# Sparse visual models for biologically inspired sensorimotor control

**Li Yang   and   Marwan Jabri**
OGI School of Science & Engineering
Oregon Health and Science University
20000 N.W. Walker Road, Beaverton, OR 97006
{lilyzhou, marwan}@ece.ogi.edu

## Abstract

Given the importance of using resources efficiently in the competition for survival, it is reasonable to think that natural evolution has discovered efficient cortical coding strategies for representing natural visual information. Sparse representations have intrinsic advantages in terms of fault-tolerance and low-power consumption potential, and can therefore be attractive for robot sensorimotor control with powerful dispositions for decision-making. Inspired by the mammalian brain and its visual ventral pathway, we present in this paper a hierarchical sparse coding network architecture that extracts visual features for use in sensorimotor control. Testing with natural images demonstrates that this sparse coding facilitates processing and learning in subsequent layers. Previous studies have shown how the responses of complex cells could be sparsely represented by a higher-order neural layer. Here we extend sparse coding in each network layer, showing that detailed modeling of earlier stages in the visual pathway enhances the characteristics of the receptive fields developed in subsequent stages. The yield network is more dynamic with richer and more biologically plausible input and output representation.

## 1. Introduction

One of the major difficulties in robot navigation is the capability of rapidly responding to unpredictable and novel situations. The level of delegation and autonomy of robotic systems in remote missions will depend on their computational capabilities for decision-making, their adaptation and learning, and their capability to survive uncertain environments. The engineering of such systems will be more complex than ever, and design faults can be very costly. Resilience to design faults, faults due to the elements, and proper responses to novel situations will be critical. Biological systems provide solutions to similar problems, and the brain provides a rich source of computational paradigms that can be used as inspiration for revolutionary computational solutions (Mousset, Jabri et al. 2000; Jabri 2001; Wang, Jin et al. 2002). We are developing in our laboratory biologically inspired sensorimotor control systems for controlling robots and decisions systems. Our approach to visual processing is based on the modeling of the visual ventral pathway, and this is the focus of the present paper.

The mammalian cortex has evolved over millions of years to effectively cope with visual information of the natural environment. Given the importance of using resources efficiently in the competition for survival, it is reasonable to think that natural evolution has discovered efficient cortical coding strategies for representing natural visual information. Here, the notion of efficiency is based on Barlow's principle of redundancy reduction (Barlow 1994), which proposes that a useful goal of sensor coding is to transform the input in a manner that reduces the redundancy due to complex statistical dependencies among elements of the input streams. The usefulness of redundancy reduction can be understood by considering the process of image formation, which occurs by light reflecting off of independent entities (i.e. objects) in the world and being focused onto an array of photoreceptors in the retina. The activities of the photoreceptors themselves do not form a particularly useful signal to the organism because the structure present in the world is not made explicit, but rather is embedded in the form of complex statistical dependencies, or redundancies among photoreceptor activities. A reasonable goal of the visual system is to extract these statistical dependencies so that images may be explained in terms of a collection of independent events, so that means forming the sparse representation for a given image. The hope is that such a sparse coding strategy will recover an explicit representation of the underlying independent entities that gave rise to the image, which would be useful to the survival of organism.

Furthermore, sparse coding (Amari 1993) has been proven to provide superior information storage capacity compared to local (grandmother cell theory or Gnostic representation) or distributed information representations. Because only very few neurons need to be activated and that there are only a few neurons encoding an event, sparse representation have intrinsic fault-tolerance and low-power consumption potential. Fault tolerance is a critical requirement in the remote deployment of intelligent systems, which has been attributed to neural networks because of the distributed representations that develop during learning. Another

important aspect of any physical realization of computational models is the power consumed. Biologically based principles such as sparse representation may have the information processing capabilities as well as huge payoffs in power/energy minimization and optimal resource management. Also, sparse representations are important from an implementation perspective. The physical connectivity of large scale networks require strategies that exploits sparseness of networks, local connectivity, population-based encoding, and information flow in operation as well as during learning.

