MOTION RECONSTRUCTION IN NATURAL SCENES FROM CORTICAL ACTIVITY WAVES

Wenxue Wang ^{*,1} **Bijoy K. Ghosh** ^{*,2}

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

Abstract: Detection of moving targets in nature is an important problem in neuroscience. In freshwater turtles, it is believed that the visual cortex performs this job. In this paper, a set of 'moving fishes' are considered and the visual data is compressed using a set of spatial basis functions. Input to the model cortex is this compressed visual data, and the associated cortical response is generated. This paper illustrates the role of dynamic motion reconstruction in a natural scene from the cortical activity waves using a bank of autoregressive moving average processors. *Copyright ©2005 IFAC*

Keywords: Modelling, Identification, ARMA models, Biocybernetics, Computer simulation, Estimation

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 waves of activity. These activities have been experimentally observed assuming a stationary and a moving flash as an input (Prechtl *et al.*, 1997); (Senseman, 1996). A large scale model of the cortex, the NGU model (Nenadic *et al.*, 2000); (Nenadic *et al.*, 2002), has also been constructed with a software package, GENESIS (Bower and Beeman, 1998), that has the ability to simulate cortical waves with the same qualitative features as the cortical waves seen in experimental preparations. The dynamics of the activity of waves has been studied as well as estimation and detection problems by stimulating the NGU cortex model with inputs of flash patterns (Nenadic *et al.*, 2000); (Nenadic *et al.*, 2002); (Nenadic *et al.*, 2003); (Du and Ghosh, 2003). It is believed that the activity waves of a turtle visual cortex encode features of the visual scenes, viz. existence of a target and its shape, position, and velocity. The NGU model was modified by adding another type of inhibitory neurons, subpial cells, to produce the WNGU model (Ulinski *et al.*, 2003); (Wang *et al.*, 2004). The purpose of this paper is to estimate the cortical inputs from the associated neural responses by constructing an Autoregressive and Moving Average model and then reconstruct the visual inputs of natural scenes with estimates of the cortical inputs. In order to simulate cortical responses, suitable cortical inputs have to be constructed from visual inputs of natural scenes. The inputs to the cortex have to be of sufficiently low dimension and yet maintain the spatiotemporal information of the visual inputs. Cortical inputs are fed to the cortex model to produce cortical activity waves which are used to estimate the target in the visual space. The schematic diagram of the visual system is described in Fig 1. With sparse over-complete representation, natural scenes can be represented as a linear superposition of a set of sparse basis functions with coefficients. The coefficients are assumed to be the activities of retinal neurons and the cortical inputs. KL-decomposition is used to reduce the dimension of

[1] Partially supported from NSF grant ECS-9976174

[2] Partially supported from NSF grant ECS-9976174, and ECS-0323693

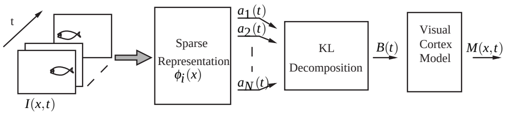

Fig. 1. System Diagram

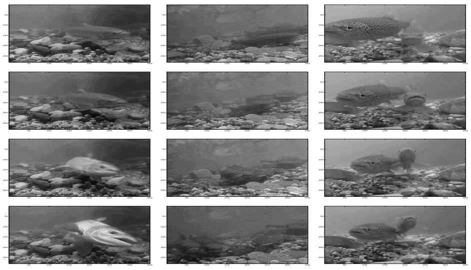

Fig. 2. 3 scenes in this study. From top to bottom are 4 sequential images of each scene. From left to right are 3 scenes.

the cortical inputs while maintaining the spatiotemporal information of the visual scenes and the reduced cortical inputs are fed to the WNGU model to produce cortical responses.

