Image Feature Extraction Using Independent Component Analysis

Jarmo Hurri, Aapo Hyvärinen, Juha Karhunen, and Erkki Oja

Helsinki University of Technology, Laboratory of Computer and Information Science
Rakentajanaukio 2 C, FIN-02150 Espoo, FINLAND
Email: {Aapo.Hyvarinen, Juha.Karhunen, Erkki.Oja}@hut.fi

ABSTRACT

In Independent Component Analysis, one tries to model the underlying data so that in the linear expansion of the data vectors the coefficients are as independent as possible. This often leads to natural features characterizing well the data. In this paper, we present some results on applying Independent Component Analysis to image data. This has become possible by using a recently developed, computationally highly efficient fixed-point learning rule. The resulting feature masks are sensitive either to lines and edges of varying thickness or to local spatial features and frequencies.

1. INTRODUCTION

Independent Component Analysis (ICA) [1] is a recently developed statistical technique which often characterizes the data in a natural way. It can be viewed as an extension of standard Principal Component Analysis (PCA), where the coefficients of the expansion must be mutually independent (or as independent as possible) instead of being merely uncorrelated [1, 4]. This in turn implies that higher-order statistics are required for determining the ICA expansion, even though the expansion itself is linear.

To this date, the ICA model has been applied almost exclusively to blind source separation and blind deconvolution. In blind source separation, one tries to separate a few independent but unknown source signals from their linear mixtures without knowing the mixture coefficients [2, 3]. However, ICA should be applicable to a much wider class of problems, such as feature extraction, data compression, and signal analysis to mention a few [1, 3]. One obvious explanation for the lack of applications of ICA in these areas is that the expansion is much more difficult to compute than standard PCA for example. No closed-form or simple numerical solution is available, and one must resort to iterative techniques. The existing batch-type techniques [1] for estimating the ICA expansion are computationally demanding. Simpler neural techniques [2, 3] have difficulties with convergence when the dimensionality of the data is higher than about 10, and they can be applied only to certain types of data.

However, we have recently developed a fast and computationally simple fixed-point rule [5, 6] for computing the independent components. Furthermore, one can prove the convergence of this learning rule theoretically, and the algorithm can be applied to general data. These new developments make it possible to apply ICA to analyses of higher-dimensional data.

In this paper, we apply ICA to image feature extraction. For this purpose, many different methods have been proposed ranging from simple fixed masks to fairly sophisticated approaches [7, 8]. In particular, the related PCA expansion has been successfully applied to extracting 'eigenfilters' that characterize textures [9]. Besides feature extraction, PCA has several other applications in image processing, such as compression of (especially multispectral) images, image rotation etc. [7, 8]. Therefore, we expect that the related but information theoretically more meaningful ICA expansion will turn out to be a useful tool in image processing.

This paper presents our first results on this topic, showing that the basis vectors of ICA indeed represent interesting features in natural images. They describe spatial frequency information, edges with different orientation etc. The basis provided by ICA is data dependent, and it is found in a completely unsupervised manner from the original image data.

Probably the only paper that is closely related to ours is [10]. There Bell and Sejnowski use their source separation algorithm for finding ICA filters from images. However, our methods differ in several ways. We use a different separating algorithm that can be applied to more general data sets, preprocess the data differently, and consider the basis vectors of ICA rather than ICA filters.

2. INDEPENDENT COMPONENT ANALYSIS

Denote by $\mathbf{x}(t) = [x_1(t), \ldots, x_n(t)]^T$ the n-dimensional t:th data vector, in our case an image window scanned into a vector of pixel gray levels. It is assumed that the data vectors $\mathbf{x}(t)$ have some common unknown zero-mean non-Gaussian statistical distribution. In Independent Component Analysis, we try to find for the data vectors $\mathbf{x}(t)$ the expansion

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) = \sum_{i=1}^{m} s_i(t)\mathbf{a}_i + \mathbf{n}(t). \qquad (1)$$

Here the vector $\mathbf{s}(t) = [s_1(t), \ldots, s_m(t)]^T$ contains the m independent components (or source signals) $s_i(t)$ for the data vector $\mathbf{x}(t)$. $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_m]$ is a constant full-rank $n \times m$ matrix, often called the mixing matrix. The vectors $\mathbf{a}_i$, $i = 1, \ldots, m$, are the basis vectors of ICA; see [1, 3]. The additive noise term $\mathbf{n}(t)$ (describing modeling errors) is usually omitted from (1), because it is in practice impossible to separate noise from the independent components without some additional a priori information.

