

CHAPITRE 11

ANALYSE DE SCENES NATURELLES PAR COMPOSANTES INDEPENDANTES

Anne Guérin-Dugué¹, Hervé Le Borgne²
L.I.S – I.N.P.G.
46 avenue Félix Viallet
F-38000 Grenoble Cedex
Mél : {guerin, hleborgn }@lis.inpg.fr

11.1. Introduction

Chez l'être humain, la reconnaissance visuelle des scènes, des objets, des visages est généralement rapide, automatique et fiable. Cette simplicité contraste avec la difficulté à modéliser, en psychologie de la vision les processus de reconnaissance visuelle et à produire, en vision par ordinateur des algorithmes de reconnaissance simples, efficaces et robustes.

A partir des travaux pionniers d'Hubert et Wiesel, une grande majorité des systèmes de reconnaissance à base de caractéristiques sont modélisés à partir de familles de filtres de type passe-bande orienté (Gabor 2D, dérivées de fonctions gaussiennes, ...) selon des stratégies de codage inspirées de celles du cortex visuel. De nombreuses études [2, 8, 14, 16] montrent que ces détecteurs corticaux pourraient résulter de l'application d'un principe de réduction de redondance par indépendance statistique de leurs activités.

Parmi les techniques proposées pour construire de tel code, l'Analyse en Composantes Indépendantes (ACI) [7, 11, 12] est une voie très prometteuse qui fournit de manière non supervisée des unités de codage assimilées en première approximation à des champs récepteurs de type « Gabor » similaires aux cellules simples du cortex visuel primaire [3, 12, 20].

Cette voie est expérimentée ici dans le contexte de l'analyse de scènes complexes. Après avoir décrit la méthodologie et l'implantation de l'ACI sur une base d'images, les détecteurs extraits sont analysés quant à leurs propriétés fréquentielles, de codage creux et de reconstruction. La propriété d'indépendance des codes permet une estimation simple et fiable des densités de probabilité des codes utiles à l'établissement de mesures de dissimilarités entre images. Enfin les potentialités de la méthode sont illustrées sur une tâche de catégorisation de scènes naturelles.

11.2. Principe de l'ACI appliquée aux images

Issue des travaux sur la séparation de sources [11], l'analyse en composantes indépendantes [4] est un outils d'analyse de données permettant de trouver une nouvelle base de représentation dans laquelle les projections des observations sur cette base fournissent des composantes indépendantes. En comparaison, une analyse en composantes principales fournit une base de représentation dans laquelle les composantes sont simplement décorréliées.

Soit $\Phi = \{\phi_i, i = 1..m\}$ cette nouvelle base. Soit $e \in \mathfrak{R}^p$ une observation. Dans cette base, soient $s_i = \langle e, \phi_i \rangle$, les composantes statistiquement indépendantes, on a:

$$e = \sum_{i=1}^m s_i \cdot \phi_i. \quad (1)$$

De nombreux critères ont été proposés pour estimer les primitives ϕ_i , comme par exemple (i) la minimisation de l'information mutuelle entre les sorties (s_1, s_2, \dots, s_m) [16], (ii) l'optimisation d'une fonction de contraste, ou

¹ Anne Guérin-Dugué est actuellement Chargé de Recherche, détachée auprès de l'INRIA Rhône-Alpes.

² Hervé Le Borgne est financé par la région Rhône-Alpes dans le cadre du projet EMERGENCE 2000 « ASCII : Architecture Sémanti-Cognitive d'Indexation d'Images ».

d'entropie utilisant les statistiques d'ordre quatre [4,10]. Nous avons utilisé ici l'algorithme appelé « Fast ICA » [10] pour ses propriétés de convergence rapide et d'extraction séquentielle des sources. Il est basé sur l'optimisation d'une fonction d'entropie (appelée entropie négative ou « negentropy ») appliquée sur les sorties. (Se référer à la présentation « The Fast ICA algorithm » d'Erkki OJA dans ce présent recueil).

En appliquant un tel modèle (eq. 1) aux images, on considère alors qu'une image (e) est une combinaison linéaire de fonctions de base (ϕ). Les coefficients de pondération forment le code ($s=[s_1, s_2, \dots, s_m]$) de l'image, comme l'illustre la figure 1.

