ISTANBUL UNIVERSITY –
JOURNAL OF ELECTRICAL & ELECTRONICS ENGINEERING

YEAR : 2003 (673-681)
VOLUME : 3
NUMBER : 1

SINGLE-FRAME SUPER-RESOLUTION BY INFERENCE FROM LEARNED FEATURES

Olcay KURŞUN[1] **Oleg FAVOROV**[2]

[1,2] School of Electrical Engineering and Computer Science
University of Central Florida
Orlando, FL 32816, USA

kursun@cs.ucf.edu favorov@cs.ucf.edu

ABSTRACT

Super-resolution is the creation of higher resolution views of pixel-based images through interpolation between the original pixels. Greater super-resolution can be achieved by taking advantage of local regularities inherent in natural images. In this paper, to learn regularities, we make use of the recently proposed SINBAD model of how the cerebral cortical network learns regularities by discovering regularity-simplifying environmental features [5, 14]. Using the regularities discovered with the SINBAD approach, we were able to predict more accurately the interpolated pixels from the ones in the original image and were able to generate visually plausible fine spatial details in the expanded image.

Keywords: *Neural Network Models, Learning Algorithms, Neurobiology*.

I. INTRODUCTION

Super-resolution refers to generation of higher resolution views of pixel-based image representations. One way of achieving super-resolution is to integrate information over multiple slightly shifted frames of the same image. In contrast, single-frame super-resolution is fundamentally a task of interpolating between the pixels in the original single frame image. Two-dimensional cubic spline interpolation [9] is probably the most commonly used interpolating technique (Adobe Photoshop, for example, uses it for image enlargements). However, cubic spline suffers from blurring of lines, edges, and fine textural details in the enlarged image.

Baker and Kanade [2] and Freeman et al. [8] propose that greater super-resolution can be achieved by taking advantage of local regularities inherent in natural images. Local groups of pixels in natural images have much less variability than they would have in randomly generated images (e.g., [6]). Such regularities can be used to predict more accurately the interpolated pixels from the ones in the original image and thus generate *visually plausible* fine spatial details in the expanded image.

Received Date : 10.04.2002
Accepted Date: 12.12.2002

Identifying regularities, however, is a difficult task, and more so for more useful, higher-order regularities. A promising approach to discovering regularities has been developed by Favorov and Ryder [5, 14], aimed at explaining how the cerebral cortical network discovers regularities in its sensory inputs and how it uses these regularities to fill-in missing information. Implemented in the form of a neural network (SINBAD network) that is modeled after the cerebral cortical network, this approach is likely to be very successful in the task of super-resolution.

In this paper we briefly describe the SINBAD approach to discovering regularities, review the general design of the SINBAD network, explain how this network can be used for single-frame super-resolution, and present encouraging results of our initial experiments on natural images.

2. SINBAD APPROACH

A major source of difficulties in learning regularities is not knowing about the existence of some of the factors contributing to them. If an environmental variable that plays an important role in some regularity is not among the variables known to the observer, but is reflected implicitly in the behaviors of some of the known variables, then the observer can still, in principle, learn the regularity. However, the regularity will now become more complex, involving all these extra variables with their implicit information about the missed key variable. This involvement of more variables and, likely, extra nonlinearities contributed by them, will make the regularity more difficult to learn. Thus, to learn regularities, it is crucial first to learn separately the identities of as many environmental factors contributing to those regularities as possible. Clark and Thornton [4] call this “trading representation against computation.”

According to Becker and Hinton [3], environmental variables that simplify regularities can be discovered through a search for different, but nevertheless highly correlated functions of any kind over non-overlapping subsets of the known variables. Such *correlated functions* must have a reason for their statistical interdependence, a causal source in the environment, and therefore these functions identify this source. That is, the correlated functions over different sets of environmental variables express a hidden environmental variable (a previously unrecognized *feature* of the environment) that is responsible for the correlation [3, 5, 12, 14].

Such hidden variables are very likely to have other effects in the environment, besides the ones that led to their recognition. And once they are recognized, it will become easier to notice their other effects. Furthermore, once a number of hidden variables are discovered, correlated functions can be searched for among *them*, thus discovering higher-order hidden variables, etc.

