IEEE MELECON 2006, May 16-19, Benalmádena (Málaga), Spain

Image processing using ICA: a new perspective

Rubén Martín-Clemente and Susana Hornillo-Mellado
Dpto. de Teoría de la Señal y Comunicaciones
Escuela Superior de Ingenieros
University of Seville
Seville, Spain
Email: ruben@us.es, susanah@us.es

***Abstract*— Independent Component Analysis (ICA) provides a sparse representation of natural images in terms of a set of oriented bases. So far, the interest on this result lay on its apparent connection to the neural processing of the mammalian primary visual cortex. In this paper we provide an analysis from a formal (not physiological) point of view. We show that ICA of a natural image is equivalent to filtering the image using a high-pass filter, followed by a sampling. This result determines, on the one hand, the sparse distribution of the independent components and, on the other hand, that the image bases resemble "edges" of the original image. Some experiments are included to illustrate the theoretical conclusions.**

I. Introduction

Independent Component Analysis (ICA) is a technique for studying multivariate data that has received great interest in the last few years [4], [5], [9]. It is succinctly as follows: let the observed multidimensional data be represented as a matrix $\mathbf{X}$ with N rows and T columns, where N is the number of variables and T is the number of observations recorded on each variable. The goal of ICA is to calculate the square matrix $\mathbf{B}$ that linearly transforms $\mathbf{X}$ into a matrix

$$\mathbf{Y} = \mathbf{B}\,\mathbf{X} \tag{1}$$

containing new variables (the "independent components") that are as independent "as possible", in the sense of maximizing or minimizing some "measure of independence".

The interest on the application of ICA to natural images started from the results presented by Bell and Sejnowski in [2], that greatly resemble the behaviour of some neurons of the mammalian primary visual cortex, the *simple cells* [1], [6], [11]. In particular, the columns of matrix $\mathbf{A} = \mathbf{B}^{-1}$, in this context known as *ICA bases*, represent oriented and localized structures, very similar to those of the receptive fields of simple cells. Besides, the independent components obtained are *sparsely distributed*, like the observed responses of these simple cells[1].

Many authors have suggested other approaches or explanations to this connection between ICA and the human visual system (see, for example, [3], [10], [8], [15]), but no one has provided a mathematical analysis to explain the results obtained by Bell and Sejnowski. This contribution is intended to fill this gap from a purely formal (not physiological) approach.

There are two main results presented in this paper. First, we show that the so-called *ICA filters* (i.e., the rows of matrix $\mathbf{B}$) can be expressed as a weighted sum of the eigenvectors of the data correlation matrix, with the weighting coefficients dependent on the corresponding eigenvalues. In the case of natural images, these ICA filters are mainly high-pass.

Secondly, we show that each independent component can be obtained by filtering the whole image using a rotated version of the corresponding ICA filter, followed by a sampling of the result. Considering the high-pass nature of the filters used, this new approach to ICA explains the results obtained by other authors.

II. Notation and preprocessing

Throughout the paper, x_{ij} will stand for the (i,j)th entry of matrix $\mathbf{X}$. Using a `Matlab`-like notation, the kth row of $\mathbf{X}$ will be denoted as $\mathbf{x}_{k:} = [x_{k1}, \ldots, x_{kT}]$, and its lth column will be $\mathbf{x}_{:l} = [x_{1l}, \ldots, x_{Nl}]^{\dagger}$ ($\dagger$ means "transpose"). *The same conventions hold for all the other matrices*. For simplicity it is supposed that the mean value has been subtracted from each row so that $\frac{1}{T}\sum_{n=1}^{T} x_{kn} = 0\ \forall k$.