2. 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 (Olshausen and 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:

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

where x denotes spatial position within the patch and the variable v represents Gaussian noise (i.i.d.) which is included in the probabilistic model structure in the images that are not well captured by the basis functions. The basis functions, $\phi_i(x)$, are trained on the set of images by adapting the probabilistic model to the statistics of the 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, 3 natural scenes were used as visual inputs to the visual system. Each of the natural scenes contains 4 sequential images of 280×350 pixels. Every image lasts 40ms and the scenes last 160ms as visual inputs. The targets in the natural scenes are moving fishes. The images are shown in Fig 2. The 12 images from these 3 scenes were used to train the set of one hundred 10×10-pixel basis functions, $\phi_i(x)$, which are shown in Fig 3 (This benefits from B. A. Olshausen's program). With this set of basis functions, any patch of natural scenes can be described as a linear superposition of the basis functions with corresponding amplitudes $a_i(t)$:

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

The basis functions $\phi_i(x)$ are thought of as retinal neurons with certain spatial features and the Olshausen's coefficients $a_i(t)$ as activities of these retinal neurons. These activities are inputs to the cortex model. To construct the cortical inputs from the natural scenes, $I_k(x,t)$ (where k indexes the natural scenes) with the learnt basis functions, every image was split into 35 blocks of 280×10 pixels, indexed by p from left to right. Each block contains 28 patches of 10×10 pixels, indexed by q from top to bottom, and every patch is denoted as $W_k^{p,q}(x,t)$ and can be represented as:

$$W_k^{p,q}(x,t) = \sum_{i=1}^{100} a_{k,i}^{p,q}(t)\phi_i(x) + v(x)$$

The associated Olshausen's coefficients, $a_{k,i}^{p,q}(t) \in \mathbb{R}^{100}$, were decomposed with KL-decomposition, and $a_{k,i}^{p,q}(t)$ can be represented by $\hat{a}_{k,i}^{p,q}(t) \in \mathbb{R}^{100}$ as

$$a_{k,i}^{p,q}(t) = V_a \hat{a}_{k,i}^{p,q}(t)$$

where V_a is a full matrix whose columns are the corresponding eigenvectors of the convariance matrix $Q \in \mathbb{R}^{100\times100}$ which is calculated as

$$Q = \sum_{k=1}^{3}\sum_{t=1}^{4}\sum_{p=1}^{35}\sum_{q=1}^{28}\sum_{i=1}^{100} a_{k,i}^{p,q}(t) a_{k,i}^{p,q}(t)^T.$$

For every 7 blocks of each natural scene, m predominant components of every coefficient, $\hat{a}_{k,i}^{p,q}(t) \in \mathbb{R}^{100}$, from 196 patches within the 7 blocks were arranged together into a column vector, $A_k^j(t) \in \mathbb{R}^{196m}$, where $j = 1...5$ and $A_k^j(t)$ consists of the coefficients of the patches $W_k^{p,q}(x,t)$ with $p = 7j-6,...,p = 7j$, and $q = 1,...,28$. The five vectors will be fed as inputs to the cortex model, and each of the five vectors goes to 40 of the 200 LGN neurons in the visual cortex model. However, the dimensions of the cortical inputs are too high and they have to be converted to appropriate inputs of lower dimension. KL-decomposition was used again to decompose the high dimensional coefficient vectors $A_k^j(t)$ to obtain β-coefficient vectors, $B_k^j(t) \in \mathbb{R}^{196m}$, in β-space, of the visual inputs for every 7 blocks, while maintaining the spatiotemporal information of the visual scenes. Thus $A_k^j(t)$ can be represented by $B_k^j(t)$ as

$$A_k^j(t) = V_c B_k^j(t)$$

where V_c is a full matrix whose columns are the corresponding eigenvectors of the covariance matrix $C \in \mathbb{R}^{196m\times196m}$ which is calculated as

$$C = \sum_{k=1}^{3}\sum_{t=1}^{4}\sum_{j=1}^{5}(A_k^j(t))(A_k^j(t))^T.$$

The N β-coefficients corresponding to the first N principal components of $B_k^j(t)$ were chosen as the amplitudes of the cortical signals which last 40ms. In this study, N was set to be $4, 8, 12$, and 16. Thus each scene provides $5N$ cortical signals which last 160ms. They were appropriately scaled and shifted, and then