These new variables, as prominent environmental factors, will have inferential significance for other variables and thus will provide inferential *links* between the variables used to determine their states, and the variables whose states can be inferred from them. Thus, placing the newly derived variables in-between the original ones will break down the complex inferential relations among the original variables into simpler inferential relations, from the original variables to the derived ones, and from the derived variables to other original ones. The more inferentially significant variables are added to the repertoire (deriving them from the original and the already derived ones), the more distant inferential relations will be broken down into ones that are simpler and easier to learn.

Thus, the general approach should aim to derive as many regularity-simplifying variables as possible, and learn as many ways as possible to infer each variable from the other ones. By thus expressing each variable in many different ways in terms of other variables, which in turn are expressed in terms of yet other variables, etc., this approach will construct a rich web of inferential relations. In this web all the discovered inferential relations will be tied together into a single functional entity – *an inferential model* of the observed environment.

3. SINBAD CELL

Favorov and Ryder [5, 14] proposed that the search for regularity-simplifying environmental variables is performed in the cerebral cortex by the dendritic trees of individual pyramidal cells (the main type of neurons there). According to the SINBAD model, the basic function of each pyramidal cell is (1) to discover and represent one of the regularity-simplifying environmental variables, and (2) to learn to infer the state of its variable from the states of other variables, represented by other pyramidal cells. A network of such cells – each cell just attending to representation of its variable – can function as a sophisticated and useful inferential model of the outside world.

In the SINBAD model of pyramidal cells, several dendrites of a cell teach each other to produce correlated outputs to their different inputs. As a result, the cell as a whole tunes to the environmental variable that is responsible for correlation. Since each dendrite should be capable of learning functions over its inputs that are likely to be nonlinear, dendrites are viewed as functional analogs of error backpropagation networks [13], and a pyramidal cell is modeled as a set of several backprop nets whose outputs are added together to produce the cell's output. The cell's output is also used as the training signal for each dendrite.

While real pyramidal cells have 5-8 dendrites, for simplicity the model cells so far have been given only 3 dendrites (Figure 1). Two of the dendrites, representing *basal* dendrites of pyramidal cells, are given *afferent inputs* carrying information about the states of environmental variables. The third dendrite, representing the *apical* dendrite of pyramidal cells, is given both afferent inputs and *lateral inputs* from other SINBAD cells.

