Sparse-Dispersed Coding and Images Discrimination with Independent Component Analysis

Hervé Le Borgne, Anne Guérin-Dugué[1]
Laboratory of Images and Signals
LIS – INPG, 46 avenue Félix Viallet, F-38031 Grenoble cedex
{hleborgn, anne.guerin}@lis.inpg.fr

ABSTRACT

Independent Component Analysis applied to a set of natural images provides band-pass-oriented filters, similar to simple cells of the primary visual cortex. We applied two types of pre-processing to the images, a low-pass and a whitening one in a multiresolution grid, and examine the properties of the detectors extracted by ICA. These detectors composed a new basis function set in which images are encoded. On one hand, the properties (sparseness and dispersal) of the resulting coding are compared for both pre-processing strategies. On the other hand, this new coding by independent features is used for discriminating natural images, that is a very challenging domain in image analysis and retrieval. We show that a criterion based on the dispersal property enhances the efficiency of the discrimination by selecting the most dispersed detectors coding the image database. This behaviour is well enhanced with whitened images.

1. INTRODUCTION

Properties of visual receptive fields in mammalian primary visual cortex has been extensively studied (e.g. [3, 4, 8]). From these works, most of the theories of sensory coding have proposed models to effective internal representation by redundancy reduction [1, 5, 10, 15]. Among systems that produce such effects, Independent Components Analysis provides Gabor-like detectors, similar to simple cells of the primary visual cortex, whose activities are statistically independent. The resulting coding has two main properties (sparseness and dispersal), describing how this new basis set encodes images.

On one hand, the models have created great interest, suggesting that the underlying statistical principles may be the same as those determining the structure of the cortical visual coding. On the other hand, our own daily experience, and psychophysic experiments show that image recognition and understanding is a very fast and robust process. More precisely, considering a task of scene categorisation, psychological experiments show that human subjects rapidly capture the context of the scene before recognising its individual parts [13]. Then, it is interesting to gather these two research domains to evaluate the potentiality of the independent features for a discrimination task among natural images, considering that the detectors are automatically built from the images in an unsupervised way.

In the following, we explain the methodology for learning independent components from images (section 2), the categorisation task among four semantic contexts of natural images (section 2.1), and the pre-processing strategies (section 2.2). The resulting features are analysed as 2D FIR filters by a Gabor-like modelling (section 3) to analyse the relation between the spectral properties of the images and of the detectors coming from them. A quantitative appreciation of the sparseness and dispersal of the resulting coding is also compared for the two pre-processing conditions (section 4), allowing an effective selection of filters for the discrimination task (section 5) between four semantic contexts of natural images.

2. LEARNING INDEPENDENT COMPONENTS FROM IMAGES

2.1. Database

The image database used for this study consists on a collection of 200 natural images extracted from the COREL database. In average, the amplitude spectrum of natural images falls with the spatial radial frequency as $1/f^{\alpha}$, with a falloff factor α between 0.9 and 1.2. See for example [17]. This factor can be distinguished according to the orientation of spatial frequencies. Considering its variation versus orientation, different shapes of amplitude spectra can be considered corresponding to different semantic categories [14]. Here, we considered four categories of 50 images, urban scenes, indoor scenes, closed landscapes (mountains, valleys, forests…) and opened landscapes (fields, beaches, deserts…).

2.2. Image Pre-processing Strategies

In order to compare the features extracted by ICA, we have implemented two multiresolution (3 levels) pyramids. The first one is a low-pass pyramid based on a 6th order lowpass Butterworth filter with a cutoff frequency (0.4), and the second one is a band-pass-

