PERGAMON

Neural Networks 15 (2002) 337–347

Neural Networks

www.elsevier.com/locate/neunet

Contributed article

Learning generative models of natural images

Jiann-Ming Wu*, Zheng-Han Lin

Department of Applied Mathematics, National Donghwa University, Hualien, Taiwan, ROC

Received 7 May 2001; revised 17 December 2001; accepted 17 December 2001

Abstract

This work proposes an unsupervised learning process for analysis of natural images. The derivation is based on a generative model, a stochastic coin-flip process directly operating on many disjoint multivariate Gaussian distributions. Following the maximal likelihood principle and using the Potts encoding, the goodness-of-fit of the generative model to tremendous patches randomly sampled from natural images is quantitatively expressed by an objective function subject to a set of constraints. By further combination of the objective function and the minimal wiring criterion, we achieve a mixed integer and linear programming. A hybrid of the mean field annealing and the gradient descent method is applied to the mathematical framework and produces three sets of interactive dynamics for the learning process. Numerical simulations show that the learning process is effective for extraction of orientation, localization and bandpass features and the generative model can make an ensemble of a sparse code for natural images.

Keywords: Neural networks; Cortical maps; Elastic net; Potts model; Self-organization; Unsupervised learning; Natural images

1. Introduction

Analysis of natural images has been recently considered as an essential task toward exploring the formation of receptive fields in visual cortex. The task in previous works (Olshausen & Field, 1996; Hateren van & Ruderman, 1998; Hateren van & van der Schaaf, 1998; Hyvarinen & Hoyer, 2000) is realized by employing tremendous patches sampled from natural images for emulating a stimulating environment to an artificial vision system, simulating the underlying mechanism of the formation of receptive fields by an unsupervised learning process, and then explores the biological plausibility of the learning process based on comparisons between the properties of real and artificial receptive fields.

Recently, typical approaches (Hateren van & Ruderman, 1998; Hateren van & van der Schaaf, 1998; Hyvarinen & Hoyer, 2000) to analysis of natural images have been developed based on the unsupervised learning process of independent component analysis(Lin, Cowan, & Grier, 1997a,b; Hyvärinen & Oja, 1997; Hyvarinen, 1999). The target is to seek a set of independent filters or a basis for internal representations of natural images under the conjecture of existing statistical dependency among components of patches of natural images. Once projected on independent filters, image patches are expected to form independent components whose joint distribution tends to be identical to the product of all corresponding marginal distributions. It has been stated in previous works (Olshausen & Field, 1996; Hateren van & Ruderman, 1998; Hateren van & van der Schaaf, 1998; Hyvarinen & Hoyer, 2000) that the obtained independent filters possess similar features of localization, orientation and bandpass to those of receptive fields in visual cortex.

The other possible approaches to analysis of natural images are based on self-organizing algorithms (Durbin & Willshaw, 1987; Durbin & Mitchison, 1990; Kohonen, 1982; Liou & Wu, 1996), which aim to seek a set of radial basis receptive fields as well as their ordering on a cortex-like lattice. The issue addresses on the coherent mapping via a dimensional reduction framework and topology preservation. Typical self-organizing algorithms, including the Kohonen self-organizing algorithm (Kohonen, 1982) and the elastic net algorithm (Durbin & Willshaw, 1987; Durbin & Mitchison, 1990), are effective for exploration of the formation of ocular dominance stripes and orientation modules in visual cortex (Durbin & Mitchison, 1990; Piepenbrock, Ritter, & Obermayer, 1997, 1998; Yulle, Kolodny, & Lee, 1996). Unfortunately, these applications deal with artificial stimulus instead of real stimulus like natural images, where each artificial stimuli is numerically encoded with a distinct line feature, probably including

* Corresponding author. Tel.: +886-3-866-2500x21126; fax: +886-3-866-2532.

E-mail address: jmwu@server.am.ndhu.edu.tw (J.-M. Wu).

0893-6080/02/$ - see front matter
PII: S0893-6080(02)00018-7

the orientation, location coordinates, and the stereo information. Given such a set of artificial stimulus, the self-organizing algorithm is employed to seek a set of cortical points as well as their ordering on a cortex-like map, and then the distribution of the obtained cortical points on the cortex-like map is compared with the topology of ocular dominant stripes and orientation modules discovered in visual cortex for biological comprehension. In comparison with real stimulus like natural images, components of the artificial stimulus are statistically independent and in smaller size. In the self-organizing algorithm, it is well known that the similarity measure for the adaptation of cortical points is extensively based on the Euclidean distance. Extending the self-organizing algorithms to real stimulus like natural images faces on the suitability of the Euclidean measure under the assumption of statistical dependency among a large size of components. It has been pointed out that the self-organizing algorithm results in non-faithful representations (Lin et al., 1997a,b) when applied to real stimulus. To overcome this difficulty and achieve faithful representations for natural images, this work employs the Mahalanobis distance for similarity measure between two real stimuli. Following the motivation, we developed a novel unsupervised learning process, called natural elastic net, for analysis of natural images. The natural elastic net is well equipped with a compensating process for statistical dependency among components of real stimulus, and can be considered as a combinatorial framework of the self-organizing algorithm and independent component analysis for a sparse code of real stimulus.

The derivation of the natural elastic net starts with a generative model, which is used to characterize the distribution of natural images. The generative model is a flip-coin stochastic process containing M^2 multivariate disjoint Gaussian distributions. The number of disjoint distributions is designed to coincide with neural nodes on the cortex-like map. According to a set of prior probabilities, each time the stochastic process randomly selects one of M^2 disjoint distributions and uses it to generate a stimuli. The parameter of the generative model contains M^2 local means or cortical points and a common covariance matrix. Ideally, these cortical points can be sparsely distributed among the space of real stimulus such that each cortical point is located at the center of a natural cluster of stimulus, and the structure of statistical dependency among components of real stimulus can be effectively captured by the covariance matrix. The distribution of images produced by the generative model is general enough to characterize real stimulus, and is indeed beyond a single multivariate Gaussian distribution.

Following the maximal likelihood principle, the fitness of the generative model to real stimulus can be further expressed as a quantitative objective via the formulation of Potts encoding (Peterson & Söderberg, 1989; Wu & Lin, 2000; Wu & Chiu, 2001). The estimation of parameters in the generative model thus turns to be a constrained optimization. Under the definition of the generative model, a real stimuli is assumed to be a result of one and only one disjoint distribution, so its unique unknown source can be expressed as a Potts neural variable, which is a standard unitary vector with M^2 binary elements. With the Potts encoding, following the maximal likelihood principle, the goodness-of-fit of the generative model to real stimulus is formulated to be proportional to a log likelihood function or an objective function subject to a set of constraints. To achieve a coherent mapping, the objective function is further combined with the minimal wiring criterion proposed by Durbin and Willshaw (1987) to form a mathematical framework for the natural elastic net. The mathematical framework is composed of two sets of continuous variables and one set of discrete variables respectively encoding cortical points, a covariance matrix and membership vectors. It is exactly a mixed integer and linear programming (MILP). A hybrid of the mean field annealing and the gradient descent method is applied to the optimization of the mathematical framework to overcome computational difficulty in solving a MILP. As a result, three sets of interactive dynamics are derived for the learning process, and the evolution of the interactive dynamics is well modulated by an annealing process in analogy with the physical annealing to achieve reliable self-organization. At each temperature, the learning process employs a set of mean field equations to trace the mean configuration of discrete variables under thermal equilibrium subject to maximal entropy and minimal mean energy; then based on the mean configuration, the other two sets of dynamics update the instance of continuous variables toward the minimum of the mean energy. The elastic net algorithm of Durbin and Willshaw can be proved as a special case of the learning process developed in this work.