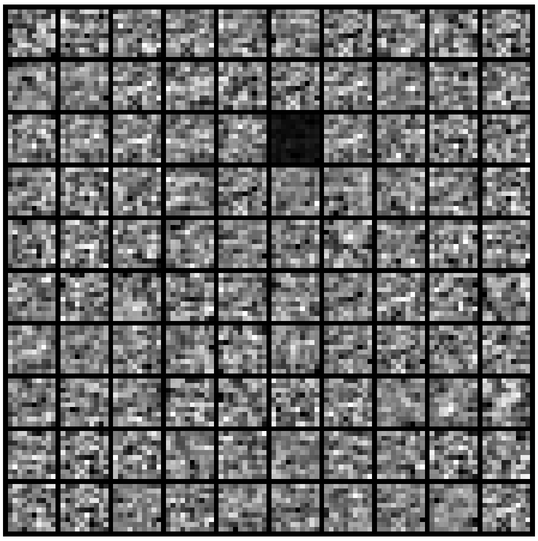

Fig. 3. Over-Complete Basis Functions

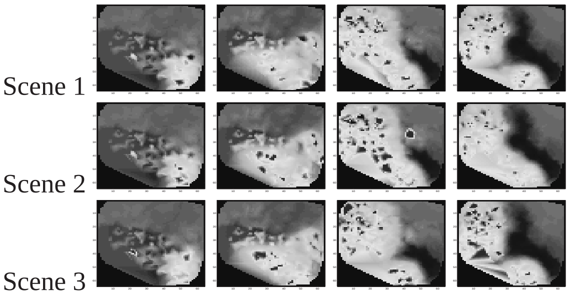

Fig. 4. Frames selected from Cortex movies with 3 natural inputs. From top to bottom are the frames of cortical movies in response to 3 different natural stimuli. From left to right are the cortical movie frames at 50, 150, 280 and 350 ms respectively

fed to the model cortex via LGN neurons. Fig 6, Fig 7, and Fig 8 show the normalized amplitudes of cortical inputs in the case $N = 12$. With the dimension-reduced cortical inputs, the associated cortical response of a large number of pyramidal cells was generated for 1200ms with WNGU cortex model for every visual scene. Some frames of cortical activity waves are shown in Fig 4.

3. RECONSTRUCTION OF THE SCENES FROM THE RESPONSE OF THE TURTLE VISUAL CORTEX

So far, we have explained that natural stimuli induce waves of activity in the model cortex. We represent this wave as a spatiotemporal signal $M(x,t)$. The activities of the cells (pyramidal) of the cortex encode features of the input visual field such as the intensity profile. A more important issue is the inverse problem: Decoding the features of the visual inputs from the activity waves of the model cortex. One problem is to reconstruct the visual scenes from the responses of the turtle visual cortex. The cortical inputs which maintain the spatiotemporal information of the visual scenes directly activate LGN-Pyramidal conductances, which subsequently induce the origination and propagation of the cortical wave. In this paper, the reconstruction of the scenes from the responses of the turtle visual cortex consists of two processes: Estimation of the β-coefficients which were fed to LGN neurons of the cortex from the responses of the visual cortex and reconstruction of visual scenes from the estimated β-coefficients, which are discussed below.

3.1 Estimation of the β-coefficients from the responses of the visual cortex

In the visual cortex model, the activities $r_m(t)$ of 679 pyramidal cells, where m indexes the pyramidal cells, encode the features of the visual inputs. In other words, the activities of pyramidal cells directly encode the amplitudes of the β-coefficients which were fed to the cortex model. So the first step in reconstruction of the visual images is to estimate the amplitudes of these β-coefficients. The responses of pyramidal cells were filtered with a second order low pass filter and the smooth activities were used to estimate the β-coefficients. Because the dimension of the activities, 679, is too high to estimate the β-coefficients, the activities were clustered locally. The cortical space was subdivided evenly into 8×8=64 small square patches and the average activity of the cells in every patch was obtained. Excluding those patches without any pyramidal cells or with average activity close to zero, 45 average activities from other patches were used in the estimation of the β-coefficients. An ARMA model (Goodwin and Sin, 1984), with neural activities as input and β-coefficient cortical signals as output with 200ms time delay, was chosen for the estimation of the β-coefficients . The 200ms time delay makes the ARMA model causal. In this study, the ARMA model used is 2^{nd} order and is described below:

$$\begin{aligned} y(t) = &-A_1 y(t-1) - A_2 y(t-2) \\ &+ B_1 u(t-1) + B_2 u(t-2) \end{aligned}$$

where $y(t)$ and $u(t)$ are output and input respectively. In this paper, for each of the 3 scenes, we have:

$$\begin{aligned} \hat{\beta}_k(t) = &-A_1 \hat{\beta}_k(t-1) - A_2 \hat{\beta}_k(t-2) \\ &+ B_1 R_k(t-1) + B_2 R_k(t-2) \end{aligned}$$

where k indexes the visual stimuli, $\hat{\beta}_k(t)$ is the estimate of β-coefficient vector and $R_k(t)$ is the average activity vector of 45 dimensions. The parameter matrices A_1, A_2, B_1, and B_2 were trained using the Matlab Identification Toolbox. 2 examples of the estimates of the β-coefficient signals are shown in Fig 5. The estimates of the β-coefficients were obtained by taking the means of the estimated β-coefficient signals within the interval from $200 + 40(j-1) + 6$ ms to $200 + 40j - 5$ ms where $j = 1, 2, 3,$ and 4. The estimates of the β-coefficients for the case N=12 are shown with the actual β-coefficients in Fig 6, Fig 7, and Fig 8.

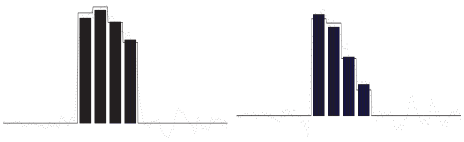

Fig. 5. 2 examples of the estimates of the β-coefficient signals. The solid lines are the actual β-coefficient signals and the dotted lines the estimated ones. The black bars are the estimate of the corresponding β-coefficients by taking the means of the estimated β-coefficient signals within the interval from $200+40(j-1)+6$ ms to $200+40j-5$ ms where $j=1,2,3,$ and 4. The width of the bars represents the interval where the means are obtained

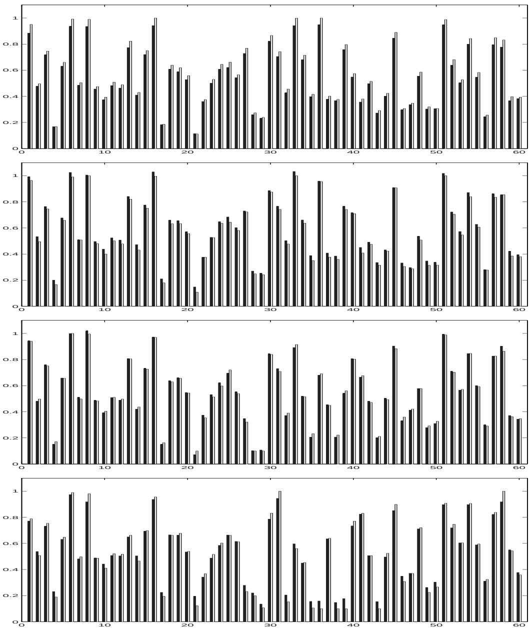

Fig. 6. Scene 1. The estimates of the β-coefficients for the case in which N=12 of the β-coefficients of every 7 blocks were fed to the visual cortex. From top to bottom are 4 frames of images in scene 1. The black bars are the estimated β-coefficients and the blank bars are the actual β-coefficients

3.2 Reconstruction of the visual scenes from the estimated β-coefficients

Above, the estimation of the β-coefficients from the average activities of the pyramidal cells was discussed. In this section, the estimated β-coefficients were used to reconstruct the visual scenes which had been used as visual inputs. Simply, the estimated β-coefficients were scaled and shifted inversely, and then the Olshausen's coefficients were estimated by taking 2 linear superpositions of the eigenvectors obtained in the 2 KL decompositions with the estimated β-coefficients. Then the scenes were reconstructed by the linear superposition of the set of the sparse basis functions with the estimated Olshausen's coefficients. Fig 9, Fig 10, and Fig 11 show the images reconstructed from estimated β-coefficients compared with the images represented by the sparse basis functions; and images reconstructed from the actual β-coefficients of N principal components for the case N=12. And the reconstructed images with estimated β-coefficients in all cases of N=4,8,16 are shown in Fig 12, Fig 13, and Fig 14.