# 2. Visual processing framework

For sensorimotor control, individual landmarks and goal locations must be extracted from complex visual scenes, and objects and their spatial relationships must be identified. Generally, the real world almost always contains more information than we can process at any given point in time, so we must learn to use it iteratively, searching for the most relevant information at any given point in time. On the other hand, findings from neurophysiology, psychophysics, and fMRI (Reynolds, Chelazzi et al. 1999; Kanwisher and Wojciulik 2000; Reynolds, Pasternak et al. 2000) all point to the roles of attention and stimulus salience in biasing the competition of neurons in ventral stream to facilitate object recognition. When subjects are instructed to attend (or choose voluntarily to attend) to a stimulus at a particular location or with a particular feature, this generates signals within areas outside visual cortex, such as parietal cortex, frontal eye field (FEF), prefrontal cortex, and amygadala. These signals are then fed back to extrastriate areas, where they bias the competition in these areas in favor of neurons that respond to the features or location of the attended stimulus. As a result, neurons that respond to the attended stimulus remain active while suppressing neurons that respond to the ignored stimuli. In other words, neuronal responses are now determined by the attended stimulus. In absence of attention control, the most salient element in the scene might dominate neuronal responses. By virtue of these biological findings, we proposed the visual processing system for sensorimotor control illustrate in Fig. 1.
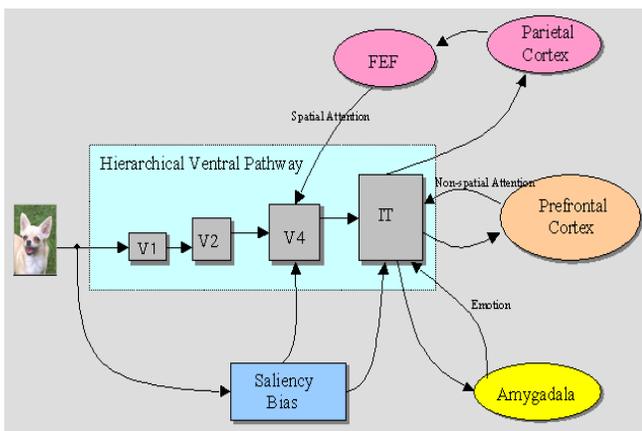


Figure 1. Bottom-up and top-down visual information fusion.

In this visual processing system, the bottom-up salience and top-down attention can complementarily filter out unwanted information from typically cluttered real-world scenes and to focus on what is important in a given situation. This will largely reduce the computational complexity and simplified the object recognition process.

# 3. Model

## 3.1 Visual ventral pathway

Much of the mammal cortex is devoted to visual processing. In the macaque monkey at least 50% of the neocortex appears to be directly involved in vision. The function of visual cortex is dependent on the organization of its connections, the types of synapses they form, and how postsynaptic neurons respond to and integrate synaptic inputs. Roughly, the visual cortex is divided into 5 separate areas, V1, V2, V3, V4, and V5/MT (Zeki 1999). Each of these areas is further subdivided and sends information to any of 20 or more other areas of the brain that process visual information (Hubel 1995). Physiological and anatomical studies suggest that organizing principles in visual cortex is forming an economical representation of the visual world. This representation is formed through a modular analysis that is both parallel and hierarchical. This general arrangement is subdivided into two parallel pathways as Fig 2. Cells in dorsal MST are particularly sensitive to small moving objects or the moving edge of large objects. These cellular characteristics make the dorsal pathway especially able to quickly detect novel or moving stimuli.
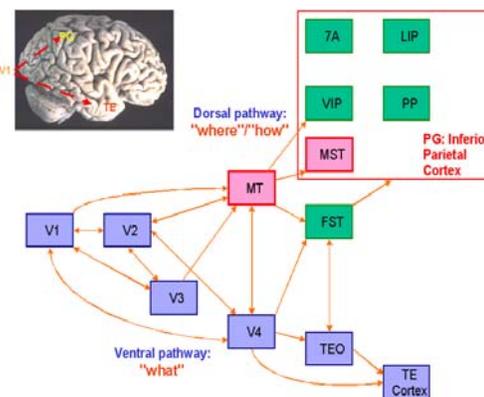


Figure 2. The Visual Pathway

Neurophysiologic and neuropsychological evidences show that the visual ventral pathway (LGN parvo layers → V1 layer 4Cβ → V1 interblobs → V2 interstripes → V4 → IT) identifies what we see. Damage to the ventral pathway will induce disorders of object recognition. Common examples of such disorders include visual agnosia, or the inability to identify objects in the visual realm, and prosopagnosia, a

subtype of visual agnosia that affects specifically the recognition of once familiar faces (Palmer 1999). Although each pathway is somewhat distinct in function, there is intercommunication between them.

Moving through the ventral stream, one can conceive a hierarchy of neurons with the steady increase of receptive field sizes. At corresponding eccentricities near the fovea receptive fields in V2 are (in linear dimensions) 2--3 times larger than in V1; in V4 perhaps 5-6 times larger; cells in IT (inferotemporal cortex) have receptive fields that can include the entire central visual field, on both sides of the vertical midline (Lennie 1998). These large receptive fields are presumably necessary to recognize large complex objects and may mediate the ability to recognize objects of any size as the same, regardless of their retinal location.