When applying the new learning process to tremendous patches sampled from natural images, we have a set of sparsely distributed cortical points as internal representations, a covariance matrix for similarity measure, and an ordering structure of cortical points on a cortex-like lattice. By numerical simulations, it is shown that the cortical points or local means obtained by the new learning process make an ensemble of features of localization, orientation and bandpass as receptive fields discovered in visual cortex.

This article is organized as follows. The generative model and the mathematical framework for the natural elastic net are developed in Section 2, the three sets of interactive dynamics are derived in Section 3. Numerical simulations for learning the generative model of natural images and our conclusions are presented in Sections 4 and 5 respectively.

2. The natural elastic net

2.1. A generative model for natural images

The typical self-organizing algorithm, such as the Kohonen

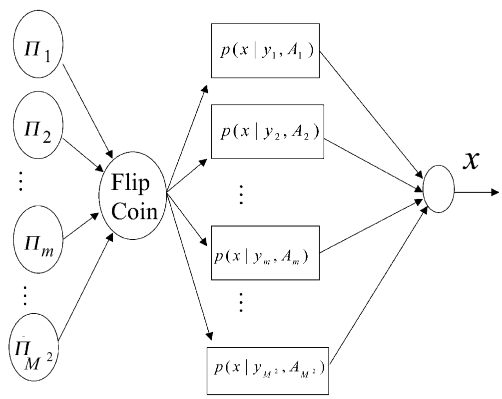

Fig. 1. The generative model.

algorithm (Kohonen, 1982) and the elastic net algorithm (Durbin & Willshaw, 1987; Durbin & Mitchison, 1990), is used to construct a coherent mapping from a parameter space R^d to a two dimensional $M \times M$ lattice based on a set of training stimulus $\{x_i \in R^d, 1 \leqq i \leqq N\}$, where d, M and N are positive integers with $N \gg M^2$. On the lattice, each node is attached with a cortical point belonging to R^d and these cortical points $\{y_k \in R^d, 1 \leqq k \leqq M^2\}$, are adapted to form effective internal representations for training stimulus. The mapping is defined by a non-overlapping Voronoi partition $\{\Omega_j\}$ into the parameter space, where $\Omega_j = \{x | \arg\ \min_k \|x - y_k\| = j, x \in R^d\}$ and $\|x\|$ denotes the Euclidean length of a vector x. It follows that the union of all regions is the space R^d; the intersection of any two distinct regions is an empty set; each stimuli is mapped to one and only one internal region. Such a mapping is coherent if nearby stimulus in the parameter space are mapped to nodes as closely as possible on the lattice and the utilization of nodes is extensively maximized. By a coherent mapping, the topology relation within training stimulus is expected to be reserved and turn visible on the lattice. The self-organizing algorithm has been applied to the exploration of ocular dominance bands and orientation modules in visual cortex (Durbin & Mitchison, 1990; Piepenbrock et al., 1997; Yulle et al., 1996).

Since the use of the Euclidean distance may not be suitable for the case of real stimulus, which are likely to have statistically dependent components, we propose the following generative model to derive the natural elastic net for analysis of real stimulus.

The generative model is a flip-coin process directly operating on M^2 disjoint multivariate Gaussian distributions as shown in Fig. 1. Each time the process randomly selects one distribution according to the prior probabilities $\{\pi_k, 1 \leqq k \leqq M^2\}$, and then triggers the selected distribution to generate a training stimuli. Let π_k be $1/K$ for all k and $K = M^2$, indicating an equal selection among disjoint distributions. Assume that all of training patches sampled from natural images are results produced by the generative model and each individual distribution is as the following multivariate Gaussian distribution

$$P_k(x) = P(x|y_k, A_k) = \frac{1}{(2\pi)^{d/2}\sqrt{|A_k^{-1}|}} \exp\left(\frac{(x-y_k)^t A_k (x-y_k)}{2}\right) \quad (1)$$

where y_k and A_k^{-1} are the local mean and the covariance matrix respectively, $|A^{-1}|$ denotes the determinant of the inverse of the matrix A, and x^t denotes the transpose of vector x. To facilitate the forthcoming computation, the covariance matrix A_k^{-1} of each individual distribution is set to be a common covariance matrix A^{-1}.

2.2. The mathematical framework

The Potts encoding is employed to facilitate the optimization of the generative model. According to the flip-coin process, each stimuli x_i is generated by one and only one individual distribution. Let the Potts neural variable $\delta_i = [\delta_{i1}, \ldots, \delta_{iK}]^t$ denote the membership of x_i with $\delta_{ik} \in \{0,1\}$ for all k and $\sum_k \delta_{ik} = 1$. Therefore each δ_i belongs the set of $\{e^k, 1 \leqq k \leqq K\}$, where e^k is a standard unitary vector with the kth element one and the others zero. If δ_i is e^k, it is said that the ith training stimuli is generated by the kth individual distribution. The following local log likelihood function can then quantitatively measure the fitness of the individual distribution P_k to all of training stimulus whose membership vectors are e^k.

$$l_k = \log \prod_{\{i:\delta_i = e^k\}} P_k(x_i) = \sum_{\{i:\delta_i = e^k\}} \log P_k(x_i) \quad (2)$$

By summing up all l_k, we have the following log likelihood function

$$\begin{aligned} l &= \sum_k l_k = \sum_k \sum_{\{i:\delta_i = e^k\}} \log P_k(x_i) = \sum_i \sum_k \delta_{ik} \log P_k(x_i) \\ &= \sum_i \sum_k \delta_{ik} \left(-\frac{1}{2}(x_i - y_k)^t A (x_i - y_k) - \frac{1}{2} \log |A^{-1}| - \frac{d}{2}\log(2\pi) \right) \\ &= -\frac{1}{2} \sum_i \sum_k \delta_{ik} (x_i - y_k)^t A (x_i - y_k) - \frac{N}{2} \log |A^{-1}| - \frac{Nd}{2}\log(2\pi) \end{aligned} \quad (3)$$

By neglecting the last constant term, reversing the sign and using the fact $\log |A^{-1}| = -\log |A|$,we obtain the first objective for the natural elastic net as follows

$$E_1 = \frac{1}{2} \sum_i \sum_k \delta_{ik} (x_i - y_k)^t A (x_i - y_k) - \frac{N}{2} \log |A| \quad (4)$$

The maximization of l is equivalent to the minimization of E_1 subject to the constraint of $\delta_i \in \{e^k, 1 \leqq k \leqq K\}$ for all i.