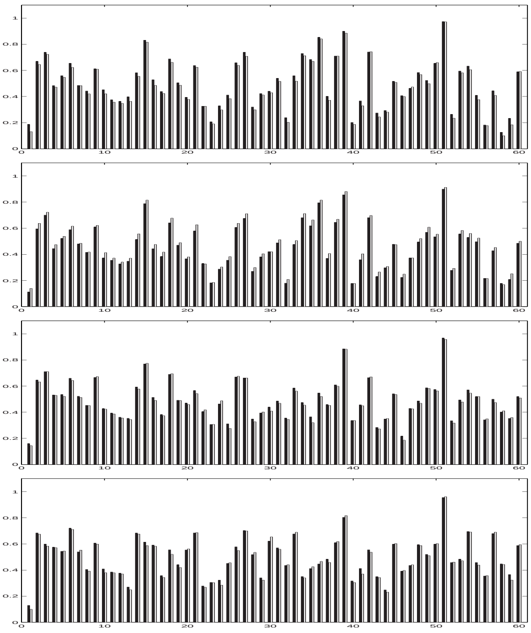

Fig. 7. Scene 2. The estimates of the β-coefficients for the case in which N=12 of the β-coefficients of every 7 blocks were fed to the visual cortex. From top to bottom are 4 frames of images in scene 2. The black bars are the estimated β-coefficients and the blank bars are the actual β-coefficients

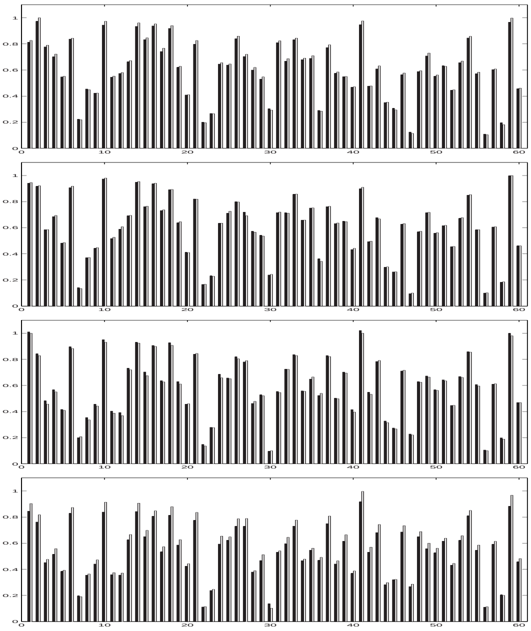

Fig. 8. Scene 3. The estimates of the β-coefficients for the case in which N=12 of the β-coefficients of every 7 blocks were fed to the visual cortex. From top to bottom are 4 frames of images in scene 3. The black bars are the estimated β-coefficients and the blank bars are the actual β-coefficients

4. RESULTS

In the above section, the estimation of the β-coefficients and the reconstruction of the visual scenes were discussed. Fig 15 shows the relative errors of the scenes

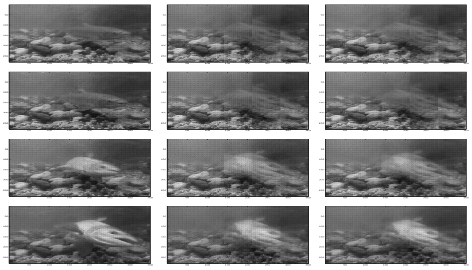

Fig. 9. Scene 1. From left to right are the images represented by the sparse basis functions, images reconstructed from the actual β-coefficients of N principal components and the reconstructed images with estimated β-coefficients for the case of N=12 from top to bottom. From top to right are 4 frames of images in scene 1

