 columns are the temporal coefficients. 34 of 45 coefficient matrices from the 6th block to the 39th block will be fed as cortical inputs to cortex model, each of which to 6 of LGN neurons (the 39th to the right-most 3 LGN neurons). However the dimensions of the cortical inputs are too high and they have to be converted to appropriate inputs of low dimension. We here use KL-decomposition technique with sliding time window of constant length of 10 ms to decompose the high dimension coefficient matrix $A_k^p(t)$ and get β -strands, $B_k^p(t)$, in β -space, of visual inputs for every block while maintaining the spatiotemporal information of visual scenes. The 3 β coefficients corresponding to the first 3 principal components are chosen as temporal cortical signal. They are scaled properly and fed to the model cortex via LGN neurons. The 3 temporal β signals of the block which provides cortical input to model cortex go to 6 LGN neurons and each of these β signals goes to two of the 6 LGN neurons respectively. The temporal cortical inputs to model cortex are shown in Fig 2. The magnitudes of β coefficients are color-coded. With the dimension reduced cortical inputs through KL-decomposition, the associated cortical response of a large number of pyramidal cells are generated with WUNG (including subpial cells) cortex model for every visual input. Some frames of cortical activity waves are shown in Fig 4.

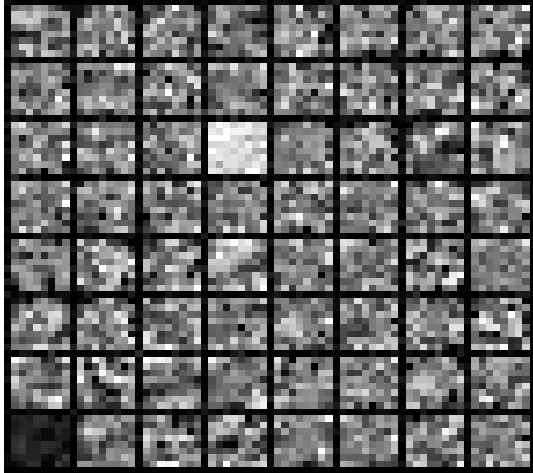


Fig. 3. Over-Complete Basis Functions

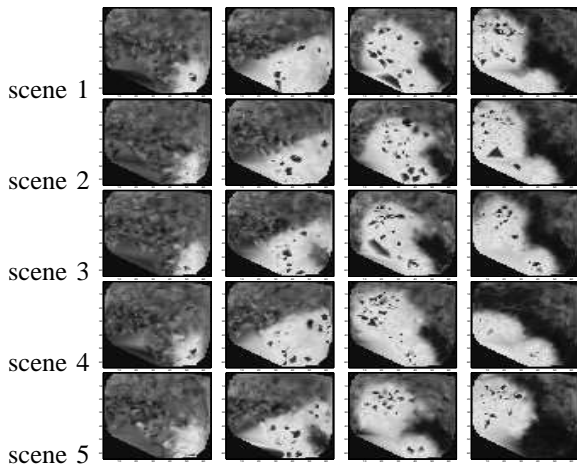


Fig. 4. Frames selected from Cortex movies with five natural inputs. From top to bottom are the frames of cortical movies in response to five different natural stimuli. From left to right are the cortical movie frames at 60, 180, 300 and 550 ms respectively

III. COMPUTATION OF MOTION FIELD FROM THE RESPONSE OF THE TURTLE VISUAL CORTEX

So far, we have described that natural stimuli induce waves of activity in the model cortex. We represent this wave as a spatiotemporal signal $M(x, t)$. In this section, we argue that the activity of the cells (pyramidal) of the cortex encodes features of the input visual field such as the motion field. The cortical input which maintains the spatiotemporal information of the visual scene directly activates LGN-Pyramidal conductances of pyramidal cells which subsequently induce the origination and propagation of the cortical wave. The LGN-Pyramidal conductance of a pyramidal cell is defined as the sum of all synaptic conductances from LGN neurons to this pyramidal cell via varicosities. There are no feedback