In the ICA model (1), the number of independent components m is often fixed in advance. In any case, $m \leq n$, and often $m = n$. The expansion (1) is determined by requiring that the coefficients $s_i(t)$, $i, \ldots, m$, are mutually independent (or in practice as independent as possible). Then the basis vectors $\mathbf{a}_i$ are generally not mutually orthogonal. This can be compared with standard PCA, where the form of the expansion is the same but the basis vectors $\mathbf{a}_i$ must be mutually orthogonal and the coefficients $s_i(t)$ have maximal variances. In many cases, the orthogonality requirement of PCA is a somewhat unnatural technical constraint, whereas the independence requirement of ICA is plausible from an information-theoretic point-of-view.

The independent components are found by determining an $m \times n$ separating matrix $\mathbf{B}$ so that the m-vector

$$\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t) \qquad (2)$$

becomes an estimate $\mathbf{y}(t) = \hat{\mathbf{s}}(t)$ of the independent component vector $\mathbf{s}(t)$. The task becomes easier, if the data vectors $\mathbf{x}(t)$ are preprocessed by whitening (sphering) them:

$$\mathbf{v}(t) = \mathbf{V}\mathbf{x}(t). \qquad (3)$$

Here $\mathbf{v}(t)$ denotes the t:th whitened vector, and $\mathbf{V}$ is an $m \times n$ whitening matrix. Whitening can be done in many ways [3]. Standard PCA is often used, because one can then simultaneously optimally compress the data vectors into an m-dimensional signal subspace (in the mean-square error sense) and filter out some of the possible noise. The PCA whitening matrix is of the form $\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}^T$, where the columns of the matrix $\mathbf{E}$ contain the PCA eigenvectors, and the diagonal matrix $\mathbf{D}$ has the corresponding eigenvalues as its elements.

After prewhitening the subsequent separating matrix, denoted here for clarity by $\mathbf{W}^T$, can be taken orthogonal: $\mathbf{W}^T\mathbf{W} = \mathbf{I}_m$. The separating equation is then

$$\mathbf{y}(t) = \mathbf{W}^T\mathbf{v}(t). \qquad (4)$$

As a separating criterion, the kurtoses $\mathrm{E}\{y_i(t)^4\} - 3[\mathrm{E}\{y_i(t)^2\}]^2$ of the components $y_i(t)$ of the vector $\mathbf{y}(t)$ are especially suitable. It can be shown that the independent components are found from the local maxima of the modulus of the kurtosis for prewhitened data. The fixed-point algorithms introduced in [5, 6] can be used for finding these local maxima very efficiently.

3. METHODS

The n-dimensional data vectors $\mathbf{x}(t)$ were obtained by first taking $n^{1/2} \times n^{1/2}$ sample subimages from the available image database. After this, the mean of its elements was subtracted from each raw data vector $\mathbf{x}(t)$, which was then normalized to unit length. This preprocessing makes the data vectors approximately zero mean also in the standard statistical sense. The preprocessed vectors were then whitened using standard PCA so that the resulting vectors $\mathbf{v}(t)$ had $n-1$ components (one of the components becomes insignificant because of the subtracted mean).

After this, the generalized fixed-point algorithm described in full detail in [6] was applied to finding the independent components of the whitened data vectors $\mathbf{v}(t)$. In this algorithm, one first chooses some initial values for the columns $\mathbf{w}_i$ $(i = 1, \ldots, m)$ of the matrix $\mathbf{W}$ (rows of the orthogonal separating matrix $\mathbf{W}^T$). The key step in the generalized fixed-point algorithm is to compute a new $(k+1)$:th estimate for $\mathbf{w}_i$ using the iteration

$$\mathbf{w}_i^*(k+1) = \mathrm{E}\{\mathbf{v}g(\mathbf{w}_i(k)^T\mathbf{v}) - g'(\mathbf{w}_i(k)^T\mathbf{v})\mathbf{w}_i(k)\}, \qquad (5)$$

$$\mathbf{w}_i(k+1) = \mathbf{w}_i^*(k+1) / \parallel \mathbf{w}_i^*(k+1) \parallel . \qquad (6)$$

Here E denotes the mathematical expectation. In practice it is replaced by sample mean computed using a large number of vectors $\mathbf{v}(t)$. The function $g(u)$ can be any odd, sufficiently regular nonlinear function, and $g'(u)$ denotes its derivative. The choice $g(u) = u^3$ directly maximizes the kurtosis criterion. In practice, it is often advisable to use a robust nonlinearity that grows less than linearly; a typical choice is $g(u) = \tanh(u)$. This also has a relationship to the kurtosis criterion. For preventing the vectors $\mathbf{w}_i$, $i = 1, \ldots, m$, from converging to the same directions, they are orthogonalized against each other. This can be done either sequentially by using a deflation type procedure, or symmetrically [5, 6].