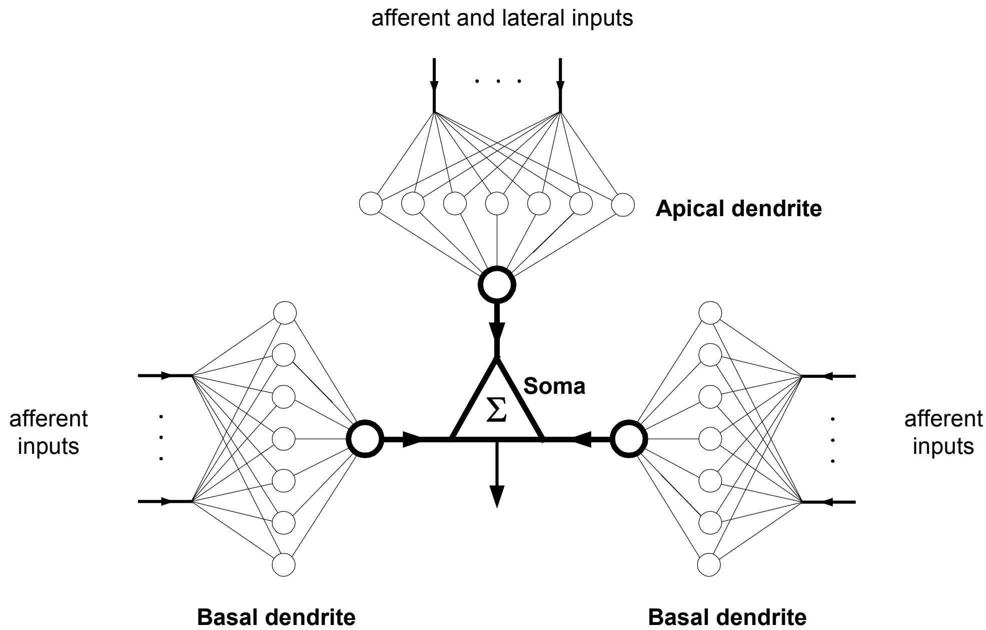

Figure 1. The SINBAD model of a cortical pyramidal cell with three dendrites connected to the soma (shown as a triangle). Each dendrite is modeled as an error backpropagation network with one output unit and a single layer of hidden units.

The activity of a hidden unit h in dendrite d is computed as a sigmoid function of the activities of its input sources:

$$H_{d,h} = \tanh(\sum_i w_{d,i,h} \cdot A_{d,i}), \qquad (1)$$

where $A_{d,i}$ is the activity of input source d,i and $w_{d,i,h}$ is the weight of its connection onto the hidden unit h of dendrite d. The activity of the output unit, i.e. the output of dendrite d, is:

$$D_d = \sum_{h=1}^{50} w_{d,h} \cdot H_{d,h}, \qquad (2)$$

where $w_{d,h}$ is the weight of the connection from the hidden unit d,h to the output unit. The outputs of the three dendrites are summated to produce the cell's output:

$$A = D_1 + D_2 + D_3. \qquad (3)$$

The cell's output A is the principal contributor to the training signal T, used to adjust the weights of connections on the three dendrites. Additional factors contributing to the training signal are: (1) the average output activity of the cell, $\bar{A}$, driving the cell to have $\bar{A} = 0$; (2) deviation of the current output activity from the average, $A - \bar{A}$, designed to expand the dynamic range of output values; and (3) lateral inhibition from other SINBAD cells, I. Thus,

$$T = A - \alpha \cdot \overline{A} + \beta \cdot (A - \overline{A}) - I, \qquad (4)$$

where α and β are scaling coefficients. Coefficient β is determined by the variability of the output activity: smaller the variability, greater the value of β. It is computed as:

$$\beta = \left[\beta_{\max} - \gamma \cdot \overline{|A - \overline{A}|}\right]^{+}, \qquad (5)$$

where $\beta_{\max}$ and γ are controlling parameters, and $[\ \cdot\ ]^{+}$ indicates that if the quantity is negative, the value is to be taken as zero.

The somal inhibition I is computed as:

$$I = \iota \cdot \sum_j w_j^{-} \cdot (A_j - \overline{A_j}), \qquad (6)$$

where ι is a scaling constant, and w_j^{-}, A_j and $\overline{A_j}$ are the somal inhibitory connection weight, activity, and average activity, respectively, of SINBAD cell j. The task of somal inhibition is to drive SINBAD cells to tune to different features of the environment, thus maximizing the number of environmental variables expressed by the network as a whole. To accomplish this task [7], these connections are made anti-Hebbian (see below).

The connections of the hidden units were adjusted according to the error backpropagation algorithm of Rumelhart et al. [13]. Specifically, the error signals δ_d are first computed for the three dendrites as:

$$\delta_d = T - 3 \cdot D_d. \qquad (7)$$

For the hidden units, δ is backpropagated as:

$$\delta_{d,h} = \delta_d \cdot w_{d,h} \cdot (1 - H_{d,h}^2). \qquad (8)$$

Connection weights are adjusted by:

$$\Delta w_{d,i,h} = \mu_i \cdot A_{d,i} \cdot \delta_{d,h} \text{ and}$$
$$\Delta w_{d,h} = \mu_h \cdot H_{d,h} \cdot \delta_d \qquad (9)$$

where μ_i and μ_h are learning rate constants for the input and hidden unit connections. Somal inhibitory connections are adjusted by:

$$\Delta w_j^{-} = \mu_s \cdot \left[-w_j^{-} + (A_j - \overline{A_j}) \cdot (A - \overline{A})\right], \qquad (10)$$

where μ_s is a learning rate constant, and w_j^{-}, A_j and $\overline{A_j}$ are the somal inhibitory connection weight, activity, and average activity, respectively, of presynaptic SINBAD cell j. This synaptic learning rule is anti-Hebbian in its effect, because it makes connection w_j^{-} track the covariance in activities of the two connected cells.

