Science in China Series C: Life Sciences

www.scichina.com
life.scichina.com

Computational model for perception of objects and motions

YANG WenLu[1,2], ZHANG LiQing[1†] & MA LiBo[1]

[1] Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;
[2] Department of Electronic Engineering, Shanghai Maritime University, Shanghai 200135, China

Perception of objects and motions in the visual scene is one of the basic problems in the visual system. There exist 'What' and 'Where' pathways in the superior visual cortex, starting from the simple cells in the primary visual cortex. The former is able to perceive objects such as forms, color, and texture, and the latter perceives 'where', for example, velocity and direction of spatial movement of objects. This paper explores brain-like computational architectures of visual information processing. We propose a visual perceptual model and computational mechanism for training the perceptual model. The computational model is a three-layer network. The first layer is the input layer which is used to receive the stimuli from natural environments. The second layer is designed for representing the internal neural information. The connections between the first layer and the second layer, called the receptive fields of neurons, are self-adaptively learned based on principle of sparse neural representation. To this end, we introduce Kullback-Leibler divergence as the measure of independence between neural responses and derive the learning algorithm based on minimizing the cost function. The proposed algorithm is applied to train the basis functions, namely receptive fields, which are localized, oriented, and bandpassed. The resultant receptive fields of neurons in the second layer have the characteristics resembling that of simple cells in the primary visual cortex. Based on these basis functions, we further construct the third layer for perception of what and where in the superior visual cortex. The proposed model is able to perceive objects and their motions with a high accuracy and strong robustness against additive noise. Computer simulation results in the final section show the feasibility of the proposed perceptual model and high efficiency of the learning algorithm.

visual perception, visual cortex, computational model, receptive fields, simple cells, complex cells

Human visual system plays an important role in perceiving environmental world. Natural scenes include very complex information such as object's form, colour, texture, spatial velocity and direction of motion, and the visual cortex has adaptively evolved into a number of functional areas for perception of varieties of such information in a very long time. These functional areas have been grouped into two main pathways: 'what' and 'where', shown in Figure 1. The former recognizes objects according to their forms and colours, and the latter responds to the spatial velocity and motion direction of objects. 'What' pathway deals with object features hierachically, starting from the primary visual cortex (V1), through V2, V4, and passing to the area TE (inferior temporal cortex). From V1, the size of receptive fields (RFs) of neurons in senior cortex becomes bigger and bigger, such as V1 (1.5°), V4 (4°), and TE (26×26°)[1]. These phenomena show that information is processed from locally to globally. The pathway of 'Where' is from 4Cα in V1 to 4B, then directly or

Received April 30, 2007; accepted April 15, 2008
doi: 10.1007/s11427-008-0074-0
†Corresponding author (email: wenluyang@online.sh.cn or zhang-lq@cs.sjtu.edu.cn)
Supported by the National Basic Research Program of China (Grant No. 2005CB724301), and National High-Tech Research Program of China (Grant No. 2006AA01Z125)

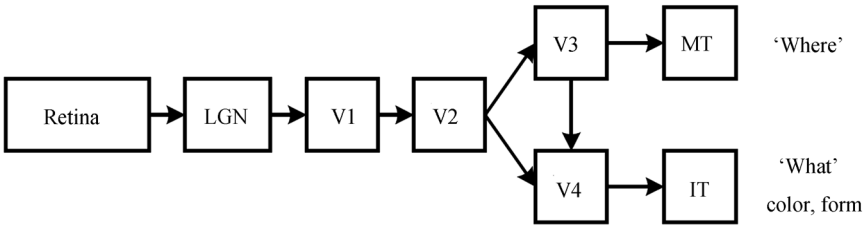

Figure 1 Two pathways: 'what' and 'where'.

indirectly through V2, V3, and to V5 (MT). Most of neurons in MT respond to motion direction and velocity of stimulus, and sensitively to parallax of two eyes. They focus on analyzing three-dimensional track of spatial motion of objects, whereas they pay no attention to the form of objects. Although MT is the lowest area for selectively analyzing motion in the hierarchical visual cortex, there are comparative proportions of the neurons in V1 and V2 sensitive to direction, speed, and parallax[2].

From this viewpoint, information processing of neurons in V1 is vital to the information analysis of superior visual cortex. Hubel and Wiesel[3] studied neural cells in V1 and found that there are two varieties of cells, called simple cells and complex cells. Simple cells have properties such as localization, orientation, and bandpass[4], and complex cells respond to the oriented bar-like stimuli, whereas insensitively respond to phase of stimulus, regardless of stimulus location in the RFs. Hubel et al.[2] further proposed the hypothesis that RFs of complex cells are composed of that of simple cells, and form RFs of hyper-complex cells. That is consistent with the biological mechanism of inform- ation processing.

On the other hand, natural scenes include infinite information, and visual system has limited neural computing resources. How to compromise the contradiction is a difficult problem. By analyzing the relationship between the statistics of natural scenes and neural responses, researchers have proposed many efficient theories and models for neural coding. For example, Barlow[5] proposed in 1961 Efficient Coding that an important constraint in information coding is that a group of neurons should make use of limited resources to code information as much as possible. Olshausen and Field[6,7] presented Sparse Coding according to the fact that probability distribution of natural scenes obeys non-Gaussian distribution and that only minor neurons respond strongly to the environmental stimulus, whereas most neurons do weakly. They proposed that natural scenes can be constructed by linear combination of many bases which are adaptively learned from natural scenes. These basis functions resemble the RFs of simple cells and the corresponding coefficients are super-Gaussian. Similar study is that statistical independence is imposed on neural responses. For instance, Bell and Sejnowski[8], van Hateren and van der Schaaf[9], Lewicki and Olshausen[10], Hyvarinen and Hoyer[11] used independent component analysis (ICA) to obtain similar results: learned basis functions are localized, oriented and bandpassed. These properties resemble the RFs of simple cells in V1 and responses of neurons are super-Gaussian.