It can be proven [6] that $\mathbf{w}_i(k)$ converges (up to the sign) to one of the columns of the matrix $\mathbf{W}$ under very mild conditions. The convergence of the fixed-point algorithms is cubic, and our experiments show that usually less than 10 iterations provide sufficiently accurate estimates. This means that the fixed-point algorithms are very fast when compared with typical gradient-based adaptive blind separation algorithms. Another advantage is that they don't require any learning parameters. The fixed-point algorithm is also much simpler than the currently best known batch algorithm introduced in [1].

From $\mathbf{w}_i$ one can compute the estimate for the corresponding basis vector $\mathbf{a}_i$ of ICA [2] using the formula

$$\hat{\mathbf{a}}_i = \mathbf{E}\mathbf{D}^{1/2}\mathbf{w}_i. \tag{7}$$

4. EXPERIMENTAL RESULTS

Our raw data consisted of 15 different images representing various natural objects or scenes (river valley, harbour, forest, frog etc.). The size of subimages was 12×12, yielding thus 143-dimensional whitened vectors $\mathbf{v}(t)$ and 143 estimated ICA basis vectors $\hat{\mathbf{a}}_i$. The subimages were taken randomly from much larger original images, and their total number was 10000.

Figures 1 and 2 show examples of typical results obtained thus far. The subimages show the 143 estimated basis vectors of ICA. Each of them is represented again as a 12×12 subimage. The subimages have been normalized so that in each of them the average gray level is 127, and the gray level range has been linearly expanded to maximum possible in the used interval $[0, 255]$. These subimages can be regarded as features describing the most significant characteristics of the analyzed image data.

Figure 1 shows the estimated ICA basis vectors when the nonlinear function in the generalized fixed-point algorithm was the sigmoid $g(u) = \tanh(u)$. It can be seen that most of the ICA basis vectors correspond to wavelet type filters that are sensitive to local features and spatial frequencies in the images. However, a part of the estimated basis vectors of ICA yield filters that are sensitive to edges and lines of varying thickness in different orientations. When different initial values were used in the fixed-point algorithm, the results were qualitatively fairly similar. However, the estimated basis vectors of ICA were somewhat different when inspected more closely. This suggests that the total number of "independent" features in the analyzed images may be larger than the dimensionality of subimages.

Figure 2 depicts the results of a similar experiment where the nonlinearity $g(u) = u^3$ was used in the generalized fixed-point iteration. Now many of the basis vectors correspond to high global spatial frequencies while others pick up some very local features. An explanation of the qualitatively different results compared with Figure 1 is that the fast growing nonlinearity $g(u) = u^3$ is sensitive to large values of u, which determine the results almost solely. This effect can be reduced by normalizing the local variances in the images, and by estimating only the most significant basis vectors of ICA, which are found by compressing the data into a lower dimensional PCA subspace. After such processing steps, the estimated basis vectors of ICA change so that they resemble more the basis vectors of Figure 1.

5. CONCLUSIONS

In this work we have applied Independent Component Analysis to natural images. This is possible using a recently developed, fast and efficient fixed-point iteration. The resulting filters are partly wavelet type, describing local spatial features and frequencies. Some of the filters describe edges and lines of varying thickness and orientation. These filters emerge in a completely unsupervised manner from the raw image data.

REFERENCES

[1] P. Comon, "Independent component analysis - a new concept?," *Signal Processing*, vol. 36, pp. 287-314, 1994.

[2] E. Oja and J. Karhunen, "Signal separation by nonlinear Hebbian learning," in *Computational Intelligence - A Dynamic System Perspective*, M. Palaniswami et al. (Eds.). New York: IEEE Press, 1995, pp. 83-97 (Invited paper).

[3] J. Karhunen, "Neural approaches to independent component analysis and source separation," in *Proc. 4th European Symp. on Artificial Neural Networks (ESANN'96)*, Bruges, Belgium, April 1996, pp. 249-266 (Invited paper).

[4] J. Karhunen and J. Joutsensalo, "Generalizations of principal component analysis, optimization problems, and neural networks," *Neural Networks*, vol. 8, no. 4, pp. 549-562, 1995.

[5] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis." Helsinki Univ. of Technology, Lab. of Computer and Information Science, Report A35, 1996. Submitted to a journal.

[6] A. Hyvärinen, "A family of fixed-point algorithms for independent component analysis." Summary submitted to *ICASSP97*.

[7] A. Jain, *Fundamentals of Digital Image Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1989.

[8] R. Gonzalez and R. Woods, *Digital Image Processing.* Reading, MA: Addison-Wesley, 1992.

[9] F. Ade, "Characterization of textures as "eigenfilters"," *Signal Processing*, vol. 5, pp. 451-457, 1983.

[10] A. Bell and T. Sejnowski, "The 'independent components' of natural scenes are edge filters," submitted to *Vision Research*, 1996.