For convenience, the data are usually transformed into uncorrelated variables by means of a *whitening* matrix, $\mathbf{W}$:

$$\bar{\mathbf{X}} = \mathbf{W}\,\mathbf{X} = \mathbf{D}^{-1/2}\,\mathbf{V}^{\dagger}\,\mathbf{X} \tag{2}$$

where $\mathbf{D}^{-1/2} = \mathrm{diag}(\lambda_1^{-1/2}, ..., \lambda_N^{-1/2})$, being $\lambda_1 \geq \ldots \geq \lambda_N$ the eigenvalues of the sample data correlation matrix, $\mathbf{R}_x = \frac{1}{T}\mathbf{X}\,\mathbf{X}^{\dagger}$; $\mathbf{V} = (\mathbf{v}_1|\ldots|\mathbf{v}_N)$ is the matrix containing, by columns, the corresponding eigenvectors ($\mathbf{R}_x\,\mathbf{v}_j = \lambda_j\,\mathbf{v}_j$). The ICA model (1) can be now written as

$$\mathbf{Y} = \bar{\mathbf{B}}\,\bar{\mathbf{X}} \tag{3}$$

where $\bar{\mathbf{B}} = \mathbf{B}\,\mathbf{W}^{-1}$. It is straightforward to show that matrix $\bar{\mathbf{B}}$ is orthogonal, i.e., $\bar{\mathbf{B}}\bar{\mathbf{B}}^{\dagger} = \mathbf{I}$ ($\mathbf{I}$ is the identity matrix).

III. FastICA revisited

In this paper we will only focus on the independent components obtained by the popular algorithm `FastICA` [9], based on the maximization of the kurtosis [8]. We have two reasons for this decision: 1.) the kurtosis is mathematically tractable and 2.) it has been proved that most ICA methods are essentially equivalent in absence of noise and outliers [8]:

[1]Similar results were found by Olshausen and Field in [13] by sparse coding.

consequently, analyzing FastICA is as valid as analyzing any other method. We first analyze the problem in which only the first independent component is calculated. The extension to the rest of independent components is discussed later on.

Let $\mathbf{y}_{1:} = [y_{11}, \ldots, y_{1T}]$ be the first independent component and define $\mathbf{y}^3_{1:} \stackrel{def}{=} [y^3_{11}, \ldots, y^3_{1T}]$. With this notation, the basic FastICA iteration can be expressed as follows:

1) $\mathbf{y}_{1:} \longleftarrow \bar{\mathbf{b}}_{1:}\,\bar{\mathbf{X}}$
2) $\bar{\mathbf{b}}_{1:} \longleftarrow \frac{1}{T}\,\mathbf{y}^3_{1:}\,\bar{\mathbf{X}}^\dagger - 3\,\bar{\mathbf{b}}_{1:}$
3) $\bar{\mathbf{b}}_{1:} \longleftarrow \bar{\mathbf{b}}_{1:}/\|\bar{\mathbf{b}}_{1:}\|$

These three steps are repeated until convergence ($\longleftarrow$ means "update"). The update logically stops when

$$\bar{\mathbf{b}}_{1:} \propto \mathbf{y}^3_{1:}\,\bar{\mathbf{X}}^\dagger \tag{4}$$

where "$\propto$" means "proportional to". Post-multiplying both sides with the whitening matrix $\mathbf{W}$, and using that $\mathbf{b}_{1:} = \bar{\mathbf{b}}_{1:}\,\mathbf{W}$, we get

$$\bar{\mathbf{b}}_{1:}\,\mathbf{W} \propto \mathbf{y}^3_{1:}\,\bar{\mathbf{X}}^\dagger\,\mathbf{W} \quad \Rightarrow \quad \mathbf{b}_{1:} \propto \mathbf{y}^3_{1:}\,\bar{\mathbf{X}}^\dagger\,\mathbf{W} \tag{5}$$

Using the definition of $\mathbf{W}$ (see (2)) in (5) we get:

$$\mathbf{b}_{1:} \propto \mathbf{y}^3_{1:}\,\bar{\mathbf{X}}^\dagger\,\mathbf{D}^{-1/2}\,\mathbf{V}^\dagger \tag{6}$$

from which we get, after some algebra, the essential relation:

$$\mathbf{b}^\dagger_{1:} \propto \sum_{n=1}^{N} \gamma_n \mathbf{v}_n \tag{7}$$

where $\mathbf{v}_n$ is the nth eigenvector of the data correlation matrix $\mathbf{R}_x$ and

$$\gamma_n = \frac{\sum_{k=1}^{T} \bar{x}_{nk}\, y^3_{1k}}{\sqrt{\lambda_n}} \stackrel{(a)}{\propto} \frac{\bar{b}_{1n}}{\sqrt{\lambda_n}} \tag{8}$$

Here, $\bar{b}_{1n}$ is the $(1,n)$th entry of $\bar{\mathbf{B}}$ and (a) follows from (4).

