

BIOLOGICAL RELEVANCE OF FEATURES EXTRACTED FROM NATURAL IMAGES

R. MUTIHAC¹, MARC M. VAN HULLE²

¹ *University of Bucharest, Department of Electricity and Biophysics, RO-76900, Romania,*

² *K.U. Leuven, Labo voor Neuro- en Psychofysiologie, B-3000 Leuven, Belgium,
{radu.mutihac, marc.vanhulle}@med.kuleuven.ac.be*

Abstract. Our contribution highlights the statistical properties and biological interpretation of the basis vectors and the filters that result from applying topographic independent component analysis (ICA) to feature extraction from randomly selected patches of natural images. The consistency of the feature sets obtained from various collections of natural image data sets applying topographic ICA supports the opinion that the statistical properties of the environmental stimuli enforce a process according to some optimization criterion, which provides a good computational model for the response properties of sensory neurons in the mammalian primary visual cortex. It is nevertheless unlikely that ICA decomposition of natural images might lead to the development of an efficient and robust image compression system.

Key words: natural images, independent component analysis (ICA), topographic maps, higher-order statistics, sensory neurons.

1. INTRODUCTION

It has widely been accepted that the sensory neurons have adapted by means of both evolutionary and developmental processes to the statistical properties of the stimuli that have mostly often been encountered in the environment. The importance of determining precise quantitative relationships between environmental statistics and neural processing is manifold. Better understanding of functional properties of neurons and neural systems, and the design of new computational models of biological neural networks based on environmental statistics are the two main issues. Secondly, finely tuned experimental design and protocols for probing biological systems can be conceived and, last but not least, improved interfaces between human beings and artificial systems can be designed, with major benefits in our interaction with the environment.

In a neurobiological context, Barlow [1] suggested that a main role of early sensor neurons is to remove the statistical redundancy in the sensory input by performing an “efficient coding” of stimuli. Such a task requires the specification of the environment, which amounts to a probability distribution over the space of

the input signals. One straightforward approach is studying the statistical properties of neural responses to natural stimulation conditions. An alternative way is to conceive statistical generative models of the input data. If the parameters of the model are estimated from natural input data, they are likely to provide deeper insight on the computational properties of the sensory neurons. There is some evidence that out of all visual images possible we see only a very small fraction [2]. If decomposing an image into independent components is one of the principal tasks of simple cells in the primary visual cortex, it should entail that the distribution of their properties is determined by the statistics of the visual environment. Olshausen and Field [3] modeling visual data with a simple linear generative model showed that the principle of maximizing sparseness (or supergaussianity) of the underlying image components explain the emergence of wavelet-like filters that resemble the receptive fields of simple cells. Running ICA on a large set of calibrated images, and comparing a series of properties of the resulting receptive fields with those of receptive fields measured in simple cells, they reported a good similarity.

In our simulations, we adopted a similar methodology and employed the topographic ICA (TICA) model introduced by Hyvärinen *et al.* [4]. TICA relaxes the independence assumption by replacing the conventional topologic ordering based on Euclidian distances of the basis vectors with a new topographic organization based on the dependence in higher-order statistics. The higher-order dependencies, which linear ICA does not remove, are used to define a topographic order such that nearby cells tend to be active (or inactive) at the same time. In this contribution we report our results in applying TICA to feature extraction from natural images. We analyzed statistically the properties of the TICA basis vectors that resulted from different sets of natural images by performing TICA decomposition. By analogy, it was conjectured to a certain extent that the topographic neighborhoods exhibit properties specific to the complex cells in the mammalian primary visual cortex (V1). Apart from emulating the features of simple cells, like the distributions for spatial frequency bandwidth, orientation tuning bandwidth, aspect ratio, and receptive field length, the topological organization allows the emergence of phase and (partial) shift invariance that characterize complex cells.