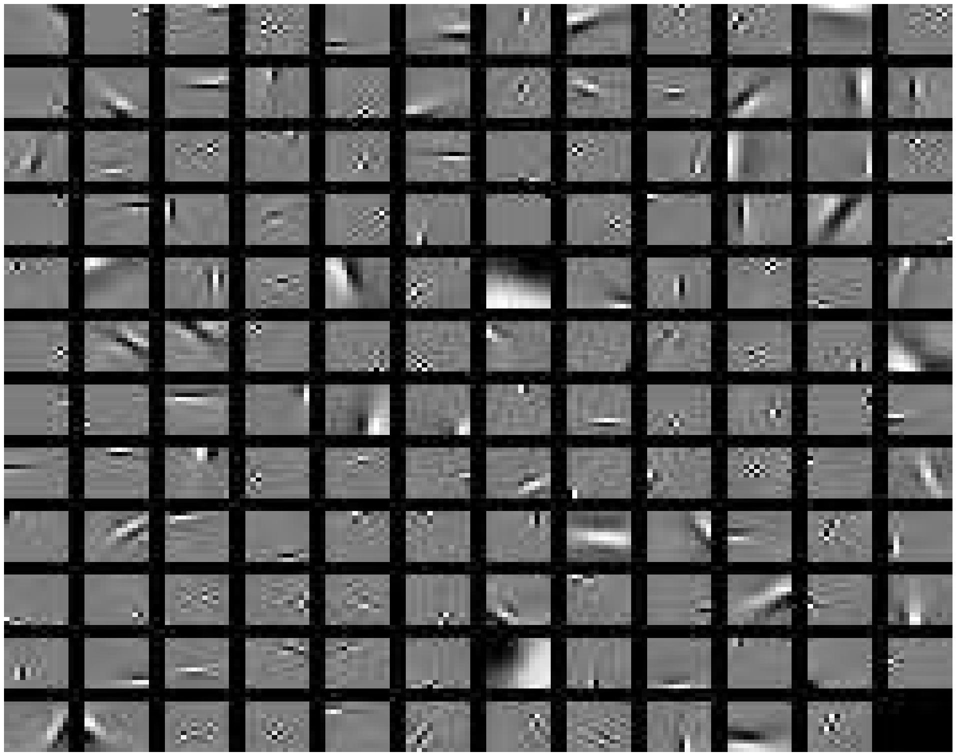

Fig. 1. *ICA basis vectors obtained using the* tanh *nonlinearity.*

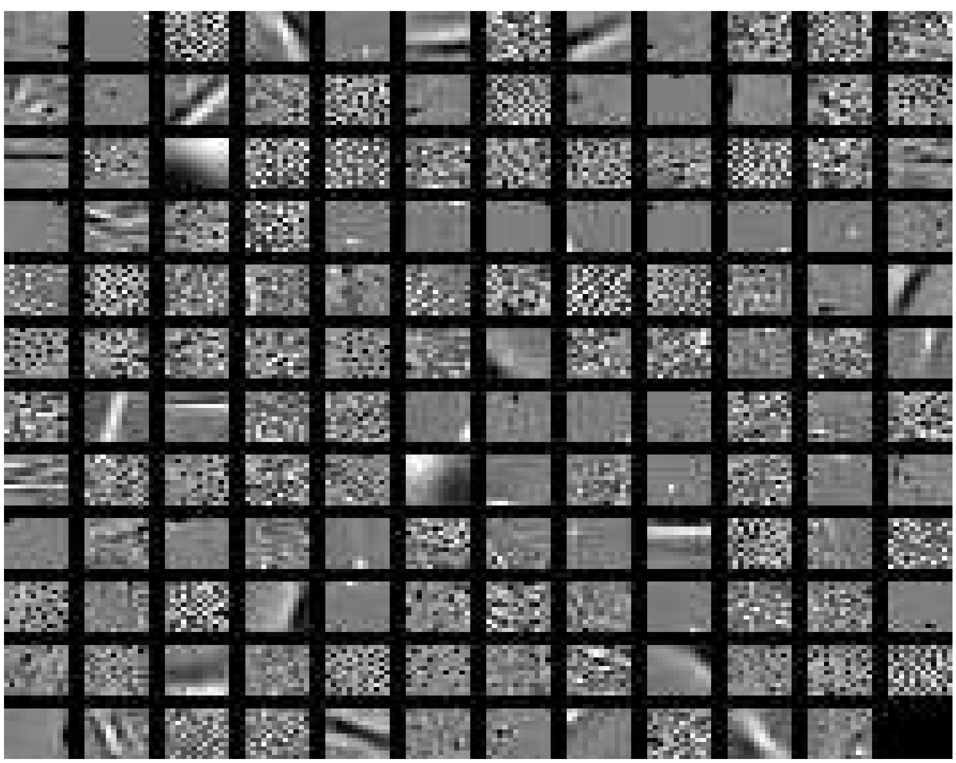

Fig. 2. *ICA basis vectors obtained using cubic nonlinearity.*