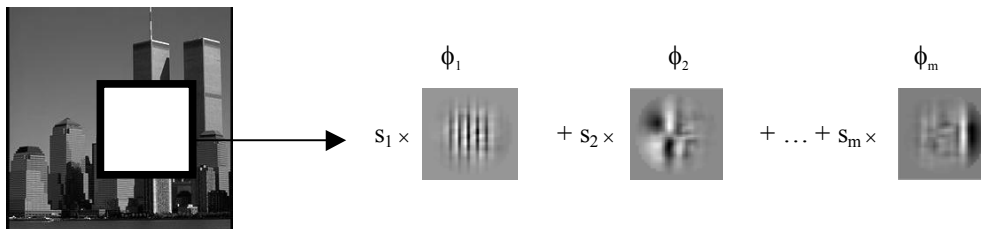


Figure 1 : Illustration de la décomposition d'une image dans la base Φ .

11.3. Méthodologie

11.3.1 Base d'images

La base d'images utilisée est constituée de 200 images 128x128 de scènes naturelles (issues de la collection d'images COREL). Ces images peuvent être interprétées comme appartenant à 4 catégories sémantiques différentes de 50 images chacune : villes, pièces d'intérieur, paysages « fermés » (forêts, montagnes, vallées, ...), paysages « ouverts » (plages, déserts, champs, ...). La base d'images est de petite taille (seulement 200 images), mais les images ont été sélectionnées pour former une base ayant une grande variabilité intra-classe.

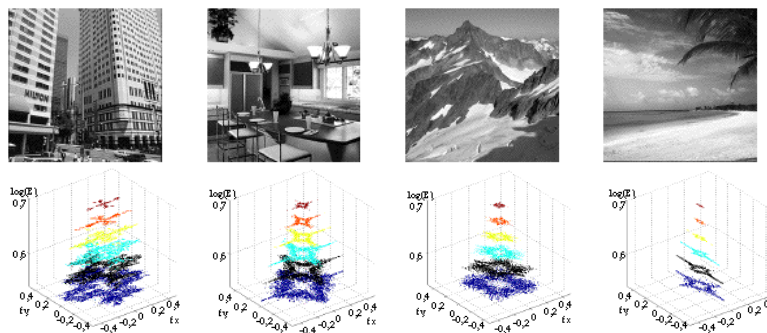


Figure 2 : Formes spectrales prototypes d'une image et son spectre d'énergie pour chaque catégorie.

Les images dans chacune des catégories ont une cohérence spectrale (cf. fig. 2) se modélisant par un gabarit suivant l'anisotropie/isotropie en orientation du spectre d'énergie dans l'espace de Fourier [15]. On remarque par exemple une forme spectrale en « croix » pouvant caractériser les scènes d'intérieur et aussi les scènes urbaines (présence d'horizontales et de verticales), ou une forme anisotrope sur les fréquences verticales pour les scènes « ouvertes » caractérisées par la présence d'une ligne d'horizon, ou une forme plutôt isotrope pour les paysages « fermés » (scènes plus texturées) [9].

Ainsi pour interpréter les résultats de l'application de l'ACI sur une collection d'images, on va réaliser 4 expérimentations séparées, à savoir une ACI par catégorie d'images. Ces expérimentations permettront d'analyser les bases de représentation obtenues en fonction des propriétés spectrales de chaque catégorie.

11.3.2 Base d'observations

Chaque ACI a été appliquée à partir d'une base d'observations extraites d'une catégorie d'images. Un prétraitement a été appliqué à la base d'images (passe-bas de type Butterworth d'ordre 4 à la fréquence spatiale de 0.2 pixel⁻¹, puis sous-échantillonnage d'un pixel sur 2) pour obtenir des images de taille 128 x 128 en ayant éliminé les informations de trop haute fréquence spatiale. Les observations d'entrée du système sont extraites des images sous forme d'images de taille $n=l \times l$ ($l=32$) extraites à des positions aléatoires dans les images. 30 images sont extraites par image, elles sont préalablement apodisées par une fenêtre de Hamming qui permet la réduction de l'effet de Gibbs qui produit artificiellement une augmentation des énergies dans les fréquences spatiales verticales et horizontales. Cette

apodisation permet un effet de focalisation au centre de l'imagette. La conséquence est que les pixels à la périphérie de l'image sont très atténués. La dimension des observations peut être réduite en supprimant les pixels périphériques et en ne conservant que les pixels centraux ayant une variance significative, soit 600 pixels. Ainsi au total, la base des observations comportent 1500 individus en 600 dimensions. Préalablement à l'application de l'ACI, la base des observations est blanchie par ACP et on fixe à ce niveau le nombre de composantes indépendantes m à extraire. Pour les expérimentations montrées ici, ce choix a été fixé arbitrairement à 30, comme résultant d'un compromis sur le temps de calcul et la complétude des caractéristiques extraites.