Remark.- Eqn. (7) indicates that the first ICA filter can be written as a linear combination of the eigenvectors of the data correlation matrix. This is not surprising *a priori*, since these eigenvectors always form an orthogonal basis of the space. Observe that *the smaller the coefficient γ_n the stronger the corresponding eigenvector is expected to be present*. In the following sections we will explore this matter when the observed data is obtained from a natural image.

IV. The eigenvalues and eigenvectors of the correlation matrix of a natural image

From now on, let us consider that the data matrix $\mathbf{X}$ is obtained from a natural image as follows: we divide the image into $\sqrt{N} \times \sqrt{N}$ patches and stack their pixels into $N \times 1$ vectors that will be the columns of $\mathbf{X}$ (i.e., the kth column of $\mathbf{X}$ corresponds to the kth patch of the image). In this case, it is well-known that, due to the great correlation between neighbouring pixels, most of the eigenvalues of the data correlation matrix, $\mathbf{R}_x$, are negligible [12]. This is the basis of the image compression. To illustrate it, let us consider the Karhunen-Loeve Transform (KLT), that provides a representation of each image patch (each column of $\mathbf{X}$) in terms of the eigenvectors of the data correlation matrix [12]:

$$\mathbf{x}_{:k} = \mathbf{V}\,\mathbf{s}_k\,, \quad k = 1, ..., T \tag{9}$$

where $\mathbf{V}$ is the matrix of eigenvectors of $\mathbf{R}_x$ and the vector $\mathbf{s}_k = [s_{1k}, ..., s_{Nk}]^\dagger$ contains the coefficients of the transformation. The KLT is characterized because the approximation of $\mathbf{x}_{:k}$ given by the $N \times 1$ vector $\tilde{\mathbf{s}}_k = [s_{1k}, ..., s_{rk}, 0, ..., 0]^\dagger$ $(r < N)$

$$\tilde{\mathbf{x}}_{:k} = \mathbf{V}\,\tilde{\mathbf{s}}_k \tag{10}$$

has minimum mean square error, and it is given by $E_r = \sum_{n=r}^{N} \lambda_n$ [12]. As the frequency content of a natural image is mainly low-pass, the eigenvectors associated to the greatest eigenvalues correspond mainly to the lower frequency content. As more eigenvectors are considered in (10), more high-pass content (more *details*) is included.

Since the smallest eigenvalues are negligible compared to the greatest ones and considering that $\|\bar{\mathbf{b}}_{1:}\| = 1$ (see step 3) of the basic FastICA iteration given in Section III), a conclusion emerges for γ_n (8): only those γ_n corresponding to the smallest eigenvalues will take significant values, leading to the approximation

$$\mathbf{b}^\dagger_{1:} \propto \sum_{n=m}^{N} \gamma_n \mathbf{v}_n\,, \quad 1 << m \leq N \tag{11}$$

This means that *the first ICA filter $\mathbf{b}_{1:}$ can be approximated by the weighted sum of the eigenvectors associated to the smallest eigenvalues of the data correlation matrix. Since they correspond to the high-frequency content of the image, the first ICA filter will have high-pass characteristics*. To see the implications of this fundamental result, we need to introduce the following innovative approach to ICA.