From the viewpoint of computing, there are many models overseas that simulate the perceptual mechanism in visual cortex. Hyvarinen and Hoyer[11] proposed a two-layer network that modeled RFs of simple and complex cells and obtained a self-organized spatio-topological map. Grimes and Rao[12] presented a bilinear generative model that studied objects and location of stimuli. Bednar et al.[13,14] applied the HLISSOM model to learning face selectivity in the visual cortex of newborn, helped understanding development of face perception, and interpreted that coaction between environment and gene comes into being the self-adaptively complex brain perception system[13]. The LISSOM model considers lateral connection between neurons in visual cortex and applies Hebbian rule to learning SOM maps of orientation selectivity, direction selectivity, and ocular dominance from motion images[14]. Serre and Stringer[15] proposed a multi-layer model that recognized objects using manual RFs and Max pooling. Rolls et al.[16] proposed a VisNet network that modeled the pathway of V1, V2, MT, MST and recognized in-plane rotation invariance. Grossberg proposed an aFILM[17] model to explain mechanisms of how to self-adaptively process lighting, spatial contrast, and surface filling. Their another model LIGHTSHA FT[18] applied one-eyed texture to perception of 3D surface and make clear how the areas of V1,

V2, and V4 convert from 2D images to 3D form. Bhatt proposed a dARTEX[19] model to interpret that coaction between laminars in cortex was able to learn and recognize object texture and form boundary.

In our country, researchers have reaped rich harvests about the mechanism of visual perception. Yang et al.[3,21] studied simple and complex cells in V1 and proposed sparse coding strategy of simple cells and the spatiotemporal coding model of complex cells. They have constructed a spatiotemporal coding model of complex cells based on the spatiotemporal filter windows of simple cells. The model is especially concerned with the coding representation in cortex of visual input. The finer structure of spatiotemporal integration coding series of complex cells in visual cortex could represent visual inputs. Tian and Lu[22] proposed a four-layer feedforward network which is characterized by modifying the Hebbian learning rule through introducing the asymmetric time window of synaptic modification found recently in neurophysiology. The model can generate by self- organization more diversified spatial-temporal response characteristics of neuronal RF. Wei[23] proposed a self-organized and adaptive model of hyper column in primary visual cortex[24], which realized hierarchical processing input pixels from retina. Mei and Zhang[25] proposed a model of simple cells which found out the relationship between ICA bases and RFs of simple cells. And the model could simulate how environments influenced development of RFs of simple cells when infant animal is at vital growing age. Chen[26] studied cell's orientation selectivity influenced by integration fields, and explained that integration fields could promote efficiency and capability of processing information in visual cortex. Shi and Shi[27] presented a spatiotemporal coding model, which made clear that neuron population would be formed by co-promotion and competition when given some stimulus.

On the basis of the idea that perception of complex information can be from that of simple information, we consider the internal sparse neural representation in V1 and further construct the superior perceptual network that perceives objects and motion direction.

1 Perceptual model

In this section, we propose a three-layer perceptual model, shown in Figure 2. The first layer, called retinal layer, is used to receive stimuli from natural environments. The second layer, called internal neural representation, is designed for representing the internal neural information and each neuron responds to stimulus through its RFs that are learned adaptively from natural scenes. And the third layer is the perceptual layer that perceives objects and motions from stimuli. We now introduce the perceptual model in detail.

In the retinal layer, each neuron represents the pixel gray value of environmental stimuli as its activity, denoted by u_{t_k}, where t_k denotes time k. The activities of neurons or pixels are considered as the input pattern to the second layer.

The second layer, called internal neural representation layer, has neurons of size $M\times N$. The neurons are connected with the first layer through their RFs in the following way $X_{t_k}^{\tau_i} = \mathcal{A}^{\tau_i} u_{t_k}$, where $\mathcal{A}^{\tau_i}$ (i=1, 2, ⋯, K) denotes the K groups of basis functions perceiving motion (Figure 4). In general, neurons fire actively when stimuli are similar to their RFs and only a small number of populations are activated by a natural image at a time. The activity is referred to as the internal neural representation. Due to the influence of natural environment, visual system evolves itself to adapt to the statistics of natural images. Therefore neural RFs can be learned from natural scenes[6,8,11].

The third layer, called perceptual layer, is used to perceive objects and motions based on internal neural representation. There are two kinds of neurons in this layer. One is used for object perception, denoted by Y_C (totally $M\times N$), and the other for motion perception denoted by Y_D ($K\times K$), respectively. The relationship between Y_C/Y_D and the second layer is described as follows:

$$\begin{aligned} &Y_C(m,n) = MAX(X^{\tau_k}(m,n)), (1 \leqslant k \leqslant K) \\ &Y_D(i,j) = MAX(X_{t_1}^{\tau_i}) \cap MAX(X_{t_2}^{\tau_j}), (1 \leqslant i,j \leqslant K), \end{aligned} \quad (1)$$

where MAX denotes an operator which selects the neuron with maximal activity as the winner one[30,31], $\cap$ indicates that $Y_D(i,j)$ fires when $MAX(X_{t_1}^{\tau_i})$ and $MAX(X_{t_2}^{\tau_j})$ are satisfied at the same time. Using the strategy of 'winner takes all', winner neuron Y_C is the one that maximizes the activities of its connected neurons in the second layer. And the winner neuron will be considered as the output of the perceptual layer. For perceiving motion, consider two stimuli at time t_1 and t_2,