11.3.3 Interprétation des résultats

A la convergence, les m fonctions de base formant les unités de codage ou détecteurs sont à utiliser comme des filtres RIF 2D (taille $l \times l = n$). La convolution de l'image $i(x,y)$ avec le filtre $\phi_i(x,y)$ fournit pour chaque position spatiale (x_0,y_0) le code $s_i(x_0,y_0)$, de l'imagette centrée en (x_0,y_0) , sur la fonction ϕ_i . On obtient 4 populations de $m=30$ filtres qui seront étudiés dans la suite, relativement à leur propriété fréquentielle, leur propriété de codage et reconstruction et enfin leur propriété de discrimination d'images.

11.4 Analyse statistique des filtres ACI

11.4.1 Modélisation

En réalisant une Analyse en Composantes Indépendantes pour chacune des quatre catégories d'images, on spécifie ainsi la famille de filtres obtenus par rapport à sa catégorie d'origine. Conformément à la méthodologie décrite précédemment, quatre populations de 30 filtres sont obtenues. En première approximation, ils sont assimilés à des filtres de type passe-bande orienté, dont on cherche le modèle de filtre de Gabor 2D le plus proche par minimisation du critère quadratique suivant sur les énergies :

$$E(u_0, v_0, \sigma_x, \sigma_y) = \frac{\iint_{u,v} |F - G(u_0, v_0, \sigma_x, \sigma_y)|^2 dudv}{\iint_{u,v} F^2 dudv} \quad (2)$$

où $F(u,v)$ est un détecteur ACI et $G(u,v)$ le filtre de Gabor paramétré par sa fréquence centrale (u_0, v_0) et son étalement spatial (σ_x, σ_y) . Cette modélisation a pour unique but, l'analyse statistique des quatre ensembles de filtres à mettre en relation avec la forme spectrale moyenne de chaque catégorie [13].

On peut qualitativement repérer plusieurs types de filtres. Pour décrire ces types, on introduit le facteur de forme ou l'anisotropie d'un filtre par le facteur de forme ϕ comme l'illustre la figure 3. Un filtre isotrope aura un facteur de forme voisin de un, contrairement à un filtre anisotrope.

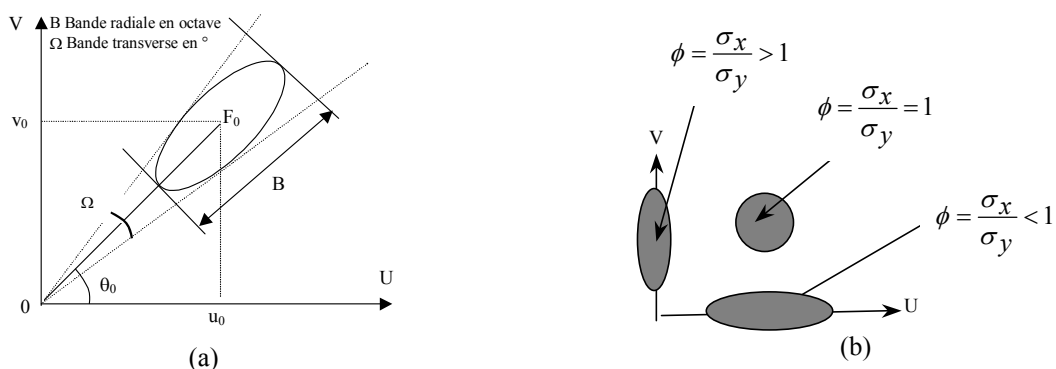


Figure 3 : (a) Illustration dans le plan de Fourier de (a) la paramétrisation d'un filtre de Gabor, (b) des filtres de type passe-bande orienté isotropes ou anisotropes.

Les types de filtres obtenus sont les suivants :

- Filtres de type passe bande orienté anisotrope vertical ou horizontal (fig. 4a.b),
- Filtres de type passe bande orienté isotrope oblique (fig. 4c),
- Filtres de type passe bande anisotrope –le filtrage n'est pas identique pour toutes les orientations dans la bande passante sélectionnée.(fig. 4d),
- Filtres anisotropes sur deux orientations orthogonales (fig. 4e).

