

Overcomplete topographic independent component analysis

Libo Ma, Liqing Zhang*

Department of Computer Science and Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China

Available online 10 March 2008

Abstract

Topographic and overcomplete representations of natural images/videos are important problems in computational neuroscience. We propose a new method using both topographic and overcomplete representations of natural images, showing emergence of properties similar to those of complex cells in primary visual cortex (V1). This method can be considered as an extension of model in Hyvärinen et al. [Topographic independent component analysis, *Neural Comput.* 13 (7) (2001) 1527–1558], which uses complete topographic representation. We utilize a sparse and approximately uncorrelated decompositions and define a topographic structure on coefficients (the dot products between basis vectors and whitened observed data vectors). The overcomplete topographic basis vectors can be learned via estimation of independent component analysis (ICA) model based on the prior assumption upon basis vectors. Computer simulations are provided to show the relationship between our model and the basic properties of complex cells in V1 cortex. The learned bases are shown to have better coding efficiency than ordinary topographic ICA (TICA) bases.

© 2008 Elsevier B.V. All rights reserved.

Keywords: Topography; Overcomplete; Independent component analysis; Complex cells

1. Introduction

Recently, linear approaches of the efficient coding hypothesis [2] have provided functional explanations for properties of simple cells in V1. Independent component analysis (ICA) [4,10,1,3,21] turned out to be equivalent to sparse coding [17], which can predict the localized, oriented, and band-pass characteristics of simple cells. It is a statistical generative model, which can approximate the statistics of natural image patches as follows:

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \sum_{i=1}^n \mathbf{a}_i s_i, \quad (1)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_m)^T$ is a vector of observed data, $\mathbf{s} = (s_1, s_2, \dots, s_n)^T$ is a vector of basis coefficients or components, and \mathbf{A} is a mixing matrix. The columns of \mathbf{A} are often called basis functions or basis vectors. In ICA, the goal is to find a linear decomposition of natural data so that the components are maximally statistically independent. However, many types of natural data, particularly

natural images, do not satisfy the basic independent assumption of ICA method. Therefore, it is important and interesting to further reveal the higher-order structures in natural data, which are not captured by the ordinary ICA model. Some recent approaches, such as topographic ICA (TICA), have been proposed to extend the ICA model to learn higher-order visual structures of the natural data by capturing nonlinear dependencies after the ICA step [7,8,13,18,20].

Topographic organization was first observed in experiments in the primary cortex [6]. This cortical property was explained by TICA model [8]. In this model, the basis vectors with strong higher-order correlations are rearranged to be near to each other. This model shows emergence of properties similar to those of complex cells in V1. However, as a result of the inherent constraint of the model, TICA is suitable for complete representation of natural images, but not for overcomplete representation (i.e., the number of components is greater than the number of observed data). The inherent constraint is the orthogonality condition of the weight matrix \mathbf{W} while approximating the likelihood, which means that the dimension of the sources cannot exceed the dimension of the data (i.e., $n \leq m$).

*Corresponding author.

E-mail addresses: malibo@sjtu.edu.cn (L. Ma), zhang-lq@cs.sjtu.edu.cn (L. Zhang).

On the other hand, several approaches for the overcomplete representation have been proposed recently [17,15,16,9]. Nevertheless, most of the overcomplete bases obtained by these methods are not topographic representation but ordinary ICA Gabor-like basis, which is similar to simple cells in V1. In this paper, we are interested in investigating whether we can achieve both topographic and overcomplete representations for natural images.

In this paper, we propose an overcomplete TICA model based on maximum a posteriori (MAP) and topographic representation. For considering the overcomplete case, a quasi-orthogonal prior probability of the mixing matrix is applied. Instead of examining the properties of components as in ordinary TICA model, we define a topography on a new variable “element”, and provide a simple form of probability density approximation of the element. Our proposed model yields overcomplete representations while ordinary ICA and TICA model only produces complete representations for natural data. Overcomplete representations provide more efficient representations than the complete case, and have been widely used in fields of computational perception and pattern recognition [15]. The coding efficiency between overcomplete TICA and ordinary is compared by computing the probability of the data given learned bases. Our model provides an improved statistic model with overcomplete topographic representation for natural images and it is promising in a wide range of fields, such as signal processing and pattern recognition.