4. SINBAD NETWORK

In our adaptation of the SINBAD network to the problem of image super-resolution, it receives its input from a 7x7 pixel window placed at various locations in the training or testing images. To enhance the contrast among the pixels, the value of the central pixel is subtracted from the values of all the other pixels and these 48 pixels are used as the original environmental variables. Thirty-two SINBAD cells are organized into a "cortical" layer (Figure 2). Upon an exposure to an image, the sensory information from the pixels is transmitted to SINBAD cells via relay cells of the "thalamic" layer, with each thalamic cell reporting the state of one of the pixels. The thalamic layer consists of 48 cells, representing the 48 pixels.

For simplicity, in this exercise we do not take advantage of the cortical topographic mapping mechanisms [5] to arrange the thalamic connections among the dendrites of SINBAD cells. Instead, we distribute thalamic connections among SINBAD dendrites randomly, and thus not as efficiently as might otherwise be possible. Specifically, each thalamic cell connects to every SINBAD cell on the apical dendrite and on one of that cell's two basal dendrites, chosen randomly for each thalamo-cortical pair. Thus, basal dendrites of different SINBAD cells have afferent connections from different combinations of thalamic cells. Furthermore, the two basal dendrites on the same SINBAD cell have connections from different thalamic cells and therefore receive explicit afferent information about *different* pixels. Consequently, in their search for correlated output functions, the two basal dendrites will have to discover and learn to respond to whatever *implicit* information they have in common, thus making this information *explicit* in the cell's output. In effect, the basal dendrites will tune the cell to one of the local image features (hidden variables) that contribute as significant factors to local image regularities.

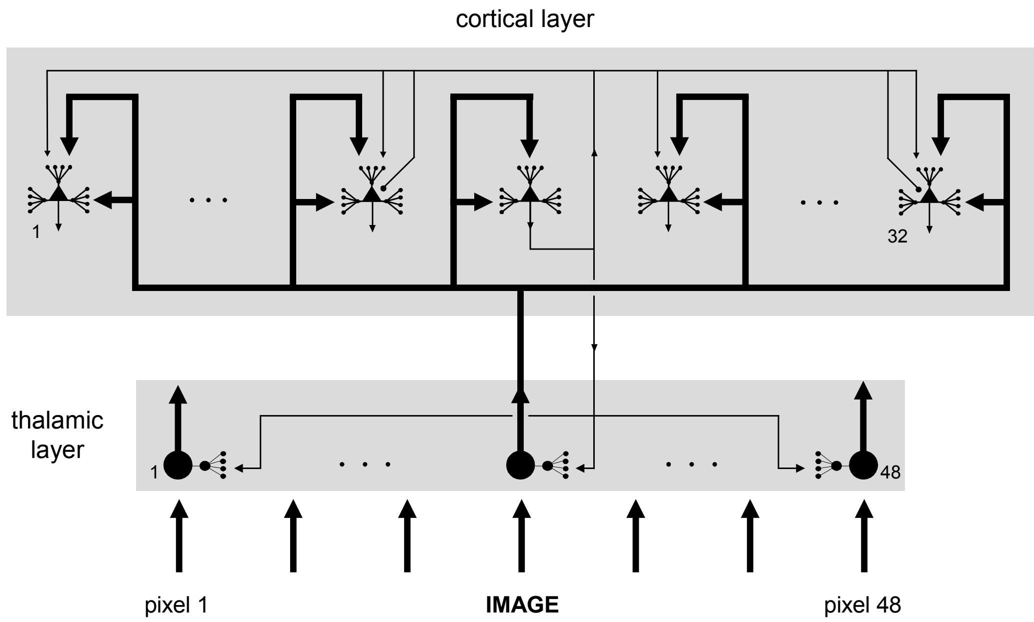

Figure 2. The connectional diagram of the SINBAD network. Thalamic and SINBAD cells are shown as solid circles and triangles, respectively, with their dendrites drawn as miniature hidden layer-to-output unit nets. For clarity, output connections are shown only for one thalamic and one SINBAD cell.