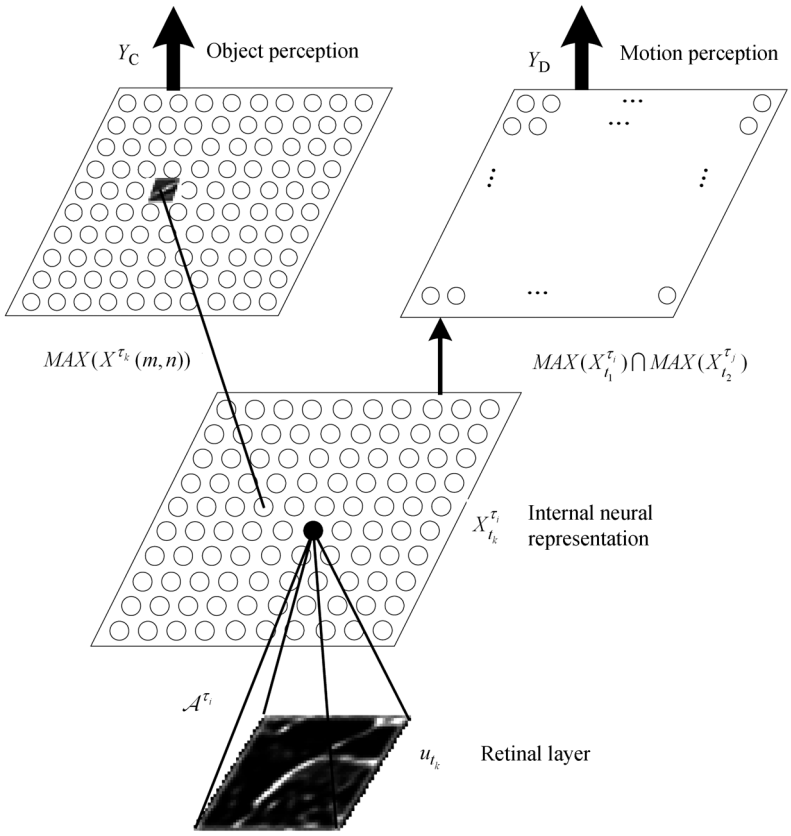

Figure 2 Perceptual model of objects and motions.

and estimate two winner neurons' positions (x_i, y_i) and (x_j, y_j) through bases $\mathcal{A}^{\tau_i}$ and $\mathcal{A}^{\tau_j}$. Then we estimate the relative distance of the two winner neurons via their Euclidian distance $\Delta d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, and the motion direction can be also obtained through the two locations at time t_1 and t_2.

2 Learning algorithm

According to the model presented in the previous section, the RFs of neurons in the internal neural representation layer should have ability to perceive motion. In order to obtain such RFs, we apply the independent component analysis (ICA) model to train the neural network from natural images. First, we introduce the learning rule based on Kullback-Leibler divergence and Natural Gradient[28,29]. Then, we provide in detail implementation of learning algorithm for internal representation and perceptual algorithm for objects and motions perception.

2.1 Learning rule

On the basis of Efficient Coding proposed by Barlow[5], Olshausen and Field[6] presented Sparse Coding for natural image representation. To this end, we apply ICA to learn basis functions from natural images. For the standard model of ICA: x=Wu, using the derivation procedure of learning rule in the literature[32], we derive a cost function from Kullback-Leibler divergence as follows:

$$R(x, \mathrm{W}) = -\frac{1}{2}\log|\det(\mathrm{WW}^T)| - \sum_{i=1}^{n} E\log q_i(x_i), \quad (2)$$

where $q_i(x_i)$ takes the Laplace probability distribution because independent components of natural images follow non-Gaussian distribution[7].

Minimizing the above cost function leads to the following natural gradient[28,29] algorithm:

$$\begin{aligned}\Delta \boldsymbol{W} &= -\eta(t)\frac{\partial R}{\partial \boldsymbol{W}}\boldsymbol{W}^T\boldsymbol{W} \\ &= \eta(t)[\mathrm{I} - \left\langle \varphi[\mathrm{x}(k)]\mathrm{u}^T(k)\mathrm{W}^T \right\rangle]\boldsymbol{W} \\ &= \eta(t)[\mathrm{I} - \left\langle \varphi[\mathrm{x}(k)]\mathrm{x}^T(k) \right\rangle]\boldsymbol{W}, \end{aligned} \quad (3)$$

where $\varphi_i(x_i) = -q_i'(x_i)/q_i(x_i)$, $\eta(t)$ is the learning rate, which approaches to zero during iteration, $\langle\cdot\rangle$ denotes the batch mean.

ICA learning algorithm generates basis functions trained from natural images, resembling the RFs of simple cells[6,7]. The standard ICA provides two inherent uncertainties. One is no order of independent components and the other is that amplitudes of ICs are normalized to unity. Thus, although the learned basis functions have similar properties as RFs of simple cells: localized, oriented, and bandpassed, they have no spatio- topological structure. This result is not consistent with the characteristic of RFs found in experimental results in physiology. Neurophysiologic experiments show that neighboring neurons have similar RFs and their orientation changes periodically in clockwise or counterclockwise. In order to achieve such a feature map, we examine the relationship between two ICs.

ICs from standard ICA are not completely independent, but two-order correlated. This result is in accord with the explanation in physiology that a group of neurons are activated together when given stimuli with some feature. From this physiological finding, we make use of two-order residual dependence to organize the ICs in the form of the self-organization map of RFs of simple cells. Such characteristic is similar to the RFs of complex cells. The result is consistent with hierarchical RFs of cells constructed by Hubel and Wiesel[3]. Now, we will derive the learning rule for self-organization map of RFs of simple cells based on Natural Gradient.