## 3.2 Hierarchical network architecture

An essential behavior of animals is the visual recognition of objects that are important for their survival. Human activity, for instance, relies heavily on the classification or identification of a large variety of visual objects (Logothetis and Sheinberg 1996). One of the major problems which must be solved by a visual system for object recognition is the building of a representation of visual information which allows recognition to occur relatively independent of size, contrast, spatial frequency, position on the retina, and angle of view, etc (Ullman, Vidal-Naquet et al. 2002). This requires that features extracted by the visual pathway create a rather complete representation of the current sensory scene using the principle of sparse coding, which means that at any one time only a small selection of all the units is active, yet this small number firing in combination suffices to represent the scene effectively.

Hubel and Wiesel proposed a model in which V1 simple cells with neighboring receptive fields feed into the complex cell with same receptive-field orientation and roughly the same positions, thereby endowing that complex cell with phase and shift invariant features. Following this, visual processing in cortex is classically modeled as a hierarchy of increasingly sophisticated representation (Fukushima 1980; Marr 1982; Biederman 1987; Poggio and Edelman 1990; LeCun, Boser et al. 1992; Riesenhuber and Poggio 1999).

Here we present a hierarchical network architecture (see Fig. 3) with sparse coding constraint to extract low level features (such as edges, orientations, spatial frequencies, and contours) for further processing in the ventral pathway, such as part-based shape representation in cortex V4 (Desimone, Schein et al. 1985; Schiller 1995; Pasupathy and Connor 1999; Pasupathy and Connor 2001), and object recognition in the inferotemporal cortex (IT) (Kobatake and Tanaka 1994; Tanaka 1996).
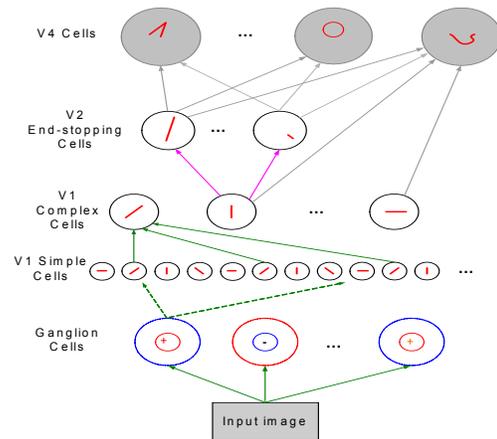


Figure 3. A hierarchical network architecture inspired by the visual ventral pathway. This paper focuses on models up to but not including V4.

## 3.3 V1 Complex cells model

V1 is located in the occipital lobe at the back of the brain. Nearly all visual information reaches the cortex via V1. The receptive fields of V1 simple cells are localized in space and time, have band-pass characteristics in spatial and temporal frequency domains, are oriented, and often sensitive to the direction of motion of a stimulus. This sort of properties encourages the notion that the purpose of the neurons in V1 is to construct economical description of the images. Independent Component Analysis (ICA) on natural images produces receptive fields like those of simple cells (Olshausen and Field 1996; Bell and Sejnowski 1997; Olshausen and Field 1997; Lee 1998; Hyvarinen, Karhunen et al. 2001).

V1 complex cells share the properties of simple cells but have two distinguishing properties of phase invariance and (limited) shift invariance. Extending ICA by combing the principle of invariant-feature subspace and the multi-dimensional ICA, the features similar to those found in complex cells emerged from maximizing the independence/sparseness between the different feature subspaces (Hyvarinen and Hoyer 2000; Hyvarinen and Hoyer 2000). A feature subspace, as any linear subspace, can always be represented by a set of orthogonal basis vectors, say $w_i(x, y), i = 1, ..., M$, where M is the dimension of the subspace. Then the value $F(I)$ of the feature F with input vector $I(x, y)$ is given by

$$F(I) = \sum_{i=1}^{m} s_i^2, \text{ and } s_i = < w_i, I >.$$

In fact, this is equivalent to computing the distance between the input vector $I(x,y)$ and a general linear combination of the weights (filters) $w_i(x,y)$ of the feature subspace. In contrast to ordinary ICA, the components $s_i$ are not assumed to be all mutually independent. Instead, it is assumed that the $s_i$ can be divided into couples, triplets or in general m-tuples, such that $s_i$ inside a given m-tuple may be dependent on each other, but the dependencies between different m-tuples are not allowed. Embedded invariant-feature subspaces in multidimensional ICA analysis, the logarithm of the likelihood $L$ of the $K$ observed image patches $I_k(x,y), k = 1,...K$ is

$$L = \log l(I_k(x,y), k = 1,...,K; w_i(x,y), i = 1,...,M)$$
$$= \sum_{k=1}^{K} \sum_{j=1}^{J} \log p(\sum_{i \in S_j} <w_i, I_k>^2) + K \log|\det W|$$

