Detection of Video Inputs Using the WUNG Model

Wenxue Wang [*,1] Bijoy K. Ghosh [2]

Washington University, Department of Electrical and Systems Engineering, Campus Box 1127, Bryan Hall 201, One Brookings Drive, Saint Louis, MO 63130

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 test detectability of video inputs using the activity pattern in the modified noise model cortex containing subpial cells. 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 inputs to the cortex. Finally, the cortical response has been displayed as a spatiotemporal signal. The paper concludes with detection of the natural visual inputs using the noise WUNG model.

Key words: Visual Cortex, Sparse Representation, KL-decomposition, Detection

1 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

* Corresponding author.
Email addresses: `ww1@zach.wustl.edu` (Wenxue Wang),
`ghosh@netra.wustl.edu` (Bijoy K. Ghosh).

[1] Research was partially supported from NSF grant ECS-9976174

[2] Research was partially supported from NSF grants ECS-9976174, and ECS-0323693

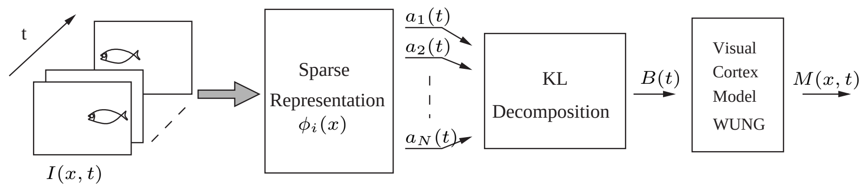

Fig. 1. System Diagram

an input. A large scale model of the cortex, the UNG model, has also been constructed with software package, GENESIS (J. M. Bower and D. Beeman,1998), that has the ability to simulate cortical waves with the same qualitative features as the cortical waves seen in experimental preparations (2000,2002). Z. Nenadic, Bijoy K. Ghosh and Philip Ulinski (2000,2002,2003) and Xiuxia Du and Bijoy K. Ghosh (2003) 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 Wang, Bijoy K. Ghosh and Philip Ulinski (2003,2004) modified the UNG model by adding inhibitory neurons, subpial cells. The purpose of this paper is to take the cortical waves from the modified WUNG model cortex and to detect different visual inputs, natural scenes. 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 sufficient 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 detect the natural visual scenes. 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 noise WUNG model to produce waves.

2 Sparse, over-complete representation and KL-decomposition

Sparse representation with an over-complete basis set was proposed to explain and exam the receptive field properties in terms of a strategy for producing a sparse distribution of output response to natural images or scenes (Bruno A. Olshausen and David J. Field, 1997). 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:

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

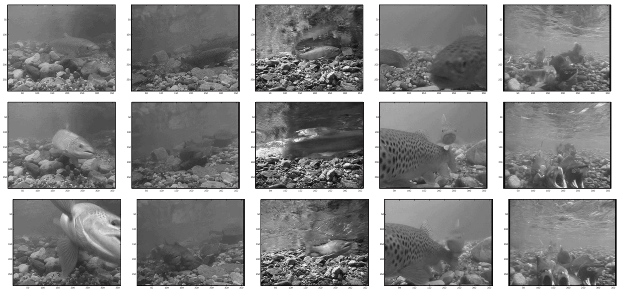

Fig. 2. Frames selected from 5 natural scenes.From left to right are those ones selected from Movie 1 through Movie 5 respectively.

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, $\phi_i(x)$, are trained from the set of images with adapting the probabilistic model to statistics of images. The basis functions 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.

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, $\phi_i(x)$, 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

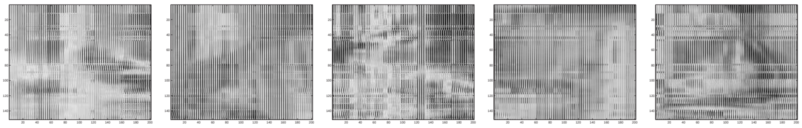

Fig. 3. Dimension-reduced cortical input signals from natural scenes 1 through 5 to Visual Cortex Model from left to right. The magnitudes of the signals are color-coded. The Y-axis is temporal axis and the X-axis indicates the location mapping between cortical signals and LGN neurons.