Assume that x_i and x_j are responses of two neurons. If x_i and x_j are spatially adjacent, $\mathrm{cov}(x_i^2, x_j^2) = \mathrm{E}\{x_i^2, x_j^2\} - \mathrm{E}\{x_i^2\}\mathrm{E}\{x_j^2\} \neq 0$. Due to the second order correlation of neighboring neuron responses, activity of each complex cell is represented approximately by square root of sum energy of activity of neighboring simple cells[11]. In other words, RFs of simple cells constitute that of complex cells, and size of RFs of complex cells is larger than that of simple cells. In the mathematical equation, that is $|y_i| = \left(\sum_{j=1}^{n} h_{ij} x_j^2\right)^{1/2}$, where h_{ij} is the ith weight between a complex cell and its connected simple cell j. Assume that the probability distribution of neuronal responses follows the Laplace function:

$$q(y_i) = \frac{1}{\sqrt{2}\sigma}\exp\left(-\frac{\sqrt{2}\,|y_i|}{\sigma}\right) = \frac{1}{\sqrt{2}\sigma}\exp\left(-\frac{\sqrt{2}}{\sigma}\sqrt{\sum_{j=1}^{n} h_{ij} x_j^2}\right), \quad (4)$$

where σ^2 is the variance of responses. We can obtain:

$$\varphi_i(y_i) = -\frac{q_i'(y_i)}{q_i(y_i)} = \left(\sum_{j=1}^{n} \frac{\sqrt{2} h_{ij} x_j}{\sqrt{\sum_{j=1}^{n} h_{ij} x_j^2}}\right). \quad (5)$$

For all y_i, rewrite in the matrix form:

$$\varphi(\mathrm{y}) = \left[\varphi(y_1), \varphi(y_2), \cdots, \varphi(y_{MN})\right]. \quad (6)$$

Substituting the right side of eqs. (6) to (3) leads to the batch updating learning algorithm with self-organization as follows:

$$\Delta W = \eta(t)[\mathrm{I} - \langle \varphi(\mathrm{y}(k))\mathrm{x}(k)^T \rangle W. \quad (7)$$

According to eq. (7), **W** is adaptively updated, resulting in the topographical map of RFs of simple cells. Neighboring RFs with the topographical map constitutes larger RFs as new superior neurons. Such neighboring RFs possess the same properties as complex cells, that is, they are sensitive in orientation, but insensitivity in phase. Now, we have obtained RFs of neurons in the second layer[6,7], which will be used for perceiving objects and motions from stimuli.

2.2 Algorithms for perception

According to the learning rule in subsection 2.1, the RFs of neurons in the internal neural representation layer can be learned from natural images. Based on the learned connected weights, we propose a computational model for perceiving objects and motions from stimuli. For object perception, only one stimulus is necessary. For motion perception, two consecutive stimuli are needed. In the next subsection, we will present algorithms of the model for perception of objects and motions.

(i) Algorithm for learning basis functions. Training data set is first preprocessed by centering and whitening. Such a preprocessing facilitates to explore high-order statistical analysis regardless of first- and second-order correlation, and another advantage is to improve learning speed of convergence. According to eq. (7), update **W** and normalize it to unity, till norm (ΔW) is less than a given threshold. The final step is to project learned basis functions from whitening subspace to original subspace, and basis functions **A** and spatial filters **W** are obtained. The algorithm is described in detail as follows:

1, Generate training data set;

2, Use PCA to center and whiten training set as the input to the learning algorithm;

3, Calculate x by equation $x = \mathbf{W}u$;

4, Update and normalize **W** by eq. (7);

5, If norm(ΔW)$\leqslant\varepsilon$ (threshold), go to step 6, otherwise go to step 3;

6, Stop and output filters **W** and bases **A**.

(ii) Algorithm for perception of objects. Perceiving objects is based on the neural representation. After basis functions for neural representation are trained, we can calculate the neural responses in the second layer given a stimulus. The perception layer is to find out the maximal activity. The object perception algorithm is as follows:

1, Stimulate the perceptual mode with u_{t_k}, and calculate responses of neurons in the internal neural representation layer: $X_{t_k}^{\tau_i}$;

2, Calculate the response of Y_C using eq. (1);

3, Find out the winner neuron with the maximal activity;

4, Output the corresponding representation of the winner neuron.

(iii) Algorithm for motion perception. For perception of motion, two consecutive stimuli are necessary. We first calculate responses of motion neurons, then find out the locations of neurons with the maximal activity with two stimuli. Next, we calculate the distance and motion direction. The procedure for motion perception is rewritten as follows:

1, Given two stimuli u_{t_1} and u_{t_2}, calculate the responses $X_{t_1}^{\tau_i}$ and $X_{t_2}^{\tau_i}$ in the second layer;

2, Calculate two maximal responses of Y_{D1} and Y_{D2} by eq. (1);

3, Find the corresponding locations of Y_{D1} and Y_{D2} as the start point and the end point, respectively;

4, Calculate the distance and direction between two points;

5, If time interval is given, motion speed can be obtained;

6, Output direction and speed as the final results.

From the description of algorithms, we see that the model has two-stages. The first stage is to learn basis functions from natural images as the RFs of simple cells. The second stage is to perceive objects and motions from stimuli. The next section will demonstrate computer simulation results.

3 Computer simulations

Basis functions for image representation can be learned from patches sampled from natural scenes published by Olshausen and Field[7]. In the learning process, connecting weights are randomly initialized. After the receptive fields for neural representation are obtained, the neural response of the network will be used for perceiving objects and motion direction.

3.1 Data preprocessing

The human visual system has been evolving with the environments in which natural scenes possess plentiful oriented information. Therefore, the visual system has the exceptional ability of perceiving information in natural images. It is reasonable to use natural images as training data. The natural images we used were published by Olshausen and Field[7]. It should be noted that more experiments have shown that similar results can be obtained from any natural images with abundant orientation information. It is convenient to make comparisons with other work by using the same image data, avoiding any specialty of training data.

We select ten natural images of size of 512×512 pixels. Training patches are generated by the method shown in Figure 3. Select randomly a number of big images, sample patches with size 10×10 pixels, and vectorize them from top left to bottom right. The arranged data is considered as one sample similar to refs. [33,34]. The total number of patch samples is 10000, denoted by **U**. Then we use principal component analysis (PCA) to center, whiten, and reduce dimensionality to 100 components. The resulting data of size 100×10000 is the training data.

3.2 Learning basis functions

First, the ICA learning rule is applied to the training data set to obtain the basis functions in the whitening subspace according to subsection 3.1. It is necessary to project them into the original image space.