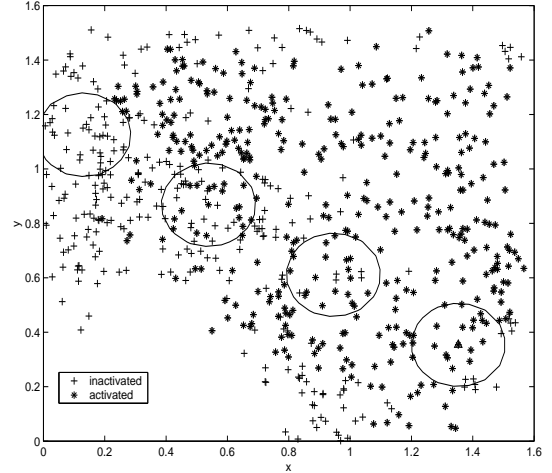


Fig. 5. The pyramidal cells of the visual cortex. Four portions of cells are chosen to compute the conductance pattern of the target cell and they locate along diagonal direction from the rostralateral pole to the caudomedial pole indexed from 1 through 4 (within the circles and the numbers of cells are 22, 23, 42 and 38 respectively). The target cell (triangle) locates at the center of the first portion at the rostralateral pole. The pyramidal cells are divided into two groups, activated (*) and inactivated (+). Activated cells are those getting from synapses from LGN neurons.

connections and recurrent connections to LGN-Pyramidal synapses which map visual space to cortical space with LGN neurons through varicosities, so it is believed that the conductance patterns of the pyramidal cells directly encode visual information such as intensity of visual stimuli and motion field in the visual space. Thus, in this paper, the computation of motion field from the response of the turtle visual cortex consists of two processes: computation of the synaptic conductance patterns of pyramidal cells from the response of the visual cortex and computation of motion field from the estimated conductance patterns, which are discussed below.

A. Computation of the conductance patterns from the response of the visual cortex

In the visual cortex model, 384 of 689 pyramidal cells get synapses from LGN neurons. Because of the density distribution of the varicosities ([8]), among these cells, those located at the rostralateral pole of the cortex get more synapses from LGN neurons and hence stronger conductance, and those located at the caudomedial pole of the cortex get less synapses from LGN neurons and weaker conductance. The conductances of the pyramidal cells at the rostralateral pole, elicited by the visual input, are strong enough to initiate the cortical wave and the conductances of the other pyramidal cells far from the pole are weak and do not initiate the wave. The cells far from the rostralateral pole spike due to the wave propagation. The wave propagation makes the information of visual space decoded with the neural activity

of different portion of the visual cortex within different time interval. Thus, to compute the conductance pattern of a cell, we may use the neural activity of different portion of the cortex within different time interval. Here we chose pyramidal cells of four circular portion of same diameter with centers evenly spaced along diagonal direction from the rostrolateral pole to the caudomedial pole to estimate the conductance of a target pyramidal cell which locates at the center of the first portion as shown in Fig 5 and the numbers of cells in these four portion are 22, 23, 42 and 38 respectively. The neural activities(only the parts we used to estimate conductance pattern) of four portions of pyramidal cells and the conductance pattern of the target cell are shown in Fig 6 and Fig 7(first column) respectively and they were already filtered with low pass filters. The visual stimuli we fed to the visual cortex model last 150 ms and hence the conductance patterns of pyramidal cells die out after about 200 ms. So here we chose the conductance patterns from 0 ms to 500 ms as target patterns and denoted it as $g_i(t)$ where i indexes the visual stimuli and the target conductances are shown in Fig 7. Regarding wave propagating time, we chose neural activities within different time interval for different portion. We denote the neural activities of the four portions with different stimuli as $R_{ij}(t)$, and the parts that we used to compute the conductance patterns are in the time interval from T_j to $T_j + T$ and they are denoted as $S_{ij}(t) = R_{ij}(t + T_j)$ where i indexes the visual stimuli, j indexes the portions of the cortex, $T_j = 50, 200, 320, 400$ for $i = 1, 2, 3, 4$ respectively, indexing the starting point of the time intervals and $T = 500ms$ is the length of the target conductance pattern that we compute. In Fig 6, only the parts of neural activities used to estimate conductance patterns are shown. To estimate the target conductance pattern $g_i(t)$ from the neural activities $S_{ij}(t)$, we constructed second order Anti-Causal Time Varying Autoregressive and Moving Average Model(ARMA)([4]) with neural activities as input and conductance pattern as output. We use ‘Anti-Causal’ because conductance is ‘cause’ and neural activities are ‘effect’ in the visual cortex. We reversed $g_i(t)$ and $S_{ij}(t)$ to $\bar{g}_i(t)$ and $\bar{S}_{ij}(t)$ with $\bar{g}_i(t) = g_i(T - t)$ and $\bar{S}_{ij}(t) = S_{ij}(T - t)$. Then the estimation $\hat{g}_i(t)$ of $\bar{g}_i(t)$ were computed from $\bar{S}_{ij}(t)$ with ARMA model. Finally we reversed the estimation $\hat{g}_i(t)$ back as the estimation $g_i^{est}(t)$ of the original conductance pattern $g_i(t)$ with $g_i^{est}(t) = \hat{g}_i(T - t)$. The ARMA model we used is 2nd order and is described as below:

$$r(t) = -A_1(t-1)r(t-1) - A_2(t-2)r(t-2) + B_1(t-1)u(t-1) + B_2(t-2)u(t-2)$$

where $r(t)$ and $u(t)$ are output and input respectively. In this paper, for each of the four portions, j , we have:

$$\hat{g}_i(t) = -A_{1j}(t-1)\hat{g}_i(t-1) - A_{2j}(t-2)\hat{g}_i(t-2) + B_{1j}(t-1)\bar{S}_{ij}(t-1) + B_{2j}(t-2)\bar{S}_{ij}(t-2)$$

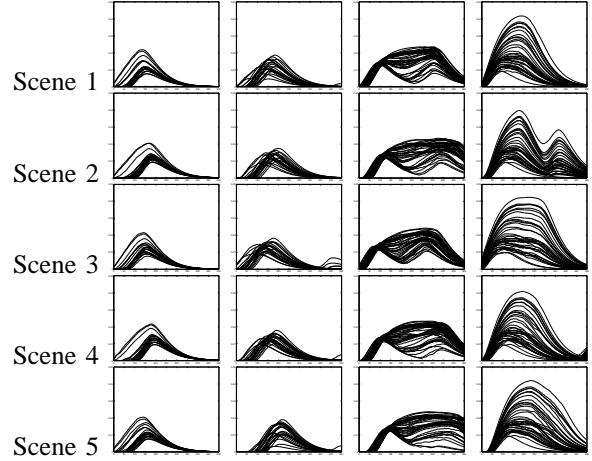


Fig. 6. The neural activities of 4 portions of pyramidal cells corresponding to 5 different visual inputs. From top to bottom are for 5 different visual inputs. From left to right, 4 columns are neural activities (only the parts of neural activities within time intervals from T_j to $T_j + T$ are shown) of the four portions of cells.

where i indexes the visual stimuli. The coefficients $A_{1j}(t)$, $A_{2j}(t)$ and $B_{1j}(t)$, $B_{2j}(t)$ with respect to every portion of the cortex were trained with Projection Algorithm([4]) based on $\bar{g}_i(t)$ and $\bar{S}_{ij}(t)$ for all $i = 1, \dots, 5$. We introduced moving window technique and trained the coefficients $A_{1j}(t)$, $A_{2j}(t)$ and $B_{1j}(t)$, $B_{2j}(t)$ within every small window separately so that the coefficients are constant within every small window. In this paper, we chose 10 ms as the length of the moving windows. The estimates of the target conductances are shown in Fig 7.