---

[1] Actually, at INRIA, IS2 project, 655 av. de l'Europe, F-38330 Montbonot Saint Martin

whitening pyramid. The whitening filter has been implemented according to a biological model of the retina of vertebrates [7]. It realises a non-linear processing as illustrated on figure 1.
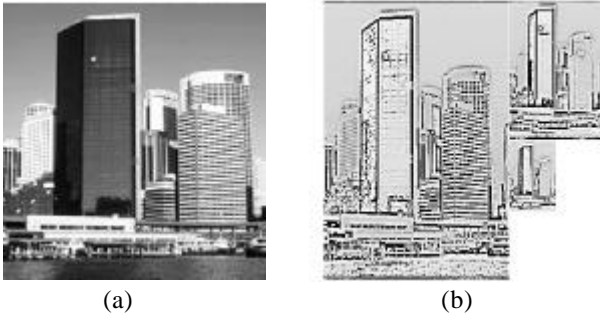

(a)   (b)

Figure 1. (a) Original image, (b) whitened image at the three resolutions

2.3 Principle

Each category in each pyramid at each resolution is separately processed in the same way (4 x 2 x 3 experiments). Then, for each pre-processed images set, a collection of patches (32 x 32) is extracted at random (30 patches per image). Each experiment is then realised with a set of 1500 patches. For one image, the random locations sequence is identical whatever the pre-processing, and different for another image. In order to minimise the anisotropy on horizontal and vertical orientations, each patch is focalised by a weighting Hamming window. Before ICA, a PCA realises a data whitening and a dimension reduction from 1024 to 50 dimensions. 90% of the total inertia is then preserved with the low-pass pre-processing, but only 65% with the whitening one. We use the "Fast-ICA" algorithm because of its fast convergence time [9]. For each experiment, 50 primitives are extracted $\{\phi_i(x,y), i=1..50\}$, assuming that each patch $I(x,y)$ is an independent combination of this set of primitives. The primitives represent the spatial patterns occurring in the different scenes such as the projection on this basis involves independent codes $\{a_i, i=1..50\}$:

$$I(x,y) = \sum_{i=1}^{50} a_i \mathbf{f}_i(x,y) \qquad (1)$$

In the following, the primitives $\phi_j(x,y)$ are considered as 2D FIR filters.

3. SIMPLE CHARACTERISATION OF «ICA FILTERS»

3.1 Principle

According to this methodology, we obtain 2x3x4 sets of primitives $\{\phi_i(x,y), i=1..50\}$ which can be compared as the spatial locations of the patches (random sequence on the image database) are identical for the two pre-processing techniques in a 2D-multiresolution grid. According to prior studies (for example [2, 11, 18]), the «ICA filters»

are, in average, close to oriented band-pass filters. In order to analyse and compare them, each filter is matched with the closest 2D Gabor function, providing thus 4 parameters characterising the filter (two frequency locations -u_0, v_0-, two spatial standard deviations -σ_x, σ_y-). More precisely, the «ICA filters» may be organised into three types, (i) oriented band-pass filters (Gabor-like filters), (ii) rather anisotropic band-pass filters, (iii) combination of two orthogonal oriented band-pass filters. Indeed, in the case (i), the modelling is well suited the filters. For the others, this modelling will extract the most salient oriented pattern from the filters. See figure 2.

	Spatial response	Fourier	Gabor
(i)			
(ii)			
(iii)			

figure 2. Example of the three types of ICA filters.

3.2 Results

In average, the filters are very similar for the two pre-processing conditions and at the three resolutions (figures 3 and 4). In fact, with the whitening filter, the spatial contrast is enhanced, and the energy is more uniformly distributed over the scales, but the spatial organisation of the local variances remains the same compared to the low-pass pre-processing. Consequently, differences between these two pre-processing techniques will then occur on the responses to the images (see § 4.). At the different resolutions, we observe also similar filters. The scale invariance in natural images is a very robust property [16].

For each image category, figure 4 shows the frequency location (u_0, v_0) of the main oriented pattern in the filter according the Gabor modelling. It was plot for filters extracted from images at resolution 128x128, pre-processed by a low-pass filter. We observe clear differences according to the categories, which have various local orientation distributions [6]. These differences provide variations on the falloff factor depending on the orientations. For the closed landscapes, the amplitude spectra is rather isotropic (isotropic α) and have more energy in high frequency (low α). For the indoor and urban scenes, $\alpha(0º)$ and $\alpha(90º)$ are less than

$\alpha(\theta)$ on the others orientations. For the open landscapes, characterising by a skyline, orientation 0° is salient and then $\alpha(0°)$ is less than $\alpha(\theta)$ on the others orientations. The location of the «ICA filters» fits these characteristics of orientations in the spectrum. At the highest resolution, noise smoothes out these differences.