Further consider the optimal order of cortical points on the lattice. When receiving a distinct stimulus, a coherent mapping derived by self-organization is expected to respond in locality. It follows that the winner has the most intensive response and those surrounding the winner respond with decreasing intensity along the distance from its own location to that of the winner on the map, where a node with a cortical point closest to the stimuli among all cortical points is the winner. Following the minimal wiring principle, two nearby cortical points on the map should be as close as possible, which leads to the following minimal wiring criterion (Durbin & Willshaw, 1987; Durbin & Mitchison, 1990) for an effective dimensional-reduction mapping,

$$E_2 = \frac{1}{2} \sum_k \sum_{j \in NB(k)} \|y_k - y_j\|_A^2 \tag{5}$$

where $NB(k)$ is a set of all nodes connecting to the kth node on the lattice and $\|y\|_A^2$ denotes the Mahalanobis square length of a vector y, which is defined by $\|y\|_A^2 = y^t A y$.

Minimizing a weighted combination of E_1 and E_2 leads to the following mathematical framework for the natural elastic net.

Minimize

$$E = E_1 + CE_2 = \frac{1}{2} \sum_i \sum_k \delta_{ik} \|x_i - y_k\|_A^2 - \frac{N}{2} \log |A| + \frac{C}{2} \sum_k \sum_{j \in NB(k)} \|y_k - y_j\|_A^2 \tag{6}$$

subject to

$$\sum_k \delta_{ik} = 1, \qquad \forall i \tag{7}$$

where C is a weighting constant.

The above mathematical framework is derived via devising a generative model for characterizing the distribution of real stimulus, using the Potts encoding to realize the log likelihood function, and combining the minimal wiring criterion to the objective function E_1. In the objective function E, the first term costs the distance between each stimuli and its representative as a quantitative measure for effective clustering; the second term maximizes the determinant of the matrix A, also the volume of a corresponding d dimensional parallel polyhedron; the last term costs the distance between any two neighboring cortical points on the lattice as a minimal wiring criterion. The distance is the Mahalanobis distance associated with the matrix A. The constraint is the unitary condition of Potts neural variables. The learning process of the natural elastic net is exactly an optimization process for the mathematical framework, which searches for parameters of the generative model suitable for training stimulus and simultaneously organizes cortical points on the cortex-like lattice following the minimal wiring principle. If the matrix A is an identity matrix, the mathematical framework is reduced to the original elastic net (Durbin & Willshaw, 1987; Liou & Wu, 1996), with a similarity measure based on the Euclidean distance, which is valid under the assumption of statistical independency among the components of training stimulus.

3. Dynamics for the natural elastic net

The above mathematical framework is a mixed integer and linear programming. The optimization task involves with discrete combinatorial variables $\{\delta_i\}$ and continuous geometrical variables $\{y_k\}$, and the matrix A. Since the energy function E is not differentiable with respect to discrete variables, the gradient descent method cannot be directly applied to the mathematical framework. By relating each membership vector δ_i to a Potts neural variable, the optimization task can be treated by a hybrid of the mean field annealing and the gradient descent, which has been successfully applied to the derivation of independent component analysis using Potts models (Wu & Chiu, 2001).

The mean field annealing is employed to search for optimal Potts neural variables near or at the global minimum of the energy function E. When given particular A and $\{y_k\}$, at each temperature, the annealing process seeks the mean configuration $\langle\delta\rangle$ of all discrete variables $\{\delta_i\}$ satisfying the condition of thermal equilibrium, which says that the probability of the system configuration is proportional to the Boltzmann distribution, such as

$$\Pr(\delta) \propto \exp(-\beta E(\delta|A, Y)), \tag{8}$$

where $E(\delta|A, Y)$ denotes the conditional energy function given A and Y, the parameter β denotes the inverse of an artificial temperature, and Y denotes a collection of all y_k. At a sufficiently low β value, probabilities of feasible configurations are likely identical, and the mean configuration is at a trivial solution with each element $\langle\delta_{ik}\rangle$ equal to $1/K$. In contrast, at a sufficiently large β value, the Boltzmann distribution is dominated by the optimal configuration, such as

$$\lim_{\beta \to \infty} \Pr(\delta) = \begin{cases} 1, & \text{if } \delta = \delta^* \\ 0, & \text{otherwise} \end{cases}$$

where

$$E(\delta^*|A, Y) = \min_\delta E(\delta|A, Y)$$

To approximate the optimal configuration, a set of mean field equations are used to track the mean configuration along the annealing process, where the β value is gradually increased from a sufficiently low value to a large one. At each β value, the mean field equations iteratively execute to reach a fixed point as an approximation to the mean configuration. The mean configuration obtained at one β value is considered as an initial instance for the relaxation at a subsequent β value. The mean field equation can be derived from the following free energy function similar to the one proposed by Peterson and Söderberg (1989)

$$\psi(A, Y, \langle\delta\rangle, u) = E(A, Y, \langle\delta\rangle) + \sum_i \sum_m \langle\delta_{im}\rangle u_{im} - \frac{1}{\beta}\sum_i \ln\left(\sum_m \exp(\beta u_{im})\right) \quad (9)$$

where u denote the set $\{u_i\}$ and each u_i is an auxiliary vector.

By setting

$$\frac{\partial\psi}{\partial\langle\delta_{im}\rangle} = 0 \text{ for all } i, m \quad (10)$$

$$\frac{\partial\psi}{\partial u_{\rm im}} = 0 \text{ for all } i, m \quad (11)$$

we have the following mean field equation for evaluating mean activations of discrete neural variables

$$u_{im} = -\frac{\partial E}{\partial\langle\delta_{im}\rangle} = -\frac{1}{2}(x_i - y_m)^t A(x_i - y_m) \quad (12)$$

$$\langle\delta_{im}\rangle = \frac{\exp(\beta u_{im})}{\sum_k \exp(\beta u_{ik})} \quad (13)$$

The mean configuration satisfying the mean field Eqs. (12) and (13) is a saddle point of the free energy (9) corresponding to a particular β value. Based on the mean configuration, we can apply the gradient descent method to derive the following updating rule for each y_m.

$$\Delta y_m \propto -\frac{\partial E}{\partial y_m} \quad (14)$$

$$\Delta y_m = \frac{1}{2}\sum_i \langle\delta_{im}\rangle(A + A^t)(x_i - y_m) + \frac{C}{2}\sum_{n\in NB(m)} (A + A^t)(y_n - y_m)$$

To zero gradient, such as $\Delta y_m = 0$, we have the following linear system.