This paper is organized as follows. First, the overcomplete TICA method is described in detail in Section 2. Some experimental results are given in Section 3. Finally, conclusions and discussions are drawn in Section 4.

2. Model

In this section, we propose a MAP method that can yield topographic bases for overcomplete sparse representation of natural images. We introduce a new variable called elements to approximate the internal state of representations. By defining a higher-order topographic structure on each element, we can approximate marginal density of the elements. The probability of mixing matrix can be derived based on the quasi-orthogonality of basis vectors in a high-dimensional space. Given whitened data vector, the posterior probability for mixing matrix can be obtained by Bayesian method. Then the overcomplete topographic basis vectors can be estimated by maximizing this posterior probability of mixing matrix.

First, in order to introduce the new variable element, it is assumed that the norms of basis vectors are set to unity and the data are prewhitened by a preprocessing step as in most ICA methods. We define element as the dot product between the i th basis vector and the whitened data vector:

$$y_i = \mathbf{a}_i^T \mathbf{z} = \mathbf{a}_i^T \mathbf{A} \mathbf{s} = s_i + \sum_{j \neq i} \mathbf{a}_i^T \mathbf{a}_j s_j, \quad (2)$$

where \mathbf{z} is the whitened data vector, s_i is the i th independent component and the second term is approximately Gaussian. In feature extraction for many types of natural signal, the ICA model is usually to explore the linear independency of components. The basic ICA method provides a rather arbitrary set of components which are maximally mutually independent. Thus, we can define sparse marginal distributions on these elements, similar to defining the sparseness on the components as in most ICA method. Therefore, any good estimator of the mixing matrix should maximize the nonGaussianity of these dot products.

Furthermore, the element vector can be described by mixing matrix and whitened data as: $\mathbf{y} = (y_1, \dots, y_n)^T = \mathbf{A}^T \mathbf{z}$. Therefore, the probability for \mathbf{z} given \mathbf{A} can be approximated as follows:

$$p(\mathbf{z}(t)|\mathbf{A}) = p(\mathbf{y}) \approx C \prod_{i=1}^n p_{y_i}(y_i), \quad (3)$$

where C is a constant. Obviously, the accuracy of the prior probability p_{y_i} is important, especially for overcomplete representations [15]. Better approximation of the prior of the elements would allow the model to capture structures in images better. Schwartz and Simoncelli have observed that, for natural images, there are significant statistical dependencies among the variances of filter outputs [19]. According to this observations, we explore the energy correlations of each element. As in ordinary TICA model, we construct a two-layer neural network. Now the response of simple cell is element y_i instead of component s_i . The simple cells are assumed to be arranged in a two-dimensional grid and the responses of simple cells are not mutually independent. Instead, the nearby simple cells have energy correlations. The topography is formally described by a neighborhood function $h(i, j)$ with indices i and j in a two-dimensional grid. Typically, if the cells are close enough to each other, $h(i, j)$ is denoted to 1. Otherwise it is assigned to 0. The second layer is composed of complex cells, which pool outputs of simple cells that are nearby on the two-dimensional grid.

To specify the model for the topography and the sparseness of elements, we define the probability of element as follows:

$$\hat{p}_{y_i}(y_i) = \exp \left(G \left(\sum_j h(i, j) y_j^2 \right) \right), \quad (4)$$

where the function $G(y)$ should be convex for nonnegative y . For example, one can use this form

$$G(y) = -\alpha \sqrt{y} + \beta. \quad (5)$$

Thus, combining Eqs. (3) and (4), leads to an approximation of the probability for \mathbf{z} given \mathbf{A} as follows:

$$p(\mathbf{z}(t)|\mathbf{A}) \approx C \prod_{i=1}^n \exp \left(G \left(\sum_j h(i, j) y_j^2 \right) \right). \quad (6)$$