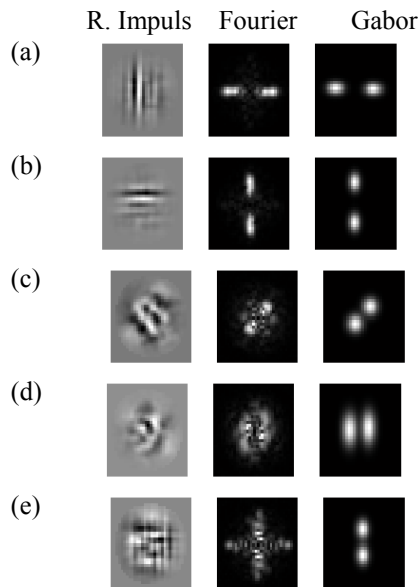


Figure 4 : Illustration pour chacun des types (voir texte) d'un filtre obtenu par ACI, de sa réponse impulsionnelle, son spectre d'énergie et le spectre d'énergie du filtre de Gabor le plus proche.

La modélisation choisie est adaptée aux filtres de type a, b ou c, et fournira le motif d'orientation principal pour les deux autres types d ou e, comme cela est illustré à la figure 4.

11.4.2 Résultats sur les propriétés fréquentielles

Via ses paramètres de positionnement et d'étalement, l'analyse statistique sur les quatre populations de filtres permet de mettre en avant 4 faits marquants :

- (i) Il y a un plus grand nombre de filtres sensibles aux orientations de référence (0° , 90°) (voir la figure 5b)
- (ii) Ces filtres sont plutôt anisotropes si leur orientation privilégiée est proche de la verticale ou de l'horizontale. La figure 6b met en évidence cette dépendance entre le facteur de forme ($\phi = \sigma_x / \sigma_y$) et l'orientation du filtre ($\text{tg}\theta_0 = v_0 / u_0$).
- (iii) Les filtres sur les orientations obliques sont plutôt sphériques ($\sigma_x = \sigma_y$), avec une largeur de bande transverse plus importante que pour les filtres sur les orientations canoniques qui sont beaucoup plus sélectifs.
- (iv) On observe également une relation décroissante entre la bande passante et la résolution du filtre (figure 6c). C'est la conséquence du spectre d'énergie des images naturelles qui est globalement en « $1/f$ ».

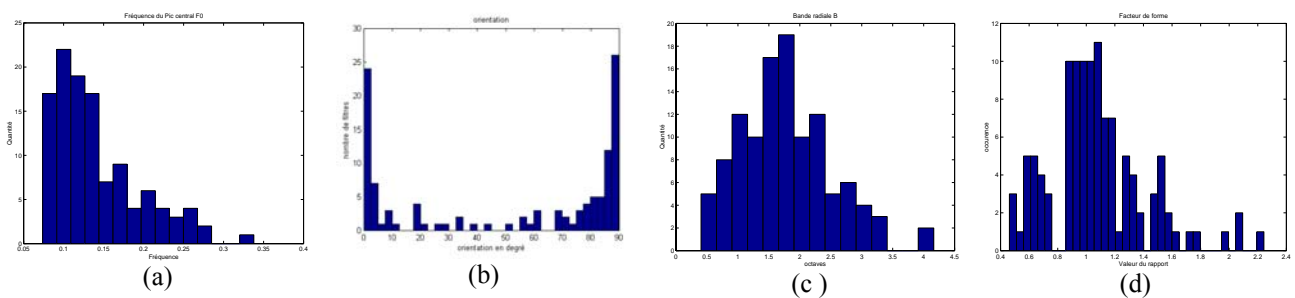


Figure 5 : Histogramme sur (a) Fréquence centrale, (b) Orientation, (c) Bande radiale, (d) Facteur de forme

Il est intéressant de noter certaines analogies (orientation, facteur de forme, bande passante) avec les cellules simples du cortex visuel [5, 6]. On trouvera dans [20] une étude comparative plus exhaustive.

La figure 6a montre pour chaque ensemble de 30 filtres, la localisation spatiale dans le plan de Fourier, du filtre de Gabor le plus proche. On remarque qu'il y a un plus grand nombre de filtres sensibles aux orientations de référence (0° , 90°), par rapport aux orientations obliques (voir aussi la statistique sur la distribution des orientations fig. 5b). Les scènes de la base étant des scènes naturelles, elles reproduisent notre environnement structuré plutôt horizontalement et verticalement. Pour les catégories « ville » et « intérieur », les localisations sont très voisines. De plus, on remarque la cohérence entre la location des filtres et les propriétés spectrales des images de chaque catégorie (cf §11.3.1).