B. The future work: Computation of the motion field from the conductance patterns

The difficulty in estimation of the motion field in the visual space from the neural activities of pyramidal cells in the visual cortex involves coordinate transformation between two spaces: visual space and cortical space. The coordinate transformation between the two spaces is complicated due to wave propagation. But we know how a certain part of visual space stimulate a certain part of pyramidal cells through varicosities by eliciting conductance pattern change of these cells directly([8]). So an alternative is to estimate the motion field with conductance patterns, defined as above, of pyramidal cells other than neural activities of pyramidal cells. As Above we have discussed how to estimate the conductance pattern of a pyramidal cell with neural activities of different portion of pyramidal cells within different time interval, and we can compute the conductance patterns of all pyramidal cell elicited directly by LGN neurons. Based on estimated conductance patterns of pyramidal cells, we may estimate the motion field of the visual space patch-wisely. For a certain patch of visual scene, we choose the conductance patterns

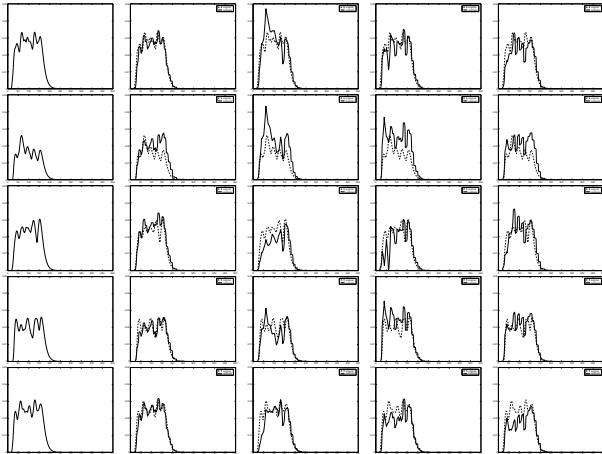


Fig. 7. the target patterns and the estimates of the target neuron with different portions of neurons' activities for 5 different visual input. The first column are the conductance patterns. From the second to the fifth columns are the estimates of the conductance patterns with four portions of cells respectively. In each figure of the estimates, two curves are the original conductance(dashed line) and the estimated conductance(solid line). From top to bottom are for 5 different visual inputs.

of the corresponding pyramidal cells elicited directly by this part of visual scene via LGN neurons through varicosities. The estimation is performed with respect to a suitable choices of basis function associated with conductance patterns of the pyramidal cell related to this patch of visual scene. Let us assume the existence of a basis function $\psi_j(x)$ associated with the j^{th} pyramidal cell, we would like to choose the functions in such a way that the motion fields of the patch of visual scenes are estimated by

$$V_i(x,t) = \sum_{j=1}^{N_p} g_{ij}(t) \psi_j(x)$$

with minimizing the error:

$$\sum_{i=1}^5 (V_i(x,t) - \sum_{j=1}^{N_p} g_{ij}(t) \psi_j(x,t))^2$$

where $(V_i(x,t))$ is the motion field of the patch of the i^{th} visual scene, $g_{ij}(t)$ is the conductance pattern of the j^{th} cell associated with the i^{th} visual scene and N_p is the number of pyramidal cells related to the patch of visual scenes. So the future work would be to find out the sets of basis functions to compute the motion fields of the visual scenes patch-wisely from the conductance patterns.

IV. RESULTS

In this paper, we discussed how to feed visual scenes as inputs to visual cortex using sparse representation with over-complete set of basis functions together with KL-decomposition while maintaining spatiotemporal information

contained in the visual inputs. Sparse representation with over-complete set of basis function is an efficient coding strategy of natural images with an probabilistic image model to generate a sparse distribution of output activity which we thought of as the activity of retinal neurons. The KL-decomposition technique then lower the dimension of the activity and provides appropriate cortical inputs to the visual cortex.

Estimating the motion field in the visual space is an important problem in Neuroscience primarily because animals have to negotiate with moving targets in order to accomplish a task. In freshwater turtles, it is believed that the visual cortex plays a role in accomplishing this task of prediction. In this paper, we tried to find a way to estimate the motion fields of visual scenes from the neural activities of pyramidal cells in the turtle visual cortex. The cortical waves always initiate at the rostromedial pole of the cortex and then propagate along the diagonal direction from rostromedial pole to caudomedial pole, and the wave propagation make the coordinate transformation between visual space and cortical space complicated as well as the connections among the cells in the cortex. We have no idea how to choose coordinate transformation functions $\Psi(x,t)$ so that the motion fields can be constructed from neural activities of pyramidal cells by