On the other hand, it is observed that, in a very high-dimensional space, the vectors are most likely orthogonal,

which is a somewhat counterintuitive phenomenon. In other words, the number of almost orthogonal directions is much larger than that of orthogonal directions. This property is called quasi-orthogonality [9,5,11,12,14]. Therefore, in a high-dimensional space even vectors having random directions might be sufficiently close to orthogonality. Based on this observation, any good estimator of mixing matrix should make different basis vectors tend to be quasi-orthogonal. In this paper, we use the prior probability of mixing matrix \mathbf{A} proposed in [9], which is derived from the quasi-orthogonality of random vectors in a high-dimensional space. It is assumed that the dot product $\mathbf{a}_i^T \mathbf{a}_j$ between basis vectors \mathbf{a}_i^T and \mathbf{a}_j are randomly and independently drawn in a high-dimensional space, then the prior probability of mixing matrix \mathbf{A} is derived as follows:

$$p(\mathbf{A}) = c_m \prod_{i < j} (1 - (\mathbf{a}_i^T \mathbf{a}_j)^2)^{m-3/2}, \quad (7)$$

where the constant $c_m = (m-1)/m\Gamma[m/2-1]/\sqrt{\pi}\Gamma[m/2+1]$. The detailed derivation can be obtained in [9].

Then the posterior probability for the mixing matrix given whitened data vector can be obtained by the Bayesian method:

$$p(\mathbf{A}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{A})p(\mathbf{A})}{p(\mathbf{z})}, \quad (8)$$

where $p(\mathbf{z})$ does not depend on \mathbf{A} . $p(\mathbf{A})$ assigns a higher probability to quasi-orthogonal matrices, since the basis vectors are in a high-dimensional space ($m > 3$). And the probability for \mathbf{z} under the model $p(\mathbf{z}|\mathbf{A})$ holds both topography of energy correlations and sparseness. Finally, we obtain the following approximation of the log-probability of the posterior for T observations $\mathbf{z}(t)$, $t = 1, \dots, T$ as follows:

$$\begin{aligned} \log p(\mathbf{A}|\mathbf{z}(t), t = 1, \dots, T) \\ \approx \sum_{t=1}^T \sum_{i=1}^n G \left(\sum_j h(i,j) y_j^2 \right) \\ + \alpha T \sum_{i < j} \log(1 - (\mathbf{a}_i^T \mathbf{a}_j)^2) + \text{const.}, \end{aligned} \quad (9)$$

where α is a constant that is related not only to c_m , but also to the approximations we have made.

Then, maximizing this log-probability of the posterior over basis vector \mathbf{a}_k using gradient ascent method yields the following learning rule:

$$\begin{aligned} \Delta \mathbf{a}_k \propto \eta \left(\sum_{t=1}^T \mathbf{z}(t) (\mathbf{a}_k^T \mathbf{z}(t)) \sum_j h(k,j) g \left(\sum_i h(i,j) (\mathbf{a}_i^T \mathbf{z}(t))^2 \right) \right. \\ \left. + \alpha T \sum_{i < j} \frac{-2 \mathbf{a}_i^T \mathbf{a}_j}{1 - (\mathbf{a}_i^T \mathbf{a}_j)^2} \mathbf{b}_k \right), \end{aligned} \quad (10)$$

where η is the learning rate. \mathbf{b}_k is the k th column vector of matrix $\mathbf{B} = [0, \dots, \mathbf{a}_j, \dots, \mathbf{a}_i, \dots, 0]$, \mathbf{a}_j is the i th column vector, and \mathbf{a}_i is the j th column vector. The function g is the

derivative of G . In each iteration after updating \mathbf{a}_k using learning rule (10), the norm of the basis vector \mathbf{a}_k needs to be set to unity. This is different from ordinary TICA algorithm, which needs to set the filter \mathbf{w}_i to be orthonormalized.

3. Simulations

To estimate the overcomplete basis functions, we use small patches taken from digital natural images of standard sets available on the internet from Hyvärinen¹ and Olshausen.² The results obtained from these two different image sets are not significantly different. The training data of 30,000 samples are generated by extracting 16×16 small patches at random positions from natural images, skipping over any patches within 8 pixels of the borders of each image. The patches are subsequently normalized such that mean pixel intensity for a given pixel across the samples is zero, and the DC component of each patch is also removed. After the dimension is reduced from 256 to 100 (i.e., projected onto the leading 100 eigenvectors of the covariance matrix), the data vectors are whitened to have unit variance. These data preprocessing can speed up training without imposing much impact on the final results obtained. The exact form of the scalar function G , which plays a similar role as the log-density of ICs in classic ICA, does not affect the estimator, as long as the overall shape is correct. Maximizing log likelihood method is used to estimate \mathbf{A} by an ordinary gradient method. Note that there is no constraint of orthogonality during each iteration other than the norms of basis vectors are set to unity.