$$\left(CN_m + \sum_i \langle\delta_{im}\rangle\right)y_{ma} - C\sum_{n\in NB(m)} y_{na} = \sum_i \langle\delta_{im}\rangle x_{ia},\ 1 \leqq m \leqq K, 1 \leqq a \leqq d$$

where N_m denotes the number of nodes in the set $NB(m)$. By solving the linear system, we have

$$[y_1, \ldots, y_K]^t = (C(H - G) + Z)^{-1} R \quad (15)$$

where both H and Z are $K \times K$ diagonal matrices with diagonal entries $H_{mm} = N_m$ and $Z_{mm} = \sum_i \langle\delta_{im}\rangle$, $1 \leqq m \leqq K$, respectively, G is a $K \times K$ adjacent matrix corresponding to the lattice with entries

$$G_{mn} = \begin{cases} 1, & \text{if the } m\text{th the node and the } n\text{th node are connected,} \\ 0 & \text{otherwise,} \end{cases}$$

and the matrix R has entries

$$R_{ma} = \sum_i \langle\delta_{im}\rangle x_{ia}, \qquad 1 \leqq m \leqq K,\ 1 \leqq a \leqq d$$

The updating rule for each element A_{ab} in the matrix A can be derived as follows

$$\Delta A_{ab} \propto -\frac{\partial E}{\partial A_{ab}} \quad (16)$$

$$\Delta A_{ab} = -\frac{1}{2}\sum_i \sum_m \langle\delta_{im}\rangle(x_{ia} - y_{ma})(x_{ia} - y_{mb}) - C\sum_m \sum_{j\in NB(m)} (y_{ma} - y_{ja})(y_{mb} - y_{jb}) + \frac{N}{2}[(A^t)^{-1}]_{ab} \quad (17)$$

Again, when $\Delta A_{ab} = 0$, we have

$$A = (W^{-1})^t \quad (18)$$

where

$$W_{ab} = \frac{1}{N}\sum_i \sum_m \langle\delta_{im}\rangle(x_{ia} - y_{ma})(x_{ia} - y_{mb}) + \frac{2C}{N}\sum_m \sum_{j\in NB(m)} (y_{ma} - y_{ja})(y_{mb} - y_{jb}) \quad (19)$$

The following step-by-step statement is the learning process for the natural elastic net toward the minimum of the objective function (6).

1. Initialize β as a sufficiently low value, $A = 0.01 \times I$ (identity matrix),
   $$y_k \approx \frac{1}{N}\sum_i x_i,\ \langle\delta_{ik}\rangle \approx \frac{1}{K}$$
2. Update $\{\langle\delta_{im}\rangle\}$ by Eqs. (12) and (13).
3. Update $\{y_m\}$by Eq. (15).
4. Update A by Eqs. (18) and (19).
5. If $\sum_i \sum_m \langle\delta_{im}\rangle^2 > \theta$ then halt, else $\beta \leftarrow \beta * (1/0.98)$, and *goto step* 2, where θ is a threshold, such as $\theta = 0.98^* N$.

4. Numerical simulations

Return to analysis of natural images using the natural elastic net. The task aims to search for a set of local means and a matrix A by the proposed learning process and explore their visual properties. Before dealing with natural images, the natural elastic net is first applied to an artificial example of 800 parameters generated by linear mixtures to verify the capability of the natural elastic net

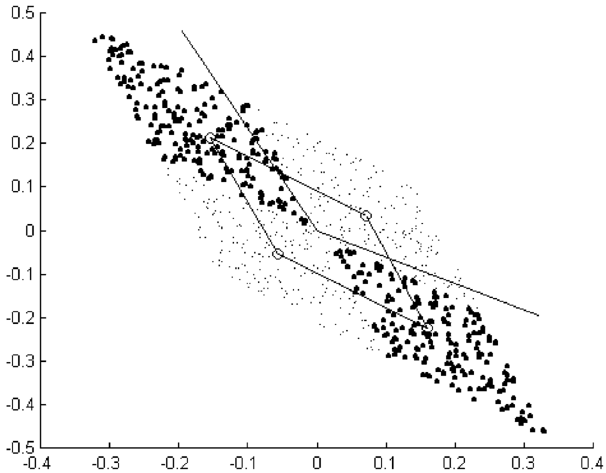

Fig. 2. 800 Training parameters, the resulting four cortical points on the lattice and the two columns of the inverse of the demixing matrix B for the first example using a 2×2 natural elastic net.

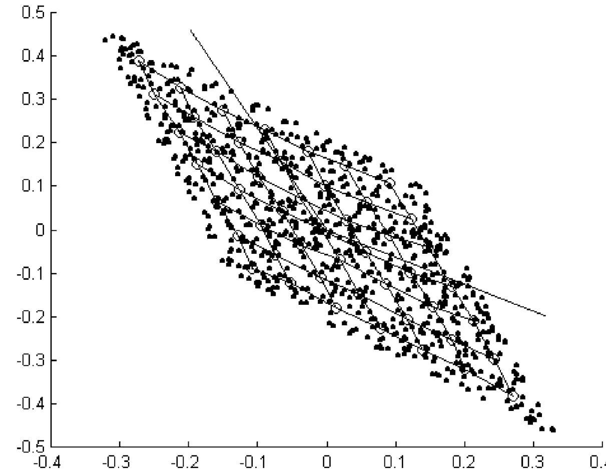

Fig. 4. The result of a 7×7 natural elastic net with 49 nodes for the first example.

in recovering independent components for faithful representations. It is expected that the natural elastic net is able to properly locate the sparsely distributed local means and generate a matrix A essentially capturing the statistical dependency within the training parameters. On the basis, the natural elastic net is applied to analysis of natural images to clarify the role of local means and the matrix A on the formation of receptive fields in visual cortex.

The learning process is implemented in MATLAB. In the following experiments, the weighting constant C in Eq. (6) is 1.8 for the first example and 7.8 for natural images, the initial β value is 1/1.5, and the scheduling factor is 0.98.

4.1. An artificial problem

As shown in Fig. 2, the 800 training parameters used in this example are generated by a linear mixture of two uniform distributions within the interval $[-0.5\ 0.5]$ via a randomly generated mixing matrix. Each training parameter contains a pair of real values and is plotted as a point in a two dimensional plane as shown from Figs. 2–4. The parameters form a parallelogram in shape whose edges are the two column vectors of the mixing matrix representing the underlying statistical dependency of two components of training parameters. A 2×2 natural elastic net is used for this example, and finally results in four cortical points and a matrix A. The demixing matrix B can be obtained by solving the relation $B^t B = A$. The four cortical points and the two columns of the inverse of the demixing matrix B are also shown in Fig. 2, where the lines connecting four cortical points denote the neighboring structure on the 2×2 lattice. Based on the similarity measure A, the four cortical points partition the parameter space into four faithful or equal regions as shown by large and small dots in Fig. 2, the matrix A well encodes the structure of statistical dependency embedded within training parameters and the cortical points are sparsely distributed over the space. The cortical points accompanied with the matrix A successfully constitute faithful representations for the training parameters.