SINBAD cells are interconnected via lateral connections, with each cell connecting to the apical dendrite of every other cell. That means that each apical dendrite receives all the afferent information available from the thalamic layer about the pixels and all the lateral information available from other SINBAD cells about image features discovered by them. In each SINBAD cell, then, the apical dendrite will have the most diverse and comprehensive information about the local image, which it will use to compute the state of the image variable chosen for that cell by its basal dendrites. Thus, the web of inferential relations among image variables is captured in the SINBAD layer specifically by the net of its apical connections.

SINBAD cells are also interconnected via inhibitory somal connections, with each cell connecting to 15 other, randomly chosen, SINBAD cells. The purpose of these connections is to make SINBAD cells tune to different image features. Finally, every SINBAD cell has a connection to every thalamic cell. This feedback system of corticothalamic connections implements Mumford's [10] idea of the thalamus being used by the cortex as a "blackboard," on which the cortex draws its interpretation of the attended subject. The web of inferential relations learned by the cortical network acts as an inferential model of the outside world, and this internal model projects its picture of the outside world back on the thalamus, so that it can be returned again to the cortex for another pass of inferential adjustment and elaboration, and so on. This will enable the cortical inferential model to fill-in holes, when they happen, in the raw picture of the world that the thalamus receives from its sensory channels. We will use this "filling-in" function of corticothalamic feedback to compute the values of the interpolated pixels in the super-resolution image.

To implement Mumford's idea, each thalamic cell is given a dendrite in the form of a backpropagation network, identical to that used to represent dendrites of SINBAD cells (eqs. 1, 2). Hidden units of this dendrite receive connections from all the SINBAD cells. When a thalamic cell receives direct information from the outside world, it's output is the average of the value of the pixel it represents and the output of its feedback dendrite. But, when the cell does not receive information from the outside world, then its output is equal to the output of its feedback dendrite. The weights of input connections to hidden units of the thalamic dendrites and of hidden unit connections to the output unit are adjusted by the error backpropagation algorithm (eqs. 7-9), using the output of the thalamic cell as the training signal T.

5. RESULTS

After initially setting all the adjustable connections to randomly chosen strengths w's, the network was exposed to grayscale images of grass, bushes, and landscapes. Each exposure lasted 7 time steps, during which the states of the SINBAD cells and the thalamic dendrites were computed iteratively from their previous states, thus giving lateral interactions among SINBAD cells time to express themselves. After the seventh time step, all the cortical and thalamic connections were adjusted according to the learning algorithms described above. On 50% of the image exposures, the eight central pixels were not shown to the network; i.e., their states were not given to the thalamic cells representing them, instead making those cells use cortical feedback to determine their outputs.

The network was trained on 2000000 image exposures. During this time the dendrites in each SINBAD cell learned to produce closely correlated outputs, indicating that they tuned to some orderly local image feature. After training, the network was exposed to images in which every 2 out of 3 pixels in each dimension were blanked out (an example is shown in Figure 3). The network's task was to infer the values of these missing pixels from the values of the remaining pixels. This inference is performed by the SINBAD layer and, since the SINBAD layer projects its representation of the outside world back on the thalamic layer, a thalamic cell that was denied the external information about the state of the pixel it represents should nevertheless represent that state more or less accurately, being informed of it by the SINBAD layer.

Thus, to fill-in the image, all the blanked out pixels were first given zero value. Next, the network's 7x7 window was scanned over the entire image in 3-pixel steps, every time centered on one of the known pixels. The eight pixels around the central one were not shown to the thalamic cells, making the network infer them from the surrounding pixels. The not-shown pixels were then assigned the values predicted by the network (i.e., the values of the thalamic cells representing those pixels). The image was scanned 7 times, during which the pixels gradually converged on the stable values.