$$V(x,t) = \sum_{j=1}^N a_j(t) \delta_j(\Psi(x,t)).$$

where $\delta_j(x,t)$ is basis function associated the j^{th} cell in cortical space. In this paper, we bypassed this difficulty by computing conductance patterns from neural activities of pyramidal cells and then try to estimate the motion fields of visual scenes with the estimated conductance patterns. To estimate the conductance pattern of a cell, we constructed a Anti-Causal Time Varying Autoregressive and Moving Average Model and used the neural activities of different portions of pyramidal cells within different time intervals with moving window technique. Fig 7 shows the estimate of conductance patterns of a target cell corresponding to different visual scenes with four portions of cells. The length of moving window in estimation is 10 ms. So here we may see another advantage of this estimation strategy that the information in visual space such as motion field could be estimated with neural activities of pyramidal cells at the different part of visual cortex at different time which can mimic the behavior of animals that they can recognize or remember objects sometime after they have seen the objects.

V. REFERENCES

- [1] J. M. Bower and D. Beeman, *The Book of Genesis*, (TELOS, Santa Clara, 1998).
- [2] Xiuxia Du and B. K. Ghosh, "Decoding the position of a visual stimulus from the cortical waves of turtles," Proceedings of the American Control Conference, (2003).

- [3] D. J. Field, "What is the goal of sensory coding?," *Neural Computation*, 6, 559–601, (1994).
- [4] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*, (PRENTICE-HALL, 1984).
- [5] P. Z. Mazurskaya, "Organization of receptive fields in the forebrain of *Emys orbicularis*," *Neurosci. Behav. Physiol.*, vol. 7, pp. 311–318, (1974).
- [6] Z. Nenadic, B. K. Ghosh and P. Ulinski, "Spatiotemporal dynamics in a model of turtle visual cortex," *Neurocomputing*, 32–33, pp. 479–486, (2000).
- [7] Z. Nenadic, B. K. Ghosh and P. Ulinski, "Modeling and estimation problems in the turtle visual cortex," *IEEE Trans. on Biomedical Engineering*, vol. 49, no. 8, pp. 753–762, (2002).
- [8] Z. Nenadic, B. K. Ghosh and P. Ulinski, "Propagating waves in visual cortex: A large scale model of turtle visual cortex," *J. Computational Neuroscience*, Vol. 14, pp. 161–184, (2003).
- [9] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," *Vision Res.*, Vol. 37, No. 23, pp. 3311–3325, (1997).
- [10] J. C. Prechtl, "Visual motion induces synchronous oscillations in turtle visual cortex," *Proc. Natl. Acad. Sci. USA* 91: 12467–12471, 1994.
- [11] J. C. Prechtl, T. H. Bullock and D. Kleinfeld, "Direct evidence for local oscillatory current sources and intracortical phase gradients in turtle visual cortex," *Proc. Natl. Acad. Sci.* 97: 877–882, 2000.
- [12] J. C. Prechtl, L. B. Cohen, P. P. Mitra, B. Pesaran and D. Kleinfeld, "Visual stimuli induce waves of electrical activity in turtle cortex," *Proc. Natl. Acad. Sci.* 94: 7621–7626, 1997.
- [13] K. A. Robbins and D. M. Senseman, "Visualizing differences in movies of cortical activity," *IEEE Trans. Visualization Compt. Graphics*, vol. 4, pp. 217–224, 1998.
- [14] D. M. Senseman, "Correspondence between visually evoked voltage sensitive dye signals and activity recorded in cortical pyramidal cells with intracellular microelectrodes," *Vis. Neurosci.*, vol. 13, pp. 963–977, 1996.
- [15] D. M. Senseman and K. A. Robbins, "Modal behavior of cortical neural networks during visual processing," *J. Neuroscience*, vol. 19, RC3, pp. 1–7, (1999).
- [16] P.S. Ulinski, Wenxue Wang and Bijoy K. Ghosh, "Generation and Control of Propagating Waves in the Visual Cortex," In *42nd IEEE Conference on Decision and Control*, (2003).
- [17] Wenxue Wang, Bijoy K. Ghosh and Philip S. Ulinski, "Cellular Mechanisms Controlling Propagating Waves in Visual Cortex," Submitted to *Journal of Computational Neuroscience*, (2004).