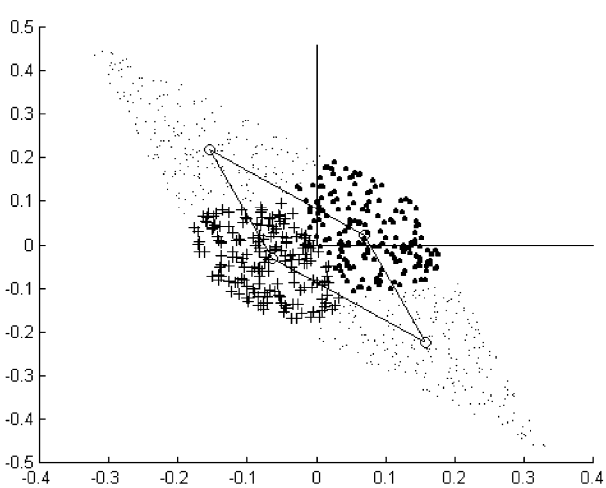

Fig. 3. 800 Training parameters, the resulting four cortical points on the lattice and the two columns of the inverse of the demixing matrix B for the first example using the simplified version of a 2×2 original elastic net.

The simplified natural elastic net is applied to the same training parameters, where the matrix A is fixed as an identity matrix. The resulting parameters are shown in Fig. 3. Although the simplified natural elastic net employs the same relaxation procedure, due to the simplified generative model, the obtained partition into the parameter space is a Voronoi partition, which captures no features about the mixing structure embedded within training parameters, and is considered as non-faithful representations. To demonstrate the reliability of the natural elastic net for the case with a larger size of cortical points, a 7×7 natural elastic net with a dynamical matrix A is applied to this example. The resulting parameters are shown in Fig. 4.

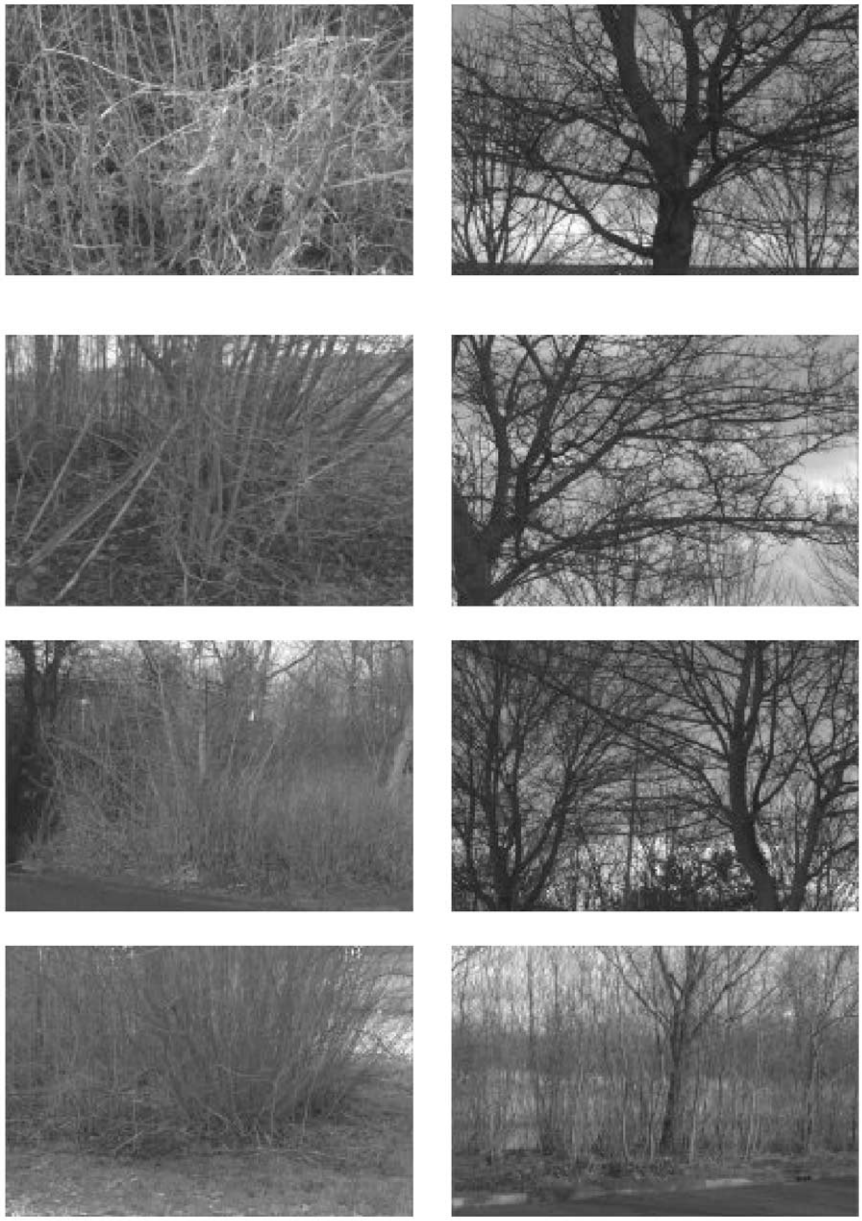

Fig. 5. Some used natural images.

4.2. Analysis of natural images

4.2.1. Natural elastic net

For analysis of natural images, a 16×16 natural elastic net is applied to a set of 20,000 patches, each containing 16×16 pixels as a random sample from natural images. Some of natural images downloaded from the homepage provided by the author of the works (Hateren van & Ruderman, 1998; Hateren van & van der Schaaf, 1998) are shown in Fig. 5. It takes about 48 h to train the natural elastic net for 20,000 patches on a personal computer. As a result, the natural elastic net has 256 cortical points ordered on a 16×16 lattice and a matrix A as parameters of the generative model for 20,000 patches of natural images.

The obtained cortical points essentially extract line features from natural images. Those cortical points with different orientations are respectively shown from Figs. 6–9 with their amplitude spectra, where darker grey values code smaller amplitudes and zero spatial frequency is at the center of each patch. In the figures, each cortical point y_i is shown by a normalized patch using the formula $y_i - \bar{y}_i/\max(y_i - \bar{y}_i)$. The orientation of the peak in the amplitude spectrum of a cortical point is orthogonal to the orientation of the corresponding line feature and the peak is distributed

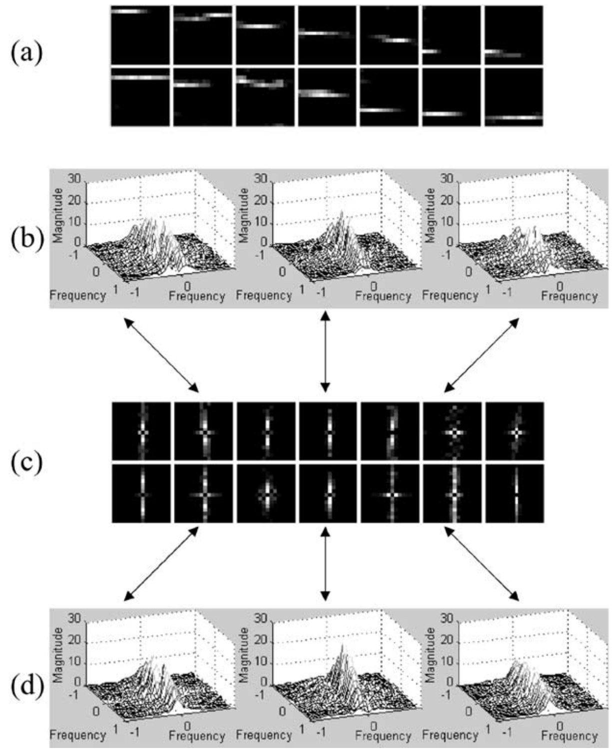

Fig. 6. The first row (a) show cortical points with horizontal orientation and the third row displays the corresponding amplitude spectra. Some 3D amplitude spectra are shown in the second (b) and fourth (d) rows.