	Resolution 64 x 64	Resolution 128 x 128	Resolution 256 x 256
Whitened			
Low-Pass			

figure 3: Examples of «ICA filters» extracted from the whitened pyramid (top), and from the low-pass pyramid (bottom) at the three resolutions

We observe a relation between the orientation of the central frequency ($\theta_0=\arctg(v_0/u_0)$) and the shape factor of the filter (σ_x/σ_y). For oblique orientations, the shape factor is nearer to 1 (isotropic filter), and for orientations near 0° and 90° respectively, the shape factor is in average respectively less and greater than 1. That is to say, the filter is more selective in the orthogonal orientation than in its preferential orientation θ_0, providing behaviour where the reference orientations (0°, 90°) are detected with accuracy. This behaviour, in agreement with neuro-biological data [3, 4], is very interesting for the discrimination task (§5). See [17] for a complete comparison with simple-cell receptive fields in the visual cortex.

4. SPARSE AND DISPERSED CODING

4.1. Definitions

According to equation (1), a set of ICA primitives constitutes a new basis of representation in which an image I(x, y) has a new code {a_i, i=1..50}. Many works have been done to describe how such a primitives set encodes images. [19] is a good review, in which Willmore discerns the two main properties.

The first property is the "sparse coding". Considering a large set of coding units, they remain inactive since the features they detect are not present in the image we encode. To encode any particular images, we use a few numbers of strongly active units, among a large set of possibilities. This type of encoding is the opposite to "distributed coding", which implies a large number of units in an image coding, and conversely uses each unit for coding a lot of images.
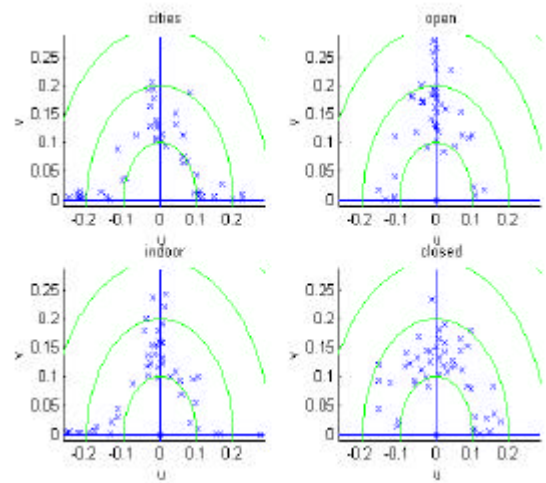

figure 4. Localisation of the central frequencies of the «ICA filters» extracted from images at resolution 128x128, pre-processed by a low-pass filter. The units are in pixel[1].

The second property is the "dispe
Considering a large set of coding units, each one has the same probability to be active. Conversely, after coding several images, all the units have contributed equally to coding. This type of encoding is the opposite to "compact coding", which is exemplified by Principal Component Analysis. In fact, in PCA, the first units encode the main part of variance of original data, while the last units do not have lot of importance.

4.2. Evaluation of the sparse and dispeprsed coding

To evaluate sparseness, Olshausen and Field [15] define three metrics (table 1: S2, S3, S4), which are explicitly maximise in their algorithm. A fourth metric (S1) can be also introduced [19] with the kurtosis of the normalised responses (peakness of the distribution). Let us call a_j the response of one basis function ϕ to one image I_j, μ and σ respectively the mean and the standard deviation of the response of that basis function to all the images (N). r_j is the centred-reduced response, $r_j=(a_j - \mu)/\sigma$. The sparseness of the code with a given set {ϕ_i} is evaluated by the average measure for all the basis functions.