2. THE ICA MODEL FOR IMAGES

The basic stationary noiseless linear ICA model assumes that $\mathbf{s}(t) \in \tilde{\mathcal{N}}^M$ and $\mathbf{x}(t) \in \tilde{\mathcal{N}}^N$ are two stochastic processes described by (column) vectors with zero mean and finite covariance, with the components of $\mathbf{s}(t)$ being statistically independent and at most one gaussian. The mixing matrix \mathbf{A} is assumed constant (for stationarity), rectangular, and full-column rank $N \times M$, having at least as

many rows as columns ($N \geq M$), and t stands as the sample index (e.g. time or point) taking discrete values

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) = \sum_{i=1}^M \mathbf{a}_i s_i(t), \quad t = 1, 2, \dots, T \quad (1)$$

The columns $\{\mathbf{a}_i\}$, $i = 1, 2, \dots, M$ of \mathbf{A} are the ICA basis vectors. Then the ICA problem can be formulated as follows: given T realizations of $\mathbf{x}(t)$, estimate both the matrix \mathbf{A} and the corresponding realizations of $\mathbf{s}(t)$. In an alternative context, the ICA decomposition (1) is equivalent with sparse coding [5]. Further on, assuming sphered (whitened) data, then most ICA algorithms are searching for a separation matrix \mathbf{W} for demixing data based on various estimation principles of independence and implemented as adaptive on-line or batch computations [6].

There is clear evidence that the distribution of natural images is nongaussian [7]. Hence it seems plausible considering a two-dimensional static monochrome image $I(x, y)$ as a linear superposition of some *features* or *basis functions* $\{a_i(x, y)\}$, $i = 1, 2, \dots, M$

$$I(x, y) = \sum_{i=1}^M a_i(x, y) s_i \quad (2)$$

where each image $I(x, y)$ has different stochastic coefficients s_i , $i = 1, 2, \dots, M$. In order to comply with the underlying ICA assumptions, the coefficients $\{s_i\}$ are assumed nongaussian and mutually independent. Estimating the model amounts to determining the values of s_i and $a_i(x, y)$ for all indexes i and points (x, y) , given a sufficient number of observations of images, such as image patches $I(x, y)$. If we restrict the study to the case where $a_i(x, y)$ form an invertible linear system, then $s_i = \langle w_i, I \rangle$, where $w_i(x, y)$, $i = 1, 2, \dots, M$ denote the inverse filters and $\langle w_i, I \rangle = \sum_{x,y} w_i(x, y) I(x, y)$ stands for the dot product. The inverse filters $\{w_i(x, y)\}$ can be identified as the *receptive fields* of the model simple cells, and the coefficients $\{s_i\}$ as their activities when presented with a given image patch $I(x, y)$. When this model is estimated with input data consisting of patches of natural scenes, the obtained filters $w_i(x, y)$ exhibit the three principal properties of simple cells in V1: they are spatially localized, oriented, and band-pass in different spatial frequency bands. Quantitative comparison of the obtained filters $\{w_i(x, y)\}$ with those measured by single-cell recordings of the macaque cortex showed a close match for most of the parameters [8].

3. THE TOPOGRAPHIC ICA MODEL

The model (2) is nevertheless inappropriate to describe the response of complex cells due to their properties of phase and (limited) shift invariance [9]. In classic ICA, the latent variables $\{s_i\}$ have no particular order and no relationship between them is assumed whatsoever. This is in compliance with the assumption of complete statistical independence of the latent variables. However, there are applications in which ICA does not completely remove the dependence between components, which may be quite informative. Since many ICA estimation methods constrain the components to be uncorrelated, it seems reasonable to preserve the uncorrelatedness in any further extension of ICA. Hyvärinen and Hoyer [10] proposed a higher-order correlation based on *energies*, which can be intuitively interpreted as simultaneous activation of the units

$$\text{cov}(s_i^2 s_j^2) = E\{s_i^2 s_j^2\} - E\{s_i^2\}E\{s_j^2\} \neq 0 \quad (3)$$