A typical result is shown in Figure 3. For a comparison we also show a cubic spline interpolation of the same image. This comparison shows that the SINBAD interpolation is visually significantly better than the cubic spline interpolation, filling the reconstructed image with realistic fine spatial details, such as sharp lines and edges. Some local distortions in the SINBAD reconstruction become apparent, when it is compared with the true image; specifically, some lines and edges are broken down into separate segments in the reconstructed image. These local distortions are due to the limited, 7x7 pixel, size of the SINBAD viewing window we used in this study. We expect that larger-size windows will alleviate this problem.

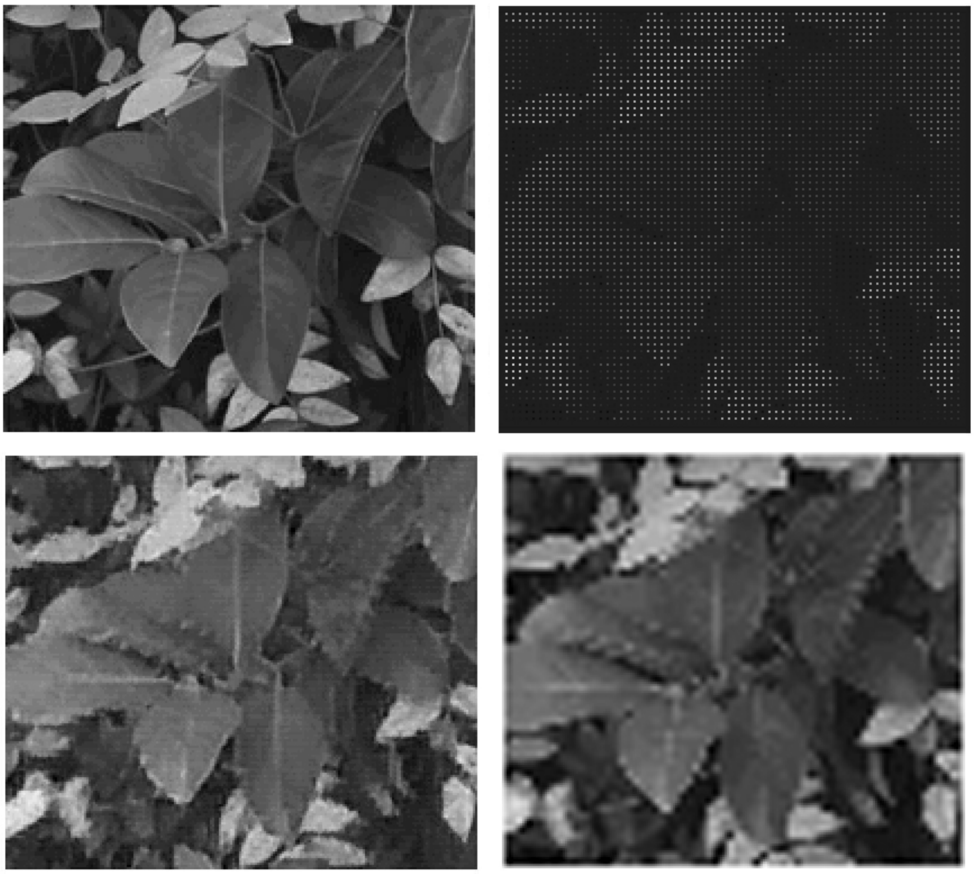

Figure 3. Image super-resolution. **A.** The original image. **B.** The reduced image in which every 2 out of 3 pixels in both directions are masked. The task of super-resolution in this exercise is to reconstruct the masked pixels and thus approximate the original image. **C.** SINBAD network's reconstruction of the reduced image. **D.** Reconstruction of the reduced image by cubic spline interpolation. Note that D image is significantly blurrier than C image.

6. CONCLUSION

These initial results of using the SINBAD network for image super-resolution are very promising. They support the idea of Baker and Kanade [2] and Freeman et al. [8] that greater super-resolution can be achieved by taking advantage of higher-order regularities present in images at various spatial scales. The main aim of this paper is to relate the SINBAD approach to discovering regularities to the problem of super-resolution and to describe the basic design of the SINBAD network in its application to image enlargement.