Figure 4 shows the spatio-organized topological map of RFs of neurons in the internal neural representation layer. Here, we consider 5×5 small blocks as one neuron's RF of the corresponding neuron in the first layer. These RFs have almost the same orientation and neighboring neurons have similar and gradually changing orientation. This topological map resembles that found in physiological experiments. To see them more clearly, we enlarge them on the right side of Figure 4. In each block, the basis function is localized, oriented, and bandpassed. That is consistent with the characteristics of

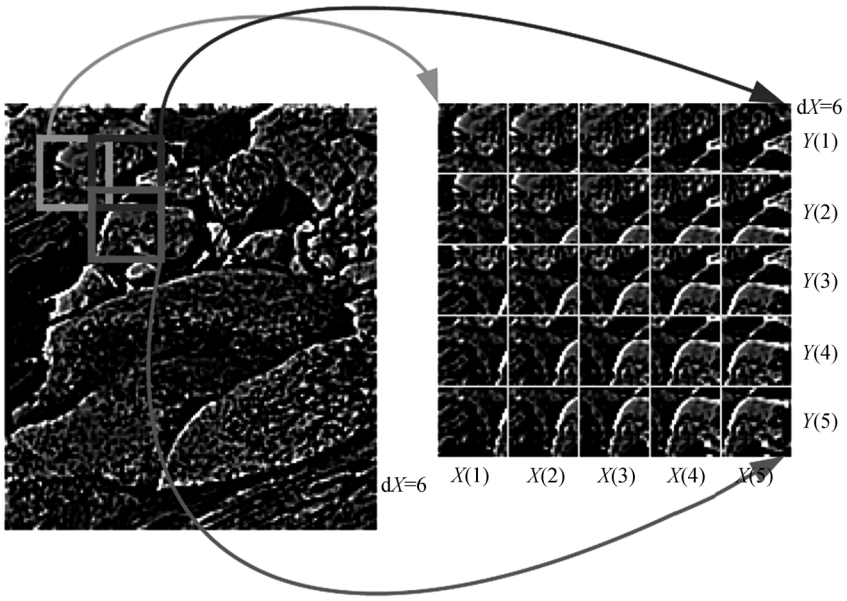

Figure 3 Sample of patches. Select randomly a pixel in big images as the coordinates of top-left, sample a patch, shift dX pixels rightward, and sample the second, till the fifth. Like the way, Then return the left and shift dY pixels downward. Totally 25 patches are sampled and vecterized as a sample.

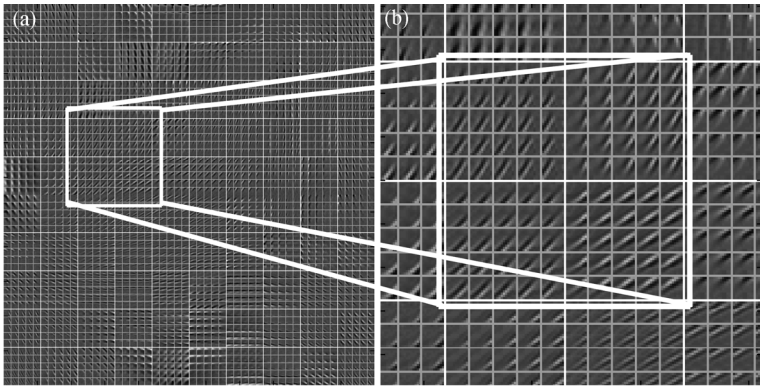

Figure 4 Spatio-organized topographical map of RFs of simple cells arranged by orientation.

RFs of simple cells found in physiological experiments[3]. Each base can be fitted well by a standard Gabor function[10].

The learning rule for perceiving objects is derived based on second-order correlation according to the neurobiological observation that neurons are activated with high-degree synchronization. Therefore, we can obtain similar RFs of complex cells, shown in Figure 5. With comparison of Figures 4 and 5, it is easily found that Y_C connected to the neuronal group in Figure 4 is to perceive orientation information from the stimulus, which is invariant to local spatial shifts. While Y_D, connected to the neuronal group in Figure 5 is to perceive motion information from the stimulus, which is invariant to variation of its contents. In the following we will present computer perceptual results of objects and motions.

3.3 Perception of objects

Because neurons respond strongly to the stimuli which resemble their RFs, for simplicity, we consider RFs of simple cells as optimal stimuli. Randomly select 1000 out of 2500 basis functions, and add Gaussian noises with different covariance to them, resulting as the testing data set in the following simulations.

Given image stimuli, some neurons in the first layer fire strongly. Adopting the strategy of 'winner takes all', the winner neuron Y_C is the one that the activity of its connected neurons in the second layer is maximized and the winner neuron is considered as the neural representation of the stimuli in the second layer. In other words,

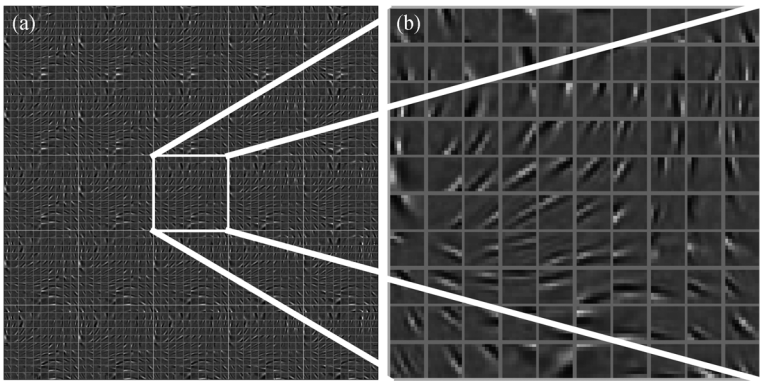

Figure 5 Spatio-organized topographical map of RFs of simple cells arranged by orientation changing.