if s_i and s_j are close on the topographic grid. In the generative model $\mathbf{x} = \mathbf{A}\mathbf{s}$ of TICA, the central issue is to define the joint density of \mathbf{s} based on the topography (i.e. simultaneous activation or inactivation of the nearby cells). The topography is generally defined by specifying a neighborhood function $h(i, j)$ that express the proximity between the components i and j . Its common form is a monotonically decreasing function of some distance measure, so that $h(i, j)$ has the form of a matrix of the model hyperparameters. Hereafter $h(i, j)$ is assumed known and fixed. The components $\{s_i\}$ of \mathbf{s} are defined by means of their variances $\{s_i^2\}$, which are assumed random variables generated according to a model specified on the basis of topography. Then the variables $\{s_i\}$ are generated mutually independent using some conditional distributions. So the dependence among the s_i 's is implied by the dependence of their variances

$$s_i = \mathbf{f}\left(\sum_{k=1}^M h(i, k) u_k\right) z_i \quad (4)$$

with $\{u_k\}$ the higher-order independent components used to generate the variances, \mathbf{f} a scalar nonlinearity, and z_i is a random variable with the same distribution as s_i , if s_i^2 is fixed to unity. The variables u_i and z_i are mutually independent.

The properties of the TICA model are discussed in detail by Hyvärinen, Hoyer, and Inki in [4]. The model is a missing variable model in which its likelihood cannot be obtained in a closed form. An approximation for the likelihood of the model results if following the derivation as in ICA [11]. An

analytical approximation was derived by assuming further simplifications, namely constant marginal densities for $\{u_i\}$ and constant conditional densities for $\{s_i\}$ as a

gaussian, and the nonlinearity is defined as $\mathbf{f}\left(\sum_{k=1}^M h(i,k)u_k\right) = \left(\sum_{k=1}^M h(i,k)u_k\right)^{\frac{1}{2}}$,

hence the log likelihood (actually an approximation of its lower bound) becomes

$$\log L(\mathbf{W}) \cong \sum_{t=1}^T \sum_{j=1}^M G\left(\sum_{i=1}^M h(i,j)(\mathbf{w}_i^T \mathbf{x}(t))^2\right) + T \log|\det \mathbf{W}| \quad (5)$$

where $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M)^T = \mathbf{A}^{-1}$, $\mathbf{x}(t)$, $t = 1, 2, \dots, T$ are the observations of \mathbf{x} . Practically, if data are sparse, convergence is achieved for almost any contrast (objective) function G that is convex for some nonnegative argument [4].

The model can be resolved by maximizing $\log L(\mathbf{W})$. The data, which are assumed zero-mean, are first whitened $\mathbf{z} = \mathbf{V}\mathbf{x} = \mathbf{V}\mathbf{A}\mathbf{s}$. If $\mathbf{V}\mathbf{A}$ is invertible, the new separating matrix becomes $\mathbf{W} = (\mathbf{V}\mathbf{A})^{-1}$ and we can constrain its rows $\{\mathbf{w}_i^T\}$ to form an orthonormal system [5], [12]. The orthonormal basis in the whitened space amounts to decorrelating the estimated components, so that their dependency in higher-order statistics remains. A simple gradient algorithm can be derived for updating the (weight) vectors $\{\mathbf{w}_i\}$ [4]

$$D\mathbf{w}_i \propto E\{\mathbf{z}(\mathbf{w}_i^T \mathbf{z})r_i\} \quad (6)$$

where $r_i = \sum_{k=1}^M h(i,k)g\left(\sum_{j=1}^M h(k,j)(\mathbf{w}_j^T \mathbf{z})^2\right)$ is a modulation factor and the function g

is the derivative of a convex function such as $G(y) = -\mathbf{a}\sqrt{y} + \mathbf{b}$. The scaling constant \mathbf{a} and the normalizing constant \mathbf{b} are determined so as to produce a probability density in compliance with the imposed constraints on $\{\mathbf{w}_i\}$. Actually, the vectors $\{\mathbf{w}_i\}$ must be normalized to unit variance and orthogonalized after every step in equation (6). If we denote the matrix $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M)^T$, the

method involving matrix square roots $\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^T)^{\frac{1}{2}}\mathbf{W}$ can be used for orthogonalization. The original mixing matrix of the unwhitened data can be computed after learning $\{\mathbf{w}_i\}$ such as $\mathbf{A} = (\mathbf{W}\mathbf{V})^{-1}$. The rows of \mathbf{A}^{-1} provide the filters (weight vectors) in the original, not whitened space.