V. ICA as a 2-D "filtering-sampling" process

Knowing the matrix $\mathbf{B}$, we can obtain the independent components as stated in (1). In this section we propose a technique to obtain these independent components *from the original image* (not from the matrix $\mathbf{X}$), consisting of a 2-D filtering followed by a sampling. This new interpretation of the independent component analysis of a natural image seems to be an unnecessary complication, but it is of great relevance for our analysis.

For convenience, we represent the kth image patch by the 2-D sequence, $x_k(n_1, n_2)$, with $n_1, n_2 = 0, ..., \sqrt{N} - 1$ (i.e. the mapping of the $N \times 1$ vector $\mathbf{x}_{:k}$ to an image). Similarly, $b_1(n_1, n_2)$ represents the first ICA filter, so that

$$y_{1k} = \sum_{i=1}^{N} b_{1i}\, x_{ik} = \sum_{n_1=0}^{\sqrt{N}-1} \sum_{n_2=0}^{\sqrt{N}-1} b_1(n_1, n_2)\, x_k(n_1, n_2) \tag{12}$$

The most-right part of this identity clearly reminds *a convolution*. To state it more properly, let us define the 2-D sequence $b^R_1(n_1, n_2)$ as $b_1(n_1, n_2)$ rotated 180° counterclockwise:

$$b^R_1(n_1, n_2) = b_1(\sqrt{N} - 1 - n_1, \sqrt{N} - 1 - n_2) \tag{13}$$

After some cumbersome algebra one finds that

$$y_{pk} = z(\sqrt{N} - 1, \sqrt{N} - 1) \tag{14}$$

where $z(n_1, n_2)$ is the 2-D convolution between the patch $x_k(n_1, n_2)$ and $b_1^R(n_1, n_2)$.

In other words, each element of $\mathbf{y}_{1:}$ is the filtering of an image patch with the corresponding ICA filter rotated $180°$ counterclockwise, followed by the sampling of the filter output at $n_1 = \sqrt{N} - 1, n_2 = \sqrt{N} - 1$. It is straightforward to show that an equivalent result will be obtained by filtering the whole image (rather than filtering isolated patches).

Finally, observe that the magnitude responses of the filters $b_1^R(n_1, n_2)$ and $b_1(n_1, n_2)$ are the same because the rotation only affects their phase response. In particular, if $b_1(n_1, n_2)$ is a high-pass filter then $b_1^R(n_1, n_2)$ will be also high-pass. In conclusion: *ICA is equivalent to a high-pass filtering and a sampling of the image*.

VI. Extension to several independent component

In practice, the `FastICA` algorithm is executed as many times as the number of desired independent components [9]. In the jth iteration it is imposed that the corresponding $\bar{\mathbf{b}}_{j:}$ has to be orthogonal to the $\bar{\mathbf{b}}_{k:}$, $k = 1, ..., j-1$ previously obtained. The jth independent component is then obtained as $\mathbf{y}_{j:} = \bar{\mathbf{b}}_{j:}\bar{\mathbf{X}} = \mathbf{b}_{j:}\mathbf{X}$. This procedure introduces a new constraint (i.e. orthogonality), but (7) and (8) are still satisfied so that the general conclusions in the preceding Sections still hold.

VII. Experiments and Discussion

Consider the natural, grey-scale image shown in Fig. 1. We divide it into 12×12 patches to compose the data matrix $\mathbf{X}$ (144×2688).

Fig. 1. The "monarch" image (504×768) (available in `http://links.uwaterloo.ca/colorset.base.html`).

Firstly, we analyze the frequency content of the ICA filters, obtained by mapping the rows of $\mathbf{B}$ into 12×12 patches. In Fig. 2 we show the magnitudes of the 2-D Fourier Transforms corresponding to some of the ICA filters. As predicted, all of them are high-pass.

In Fig. 3 we show the filtered images obtained with the first and the last ICA filter. Fig. 4 represents the independent components (i.e., the rows of matrix $\mathbf{Y}$) corresponding to the ICA filters of Fig. 2. They are clearly sparse, which is in agreement with the results obtained by other authors [7], [8].