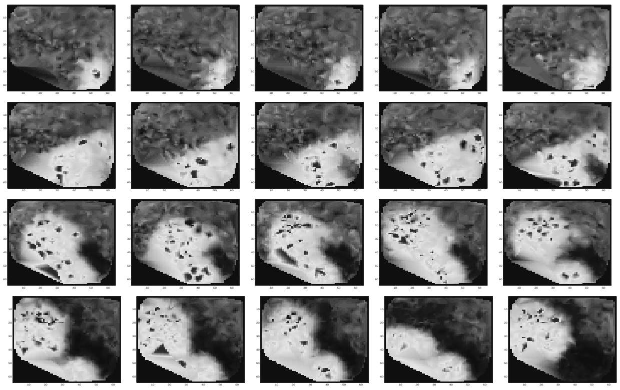

Fig. 4. Frames selected from Cortex movie with five natural inputs.From left to right are the frames of cortical movies in response to five different natural stimuli.

columns are the temporal coefficients. 34 of 45 coefficient matrices from the $6th$ block to $39th$ 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 3. The magnitudes of β coefficients are color-coded. With the dimension reduced cortical inputs through KL-decomposition, 100 activity waves $M(x,t)$, where x represents the position in cortical space, were generated with the noise WUNG cortex model for every visual input. Some frames of cortical activity waves are shown in Fig 4.

3 Detection of Natural Scenes from Cortical Waves with WUNG Model

So far, we have described how to get low dimensional cortical inputs from natural stimuli and that natural stimuli induces waves of activity in the model cortex. We represent this wave as a spatiotemporal signal $M(x,t)$. It was showed that the cortical waves elicited by visual stimuli encoded information about positions of stimuli in visual space (Z. Nenadic, Bijoy K. Ghosh and P. Ulinski, 2003). And some research on position detection of flash stimuli has been done with pre-modified UNG model by Xiuxia Du and Bijoy K. Ghosh (2003). In this paper, instead of flash inputs which were fed to LGN directly, we use natural stimuli, video scenes, which are encoded with sparse over-complete representation and reduced with KL decomposition into low dimension cortical inputs to LGN, to induce cortical waves with WUNG model. Unlike flash inputs used in previous work, natural stimuli contain more complicated information which is not just object position but also moving velocity of the object in the natural scenes. We believe that the activity of the pyramidal cells of the cortex encode features of the input visual field. In this section, we argue that we may detect the natural scenes from the cortical waves with noisy model. The noise model contains intrinsic gaussian noise generators which model biological noise existing in visual cortex and disturb cortical waves. With SDW technique developed by Xiuxia Du and Bijoy K. Ghosh (2003), the 500 cortical waves, 100 for each input, are represented with β-strands in *beta*-space and statistical mean of the β-strands for each input can be obtained easily. It is assumed that β-strands contain the same information as original cortical waves. The detection is based on these β-strands. Designate the mean of β-strands by $s_i(t)$, where $i = 1, 2, ..., 5$, for each of five natural stimuli respectively and let $r(t)$ be the β-strand of an arbitrary cortical wave in response to an unknown stimulus of the 5 ones. With noise presence, the detection is formulated as a hypothesis test problem. Let five natural stimuli correspond to five hypotheses and H_i denote the hypothesis that the cortical wave is elicited by natural scenes i, $i = 1, 2, ..., 5$. Then the hypothesis testing problem is formulated as:

$$r(t) = s_i(t) + n(t), i = 1, 2, ..., 5$$

where $n(t)$ represents the noise contained in the β-strand $r(t)$. The conditional probability density P_i under each hypothesis H_i can be calculated. With maximal likelihood criterion, we make decision that the cortical wave was induced by the natural input with maximal conditional probability.

4 Results

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] P. Z. Mazurskaya, "Organization of receptive fields in the forebrain of *Emys orbicularis*," *Neurosci. Behav. Physiol.*, vol. 7, pp. 311–318, (1974).

[4] Z. Nenadic, B. K. Ghosh and P. Ulinski, "Spatiotemporal dynamics in a model of turtle visual cortex," *Neurocomputing,*, 32–33, pp. 479–486,(2000).

[5] 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).

[6] 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).

[7] 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).

[8] 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).

[9] Wenxue Wang, Bijoy K. Ghosh and Philip S. Ulinski, "Cellular Mechanisms Controlling Propagating Waves in Visual Cortex," Submitted to *Journal of Computational Neuroscience,* (2004).

Wenxue Wang received the Master degree of Science in Institute of Systems Science, Chinese Academy of Sciences. He is now a Ph.D student in the Center for BioCybernetics and Intelligent Systems, the Department of Electrical and Systems Engineering at Washington University in St. Louis

Bijoy K. Ghosh recieved the Ph.D. degree from the Division of Applied Sciences of Harvard University. He is now a faculty member in the department of Electrical and Systems Engineering at Washington University in St. Louis where he is currently a Professor and directs the Center for BioCybernetics and Intelligent Systems.