CLASSIFICATION OF IMAGES : ICA FILTERS VS HUMAN PERCEPTION

H.Le Borgne[1], N. Guyader[1], A. Guérin-Dugué[2], J. Hérault[1]

[1]LIS – INPG, 46 av. Félix Viallet, F-38031 Grenoble Cedex 9, France
[2]CLIPS, Bat B, rue de la bibliothèque, 38041 Grenoble Cedex 9, France
{hleborgn, nguyader}@lis.inpg.fr, Jeanny.Herault@inpg.fr, Anne.Guerin@imag.fr

ABSTRACT

In this paper we compare a machine based semantic organisation of natural images with the one provided by human perception. On one hand, we have conducted a psychophysical experiment to determine a human perception space in which we have identified semantic categories. These categories and the distances between images are emphasised by analysing the human response similarities with a multidimensional scaling technique called Curvilinear Component Analysis (CCA). On the other hand, we try to perform the same scene categorisation with a computational model based on an ICA filter description.

1. INTRODUCTION

The rapid growth of multimedia databases have created new needs for various users who classify, organise, or navigate through databases. For the past few years, most researches on image databases have focused on the use of "low-level features" obtained from the raw pixel values, without any human supervision. Some discriminations between broad classes of images have been made with success [5, 12, 14] but difficulties are encountered to access at finer level of description [3]. Nevertheless, in a Content Based Image Retrieval (CBIR) paradigm, we should focus on the visual *context* of the image, that is to say its category [13].

Independent Component Analysis [8] or ICA is a non supervising way to extract low level features which present many advantages. First of all, these features are similar to the pattern of the mammalian primary visual [15] and are adapted to the data from which they are extracted [10]. Secondly, ICA features have statistically independent activities, what is congruent with most of the theories of sensory coding that have proposed models to effective internal representation by redundancy [1, 4].

In this paper, we compare the ICA filter based organisation of natural scene images with the one provided by human perception. In Part 2, we describe a psychophysical experiment for image categorisation which is an improvement of the Rogowitz and al.'s "Computer Scaling" [11], and part 3 shows how the results are computed to reveal a human perceptual space of natural scenes represented by a multidimensional technique called Curvilinear Component Analysis [3]. This organisation allows us to identify the major semantic categories according to a human perceptual space. Part 4 is dedicated to explain how ICA filters are extracted from natural images and then, how we use them to describe images. Part 5 is a synthesis which presents the human-based classification and the one we obtain with ICA filters.

2. EXPERIMENT

In the experiment, human observers judge the similarity of 105 selected images presented on a computer display. We measure the perceived similarity of each image with every other images of the base. In each trial (Figure 1), a reference image is presented with eight randomly-chosen images; the subject is asked to choose which image, among this context, is the most similar to the reference one and to give the similarity level.

2.1 Stimuli and subjects

The image database consists of a set of 105 natural images. These images have been selected in order to cover a wide range of natural environments (same types of scenes as in [5, 6]): animals, people, indoor scenes, nature as beaches or mountains, buildings…. The experiment was conducted on a monitor (luminance TIFF images), using a Matlab interface. In order to rely only on structural information, we did not consider colour. The display measured 36,5 × 27,5 centimetres and it is viewed at a distance of approximately 60 centimetres. Viewed on the display monitor, the size of each image was approximately 5,3 × 5,3 centimeters and subtended approximately 5 degrees of visual angle. Results were computed with 48 subjects with normal viewing or corrected to normal viewing [6].

2.2 Experiment description

In order to represent a wide range of possible natural scenes, we need a sufficient number of categories, as well as a sufficient number of images per category, while keeping the database size within reasonable limits for psychophysical experiments ; this explains the "weak" number of 105 images in the database. One image is presented with eight-randomly chosen images (to reduce the number of trials) among the 104 others, as in the Rogowitz's experiment [11]. The experiment is organized as follows : in a first step, which corresponds to a first screen (Figure 1), the subject is asked to select the image which is the closest in term of similarity to the reference one. This first part of each trial is limited to five seconds, which is enough to glance at the eight presented images. Thus, the association is based on global criteria, which is consistent with our global model of ICA filter description (see part 4). In a second step, the subject has to tell the proximity of the selected image to the reference one, on a scale of four levels: "very close", "close", "different", "very different". This judgement of similarity in a second step did not exist in the original Rogowitz's "Computer Scaling" experiment. It allows the subject to make "weak association" of images and "strong" ones.