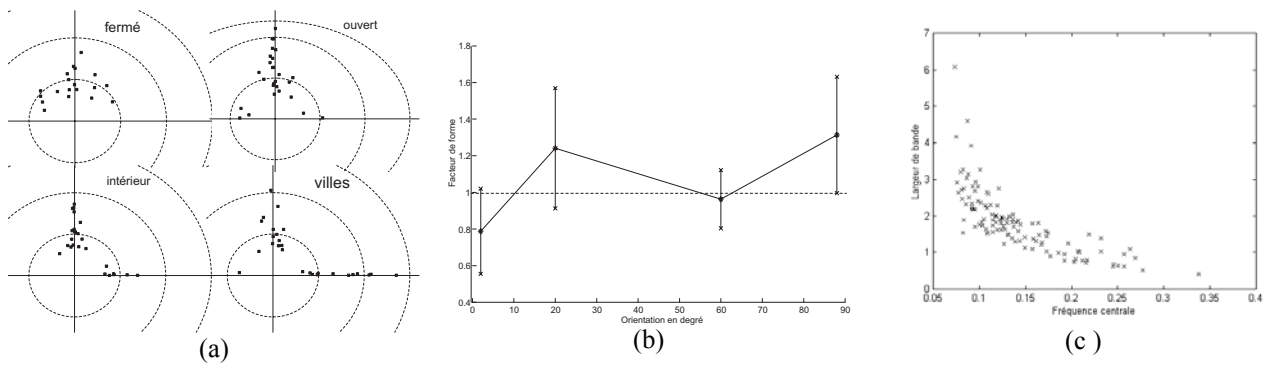


Figure 6 : (a) Localisation de la fréquence centrale , (b) Facteur de forme vs Orientation, (c) Bande radiale vs Fréquence centrale

11.5. Propriétés de codage creux et de reconstruction

Plusieurs qualificatifs sont retenus pour décrire les propriétés d'un code :

- Code *creux* par rapport à code *distribué* : Dans un codage creux, quelques unités sont actives en même temps pour coder une image, alors que pour un codage distribué, chaque image est codée par l'activité d'un nombre relativement important d'unités.
- Code *dispersé* : La dispersion d'un code est mesurée relativement au codage d'une collection importante d'images. Pour un code dispersé, toutes les unités en moyenne contribuent uniformément au codage.

Ces deux caractéristiques (dispersé – distribué) ne sont pas antinomiques. On trouvera dans [21] une discussion à ce sujet et des propositions de métrique pour juger de la qualité en ces termes, des codes fournis.

Dans la suite, nous illustrons qualitativement le caractère creux des codes fournis par l'ACI en le comparant à ceux fournis par Analyse en Composantes Principales.

V.1 Codage creux

Les études [7, 8, 16] montrent que la stratégie de codage du cortex visuel pour représenter les images naturelles résulterait d'un principe de réduction de redondance. Cette redondance est présente dans la structure même des images naturelles, sous la forme de dépendances statistiques d'ordre supérieur [16, 20]. Le cortex visuel chercherait à extraire ces dépendances statistiques afin que les images puissent être interprétées comme des événements indépendants. Ainsi, une image donnée activerait une petite partie des détecteurs corticaux seulement, correspondant à une signature de la catégorie de l'image.

Nous avons comparé le caractère des codes des images obtenus par projection sur les filtres résultant d'une Analyse en Composantes Indépendantes, avec les codes obtenus par projection sur les filtres résultant d'une Analyse en Composantes Principales sur les mêmes données. Les codes obtenus par projection des images sur une base de primitives extraites par ACP, sont simplement décorrélés entre eux. En reprenant la terminologie précédente, le code fourni par ACP n'est pas par construction dispersé, puisqu'en moyenne toutes les unités ne sont pas uniformément activées (décroissance de la variance). Par contre, par construction, il est compact, puisque par ACP, on choisit la base minimale d'unités pour traduire un maximum de variance. Par la suite, avec une simple métrique nous illustrons la différence entre ces deux codages.