(1)

where $S_j, j = 1,...,J$ denote j-th subspace; $p(\sum_{i \in S_j} s_i^2) = p_j(s_i, i \in S_j)$ gives the probability density (pdf) inside the j-th subspace of $s_i$, and $W$ is a matrix containing the filters $w_i(x,y)$ as its columns. Prewhitening the image patches $I_k(x,y)$ allows us to consider the $w_i(x,y)$ to be orthonormal, which implies that $\log|\det W|$ is zero, then

$$\Delta W \propto \frac{\partial L}{\partial W} = I(x,y) <w_i, I> g(\sum_{r \in S_{j(i)}} <w_r, I>^2) \quad (2)$$

where j(i) is the index of the subspace to which $w_i$ belongs, and $g = p'/p$ is a nonlinear function.

Since the norm of the projection of visual data on practically any subspace has a super Gaussian distribution, we need to choose the probability density $p$ in the model to be sparse, so we could use the following pdf

$$\log p(\sum_{i \in S_j} s_i^2) = -\alpha \left[ \sum_{i \in S_j} s_i^2 \right]^{1/2} + \beta \quad (3)$$

and then

$$g(u) = p'/p = -\frac{1}{2}\alpha u^{-1/2} \quad (4)$$

After training weights in V1 layer by the above algorithm, we compute the response ($CC_j$) of the j-th complex cell as

$$CC_j = \sqrt{\sum_{i \in S_j} (COV(I(x,y), rf_i))^2} \quad (5)$$

where $COV(.)$ denote the convolution of input image $I(x,y)$ with $rf_i$, and $rf_i$ is the receptive field corresponding to weights $w_i$.

## 3.4 V2 End-stopped cells model

In most respects, cells in V1 and V2 are not remarkably different. V2 is strongly reciprocally connected with V1, and end stopping seems to be more prevalent there, particularly in the pale strips. An ordinary simple cell or complex cell usually shows length summation: the longer the stimulus line, the better is the response, until the line is as long as the receptive field; making the line still longer has no effect. For an end-stopped cell, lengthening the line improves the response up to some limit, but exceeding that limit in one or both direction results in a weaker response (Hubel 1995).

Given the imaging model: $X = \sum_{i=1}^{n} a_i s_i + n$, where n is Gaussian noise. Using the response of a V1 complex cell computed in V1 layer as above, we can assume a non-negative and sparse $s_i$ (Hoyer and Hyvarinen 2002). In the language of probability theory, we wish to match as closely as possible the distribution of observed patterns arising from our imaging model $P(X/A)$ to the actual distribution of patterns observed in nature, $P^*(X)$. To assess how well this match is, we take the Kullback-Leibler (KL) divergence between the two distributions

$$KL = \int P^*(X) \log \frac{P^*(X)}{P(X/A)} dX \quad (6)$$

because $P^*(X)$ is fixed, minimizing KL amounts to maximizing $<\log P(X/A)>$. Since

$$<\log P(X/A)> = \int P^*(X) \log P(X/A) dX \quad (7)$$

the goal of learning will be to find a set of basis $A^*$ that maximize the average log-likelihood of the observed patterns under a sparse, statistically independent prior, such that $A^* = \arg\max_A <\log P(X/A)>$. We can express the objective in an energy function framework by defining $E(X, S/A) = -\log P(X/S, A)P(S)$, we have

$$E(X, A/S) = \sum_n \left[ \|X^{(n)} - AS^{(n)}\|^2 + \lambda \sum_i s_i^{(n)} \right] \quad (8)$$

where $S^{(n)}$ is the vector containing the latent variables $s_i^{(n)}$ corresponding to the n-th observed vector $X^{(n)}$, and the constant $\lambda$ defines the tradeoff between representation error and sparseness. The objective (E) was minimized by standard gradient descent with respect to $S^{(n)}$ in the short

timescale and with respect to $A$ under a longer timescale (Olshausen and Field 1997).

Learned by the above algorithm, the weights (receptive fields of cells) in V2 layer are selective for contour length in addition to being tuned to position and orientation, and exhibit end-stopping properties. It has been proposed that contour feature extraction is the ultimate purpose of end-stopping.