Figure 1. One trial of the experiment.

We expect that our experiment shows a scene organization into clusters, one cluster being ideally representative of one semantic category.

3. DATA ANALYSIS

The first step in the data analysis is to compute a matrix of distances between the 105 images from the human subjects responses, in order to use it as the entry of the multidimensional scaling data analysis algorithm. This algorithm allows us to project images into a 2-dimensions space which enhances the meaningful image categories, at least those which would be recognised in a Content Based Image Retrieval paradigm.

3.1 Experiment analysis

In each trial, a reference image i_{ref} is presented in front of eight-randomly chosen images, and the subject is picking out the image j he thinks the more similar to the reference. In a former work, we have made a similarity matrix $S(.)$, increasing $S(i_{ref},j)$ of one unit each time the couple (i_{ref},j) was enhanced by a human subject. Then, the level of similarity indicated in the second step of the experiment was used in the transformation of the similarity matrix $S(.)$ into a matrix of distance **D** (see [6] for more details). This methodology suffered some drawbacks, particularly because it did not take into account the context of the subject's choice. Even if subjects look at the whole 105 images before the experiment, they always pick out the image which the more similar to the reference one, among the eight presented. Then we can consider he associates the image j to the image i_{ref} as much as he rejects seven images (r_1, r_2,…, r_7) from i_{ref}. According to this idea, we directly compute a matrix of dissimilarity D(.), increasing D(i_{ref}, r_1), …, D(i_{ref}, r_7) of one unit at each subject's choice. Computationally speaking, this method is more interesting than the first one since each trial brings more information : seven points are computed in D(.) instead only one in the computation of S(.). This approach is more similar to the Vailaya's one in [14] than the Rogowitz's in [11]. Nevertheless we can note the second step is not used here to compute the distance matrix between image. See section 5 for discussion about this point.

3.2 Matrix of distance

The distance matrix **D** is obtained by averaging all the matrix of dissimilarity computed as described above.

D is not symmetric since the value of $D(i,j)$ designed the distance between images i and j when image i is the reference. Considering the distance matrix properties, we compute our final distance matrix as the average between itself and its transpose; the result is then symmetric. This operation rules out semantic interpretation of non-symmetries; this will be considered in future works.

This symmetric matrix is then processed by a CCA which allows us to project the human judgement into a 2-dimensions metric space (section 5.1). Results suggest semantic categories which will be used to perform classification over an independent image database (section 5.2).

4. COMPUTATIONAL MODEL

4.1 Learning independent components from images

The first part of our computational model consists of finding independent components which describe a large number of natural scene images. For that, we use a training image database, different from the experiment one, made of 13 luminance images of size 256 × 384 pixels. This database was established to well represent the semantic categories of images that human subjects had put into relief in the psychological experiment (see section 6.1). Images are processed by a whitening filter implemented according to a vertebrates biological retina model [7] which realises a non linear processing. From each image, we extract at random patches (*i.e.* small image parts of size 32 × 32 pixels) such that we hold more than 20,000 cumulative patches. In order to minimise the anisotropy on horizontal and vertical orientation, each patch is focused by a weighting Hanning window. Before an Independent Component Analysis (ICA), a principal component analysis (PCA) realises a data whitening and a dimension reduction from 1024 (32 × 32) to 200 dimensions, with a retaining of 90% of the total inertia.

In the ICA paradigm applied to images [8], we assume that each patch $I(x,y)$ is an independent combination of a set of primitives $\{\Phi_i(x,y)$, i = 1..200$\}$. The primitives represent the spatial patterns occurring in the different scenes such as the projection on this basis involves independent codes $\{a_i, i = 1..200\}$:

$$I(x,y) \quad \sum_{i\ 1}^{200} a_i \quad {}_i(x,y) \qquad (1)$$

Practically, we use the “Fast-ICA” algorithm [9] with the symmetric method, because of its fast convergence time. It provides four collections of 200 primitives, that we consider as 2D filters $\{F_i(x,y),\ i = 1..200\}$.