Soit $S=(S_1, \dots, S_{30})$, le code pour une observation (imagette $l \times l$ extraite d'une image) représentée dans sa base de détecteurs. D'un point de vue statistique, chaque composante S_i du code des images peut être vu comme une variable aléatoire indépendante. Pour une collection d'images ayant des codes creux, chaque composante est essentiellement inactive, si bien que sa densité de probabilité présente un pic important autour de zéro. Le caractère creux de ce codage peut s'évaluer en comptant le nombre de composantes S_i actives. Les moyennes des codes S étant à peu près nulles, une composante S_i est considérée comme active quand elle est située à plus d'un écart-type de sa moyenne. Pour cela on évalue pour chaque observation une variable aléatoire c ainsi calculée :

$$c = \sum_{i=1}^{30} F(S_i), \quad (3)$$

avec F la fonction à seuil ainsi définie, $F(S) = 0$ si $\mu - \sigma < S < \mu + \sigma$, et $F(S) = 1$, sinon, avec μ et σ respectivement la moyenne et l'écart-type de la variable S . La variable aléatoire c mesure donc le nombre de composantes actives dans un code. La figure 7a, respectivement 7b illustre l'histogramme de c pour des observations issues des 4 catégories pour un

codage par filtres ACI et respectivement par filtres ACP. Ces deux familles d'histogrammes exhibent des structures fondamentalement différentes.