4. EXPERIMENTAL

We collected and calibrated natural images from the Natural Resources Conservation Service [13], converted to monochrome uncompressed TIFF graphics format files with 8-bit pixel depth (i.e. 256 gray levels), and cropped down to 256×256 pixel size, with a resolution of 150 pixel/inch. Original images were randomly rotated in order to avoid any bias caused by camera orientation and/or light source position. Then square 16×16 pixel image patches were randomly selected and cropped. The image patches, which actually constituted our raw data, were subsequently converted into vectors of length 256. We selected in each study case 24 images, with both unimodal and multimodal histograms, out of which 48,000 image patches were extracted and stored as columns of a data matrix \mathbf{X} . The mean gray scale value of each image patch was subtracted and the dimensionality was reduced (e.g. low-pass filtering) by running principal component analysis (PCA) and retaining the principal 144 components with the largest variances. Consequently the size of our data matrix \mathbf{X} was fixed down to 256×144 . Then the data were sphered by zero-phase whitening filter, which equated to a multiplication of data \mathbf{X} by the inverse of its squared covariance matrix \mathbf{C} . Now data were contained in a 144-dimensional subspace spanned by the 144 most energetic basis vectors that formed an orthonormal system, but not for the original data space. For visualization purposes and in order to avoid border effects, the topography was chosen as 2D torus lattice as suggested first by Kohonen [14]. We used only one neighborhood size S_m of a 3×3 square around each unit. It meant that the neighborhood function could be expressed as $h(i, j) = 1$ if $\exists m: i, j \in S_m$ and zero otherwise. The form of function G was the simplest possible amended with a small constant ϵ for numerical stability [4], that is $G(y) = -\mathbf{a} \sqrt{\epsilon + y} + \mathbf{b}$. Then the gradient method was used to maximize the approximation of the likelihood (5) over the 48,000 image patches under the constraint of orthonormality of the 144 filters in the whitened space. Each running of the algorithm on a PC with Pentium 4 processor at 1.5 GHz took around 40 hours until reaching a similar value of the objective function in all cases.

5. RESULTS AND DISCUSSION

We used 3 sets of 24 images each. First set contained as general as possible natural images (e.g. landscapes, animals, mountains, etc), the second contained mostly rounded natural objects such as rain drops, and the third was formed almost exclusively by squared natural objects like city buildings (Fig. 1). The corresponding basis vectors obtained by TICA are shown in Fig. 2.

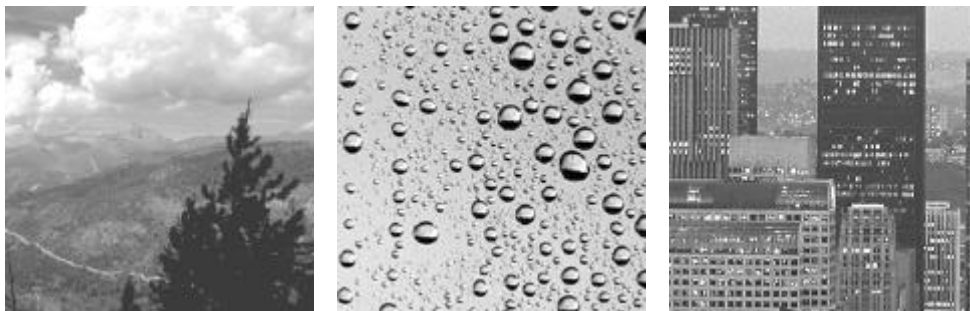


Fig. 1. – Samples from each natural image set: general (left), round shapes (middle), and square shapes (right).