4.2 Image representation with ICA filters

We now consider N ICA filters obtained as we explained before. A natural image can thus be characterised by a collection of N responses to these filters, which are considered as particular observations of random variables $\{R_i, i=1…N\}$. The energetic responses of an image $I(x,y)$ to the selected pool of filters $\{F_i(x,y), i=1..N\}$ are estimated as follow :

$$\forall i \in [1, N], r_i = (I * F_i)^2 \qquad (2)$$

Where * represents a classic convolution between image I and filter F_i. From the “valid” part of the response, we keep 2350 randomly choosen observations $\{r_i(k), k = 1...2350\}$ of each random variable R_i. Classically, we could calculate the averages of the observations r_i and compute the Euclidean distance between them, as an estimation of the distance between the corresponding images ([12, 14]). In this paper, we use histograms because it provides a more complete information that improve following task of discrimination or classifications. Moreover, we do not use the Euclidean distance as a measure of dissimilarity between images, but rather a Kullback-Leibler inspired measure defined in its symmetric version as :

$$KL_H(H_1, H_2) \quad \sum_{b\ 1}^{64} H_1(b)\log\frac{H_1(b)}{H_2(b)} \quad H_2(b)\log\frac{H_2(b)}{H_1(b)} \qquad (3)$$

Thanks to the concavity of the logarithm function, and the equal number of point in each histogram, this measure is positive when H_1 and H_2 are different, and is zero if H_1 is equal to H_2 but it does not fulfils the triangle inequality. Moreover, since the Kullback-Leibler divergence between two multidimensional distributions with independent components is the sum of the Kullback-Leibler divergences between each component, we compute the distance between two collections of P histograms as the sum of P distances between individual histograms.

5. CLASSIFICATION OF IMAGES

5.1 Finding a human perception space

We take the distance matrix **D** as described in part 3. This matrix is neither related to any known space, nor any known metric. A number of methods provide a representation of this unknown space. Here, with the CCA [4], the local topology of the input average manifold contained in the distance matrix is mapped into a 2-dimensions representation space (Figure 2). We can distinguish semantic clusters in the data, like indoor scenes, people, forests, deserts, mountains, buildings, animals, fields.

Figure 2. two-dimensions image organization with human perception matrix.

As we explained in section 3, we do not take into account the second step of the psychological experiment, because we consider that the first step of the experiment enhances “inter-class information” while the second step is connected to “intra-class information”. Since our main goal is to put in relief categories, we have focused on the information which directly leads to a clustered organisation.

5.2 Organisation and classification with ICA filters

To evaluate the power discrimination of ICA filters we drive two experiments. The first one aims at organising the same 105 images as in the above paragraph, and the second one is a classification test over a larger database (540 independent images). In both cases, we compute signatures of images according to equation 2, and distances between them according to equation 3.

In the first experiment, computation leads to a 105x105 matrix of distances between images which is mapped, via Curvilinear Component Analysis, into a 2-dimensions representation space, in order to directly compare with the human organisation. This projection shows that images with human people or animals are very bad placed in comparison with the human’s one. In fact, this is due to the high-level of semantic that contain these images. At the opposite, ICA filters only catch the global structure of the image, and are well adapted to organise scenes like man made scenes, open scenes (beaches, fields, deserts) and textured one (forests, mountains). Note that scenes of mountain are not very well classified. Figure 3 is showing the organisation provided by ICA filters when we display these last categories only.

Figure 3. image organisation with ICA filters.