We train our model on overcomplete settings. The basis is set to $2\times$ and $4\times$ overcompleteness, i.e., the dimension of components is 200 and 400, respectively. In addition, the constant α is set to 0.04 and 0.02, respectively. Learned basis vectors are shown in Figs. 1(a) and 2. The $2\times$ and $4\times$ overcomplete basis are quite well estimated, though a few basis vectors are a bit messy. With notable similarity to experimentally observed cortical topography, one can observe pinwheel singularities in the organization map. It can be seen that the locations of two neighboring basis vectors are near to each other and their orientation and frequency tend to be similar as well. However, their phases are very different.

The approximation of basis vectors is investigated in more detail by performing with different sizes of neighborhoods. For the sake of avoiding border effects, the two-dimensional torus lattice is chosen. The learned basis vectors in $2\times$ overcomplete case and neighborhoods of 5×5 are shown in Fig. 1(b). It can be seen that different sizes of neighborhoods bring somewhat different topographic organization of basis vectors. The basis vectors with a larger neighborhood yield more elongated Gabor shape

¹<http://www.cis.hut.fi/projects/ica/data/images/>

²<http://redwood.berkeley.edu/bruno/sparsenet/>

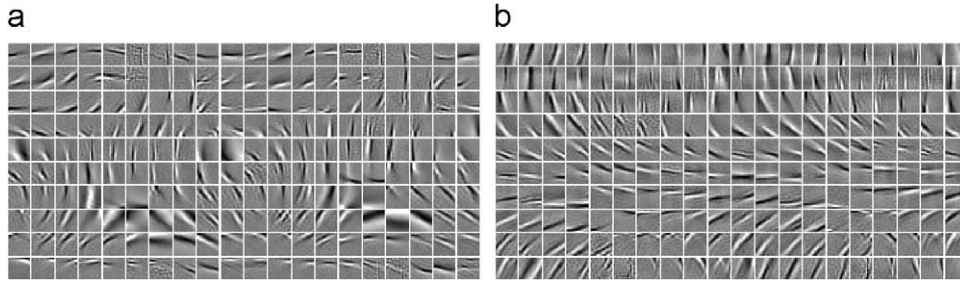


Fig. 1. The estimated basis vectors of $2\times$ overcompleteness with two neighborhood sizes. The dimension of components is 200 and 400, respectively. (a) Neighborhood size 3×3 . (b) Neighborhood size 5×5 .

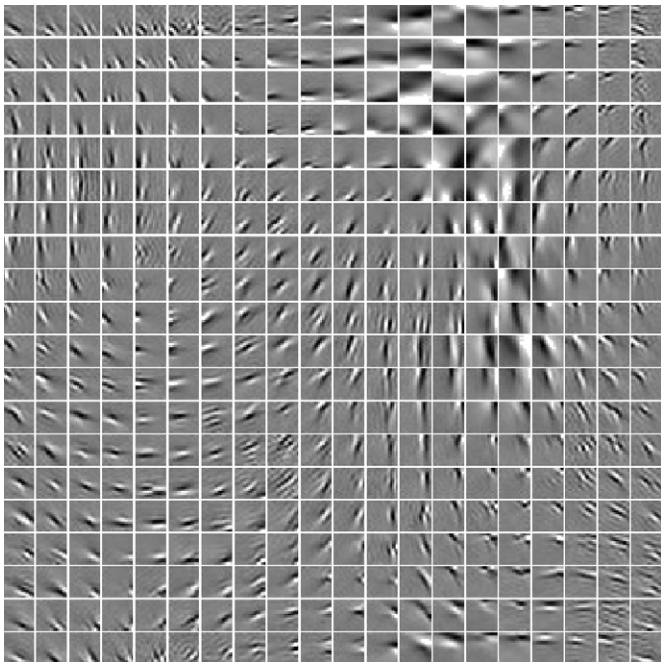


Fig. 2. The estimated basis vectors of $4\times$ overcompleteness with neighborhood size 3×3 .

and the topographic organization is more strongly affected by the orientation.