$$S_1 = \left\{ \frac{1}{N}\sum_{j=1}^{N} r_j^4 \right\} - 3 \qquad S_2 = \left\{ \frac{1}{N}\sum_{j=1}^{N}\exp(-r_j^2) \right\} - \frac{1}{\sqrt{3}}$$

$$S_3 = 0.5331 - \left\{ \frac{1}{N}\sum_{j=1}^{N}\log_{10}(1+r_j^2) \right\} \qquad S_4 = \sqrt{\frac{2}{p}} - \left\{ \frac{1}{N}\sum_{j=1}^{N}|r_j| \right\}$$

Table 1. Four sparseness measures. See 4.2 for details

For the "dispersal" property, Willmore proposes a measure based on the variance of the responses ("scree plot" method). The idea is the following. When a filter encodes a set of images, the variance of its response indicates how this filter is useful for this coding. Comparing all the variances gives the relative importance of each filter to encode the images set. Then the variances

256
128
64

functions set.

Scree plots for "city" filters

Scree plots for "closed" filters
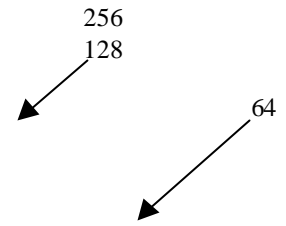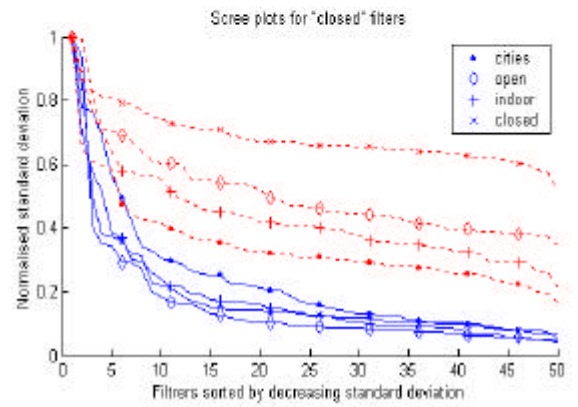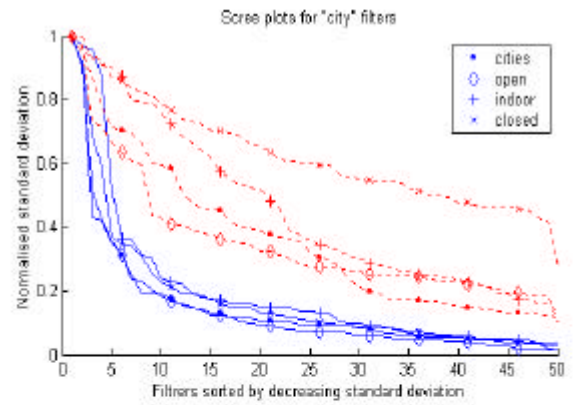
low-pass conditions (plain plot), for the city (top) and closed (bottom) filters, and images according to their category (see the legend).

We consider now how a specific set of basis functions (50 filters from categories) encodes its specific category. Indeed, we could assume that whitening could increase

dispersal of filters for images they were extracted only. In such a case, the excess of dispersal for "white filters" on figure 6 could be due to quart of data for each filter. For this reason, we observed dispersal of filters for each category of images.

Figure 7 shows the scree plots of this experiment. We have observed about the same behaviour for the four categories of filters, then we represented only the plots of filters extracted from images of cities and those from closed images. In four the cases, "white filters" are more dispersal than "raw" ones. In any rate too, filters extracted from whitened closed images are the most dispersal. In fact, there are in this set of basis function, the most atypical filters compared to the other sets.

5. DISCRIMINATION OF IMAGES WITH "ICA FILTERS"

5.1. Principles

To evaluate the discrimination power of the "ICA filters", the learning paradigm is based on a simple K-nearest neighbours classifier. The K parameter and the recognition rate are estimated by the cross-validation method. Images are characterised by their mean energy responses after projection on a given set of basis functions. The Mahanobis distance implements distances between images. Others strategies may also be used by Kullback divergence [11] between the probability densities of the responses (estimated by the product of the marginal densities). With such a strategy, the density estimate is simplified by the independence property, and we have obtained a recognition rate a little bit greater with a more time consuming method compared to the Mahanobis distance processing.