Every vector in TICA decomposition represents a direction, that is, a I -dimensional subspace. TICA applied to natural image data yields Gabor-like basis vectors for randomly selected image patches as ICA does. Supplementary, the topographic organization maps basis vectors that are similar in location, orientation, and frequency close to each other.

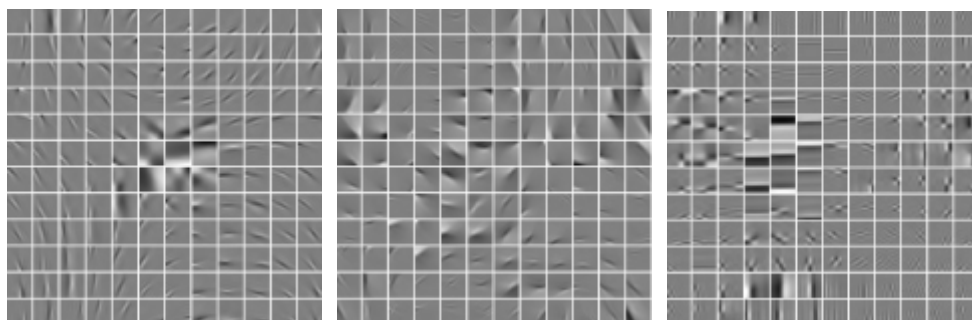


Fig. 2. – 144 -dimensional basis vectors for the general images (left), round shape images (middle), and square shape images (right).

The topographic map of the general image set appears highly structured compared to the other two basis vector sets, due to significantly more invariant features in the general images. In fact, the basis systems for the round and square image sets appear to be overdetermined, that is, the dimensionality of these spaces are likely lower than 144 . We expect that for larger neighborhoods, the preponderance of spatial frequency to decrease in favor of orientation and location. A quantitative study of the topographic organization can be carried out by computing the correlations of the energies between components of the whole data space. The model predicts a gradually vanishing of the correlations with the distance on the topographic grid. The analysis may fail because of two reasons: (i) if the basis vector system is underdetermined, the neighborhoods are too small,

and/or there are not enough data, (ii) if the image patches do not comply entirely with the model (i.e. they are not a linear superposition of invariant features).

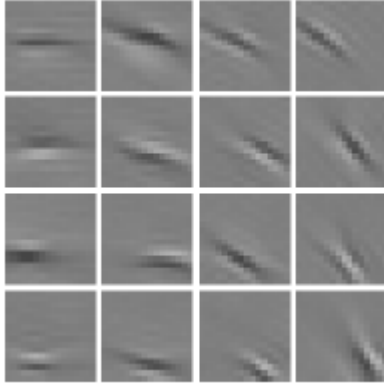


Fig. 3. – The characteristics of the basis vectors as a function of their relative position on the topographic grid.

The similarity between the neighborhoods in TICA and the receptive fields of complex cells in the visual cortex is supported by the tendency of nearby basis vectors (simple cells) in the topographic map to be of similar orientation and frequency but to have very different phases (Fig. 3).

A complete analysis of image statistics and their relation to coding in the visual system must include time. Data collected by retina have important temporal structure due to both the observer's self-motion and the movement of objects in the real world. Moreover, the neurons have significant temporal response characteristics, which are difficult or impossible to be clearly separated from their spatial characteristics. So far, the results on the relationship between the environmental statistics and neural processing models have mainly served to explain the neural functions rather than predicting aspects of sensory processing not observed yet. It is our strong belief that in near future inferences would be performed by means of statistical neuronal models conditional on maximizing some new forms of independence of its subspaces.