We analyze in more detail the properties of basis vectors by fitting a Gabor function to each basis vector (using the least squares procedure). Fig. 3 shows the distribution of parameters obtained by fitting Gabor functions to complete, $2\times$ and $4\times$ overcomplete basis vectors. It can be seen that the orientation is quite independent of frequency. With the increasing of the level of overcompleteness, the scattering points in the plot of location, spatial frequency and orientation become denser and more uniform. And the distribution of phase becomes more uniform for overcomplete cases.

We also examine the local parameter correlations and population properties in terms of Gabor parameters, including spatial position, orientation, frequency, and phase. Fig. 4 shows the scatterplots of each pair of two nearby basis vectors for one single parameter with $4\times$ overcompleteness. One can observe that the location,

orientation, and frequency of two nearby basis vectors are strongly correlated, while the phases are not correlated. They are random points in the phase plane.

Fig. 5 shows the population maps of fitted Gabor parameters with $4\times$ overcompleteness for the same parameter. As for orientation map, one can observe gradual changes of color from red to blue, which represent the changes of orientation. Especially, the pinwheels structure is observed for overcomplete cases. From the frequency map, one can find the spot of low-frequency basis, which can also be observed in Fig. 2. In the phase map, one can see that the distribution of phases is random indicating there is no phase structure in the basis vectors in this model. It is closely related to complex cells in V1 in that it is based on location, frequency, and orientation and independent of phase.

In order to examine the quasi-orthogonality of the learned basis vectors, we calculate the minimum angle between each basis vector and the rest basis vectors in the whitened space. If this minimum angle is close to 90° , the given basis vector is said to be quasi-orthogonal to the rest basis vectors. The minimum angles can be calculated as follows: take the absolute value of matrix $\mathbf{A}^T\mathbf{A}$, and set its diagonal elements to be zero. Then, the arccos value of the maximum of each column vector of this absolute value matrix is the minimum angle. The minimum angles derived from ordinary TICA and our overcomplete TICA in $2\times$ and $4\times$ overcomplete case are shown in Fig. 6. We can see that all the angles derived from TICA model are 90° for its orthogonality conditions. Whereas, all the minimum angles derive from our model are above 60° , which demonstrates good quasi-orthogonality. This is consistent with the relaxation of orthogonality conditions in a high-dimensional space.

We also compare coding efficiency of the basis functions learned by ordinary TICA and our OTICA method. In this paper, the estimating method of coding length in [16] is used to estimate the coding efficiency. The objective function $P(\mathbf{x}|\mathbf{A})$ for the probability of the data is a natural measure for estimating coding cost. For noiseless model, the number of bits required to encode the pattern is given by