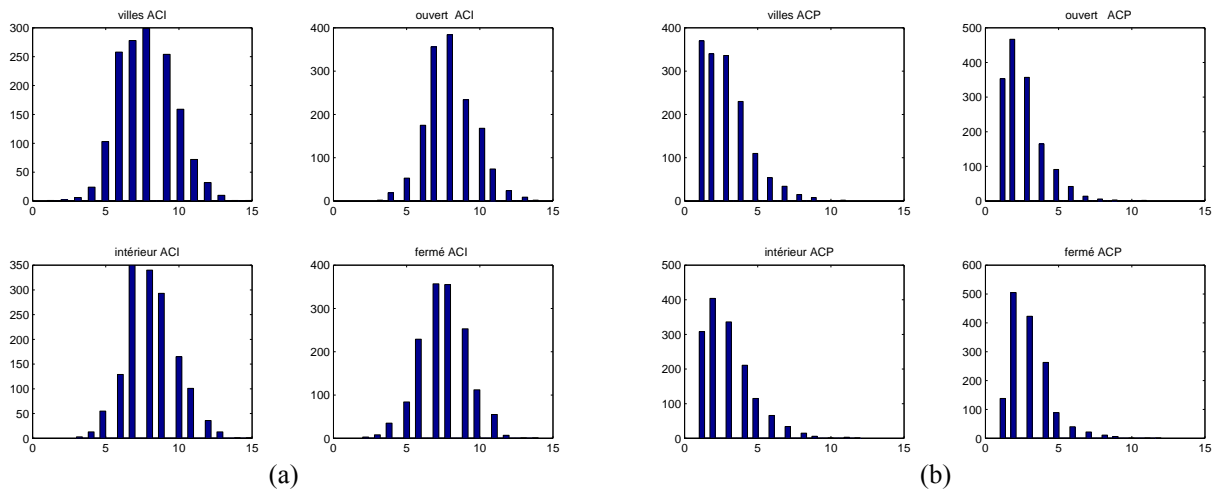


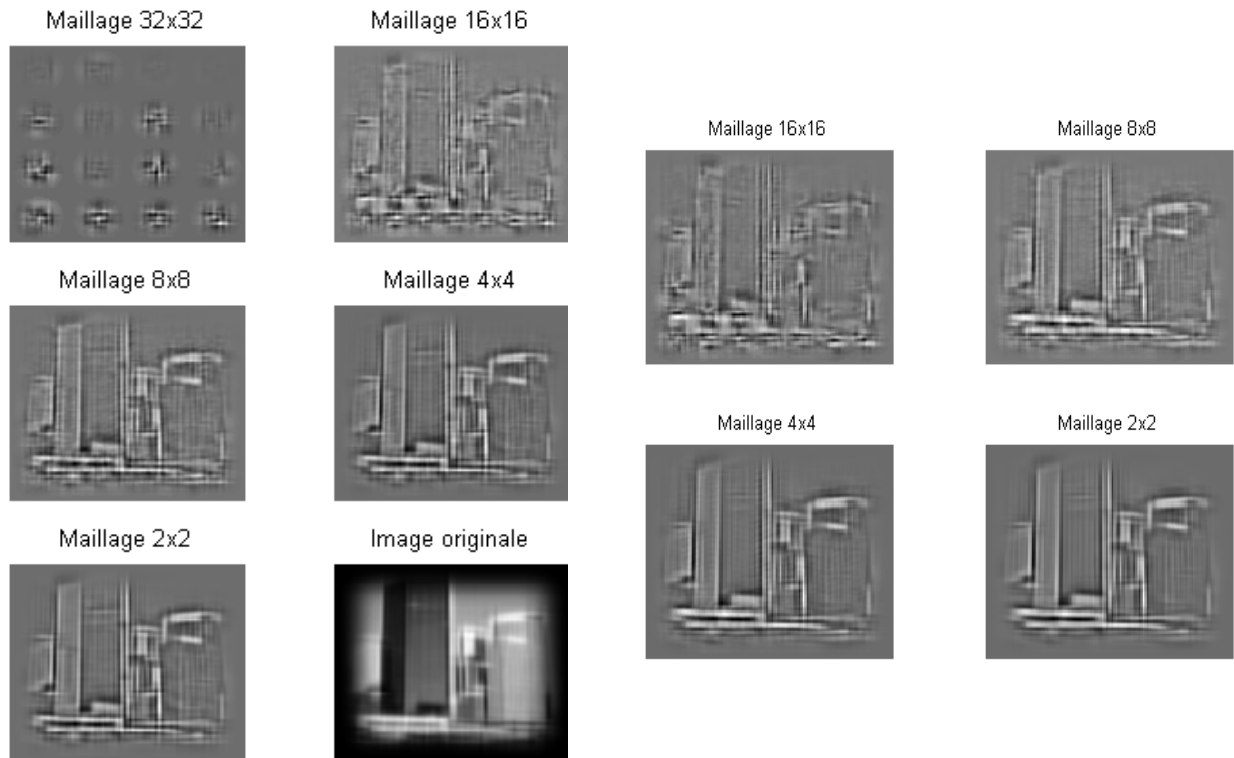
Figure 7 : Histogramme de c illustrant la structure des codes (a) ACI, (b) ACP

Pour les codes ACP, les unités de codage ont majoritairement une activité équivalente, donc proche de la moyenne des activités. En d'autres termes dans un code, il y a peu de composantes actives et un nombre plus important de composantes inactives. Au contraire, pour les codes ACI, il y a majoritairement cinq à dix unités (sur 30 détecteurs) qui sont plus actives que les autres pouvant porter l'information pertinente. Cela montre un caractère creux pour les codes ACI avec une dispersion plus grande sur les valeurs pour les codes ACI par rapport aux codes ACP.

Cette propriété sera illustrée au paragraphe suivant par la reconstruction des images suivant leur codage dans les deux bases de détecteurs.

V.2 Reconstruction

Ce caractère creux du code obtenu par projection d'une image sur une collection de filtres ACI, permet d'espérer la reconstruction de cette image avec un nombre restreint de primitives.



(a) (b)

Figure 8 : Reconstruction d'une image à l'aide de primitives ACI, (a) avec tous les détecteurs (b) avec seulement les détecteurs les plus actifs

La figure 8a, respectivement 9a illustre la reconstruction d'une image complète à partir de sa décomposition dans la base des détecteurs ACI et respectivement ACP. En tout premier lieu, notons que le critère optimisé par l'ACP est l'erreur quadratique moyenne de reconstruction. Deuxièmement, les primitives obtenues par Analyse en Composantes Indépendantes étant des filtres de type passe-bande, la valeur moyenne de l'image ne peut pas être restituée et on ne pourra reconstruire que la part des hautes fréquences spatiales de l'image (d'où le rendu visuel des figures 8a-8b). L'erreur quadratique a été calculée par rapport à la version passe haut de l'image.

Ces reconstructions d'image sont réalisées à partir d'une image découpée en blocs de taille 32 x 32. Les lieux de ces blocs sont espacés de façon régulière, mais avec plus ou moins de recouvrement. La valeur du maillage indiqué sur les figures représente l'écart entre 2 centres de bloc à reconstruire. Chaque bloc ainsi extrait est projeté sur la base de filtre ACI de la catégorie d'image correspondante et sur la base de filtres ACP. Enfin, une fois le code obtenu, l'image est reconstruite en superposant les primitives pondérées par les composantes du code ainsi obtenu.

Les filtres ACP et ACI sont situés dans le plan de Fourier pour les fréquences spatiales moyennes jusqu'à environ 10 cycles par image. La reconstruction s'effectuera donc dans cette gamme de fréquences, d'où l'aspect flou des images à la figure 8a pour les filtres ACP. Pour la reconstruction ACI, il y a en plus la perte des informations en basses fréquences par le caractère passe-bande des filtres. Les frontières (lignes, coins) sont donc rehaussées.