Even if the basis functions characterise the images locally, the describing features for each image are composed of the mean squared response for each function. By such a description, the spatial organisation of the images is lost, but we know that human subjects are efficient for this categorisation task without using information on the spatial organisation of the scene and without recognising the objects in the scene.

Our test hypothesis is that the most dispersal filters to encode the image database would be efficient to categorise the database. For a basis function associated with a great "dispersal factor", the coding values resulting from projection of images have great fluctuations around its mean value. Thus this function is well adapted for a particular subset of images, and not adapted for another subset. For all the functions, how do these subsets fit with the 4 learning categories ? To answer this question, a selection strategy is built from the "dispersal factor" used as a threshold. Let us consider the inverse function of the "scree plot" (figure 6) where the abscise is the "dispersal factor" and the ordinate, the numeral of the sorted filters

which corresponds to the number of filters having a "dispersal factor" superior or equal to a given one (figure 8). Starting from 0, all the filters are selected, up to one, the number of selected filters decreases. The less dispersal filters are rapidly suppressed, while the most dispersal ones remain selected. With this methodology, we have compared the two pre-processing conditions at the three resolutions.

5.2. Results

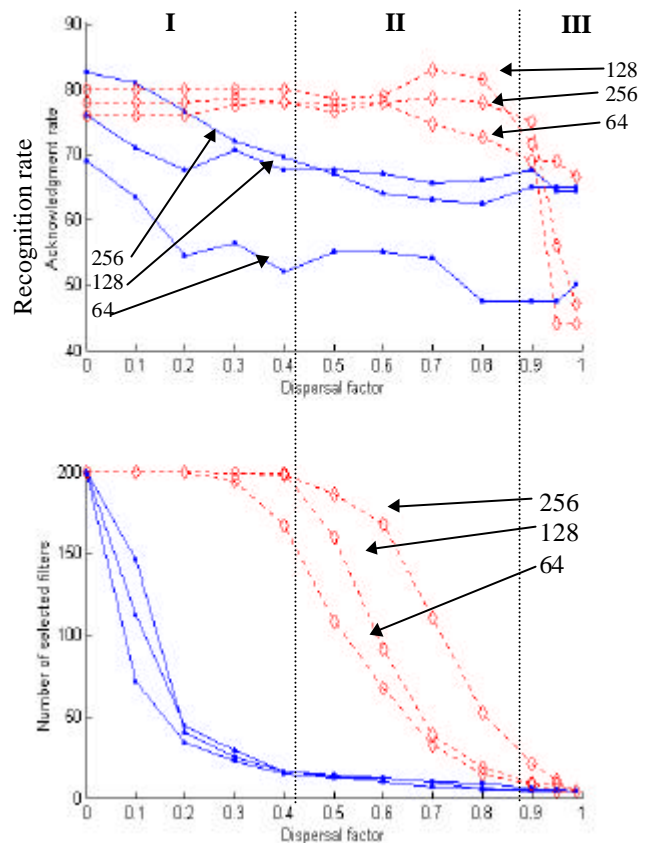We can notice three behaviours according to the dispersal factor (region I, II, III in figure 8).

figure 8: Recognition rate (top) and corresponding number of filters (bottom) according to dispersal factor, for three resolutions, and both "white filters" (dotted) and "raw filters" (solid). 3 regions (I, II, III) are discerned.

From 0 to 0.4 (region I), few "white filters" are removed while lot of "raw filters", having a low "dispersal factor", are eliminated. For a "dispersal factor" of 0.4, more than 150 "white filters" are retained, while less than 20 "raw filters" were kept. As a consequence, recognition rate of classification with "raw filters" is decreasing in 5% for resolution 128 (from 75% to 70%), and 15% for resolution 64 (from 70% to 55%). At the opposite, classification with "white filters" is contending between 75% and 80% for the three resolutions.

627