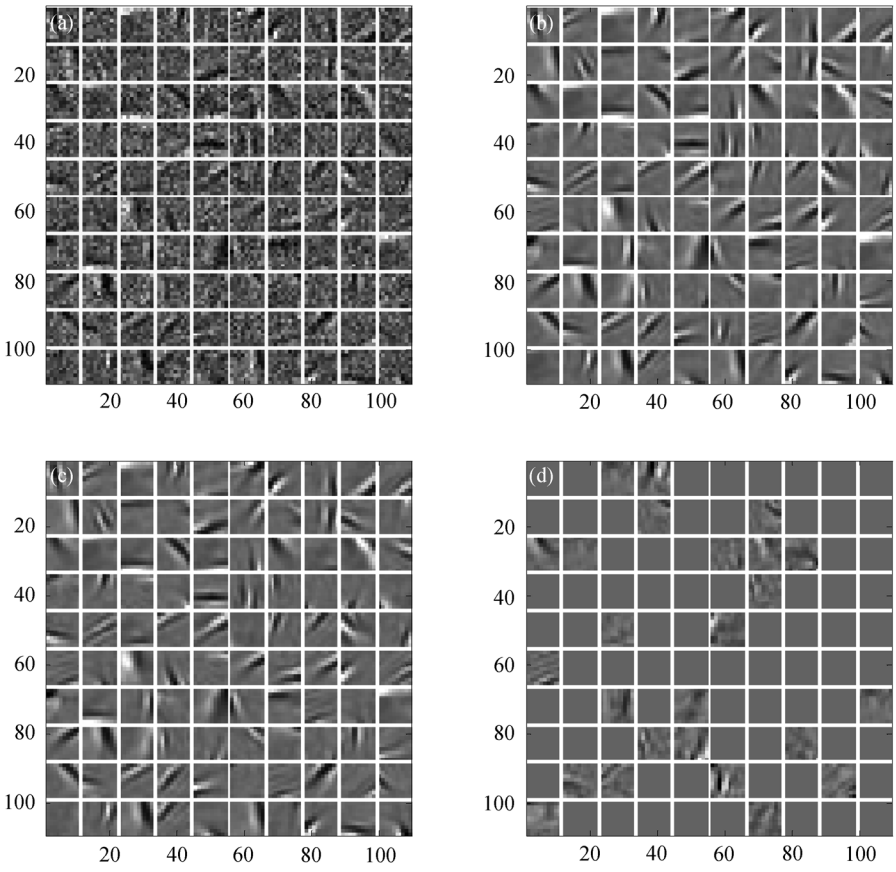

Figure 6 Perception of objects. (a) Stimuli; (b) receptive fields; (c) perception fields; (d) differences.

its RF is the resulting oriented bar-like stimulus. For example, Figure 6 shows the subset of the testing data, RFs of neurons, optimal stimuli, and the corresponding final fitted RFs.

In Figure 6, (a) is the subset of testing data with additive noises to optimal stimuli. The average of SNR (SNR=10log(S^2/N^2), S^2 is the power of signal, and N^2 the power of noise) is –0.17 dB. The testing data is preprocessed as follows. First, we randomly select subsets of 2500 basis functions as stimuli in the testing phase, and add Gaussian noises with 20% peak value of basis functions to form the testing data set. For example, Figure 6 shows 100 samples of stimuli in Figure 6(a). Figure 6(b) represents RFs, which are optimal stimuli to which neurons strongly respond and Figure 6(c) denotes subsets of RFs perceived. Figure 6(d) shows the differ-

ence between perceived results and optimal stimuli. If the difference is not zero, the corresponding object is not exactly perceived with replacement of the neighboring oriented bar. That is because noise shifts the right orientation to the neighboring, as shown in Figure 7. When given a stimulus, responses of all neurons are calculated and the neuron with the maximal activity is selected. Its representation is considered as the resulting perception. From Figure 6, the perceptual model can still achieve high accuracy of object perception in the case of SNR –0.17 dB.

To evaluate robustness of the proposed model against noise, we test the perception performance with different SNRs. The perception results in Figure 7 show that the model is able to perceive contents of stimuli almost perfectly with lower noise. It is clear to see that with the increase of noise, perception of contents and phase becomes difficult. If the model only perceives contents regardless of phase, the result is very promising even in the case of SNR −5 dB. Otherwise, noises cause large offset of phase.

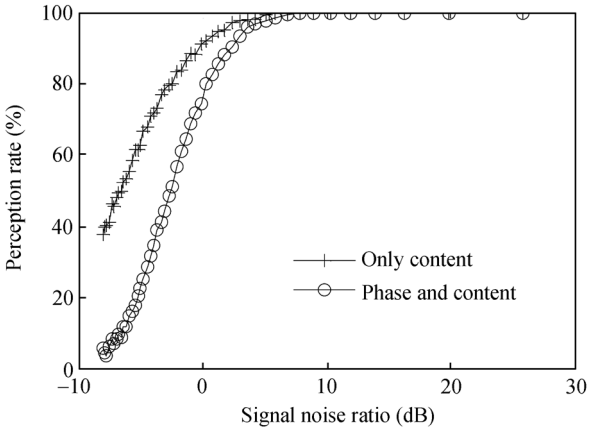

Figure 7 Accuracy of object perception.

3.4 Motion perception

The method of generating testing data is the same as mentioned in subsection 3.3. For motion perception, two stimuli are needed. When the model is stimulated by two consecutive images, two neurons, located at different places and sensitive to the same feature, will be activated in succession. Using these perceived locations, we can estimate the motion direction of objects. Perception results are shown in Figure 8. For example, the location of the first stimulus is (5,1) and the second is (4,3). So, the motion direction is upright-ward, the angle is 26°, and speed is 2.2 pixels per sampling time. The bottom-right in Figure 8 shows the RFs of winner neurons which have the same features as stimuli.