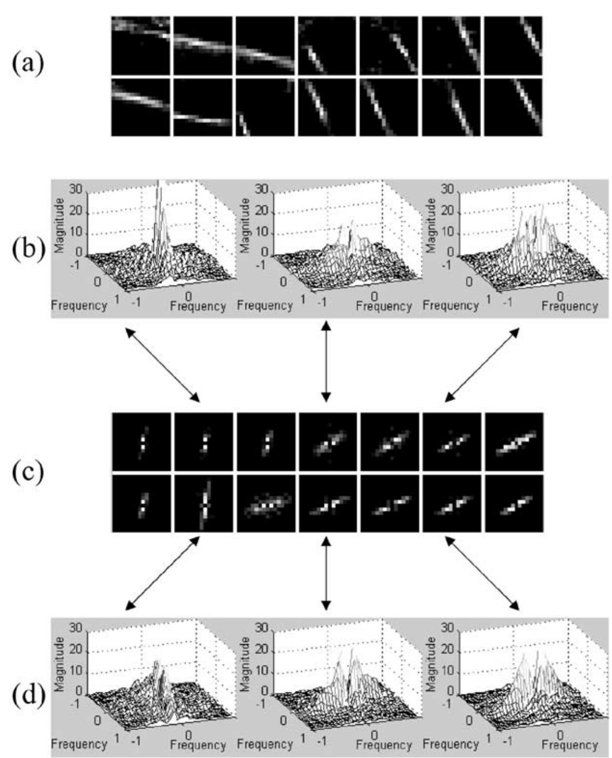

Fig. 7. The first row (a) show cortical points with diagonal orientation and the third row displays the corresponding amplitude spectra. Some 3D amplitude spectra are shown in the second (b) and fourth (d) rows.

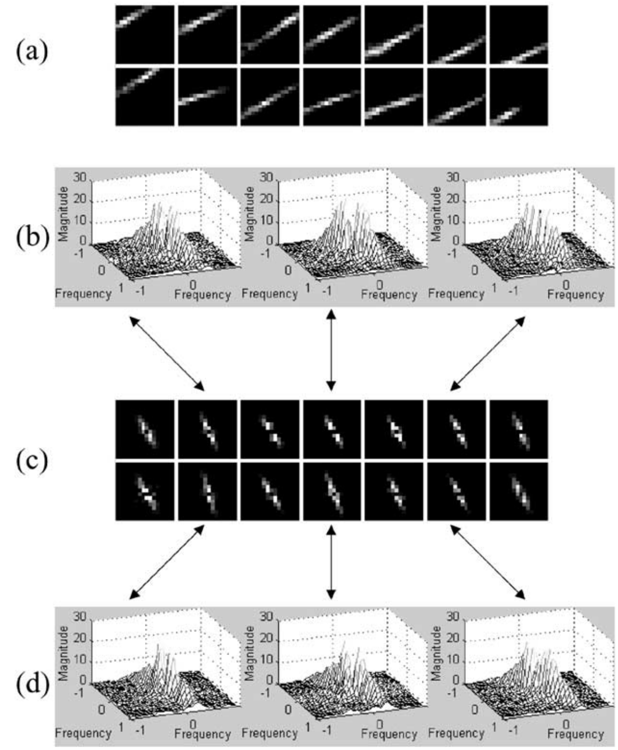

Fig. 8. The first row (a) show cortical points with right diagonal orientation and the third row displays the corresponding amplitude spectra. Some 3D amplitude spectra are shown in the second (b) and fourth (d) rows.

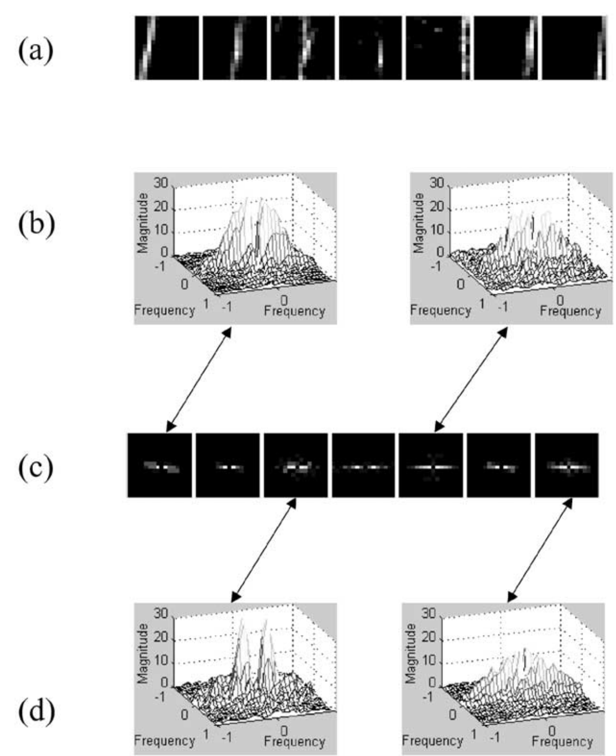

Fig. 9. The first row (a) show cortical points with vertical orientation and the third row displays the corresponding amplitude spectra. Some 3D amplitude spectra are shown in the second (b) and fourth (d) rows.

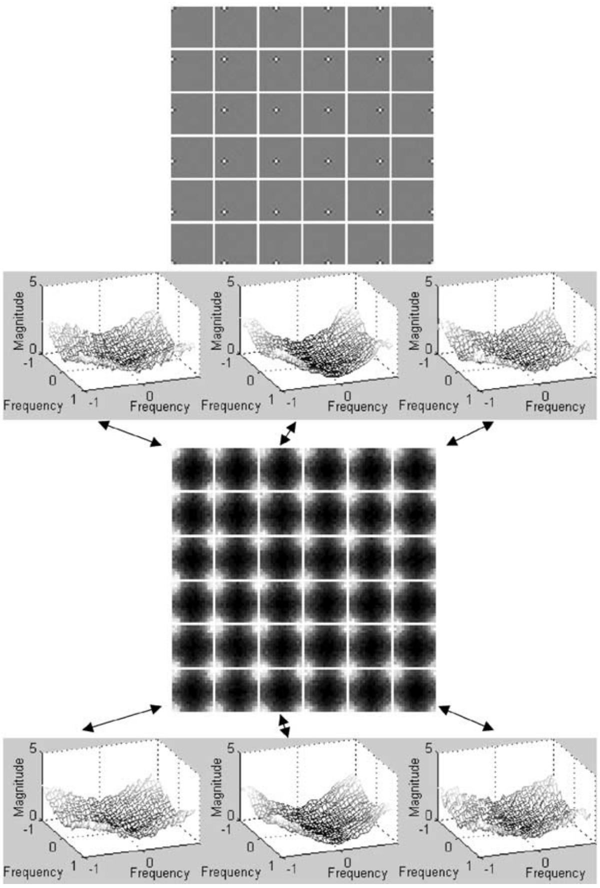

Fig. 10. Some consequent rows of the covariance matrix obtained by the learning process.