In order to generalise these results, we classify a larger independent set of images. This set consists on a collection of 540 natural images (256 x 256 pixels, and 256 grey-level values) extracted from several databases, and reclaimed on the world wide web. We consider four categories containing about 135 images each : urban outdoor scenes, indoor scenes, open landscapes (fields, beaches, deserts…) and closed landscapes which contains textured scenes without preferential direction (mountains, valleys, forests…). Image labels were established by a human judgement [6]. We use a K Nearest Neighbours (KNN) classifier ; the evaluation of performance has been estimated in a "Leave-one-out" process with an optimal value K among {3, 4,…10}. This experiment gives a recognition rate of 85% which confirm the good performances of ICA filters for classification of natural scenes with no presence of people or animals.

6. CONCLUSION AND FUTURE WORKS

In this paper, we had conducted a psychological experiment which had put into relief semantic categories of images. On the other hand, the organisation provided by ICA filters enhances the same categories excepted for images with a semantic "alive subject" (animal or people) that are not as well classified as the others. An experiment of classification over an independent base had confirmed the good power of discrimination of ICA filters for the images of scenes without "alive subject". Future researches will investigate a way to classify these kinds of images. Another way of research will be to combine the "intra-class" and "inter-class" information in the computation of the "human matrix of distance".

7. AKNOWLEDGEMENT

The Authors thank A. Chauvin, S. Donadieu, C. Marendaz and C. Peyrin from LPSC and C. Berrut from CLIPS for fruitful discussions. They also thank professor Erkki Oja and J. Laaksonen for the welcome in the Laboratory CIS of Helsinky, and the discussions with the Computational Neuroscience Group leaded by A. Hyvärinen. This work is partially funded by ELESA and IMAG (SCoPIe project on image indexing). The Rhône-Alpes region funds the first author.

8. REFERENCES

[1] Barlow H.B. Unsupervised Learning Neural Computation, vol. 1, pp. 295-315, 1989.
[2] Cox I.J., Ghosn J., Miller M.L., Papathomas T.V., Yianilos P.N., "Hidden Annotation in Content Based Image retrieval", Proc. of CVPR'97, 1997.
[3] Demartines P. and Hérault J., "Curvilinear Component Analysis : a Self-Organizing Neural Network for Non-Linear Mapping of Data Sets", IEEE Trans. On Neural Networks, 8, 1, 148-154, 1997.
[4] Field D.J. (1994). What is the Goal of Sensory Coding ?, Neural Computation, vol. 6, pp. 559-601.
[5] Guérin-Dugué A., Oliva A., Classification of Scene Photographs from Local Orientations Features, Pattern Recognition Letters, 21, pp 1135-1140, 2000.
[6] Guyader N., Le Borgne H., Hérault J. and Guérin-Dugué A., "Towards the introduction of human perception in a natural scene classification system", Proc. of IEEE NNSP'02, Martigny, Switzerland, 2002.
[7] Hérault J., "De la rétine aux circuits neuromorphiques", chap3, in "Les systèmes de vision" J.M. Jolion ed., IC2 col., Hermes, Paris, 2001.
[8] Hyvärinen A., Karhunen, J., Oja, E., 2001. Independent Component Analysis, John Wiley & Sons.
[9] Hyvärinen A., Oja. E.,. A Fast fixed-point algorithm for ICA, Neural Computation, vol 9, no 7, pp. 1483-1492, 1997
[10] Le Borgne H., Guérin-Dugué A., "Sparse-Dispersed Coding and Images Discrimination with ICA", Third international conference on ICA, San Diego, USA, 2001.
[11] Rogowitz B., Frese T., Smith J., Bouman C.A., and Kalin E., "Perceptual image similarity experiments", Human Vision and Electronic Imaging III, Proc. of SPIE, vol. 3299, San Jose, CA, January 26-29, 1998.
[12] Szummer M., Picard R.W., Indoor-Outdoor image classification, , ICCV'98, 1998.
[13] Torralba A., Oliva A. Statistics of Natural Image Categories, (accepted pending revisions), Network: Computation in Neural Systems. 2003
[14] Vailaya A., Jain A., Zhang H.J., On image classification : City images vs Landscapes, Pattern Recognition, vol. 31, n°12, pp. 1921-1935, 1998.
[15] Van Hateren J.H., Van der Schaaf A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex, Proc. of the Royal Soc. of London, series B, vol 265, pp. 359-366.