Figure 9 shows the perception accuracy of motion with respect to the change of SNRs. Like perception of contents, the proposed model can perceive the motions almost perfectly in the case of lower noise. However, the performance degenerates when SNR becomes small. From Figure 7, if there are errors of content position perception at the first and second stimulus, then the accuracy of motion perception is based on the first two position perceptions. Therefore, performance becomes worse rapidly. Figure 7 shows the simulation results. The line with circle denotes motion accuracy with perceiving phase and contents, and the line with plus denotes motion accuracy with only contents perception. Thus, we draw a conclusion that the proposed model can achieve better motion perception accuracy if the ignoring phases error. This is because the simple cell is only activated by its matched feature, while activities of the complex cell depend on the neural response of simple cells. Accuracy of motion perception first depends on that of contents perception. Due to noise, phase of neuron may not be estimated correctly. Consequently, the model with phase estimation may degrade performance of motion perception.

4 Discussion and conclusion

Perception of objects and motions in the visual scene is one of the basic problems in the visual system. There exist ‘What’ and ‘Where’ pathways in the superior visual cortex, starting from the simple cells in the primary visual cortex. The former is able to perceive ‘What’: contents of objects such as shapes, color, and texture, and the latter perceives ‘Where’, for example, velocity and direction of spatial movement of objects. Based on the physiological mechanism, this paper has proposed a computational and hierarchical network that models perception functions in superior cortex and simulates two visual pathways. Computer simulations have confirmed three main results: (1) The proposed model provides basis functions with Gabor-like characteristics such as localization, orientation, and bandpass. These features have been found in physiological experiments; (2) Learned topographical feature map is gradually changing orientation and in accord with similar results of physiological experiments; (3) Modeling ‘What’ and ‘Where’ pathways for perception of objects and motions in visual scenes. Computer simulation results show that

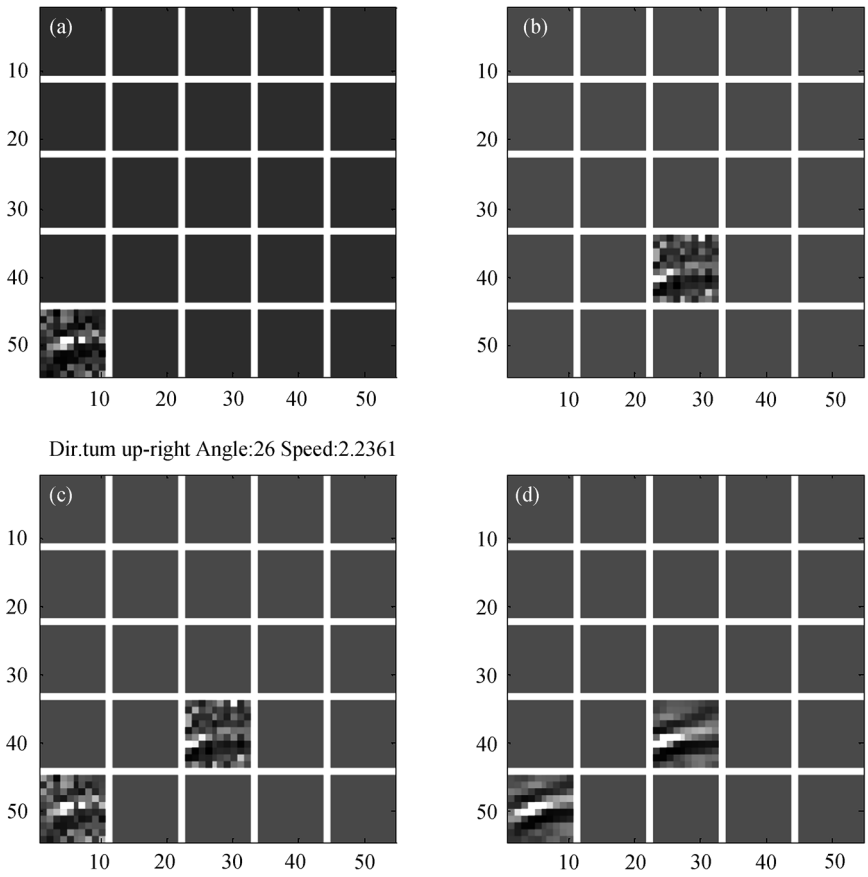

Figure 8 Motion perception. (a) The first stimulus; (b) the second stimulus; (c) both stimuli (direction: turn up-right; angle; 26°; speed: 2.2361 pixels/time); (d) the best RFs fitted.

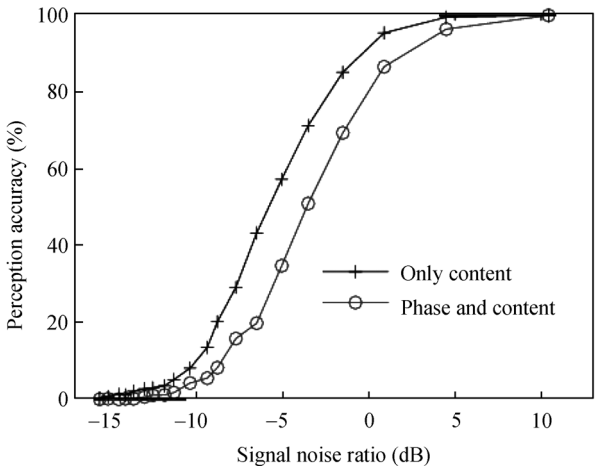

Figure 9 Accuracy of motion perception.

the proposed model successfully simulates the mechanism of two pathways and provides satisfactory results, verifying the efficient performance of the proposed model and algorithms.

Grimes and Rao[12] proposed a bilinear generative model to study the translation invariance by minimizing reconstruction errors. Their model provided horizontal and vertical translation invariance features. However, our model has three different points from theirs: (1) Different learning rule. Our goal is to force responses of neighboring neurons to be sparse and statistically independent. Our learning algorithm is simpler than theirs. (2) Our model provides objects and motion direction in visual scenes. And direction is omnidirectional. (3) Our model provides spatio-topographical maps of RFs of simple cells, whereas theirs cannot.