within the scope from low frequency to high frequency. These line features are similar to those of the bandpass filters obtained in previous works (Hateren van & Ruderman, 1998; Hateren van & van der Schaaf, 1998; Hyvarinen & Hoyer, 2000). However the line features in previous works are associated with the demixing structure of independent components analysis, but in this work are recognized as local means of the generative model. The cortical points of the natural elastic net effectively equi-partition the space of patches into non-overlapping faithful regions based on a similarity measure of the matrix A, and are regarded as centers of these internal regions.

The matrix A of the natural elastic net encodes no property of line features, although the natural elastic net is developed under the statistical dependency assumption similar to independent component analysis. Some rows of the obtained matrix A with their amplitude spectra are shown in Fig. 10. They describe a type of statistical dependency, but represent no line features. The corresponding amplitude spectra are dominated by high frequency responses. By numerical simulations, the natural elastic net relates line features of natural images to local means instead of the demixing structure or independent filters.

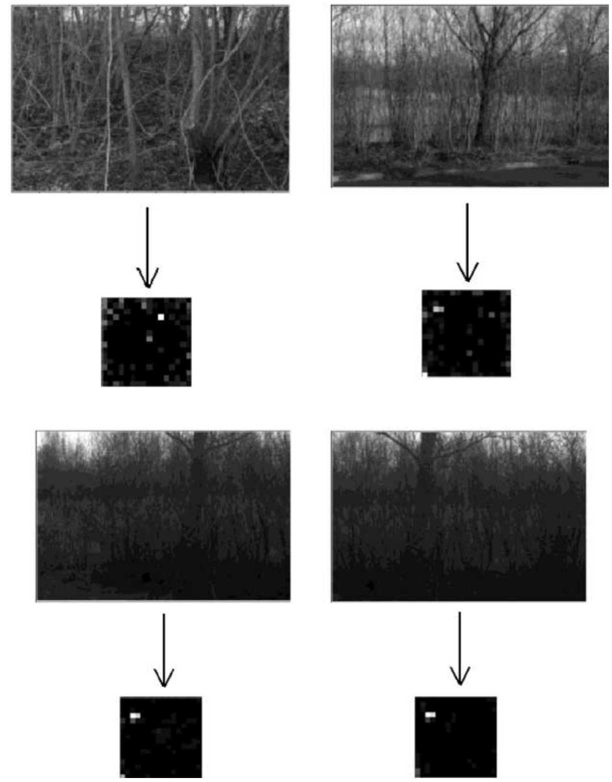

Fig. 11. Four natural image and their cumulated probabilities of all cortical points.

Based on the obtained generative model, the probability response P_k (x) of each cortical point y_k for a stimuli x in Eq. (1) can be normalized as follows

$$u_k(x) = \frac{\exp(-\|x - y_k\|_A)}{\sum_j \exp(-\|x - y_j\|_A)} = \frac{\exp(-\|z - z_k\|)}{\sum_j \exp(-\|z - z_j\|)}$$

where $z = Bx$, $z_k = By_k$ and $A = B^t B$. If the cumulate probability of a natural image associated with each cortical point y_k is defined as the sum $\sum_x v_k(x)$ over all regularly decomposed patches of the image, it may be considered as a quantitative consciousness of a natural image. Each of four natural images in Fig. 11 is decomposed into many 16×16 patches such that the 256 cumulate probabilities of each image are calculated to form a 16×16 cumulate matrix. The four natural images and their corresponding cumulate matrices are shown in Fig. 11. By human vision, the natural image in the left-upper corner is significantly different from the others, which have been selected to be similar to each other. It is observed that the similarity relation in the space of the full natural images is reserved in the space of cumulate matrices by some means. Quantitative discriminant analysis of the cumulate matrix may be further related to the task of image classification.

Further defining spatial frequency bandwidth as the full

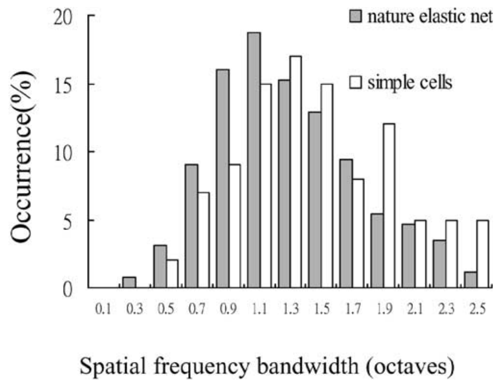

Fig. 12. A quantitative comparison of cortical points of nature elastic net and those in macaque simple cells, where the horizontal axis denotes spatial frequency bandwidth and the vertical axis denotes occurrence in percentages.

width at half maximum (FWHM) of each cortical point along the orientation of the peak in the amplitude spectra, which follows the definition in the work (Hateren van & Ruderman, 1998), we can draw the distribution of FWHM of all z_k as in Fig. 12. The distribution significantly coincides with the measurement of cortical receptive fields in macaque simple cells (DeValois, Yund, & Hepler, 1982a; DeValois, Albrecht, & Thorell, 1982b), which may present a possible biological plausibility to the natural elastic net.

The simplified natural elastic net with fixing the matrix A as an identity matrix is also applied to analysis of natural images. The simulation condition is the same as in the previous experiment. The final 16×16 cortical points are shown in Fig. 13. The simplified natural elastic net extracts features of coarse lines as compared with the line features shown from Figs. 6–9. The dynamical covariance matrix of the natural elastic net dose not directly relate to the line features, but it affects the fineness of the extracted line features by some means.

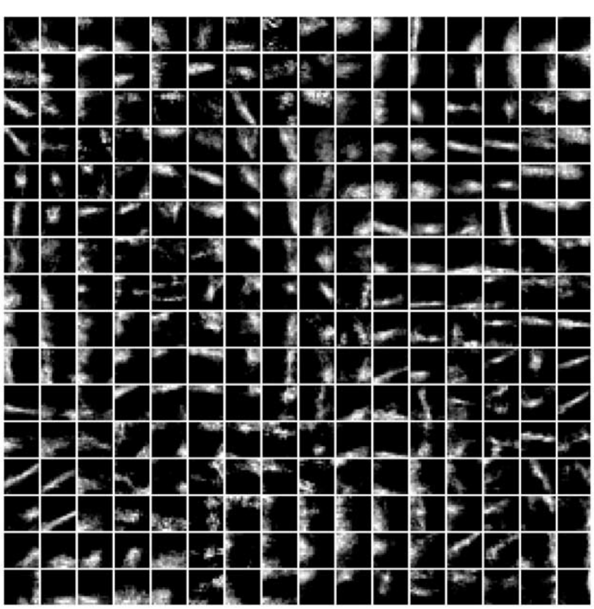

Fig. 13. The cortical points obtained by the simplified natural elastic net.