$$\# \text{ bits} \geq -\log_2 P(\mathbf{x}|\mathbf{A}). \quad (11)$$

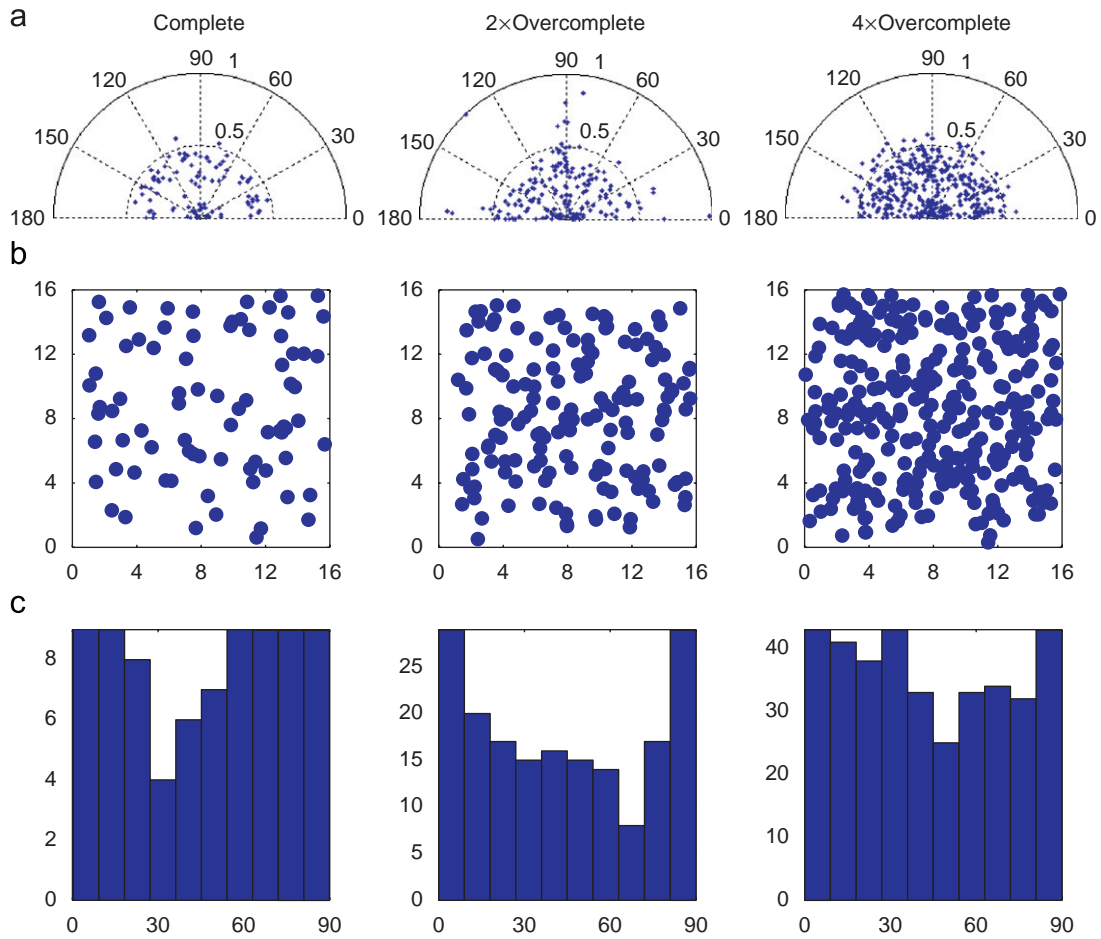


Fig. 3. The distributions of some parameters derived by fitting Gabor function. The leftmost column is a complete case, the middle column is 2x overcomplete case and the rightmost column is 4x overcomplete case. (a) Joint distribution of orientation and spatial frequency (plotted in the upper-half plane). (b) Center location within a patch. (c) Histogram of phase (mapped to range 0–90°).

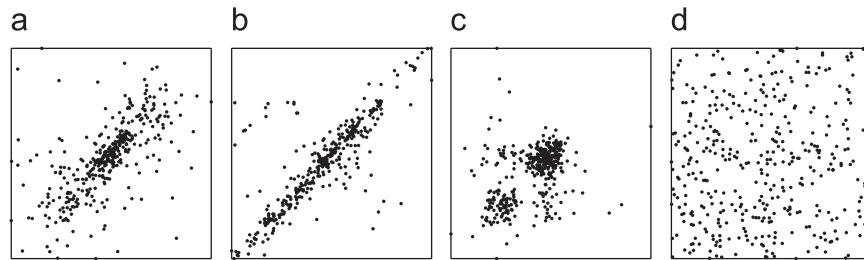


Fig. 4. Distributions of pairs of two adjacent basis vectors. Gabor parameters: (a) location; (b) orientation; (c) frequency; and (d) phase with 4x overcompleteness.

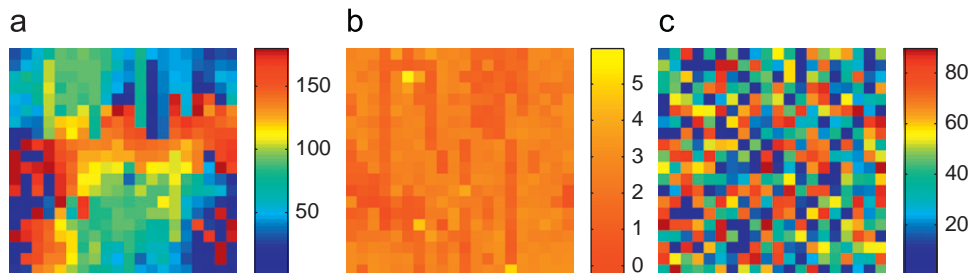


Fig. 5. Population maps of Gabor parameters: (a) orientation; (b) frequency; and (c) phase with 4x overcompleteness.