## 4. Experiments and results

In the V1 layer, the image patches (16×16 pixels) for training were randomly sampled from twelve monochrome natural images involving trees, leafs, and animals, and so on. The training set $X = \{I_k, k = 1,...,K\}$ was pre-whitened by: (a) subtracting the mean gray-scale value from the data, this removes the first order statistics; (b) the dimension of the data was then reduced by the principle component analysis (PCA) with the largest variances; the PCA filter is $W_P = D^{-1/2}E^T$, where we have $EDE^{-1} = \langle XX^T \rangle$, and $D$ is the diagonal matrix of eigenvalues, and the columns of $E$ is the eigenvectors of the covariance matrix. Using random initial values for $W$, the likelihood of 50,000 such observations was maximized under the constraint of orthonormality of the filters in the whitened space. Using the learning rule in Eq. (2), we learned 40 complex cells (subspaces) with the subspace dimension of 4. Next, we computed the responses of complex cells using Eq. (5). This process took 3 hours running MATLAB on a Dell Precision Station (530 MT, 2GHZ and 4GB). Fig. 4 shows the response of one of these complex cells by testing with a grayscale dog image. It can be seen that the basis vectors

associated with a single complex cell all have approximately the same orientation and frequency. Their locations are not identical, but close to each other. The phases differ considerably. Note: the responses of V1 cells are very sparse.

We first computed 200 complex cell responses ($CC$) from five natural images, where $CC = \{CC_{i_j}, i = 1,...,5; j = 1,...,40\}$ in the V1 layer. Then, the training set $X = \{x_n, n = 1,...,N\}$ for V2 layer was obtained by randomly extracting 24×24 patches (recall that the receptive field of a V2 cell is typically 2~3 times larger than that of a V1 cell) from 50 complex cell responses among CCs. Using 20,000 such patches, we trained the weights in the V2 layer using the methods described in (Olshausen and Field 1997) and (Hoyer and Hyvarinen 2002). Combining the sparse coding and non-negative constraint, after 40 iterations, the learned 288 weights/receptive fields of V2 cells are shown in Fig. 5. This process took 6 hours by running MATLAB on the same computer mentioned above. Visually, the basis patterns are in different position, different orientation, and different length. Moreover, for characterizing the learned V2 cell receptive fields, we approximated them in the parameter space as done in (Hoyer and Hyvarinen 2002). The main results are shown in Fig. 6, which shows a richer tuning of orientation and length than what has been reported before. This kind of length tuning, or the property of end-stopped cell, is very interesting for visual features representation. As pointed out in (Hoyer and Hyvarinen 2002), the necessity for different length basis patterns comes from the fact that long basis patterns simply cannot code short (or curved) contours and short basis patterns are inefficient at representing long, straight contours.
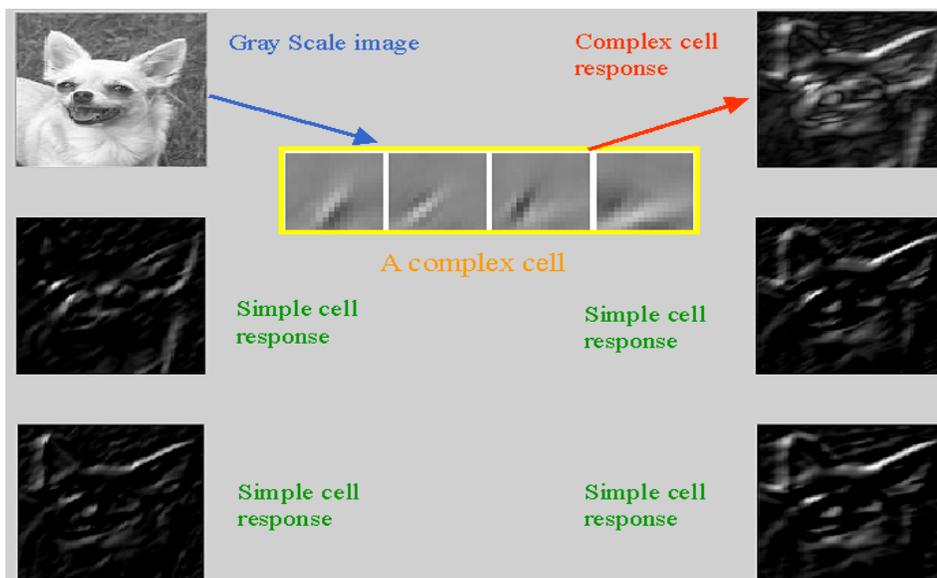