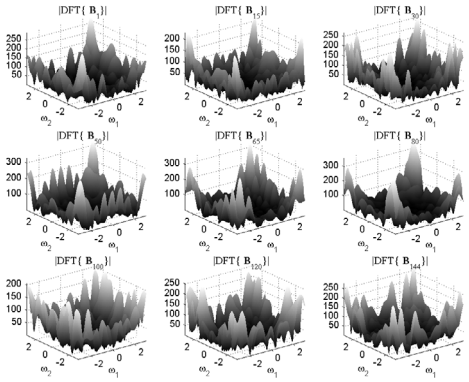

Fig. 2. Some of the 12×12 ICA filters (1, 15, 30, 50, 65, 80, 100, 120 and 144) obtained for the "monarch" image.

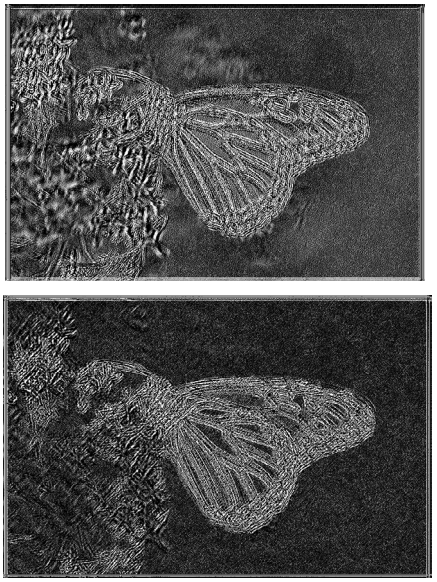

Fig. 3. The "monarch" image filtered with the first and the last "ICA filter".

This can be now explained as follows: our previous derivation indicates that ICA mainly performs a high-pass filtering of the image data. Carrying out a high-pass filtering of a natural image, only the edges are enhanced whereas most of the image is attenuated or even eliminated[2]. So, when we sample the filter output it is expected that most samples of the filtered image are small: only those samples at the edges will take significative values and the resulting independent component will be, consequently, "sparse". Furthermore, observe that each independent component has *one* very large element. This is expected because such a solution maximizes the kurtosis, and we can not forget that this is the objective of the FastICA algorithm [8][3].

[2]Consequently, the filtered image will have a very sparse distribution, characterized by a sharp histogram centered around zero

[3]It is easy to show that the unity-variance signal with maximum kurtosis is the signal in which all its elements vanish excepting one.

It also follows from above that the "large values" of the independent components are not randomly placed: they correspond to "edges" of the image. From (1) we obtain that each image patch $\mathbf{x}_{:k}$ can be decomposed in terms of a set of basic building blocks or *ICA bases*:

$$\mathbf{x}_{:k} = \sum_{j=1}^{N} y_{jk}\, \mathbf{a}_{:j} \tag{15}$$

where $\mathbf{a}_{:j}$ is the jth column of matrix $\mathbf{A} = \mathbf{B}^{-1}$. As we said, in each independent component $\mathbf{y}_{j:}$, $j = 1, ..., N$, there is one dominant value, say y_{jd_j}. The index d_j is different from one independent component to another, due to the orthogonal relation among the ICA filters, meaning that:

$$\mathbf{x}_{:d_j} \simeq y_{jd_j}\, \mathbf{a}_{:j} \tag{16}$$

That is, each ICA basis, $\mathbf{a}_{:j}$ is approximately proportional to an image patch $\mathbf{x}_{:d_j}$ that correspond to an "edge". Several authors had reported that the ICA basis resemble "edges" [7], [8] but, to the best of our knowledge, no purely mathematical arguments had been proposed to explain it so far (even though that the explanation is really simple, as we have shown). To illustrate this point, in Fig. 5 we show the ICA bases corresponding to the "monarch" image. Most of them are like patches of the image corresponding to the edges, due to the dominant element present in each independent component.