Hyvarinen et al.[11] considered the second-order correlation of responses of simple cells, but the Topo-ICA model cannot produce overcomplete basis functions because of constrains of orthogonality. However, complex cells in their model had weakly smooth responses to similar orientation because when the size of RFs increases, the translation invariance will disappear in their perception model. Our model provides real invariance of object perception because each complex connected to a

group of simple cells with similar orientation. On the other hand, our learning algorithm is based on Natural Gradient and has higher computing efficiency.

Now, the proposed model is able to perceive objects and motions in visual scenes. But it is limited to simple stimuli and translation. Our future work will focus on extending the model to perceive complex objects and motions such as that in videos and image sequences, to perceive information nonlinearly transformed in scaling and rotating visual scenes.

1 Sabine K, Ungerleider L G. Mechanisms of visual attention in the human cortext. Ann Rev Neurosci, 2000, 23: 315－341
2 Shou T. Brain Mechanism of Visual Information Processing (in Chinese). Shanghai: Shanghai Scientific & Technological Education Publishing House, 1997. 145－149
3 Hubel D H, Wiesel T N. Receptive fields and functional architecture of monkey striate cortex. J Physiol, 1968, 195(1): 215－243
4 DeValois R L, Albrecht D G, Thorell L G. Spatial frequency selectivity of cells in macaque visual cortex. Vision Res, 1982, 22: 545－559
5 Barlow H. Redundancy reduction revisited. Network, 2001, 12(3): 241－253
6 Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 1996, 381: 607－609
7 Olshausen B A, Field D J. Sparse Coding with an overcomplete basis set: A strategy employed by V1? Vision Res, 1997, 37: 331－325
8 Bell A J, Sejnowski T J. The independent ‘components’ of natural scenes are edge filters. Vision Res, 1997, 37: 3327－3338
9 van Hateren J H, van der Schaaf A. Independent component filters of natural images compared with simple cells in primary visual cortex. Proc R Soc Lond B Biol Sci, 1998, 265(1394): 359－366
10 Lewicki M S, Olshausen B A. Probabilistic framework for the adaptation and comparison of image codes. J Opt Soc Am A Opt Image Sci Vis, 1999, 16(7): 1587－1601
11 Hyvarinen A, Hoyer P O. A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. Vision Res, 2001, 41(18): 2413－2423
12 Grimes D B, Rao R P N. Bilinear sparse coding for invariant vision. Neural Comput, 2005, 17(11): 47－73
13 Bednar J A, Miikkulainen R. Learning innate face preferences. Neural Comput, 2003, 15(7): 1525－1557
14 Bednar J A, Miikkulainen R. Joint maps for orientation, eye, and direction preference in a self-organizing model of V1. Neurocomputing, 2006, 69(10-12): 1272－1276
15 Serre T, Wolf L, Bileschi S, et al. Robust object recognition with cortex-like mechanisms. IEEE Trans Pattern Anal Mach Intell, 2007, 29(3): 411－426
16 Rolls E T, Stringer S M. Invariant global motion recognition in the dorsal visual system: A unifying theory. Neural Comput, 2007, 19(1): 139－169
17 Grossberg S, Hong S. A neural model of surface perception: Lightness, anchoring, and filling-in. Spat Vis, 2006, 19(2-4): 263－321
18 Grossberg S, Kuhlmann L, Mingolla E. A neural model of 3D shape-from-texture: Multiple-scale filtering, boundary grouping, and surface filling-in. Vision Res, 2007, 47(5): 634－672
19 Bhatt R, Carpenter G, Grossberg S. Texture segregation by visual cortex: Perceptual grouping, attention, and learning. Boston University: Technical Report CAS/CNS-TR-2006-007, 2007
20 Yang Q, Qi X, Wang Y. The Spatiotemporal coding properties of complex cell in the visual cortex. Acta Biophys Sin (in Chinese), 2000, 16(2): 280－287
21 Yang Q, Qi X, Wang Y. The sparse coding strategy employed by simple cells in visual cortex V1. Chin J Comput Phys (in Chinese), 2001, 18(2): 143－146
22 Tian J, Lu H M. Self-organization model on receptive field of neuron with asymmetric time window of synaptic modification. Chin Sci Bull, 2001, 46(12): 1033－1036
23 Wei H. A som network model for feature extraction by hyper columns architecture of primary visual cortex. J Zhejiang Univ (Engineer Sci) (in Chinese), 2001, 35(3): 258－263
24 Yang X. Neural mechanism of vision. Shanghai: Shanghai Scientific & Technological Education Publishing House, 1996. 215－267
25 Mei J, Zhang L. A new model of simple cell in the visual cortex. Acta Biophys Sin (in Chinese), 2003, 19(1): 58－62
26 Chen G. Influence of stimulation in integration field on orientation selectivity of neurons in primary visual cortex (in Chinese). Dissertation for the Doctoral Degree. Shanghai Institutes for Biological Sciences, 2005. 48－49
27 Shi Z W, Shi Z Z. Computational model of time coding. Acta Biophys Sin (in Chinese), 2006, S(22): 110
28 Zhang L, Cichocki A, Amari S. Natural gradient algorithm to blind separation of over-determined mixture with additive noises. IEEE Signal Proc Lett, 1999, 6(11): 293－295
29 Zhang L, Cichocki A, Amari S. Self-adaptive blind source separation based on activation function adaptation. IEEE Trans Neural Netw, 2004, 15(2): 233－244
30 Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nat Neurosci, 1999, 2(11): 1019－1025
31 Poggio T, Bizzi E. Generalization in vision and motor control. Nature, 2004, 431(7010): 768－774
32 Amari S, Cichocki A, Yang H H. A new learning algorithm for blind signal separation. Adv Neural Inf Process Syst, 1996, 8: 757－763
33 van Hateren J H, Ruderman D L. Independent component analysis Of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. Proc R Soc Lond B Biol Sci, 1998, 265(1412): 2315－2320
34 Szatmary B. Independent component analysis of temporal sequences subject to constraints by LGN inputs yields all the three major cell types of the primary vi sual cortex. J Comput Neurosci, 2001, 11: 241－248