Figure 4. Give an example for the learned complex cell and its response in V1 layer. Every subspace of four basis vectors corresponds to one complex cell. For comparison, we also give the responses of four basis vectors that are the receptive fields of simple cells.
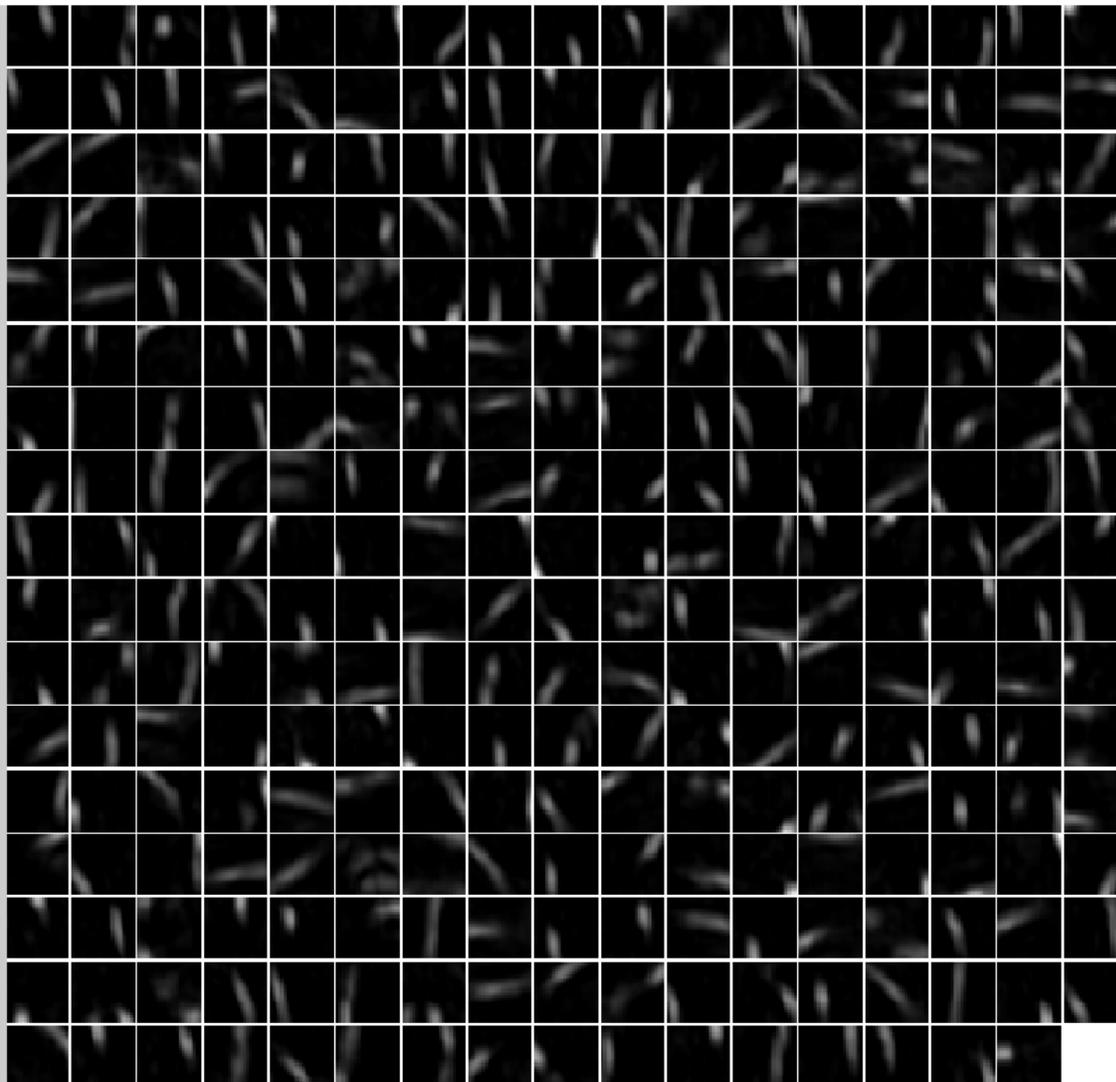
Figure 5. The 288 receptive fields learned in the V2 layer. They are in different position, different orientation and different length.