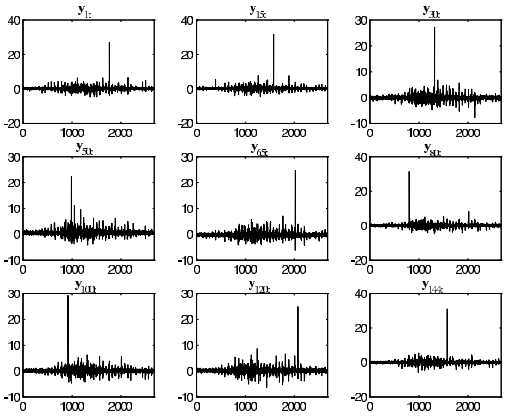

Fig. 4. Some of the independent components corresponding to the "monarch" image.

VIII. Conclusions

We have analyzed the results obtained by performing Independent Component Analysis (ICA) to natural images. In particular, we have focused on the solutions given by the popular algorithm FastICA, which is based on the maximization of the kurtosis. It is shown that the ICA filters that generate the independent components have mainly high-pass characteristics. It has been proven that this property explains the fact that only a small portion of the elements of each independent component take significant values that correspond to the "edges" of the image. To the best of our knowledge, only physiological arguments had been given to explain this matter so far. It has been proven also that the presence of one dominant value, different for each independent component, is responsible of the fact that the ICA basis are similar to the edges of the original image. The latter is a new result, unpublished in the existing literature: so far, it had been shown in previous works only that the ICA bases resemble "edges" and no relation with the image had been put in manifest.

Fig. 5. "ICA bases" (12×12) corresponding to the "monarch" image.

References

[1] Barlow, H. B.: *Sensory communication. Possible principles underlying the transformation of sensory messages*. MIT Press (1961) 217–234

[2] Bell, A. J., Sejnowski, T. J.: *The independent component of natural images are edge filters*. Vision Research, vol. 37, no. 23 (1997) 3327–3338

[3] Caywood, M., Willmore, B., Tolhurst, D.: *Independent Components of Color Natural Scenes Resemble V1 Neurons in Their Spatial and Color Tuning*. Journal of Neurophysiology, vol. 91 (2004) 2859–2873

[4] Cichocki, A., Amari, S. I.: *Adaptive blind signal and image processing*. John Wiley & Sons (2002)

[5] Comon, P.: *Independent component analysis, a new concept?*. Signal Processing, vol. 36, no. 3 (1994) 287–314

[6] Field, D. J.: *Relations between the statistics of natural images and the response properties of cortical cells*. Journal of the Optical Society of America A, vol. 4 (1987) 2379–2394

[7] A. Hyvärinen, M. Gutmann and P.O. Hoyer. *Statistical model of natural stimuli predicts edge-like pooling of spatial frequency channels in V2*. BMC Neuroscience, (2005) 6–12

[8] Hyvärinen, A., Karhunen, J., Oja, E.: *Independent component analysis*. John Wiley & Sons (2001)

[9] Hyvärinen, A., Oja, E.: *A fast fixed-point algorithm for independent component analysis*. Neural Computation, vol. 6 (1997) 1484–1492

[10] Hyvärinen, A., Hoyer, P. O.: *A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images*. Vision Research 41 (2001) 2413–2423

[11] Hubel, D. H., Wiesel T. N.: *Receptive fields, binocular interaction and functional architecture in cat's visual cortex*. Journal of Physiology, vol. 160 (1962) 106–154

[12] Jain, A.: *Fundamentals of Digital Image Processing*. Prentice Hall (1989)

[13] Olshausen, B. A., Field, D. J.: *Sparse coding with an overcomplete basis set: a strategy employed by V1?* Vision Research, vol. 37 (23) (1997) 3311–3325

[14] Olshausen, B. A.: *Principles of image representation in visual cortex*. The Visual Neurosciences, L.M. Chalupa, J.S. Werner, Eds. MIT Press (2003) 1603–1615

[15] van Hateren, J. H., van der Schaaf, A.: *Independent component filters of natural images compared with simple cells in primary visual cortex*. Proc. Roy. Soc. Lond. B 265 (1998) 359–366