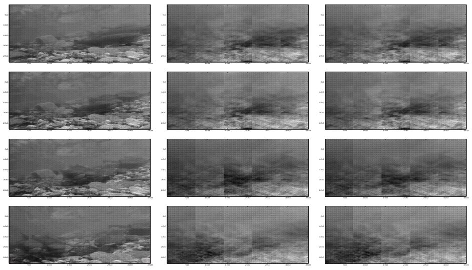

Fig. 10. Scene 2. From left to right are the images represented by the sparse basis functions, images reconstructed from the actual β-coefficients of N principal components and the reconstructed images with estimated β-coefficients for the case of N=12 from top to bottom. From top to right are 4 frames of images in scene 2

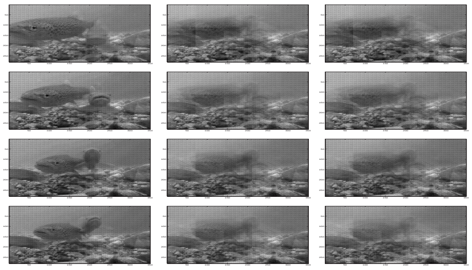

Fig. 11. Scene 3. From left to right are the images represented by the sparse basis functions, images reconstructed from the actual β-coefficients of N principal components and the reconstructed images with estimated β-coefficients for the case of N=12 from top to bottom. From top to right are 4 frames of images in scene 3

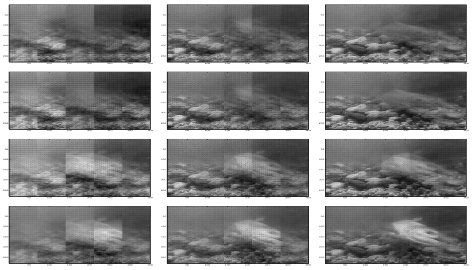

Fig. 12. Scene 1. From left to right are the reconstructed images with estimated β-coefficients in all cases of N=4, 8, 16. From top to right are 4 frames of images in scene 1

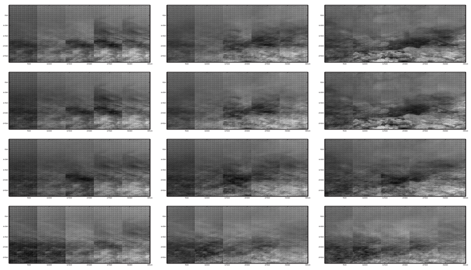

Fig. 13. Scene 2. From left to right are the reconstructed images with estimated β-coefficients in all cases of N=4, 8, 16. From top to right are 4 frames of images in scene 2

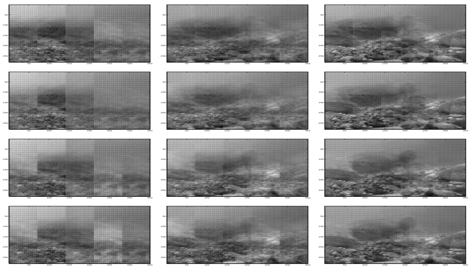

Fig. 14. Scene 3. From left to right are the reconstructed images with estimated β-coefficients in all cases of N=4, 8, 16. From top to right are 4 frames of images in scene 3