*(a)*

*(b)*
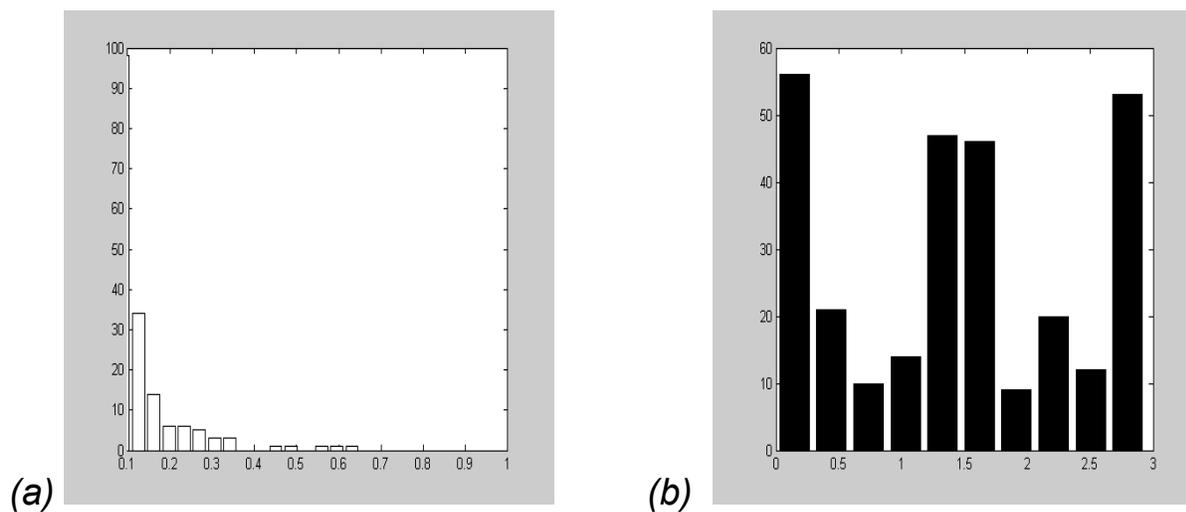
Figure 6. (a) Distribution of the receptive fields lengths in V2 layer, which are normalized by the width of the sampling window; (b) Distribution of the receptive field orientations (from 0° to 180°: label 0~3 in the horizontal axis) in V2 layer.

# 5. Discussion

We presented in this paper a hierarchical network architecture inspired by the mammalian ventral pathway to sparsely represent visual features for use in sensorimotor control. This sparse representation provided intrinsic low power and fault-tolerant computing substrate to sensorimotor control systems. By unsupervised learning algorithms, the learned visual models made the sensorimotor control systems to automatically adapt to uncertain and novel environment. We also show that in such a model, V2 cell receptive fields develop end-stopping properties. According to Hubel and Wiesel, the optimal stimulus for an end-stopped cell is a line that extends for a certain distance and no further. For a cell that responds to edges and is end-stopped at one end only, a corner is ideal; for a cell that responds to slits or black bars and is stopped at both ends, the optimum stimulus is a short white or black line or a line that curves so that it is appropriate in the activating region and inappropriate (different by 20 to 30 degrees or more) in flanking regions. We can thus view end-stopped cells as sensitive to corners, to curvature, or to sudden breaks in line. These contours are very crucial for shape representation in cortex V4 (Gallant, Braun et al. 1993; Wilkinson, James et al. 2000; Pasupathy and Connor 2001), thus they are very important for object representation and recognition in IT.

Our approach is related to Hoyer's contour coding network (Hoyer and Hyvarinen 2002). However, Hoyer computed the complex cell responses by a simple energy model, therefore the receptive fields in his V1 layer are fixed, or pre-calculated. In contrast, our approach uses the end-to-end learnt receptive fields, and thus represents the natural image sparsely and sufficiently (see Fig. 4). Also, the property of the receptive fields and their sizes in our architecture are richer and more in-line with the diversity known from biology. Note that repeating Hoyer's experiments using 100,000 image patches and 100 iterations took 2 days on the same computer mentioned above, the selective resulting basis patterns are shown in Fig. 7. Practically, using the responses of V2 cells in our architecture, we have trained the
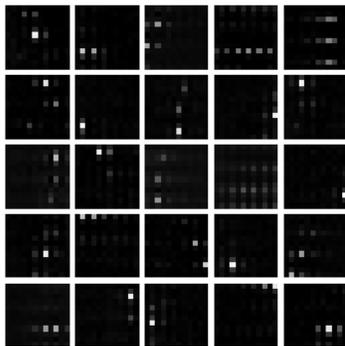
V4 layer and obtained some interesting results, such as object parts. However, the responses of V2 cells produced by Hoyer's model are too weak to be further used in high layers.

Our study is also related to the predictive coding model of (Rao and Ballard 1999), in which, the feedback connections from a higher- to a lower- order visual area carry predictions of lower-level neural activities. The feedforward connections carry the residual errors between the predictions and the actual lower-level activities. They proposed that end-stopping cell stopped responding when the stimulus length was increased because then it could be predicted and there were no residual errors.

## Acknowledgements

## References

Amari, S. (1993). Neural representation of information by sparse encoding. *Brain Mechanisms of Perception and Memory from Neuron to Behavior*. T. O. e. al, Oxford University Press: 630-637.