6. CONCLUSIONS

All variants of ICA are seeking for statistical independence of the estimates and describe the images in terms of linear superposition. Yet natural images are not formed by sums of independent components since image formation often obeys the rules of occlusion rather than addition of light coming from different objects. Analysis of statistical relationships in images reveals nonlinear dependencies across space as well as across scale and orientation [15]. Still image decomposition by various forms of ICA performs image invariant-feature extraction, which we proved to be statistically relevant in mimicking early visual processes.

The main utility of topography is visualization [14], which shows the connections between components and possibly adding some information. Classic ICA applied to natural images yields a linear decomposition into Gabor-like linear features that resemble the receptive fields of simple cells. Topographic ICA organizes image features in compliance with the defined topography. Moreover, the topographic neighborhoods resemble complex cells in their response by exhibiting phase and shift invariance [16]. In this way TICA shows simultaneous emergence of complex cell properties and topographic organization following the same principle of defining topography by simultaneous activation (or inactivation) of neighbors.

ACKNOWLEDGMENTS

R.M. is supported by a postdoc grant from the European Community, FP5 (QLG3-CT-2000-30161). M.M.V.H. is supported by research grants received from the Fund for Scientific Research (G.0185.96N), the National Lottery (Belgium) (9.0185.96), the Flemish Regional Ministry of Education (Belgium) (GOA 95/99-06; 2000/11), the Flemish Ministry for Science and Technology (VIS/98/012), and the European Community, FP5 (QLG3-CT-2000-30161 and IST-2001-32114).

REFERENCES

1. Barlow, H.B., Possible principles underlying the transformation of sensory messages. In Rosenblith, W.A. (ed.): *Sensory Communication*. MIT Press, Cambridge, MA (1961) 217-234.
2. Ruderman, D.L., Bialek, W., Statistics of natural images: Scaling in the woods. *Phys. Rev. Lett.* **73**, 6 (1994) 814-817.
3. Olshausen, B.A., Field, D.J., Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381** (1996) 607-609.
4. Hyvärinen, A., Hoyer, P.O., Inki, M., Topographic independent component analysis. *Neural Comput.* **13**, 7 (2001) 1527-1558.
5. Comon, P., Independent component analysis, A new concept? *Signal Processing* **36**, 3 (1994) 287-314.
6. Mutihac, R., van Hulle, M.M., Neural Network Implementations of Independent Component Analysis. *Proc. 2002 IEEE Workshop on Neural Networks for Signal Processing XII (NNSP 2002)*, Boulard, H., Adali, T., Bemgio S., Larsen, J., Douglas, S. (eds.), Martigny, Switzerland (2002) 505-515.
7. Simoncelli, E.P., Olshausen, B.A., Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24** (2001) 193-216.
8. van Hateren, J.H., van der Schaaf, A., Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Royal. Society ser. B*, **265** (1998) 359-366.

9. Pollen, D., Ronner S., Visual cortical neurons as localized spatial frequency filters. *IEEE Trans. on Systems, Man, and Cybernetics*, **13** (1983) 907-916.
10. Hyvärinen, A., Hoyer, P.O., Topographic independent component analysis as a model of V1 organization and receptive fields. *Neurocomput.* **38-40** (2001) 1307-1315.
11. Pham, D.-T., Garrat, P., Jutten, C., Separation of a mixture of independent sources through a maximum likelihood approach. *Proc. EUSIPCO* (1992) 771-774.
12. Cardoso, J.-F., Laheld, B.H.: Equivariant adaptive source separation. *IEEE T. Signal Proces.* **44**, 12 (1996) 3017-3030.
13. Natural Resources Conservation Service (NRCS) available at <http://photogallery.nrcs.usda.gov>,
14. Kohonen, T., *Self-Organizing Maps*. Springer-Verlag, Berlin (1995).
15. Simoncelli, E.P., Schwartz, O., Image statistics and cortical normalization models, in Kearns, M.S., Solla, S.A., Kohn, D.A. (eds.) *Advances in Neural Information Processing* **11** (1999) 153-159.
16. Hyvärinen, A., Hoyer, P.O., Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput.* **12**, 7 (2000) 1705-1720.