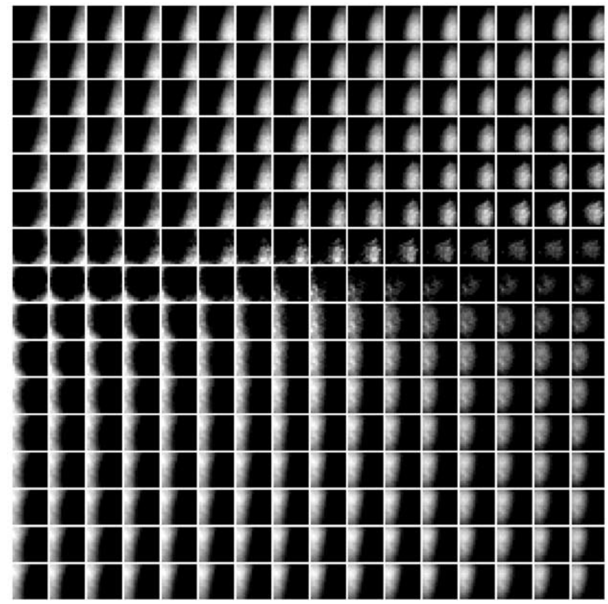

Fig. 14. The cortical points obtained by the Kohonen self-organizing algorithm.

4.2.2. Kohonen self-organizing algorithm

We further apply the Kohonen self-organizing algorithm to the same training set of patches of natural images. The used Matlab program is the SOM Toolbox 2.0 downloaded from the homepage provided by Kohonen (1982), and the size of the lattice is also 16×16. The overall cortical map generated by the Kohonen self-organizing algorithm is shown in Fig. 14, where the corresponding spectral analysis is not presented. It is observed that the Kohonen self-organizing algorithm fails to extract line features from tremendous random patches of natural images.

4.3. Discussions

We have applied the natural elastic net to estimate the generative model of natural images and have obtained piece-wise multivariate Gaussian distributions characterized by a set of local means and the common covariance matrix. In the first experiment, the natural elastic net has successfully captured the statistical dependency within training parameters via a dynamical matrix A. Besides following the strategy in previous works (Olshausen & Field, 1996; Hateren van & Ruderman, 1998; Hateren van & van der Schaaf, 1998; Hyvarinen & Hoyer, 2000) to resolve the structure of statistical dependency, the natural elastic net simultaneously stresses the importance of local means or radial basis receptive fields. The natural elastic net is a whole process of the clustering analysis and independent component analysis; it can be used to clarify the role of local means and the statistical dependency among image pixels on the formation of receptive fields in visual cortex. By numerical simulations, it is observed that the local means are internal representations with line features; although the matrix A has captured the underlying structure of statistical dependency among pixels of patches of natural images, it depicts no line features as receptive fields in visual cortex. The estimated parameters of the generative model essentially serve as effective internal representations of patches of natural images, and their reliability is supported by the circumspect arrangement of the generative model, the solid derivation of the parameter estimation and the accuracy of the neural relaxation based on a hybrid of the mean field annealing and the gradient descent methods.

5. Conclusions

Based on piecewise multivariate Gaussian distributions, we have proposed a generative model for characterizing real stimulus, such as patches of natural images. The fitness of the generative model to all training patches is further combined with the minimal wiring criterion to constitute an optimization framework for deriving the learning process of the natural elastic net. By applying a hybrid of the mean field annealing and the gradient descent method to the mathematical framework, we have three sets of interactive dynamics for the learning process. It is a generalized version of the elastic net proposed by Durbin and Willshaw (1987) with an extensive enhancement on computational accuracy, and has the following properties for analysis of real stimulus.

1. The underlying distribution of the generative model is essential and general for describing real stimulus with statistical dependent components.
2. The parameters within the generative model can be effectively determined by collective decisions of neural networks via the neural relaxation based on a hybrid of the mean field annealing and the gradient descent method.
3. Subjective to the ordering criterion of self-organization and the maximal likelihood principle, the local means and the covariance matrix of the generative models constitute faithful representations for real stimulus of an unsupervised learning task.
4. For analysis of natural images, the natural elastic net provides an alternative biological plausibility for the formation of receptive fields in visual cortex beyond the previous works based on independent component analysis. The extracted line feature is associated with the local mean of the generative model instead of the mixing structure of statistical dependent components.

References

DeValois, R. L., Yund, E. W., & Hepler, N. (1982a). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 531–544.

DeValois, R. L., Albrecht, D. G., & Thorell, L. G. (1982b). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 545–559.

Durbin, R., & Willshaw, D. (1987). An analogue approach to the travelling salesman problem using an elastic net method. *Nature*, *326*, 689–691.

Durbin, R., & Mitchison, G. (1990). ph A dimension reduction framework for cortical maps. *Nature*, *343*, 644–647.

Hateren van, J. H., & Ruderman, D. L. (1998). Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of Royal Society of London B, Biology*, *265* (1412), 2315–2320 Dec 7.

Hateren van, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of Royal Society of London B, Biology*, *265* (1394), 359–366 Mar 7.

Hyvärinen, A., & Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, *9*, 1483–1492.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, *10* (3), 626–634.

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

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59–69.

Liou, C. Y., & Wu, J. M. (1996). Self-organization using potts models. *Neural Networks*, *9* (4), 671–684.

Lin, J. K., Cowan, J. D., & Grier, D. G. (1997a). Source separation and density estimation by faithful equivalent SOM. In M. C. Mozer, *Advances in Neural Information Processing Systems* (pp. 536–542). vol. 9. Cambridge/London: MIT Press/Bradford Books.

Lin, J. K., Cowan, J. D., & Grier, D. G. (1997b). Faithful representation of separable distributions. *Neural Computation*, *9*, 1305–1320.

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

Peterson, C., & Söderberg, B. (1989). A new method for mapping optimization problems onto neural network. *International Journal of Neural Systems*, *1*, 3.

Piepenbrock, C., Ritter, H., & Obermayer, K. (1997). The joint development of orientation and ocular dominance: role of constraints. *Neural Computation*, *9* (5), 959–970.

Wu, J. M., Lin, Z. H. (2000). Natural elastic nets for faithful representations. International Computer Symposium, Taiwan.

Wu, J. M., & Chiu, S. J. (2001). Independent component analysis using Potts models. *IEEE Transactions on Neural Networks*, *12* (2) March.

Yulle, A. L., Kolodny, J. A., & Lee, C. W. (1996). Dimension Reduction, generalized deformable models and the development of ocularity and orientation. *Neural Networks*, *9* (2), 309–319.