The network design described here is a minimal one, and in future we plan to expand it by incorporating additional features of the brain's visual information processing. The planned additions include the following. First, we should incorporate visual data preprocessing, performed in the retina, which optimizes the input reaching the cortex by decorrelating information carried by individual channels [1].

Next, the second stage of visual preprocessing is performed by the input layer of the primary visual cortex, resulting in extraction of such primitive features as local lines and edges [11]. By performing these operations on our image data prior to their delivery to the SINBAD layer will free SINBAD cells to learn higher-order visual features, enabling the network to make more insightful inferences.

Finally, just as the cortex is made up of multiple areas that perform progressively higher-level processing, the SINBAD network also should be expanded to include more than one layer of SINBAD cells. This will enable higher-level SINBAD cells to learn progressively higher-order visual features of significance to image regularities, thus enhancing the network's super-resolution capabilities.

REFERENCES

[1] Attick JJ, Redlich AN., "What does the retina know about natural scenes". *Neural Computation* 4: 196-210, 1992.

[2] Baker S, Kanade T., "Super-resolution: reconstruction or recognition?" *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, IEEE, Baltimore, Maryland, 2001

[3] Becker S, Hinton GE "A self-organizing neural network that discovers surfaces in random-dot stereograms." Nature 355: 161-163, 1992.

[4] Clark A, Thornton C., Trading places: computation, representation, and the limits of uninformed learning. *Behavioral and Brain Sciences* 20: 57-90, 1997.

[5] Favorov OV, Ryder D., "SINBAD: A neocortical mechanism for discovering environmental variables and regularities hidden in sensory input." *Biological Cybernetics* (submitted), 2002.

[6] Field DJ "What is the goal of sensory coding." *Neural Computation* 6:559–601,1994.

[7] Foldiak P., "Forming sparse representations by local anti-Hebbian learning." *Biological Cybernetics* 64: 165-170, 1990.

[8] Freeman WT, Jones TR, Pasztor EC., "Example-Based Super-Resolution." *Mitsubishi Electronic Research Laboratory*, TR-2001-30, 2001.

[9] Keys R., "Bicubic interpolation." *IEEE Trans. Acoust. Speech, Signal Processing* 29:1153–1160, 1981

[10] Mumford D., "On the computational architecture of the neocortex. *Biological Cybernetics"* 65: 135-145, 1991

[11] Olshausen BA, Field DJ., "Emergence of simple-cell receptive field properties by learning a sparse code for natural images." *Nature* 381: 607-609, 1996

[12] Phillips WA, Singer W., "In search of common foundations for cortical computation." *Behavioral and Brain Sciences* 20: 657-722, 1997

[13] Rumelhart DE, Hinton GE, Williams RJ., "Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, PDP Research Group (eds) Parallel distributed processing: explorations in the microstructure of cognition",*MIT Press*, Cambridge, Mass, 1: 318-362, 1986

[14] Ryder D, Favorov OV., "The new associationism: a neural explanation for the predictive powers of cerebral cortex." *Brain and Mind* 2: 161-194, 2001

Olcay Kursun received the degree of Master of Science in Computer Science in 2001 from the University of Central Florida, the degrees of Master of Engineering and Bachelor of Science in Computer Engineering from Bogazici University, Istanbul, Turkey. Currently, he has been enrolled in PhD study at the University of Central Florida, where he taught Formal Languages and Automata Theory, and Discrete Computational Structures. His current interests are unsupervised learning with particular emphasis on neural network models, pattern recognition and artificial intelligence.

Oleg Favorov, PhD, received his undergraduate degree in Physical Anthropology from Moscow State University in 1977. He received a doctorate in Physiology at the University of North Carolina, Chapel Hill, in 1986. He did postdoctoral work at Rockefeller University, New York. He subsequently returned to the University of North Carolina, as an Assistant Professor in the Department of Biomedical Engineering. He has been a member of the faculty of the University of Central Florida since 1999. His research interests are functional organization of cerebral cortical networks and their contributions to sensory information processing, cerebral cortex as a computational system.