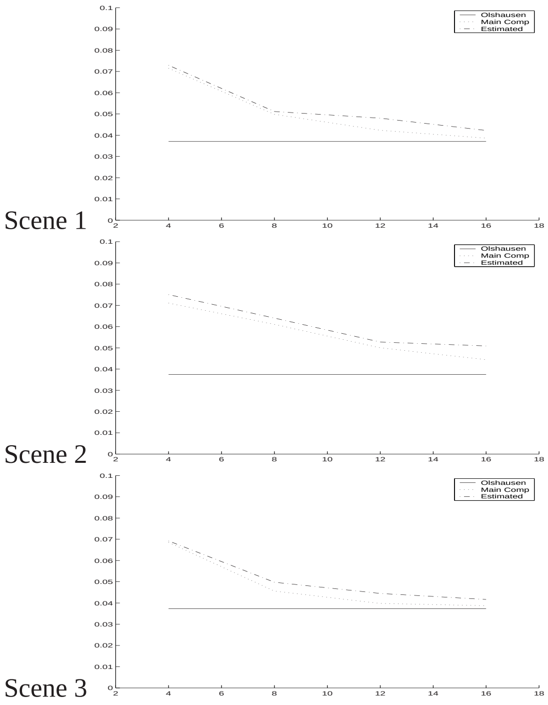

Fig. 15. The relative errors of the scenes represented by the sparse basis functions(Olshausen), the scenes reconstructed from the actual β-coefficients of N principal components(Main Comp) and the reconstructed scenes with estimated β-coefficients(Estimated) to the original scenes versus the choices of N=4,8,12,16. The solid lines are for Olshausen's representation, the dotted lines are for actual Main Comp representation and the dashdot lines are for Estimated scenes. From top to bottom are scene 1, scene 2 and scene 3

represented by the sparse basis functions, the scenes reconstructed from the actual β-coefficients of N principal components, and the reconstructed scenes with the estimated β-coefficients to the original scenes versus the choices of N=4, 8, 12, and 16. The relative error of an objective image is defined as the matrix norm of the difference between the objective image and the original image normalized by the matrix norm of the original image. The relative error of a scene is defined as the mean of the relative errors of all frames of images contained in this scene. The β-coefficients were estimated very well whatever the choice of N is. This point can be seen in Fig 5, Fig 6, Fig 7, and Fig 8(The other cases were not included here). The point can also be seen by comparing the relative errors of the scenes reconstructed from the actual β-coefficients of N principal components and the reconstructed scenes with the estimated β-coefficients to the original scenes versus the choices of N in Fig 15 where these two lines are very close for all 3 scenes. The ARMA model can estimate the β-coefficients from the responses of the visual cortex model. The quality of the reconstructed scenes is improved as the number N increases. Information loss occurred since only part of β-coefficients were fed to the visual cortex. Information loss also occurred in sparse representation. The contrast of the object in the image to the background affects the reconstruction of the scenes. Overall, the features of visual stimuli can be reconstructed very well from the activity waves of the visual cortex with an ARMA model. This study also illustrates that the visual cortex plays an important role in encoding information on visual inputs.

REFERENCES

Bower, J. M. and D. Beeman (1998). *The Book of Genesis*. Allan M. Wylde. Santa Clara.

Du, X. and B. K. Ghosh (2003). Decoding the position of a visual stimulus from the cortical waves of turtles. *Proceedings of the American Control Conference* pp. 477–482.

Goodwin, G. C. and K. S. Sin (1984). *Adaptive Filtering Prediction and Control*. PRENTICE-HALL. Englewood Cliffs, NJ.

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

Nenadic, Z., B. K. Ghosh and P. Ulinski (2002). Modeling and estimation problems in the turtle visual cortex. *IEEE Trans. on Biomedical Engineering* **49**, 753–762.

Nenadic, Z., B. K. Ghosh and P. Ulinski (2003). Propagating waves in visual cortex: A large scale model of turtle visual cortex. *J. Computational Neuroscience* **14**, 161–184.

Olshausen, B. A. and D. J. Field (1997). Sparse coding with an overcomplete basis set: A strategy employed by v1?. *Vision Res.* **37**, 3311–3325.

Prechtl, J. C., L. B. Cohen, P. P. Mitra, B. Pesaran and D. Kleinfeld (1997). Visual stimuli induce waves of electrical activity in turtle cortex. *Proc. Natl. Acad. Sci.* **94**, 7621–7626.

Senseman, D. M. (1996). Correspondence between visually evoked voltage sensitive dye signals and activity recorded in cortical pyramidal cells with intracellular microelectrodes. *Vis. Neurosci.* **13**, 963–977.

Ulinski, P. S., W. Wang and B. K. Ghosh (2003). Generation and control of propagating waves in the visual cortex. *Proceedings of 42nd IEEE Conference on Decision and Control* pp. 6429–6434.

Wang, W., B. K. Ghosh and P. S. Ulinski (2004). Integrative physiology of subpial cells. *Submitted to Journal of Computational Neuroscience*.