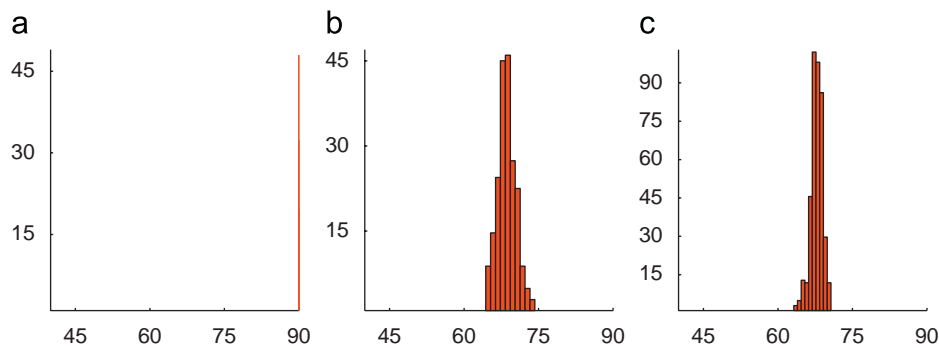


Fig. 6. The minimum angles between the learned basis vectors. (a) Complete case derived from TICA model. (b and c) $2\times$ and $4\times$ overcomplete case derived from our OTICA model.

Table 1
Estimated coding length using $P(x|A)$

Method	Coding length
TICA with complete bases	26.99 ± 0.43
OTICA with complete bases	15.20 ± 0.22
OTICA with $2\times$ overcomplete bases	7.17 ± 0.11
OTICA with $4\times$ overcomplete bases	5.41 ± 0.09

The estimated coding costs are calculated from 100 sets of 1000 randomly sampled 16×16 patches of natural images. Table 1 shows the estimated coding cost of TICA and OTICA model. We can see that our overcomplete TICA method provides significantly efficient representations.

4. Conclusion

We introduce a new generative model based on a quasi-orthogonal prior on the basis vectors for estimating overcomplete topographic basis in a high-dimensional space. It is an extension of TICA method proposed by Hyvärinen [8], which provides a complete basis of natural images. Our proposed model is able to provide an overcomplete topographic representation by examining the properties of dot products between basis vectors and whitened data vectors and maximizing the nonGaussianity and quasi-orthogonality of these dot products.

Our method is different from ordinary TICA and ordinary overcomplete ICA in several ways. First, we examine the properties of dot products between basis vectors and whitened data vectors in a high-dimensional space. Second, instead of examining the components, we define a topographic order for the elements and employ a simple form of probability density approximation of elements with an energy of the neighborhood term. And the mixing matrix A can be directly estimated by maximizing the log-probability for posterior of the mixing matrix given whitened data. Finally, we need not impose the orthogonal constraint on the mixing matrix besides the unity normalization of basis functions in each iteration. Thus maximizing the log-probability for posterior of

mixing matrix given whitened data, we can estimate different levels of overcomplete topographic basis vectors.

An important aspect of our model is the definition of a new variable “element” to avoid the orthogonal constraint on the mixing matrix. The main advantage is that, by introducing element, we do not need to estimate the component s_i in every step, as in most algorithms. Furthermore, each element can be decomposed into the sum of the component and the effects of nearby components in two-dimensional topography. And the weighted summation of nearby components is approximately Gaussian. Therefore, we can use this element to estimate the bases of the model. The element can be seen as the response of simple cell including the effects of nearby simple cells.

Simulations on natural image data show that our proposed overcomplete TICA produces a Gabor-like linear basis vectors and shows simultaneous emergence of complex cell properties including location, frequency, orientation, and phase. The results of coding efficiency experiments demonstrate that the overcomplete TICA model provides much more efficient representations than ordinary TICA method. It is promising in a wide range of fields, such as signal processing and pattern recognition.

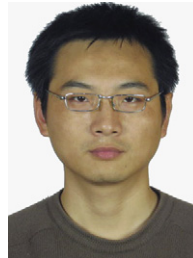
Acknowledgments

The work was supported by the National Basic Research Program of China (Grant no. 2005CB724301) and the National High-Tech Research Program of China (Grant no.252006AA01Z125).