Barlow, H. (1994). What is the computational goal of the Neocortex? *Large-Scale Neuronal Theories of the Brain*. C. K. a. J. L. Davis.

Bell, A. J. and T. Sejnowski (1997). The independent components of natural scenes are edge filters. *Vision Research,* 37(23).

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review,* 94(2): 115-147.

Desimone, R., S. J. Schein, et al. (1985). Contour, color and shape analysis beyond the striate cortex. *Vision Research,* 25(3): 441-452.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics,* 36: 193-202.

Gallant, J. L., J. Braun, et al. (1993). Selectivity for polar, hyperbolic, and cartesian gratings in Macaque visual cortex. *Science,* 259: 100-103.

Hoyer, P. O. and A. Hyvarinen (2002). A multi-layer sparse coding network learns contour coding from natural images. *Vision Research,* 42(12): 1593-1605.

Hubel, D. H. (1995). *Eye, Brain, and Vision*, Scientific American Library.

Figure 7. A selective set of basis function learned by Hoyer's network.

Hyvarinen, A. and P. Hoyer (2000). Emergence of topography and complex cell properties from natural images using extensions of ICA. *Advances in Neural Information Processing Systems* 12 (NIPS*99): 827-833.

Hyvarinen, A. and P. O. Hoyer (2000). Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation,* 12(7): 1705-1720.

Hyvarinen, A., J. Karhunen, et al. (2001). *Independent Component Analysis*, John Wiley & Sons, Inc.

Jabri, M. A. (2001). Biological computing for robot navigation and control (*Technical Report 2001-A*). Beaverton, Oregon, OGI school of Science & Engineering, Oregon Health & Science University.

Kanwisher, N. and E. Wojciulik (2000). Visual Attention: Insights from Brain Imaging. *Nature Reviews Neuroscience,* 1: 91-100.

Kobatake, E. and K. Tanaka (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the Macaque cerebral cortex. *Journal of Neurophysiology,* 71(3): 856-867.

LeCun, Y., B. Boser, et al. (1992). Handwritten digit recognition with a back-propagation network. *Neural Networks, current applications*. L. P. G. J.

Lee,T.-W. (1998). *Independent Component Analysis_Theory and Application*. Boston, Kluwer Academic Publishers.

Lennie, P. (1998). Single units and visual cortical organization. *Perception,* 27: 889-935.

Logothetis, N. K. and D. L. Sheinberg (1996). Visual object recognition. *Annual Review of Neuroscience,* 19: 577-621.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco, W H Freeman & Co.

Mousset, E., M. A. Jabri, et al. (2000). *Gaze-shifting in humans and humanoids*. IEEE Humanoids 2000 conference, MIT, IEEE Robotics and Automation Society.

Olshausen, B. A. and D. J. Field (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature,* 381(13): 607-609.

Olshausen, B. A. and D. J. Field (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research,* 37: 3311-3325.

Palmer, S. E. (1999). *Vision Science-Photons to Phenomenology*. Cambridge, Massachusetts, The MIT Press.

Pasupathy, A. and C. E. Connor (1999). Responses to contour features in Macaque area V4. *Journal of Neurophysiology,* 82: 2490-2502.

Pasupathy, A. and C. E. Connor (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *Journal of Neurophysiology,* 86: 2505-2519.

Poggio, T. and S. Edelman (1990). A network that learns to recognize 3D objects. *Nature,* 343: 263-266.

Rao, R. P. N. and D. H. Ballard (1999). Predictive Coding in the Visual Cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience,* 2(1): 79-87.

Reynolds, J. H., L. Chelazzi, et al. (1999). Competitive Mechanisms Subserve Attention in Macaque Areas V2 and V4. *The Journal of Neuroscience,* 19: 1736-1753.

Reynolds, J. H., T. Pasternak, et al. (2000). Attention Increases Sensitivity of V4 Neurons. *Neuron,* 26: 703-714.

Riesenhuber, M. and T. Poggio (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience,* 2: 1019-1025.

Schiller, P. H. (1995). Effect of lesion in visual cortical area V4 on the recognition of transformed objects. *Nature,* 376(27): 342-344.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience,* 19: 109-139.

Ullman, S., M. Vidal-Naquet, et al. (2002). Visual features of intermediate complexity and their use in classfication. *Nature Neuroscience,* 5(7): 682-687.

Wang, X., J. Jin, et al. (2002). Neural network models for the gaze shift system in the superior colliculus and cerebellum. *Neural Networks,* 15(7): 811-832.

Wilkinson, F., T. W. James, et al. (2000). An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. *Current Biology,* 10: 1455-1458.

Zeki, S. (1999). *Inner Vision*. New York, Oxford University Press.