References

- [1] S. Amari, A. Cichocki, H. Yang, A new learning algorithm for blind signal separation, *Adv. Neural Inf. Process. Syst.* 8 (1996) 757–763.
- [2] H.B. Barlow, Redundancy reduction revisited, *Network Comput. Neural Syst.* 12 (2001) 241–253.
- [3] A. Bell, T. Sejnowski, The ‘independent components’ of natural scenes are edge filters, *Vision Res.* 37 (23) (1997) 3327–3338.
- [4] P. Comon, Independent component analysis—a new concept?, *Signal Process.* 36 (1994) 287–314.

- [5] R. Hecht-Nielsen, Context vectors: general purpose approximate meaning representations self-organized from raw data, *Comput. Intell. Imitating Life* (1994) 43–56.
- [6] D.H. Hubel, T.N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, *J. Physiol.* 195 (1) (1968) 215–243.
- [7] A. Hyvärinen, P.O. Hoyer, Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces, *Neural Comput.* 12 (7) (2000) 1705–1720.
- [8] A. Hyvärinen, P.O. Hoyer, M. Inki, Topographic independent component analysis, *Neural Comput.* 13 (7) (2001) 1527–1558.
- [9] A. Hyvärinen, M. Inki, Estimating overcomplete independent component bases for image windows, *J. Math. Imaging Vision* 17 (2) (2002) 139–152.
- [10] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis, *Neural Comput.* 9 (1997) 1483–1492.
- [11] P. Kanerva, *Sparse Distributed Memory*, MIT Press, Cambridge, MA, 1998.
- [12] J. Karhunen, E. Oja, L. Wang, R. Vigário, J. Joutsensalo, A class of neural networks for independent component analysis, *IEEE Trans. Neural Networks* 8 (1997) 486–504.
- [13] Y. Karklin, M.S. Lewicki, A hierarchical Bayesian model for learning nonlinear statistical regularities in nonstationary natural signals, *Neural Comput.* 17 (2005) 397–423.
- [14] S. Kaski, Dimensionality reduction by random mapping: fast similarity computation for clustering, in: *Proceedings of the International Joint Conference on Neural Networks (IJCNN'98)*, Anchorage, Alaska, 1998, pp. 413–418.
- [15] M.S. Lewicki, B.A. Olshausen, Probabilistic framework for the adaptation and comparison of image codes, *J. Opt. Soc. Am. A Opt. Image Sci.* 6 (1999) 1587–1601.
- [16] M.S. Lewicki, T.J. Sejnowski, Learning overcomplete representations, *Neural Comput.* 12 (2000) 337–365.
- [17] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: a strategy employed by V1?, *Vision Res.* 37 (1997) 3311–3325.
- [18] J. Portilla, V. Strela, M.J. Wainwright, E.P. Simoncelli, Image denoising using scale mixtures of Gaussians in the wavelet domain, *IEEE Trans. Image Process.* 12 (11) (2003) 1338–1351.
- [19] O. Schwartz, E. Simoncelli, Natural signal statistics and sensory gain control, *Nat. Neurosci.* 4 (2001) 819–825.
- [20] M. Welling, G.E. Hinton, S. Osindero, Learning sparse topographic representations with products of student- t distributions, *Adv. Neural Inf. Process. Syst.* 15 (2002) 1359–1366.
- [21] L. Zhang, A. Cichocki, S. Amari, Self-adaptive blind source separation based on activation function adaptation, *IEEE Trans. Neural Networks* 15 (2) (2004) 233–244.



Libo Ma is currently a Ph.D. candidate with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China. His research interests include computational neuroscience, image processing, and computer vision.



Liqing Zhang received his B.S. degree in mathematics from Hangzhou University, in 1983, and the Ph.D. in computer sciences from Zhongshan University, China, in 1988. He became an Associate Professor in 1990 and then a Full Professor in 1995 at the Department of Automation, South China University of Technology. He joined Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Japan, in 1997 as a Research Scientist. Since 2002, he has been working in Department of Computer Sciences and Engineering, Shanghai Jiao Tong University, China. His research interests include neuroinformatics, perception computing, and statistical learning. He has published more than 130 papers.