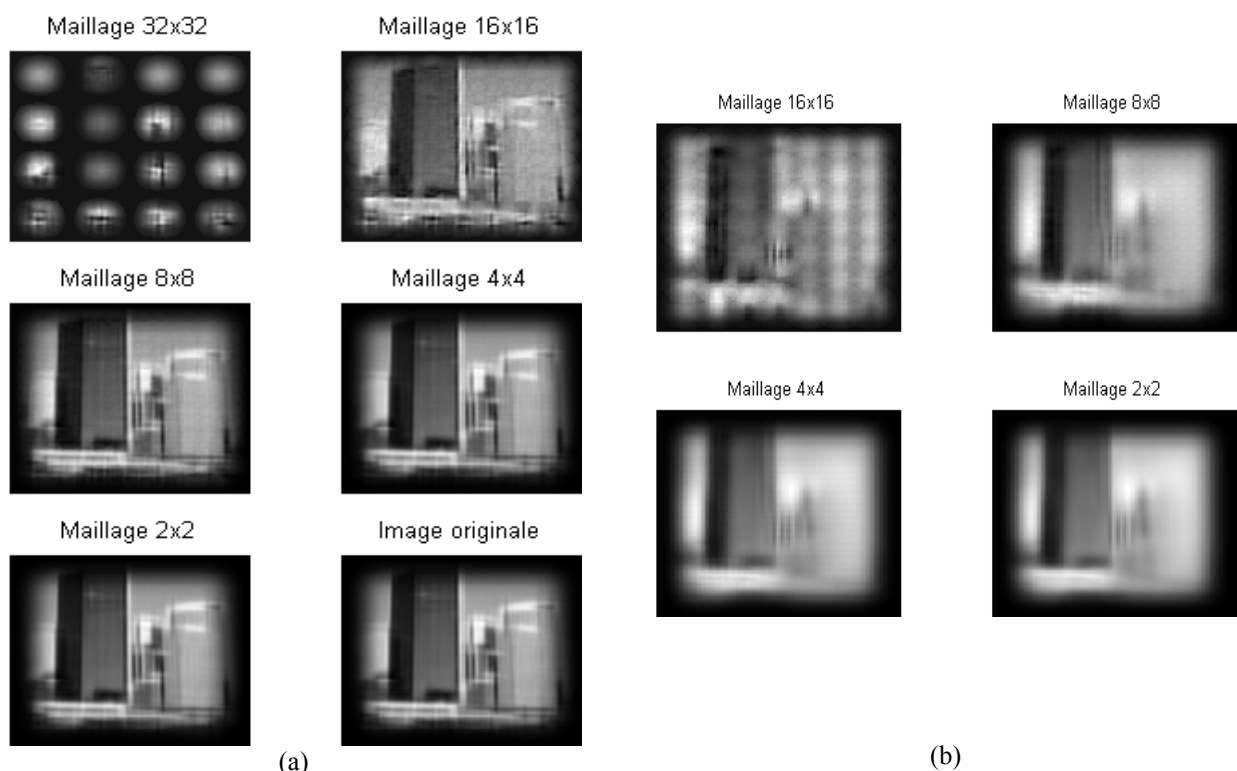


Figure 9 : Reconstruction d'une image à l'aide de primitives ACP, (a) avec tous les détecteurs (b) avec seulement les détecteurs les plus actifs

Les figures 8b et 9b illustrent le résultat de cette procédure de reconstruction en ne prenant que les détecteurs les plus actifs au sens défini au paragraphe 11.5.1. Dans le cas d'une reconstruction par filtres ACI, 25% des détecteurs sont sélectionnés et seulement 9% pour la reconstruction par filtres ACP. L'écart quadratique moyen entre ce résultat de reconstruction et celui, précédent, avec tous les filtres est plus important dans le cas ACP que dans le cas ACI (fig. 10) confirmant le rendu visuel aux figures 8a-8b d'une part et 9a-9b d'autre part.

On met en évidence ici, une conséquence du caractère creux du code ACI par rapport ACP. Dans le premier cas (ACI), un seuil sur le niveau d'activité permet de sélectionner les détecteurs les plus pertinents et les autres pouvant être négligés. Dans le second cas (ACP), une sélection sur le même critère élimine un trop grand nombre de détecteurs d'activité plus faible qui collectivement apporte une information non négligeable.

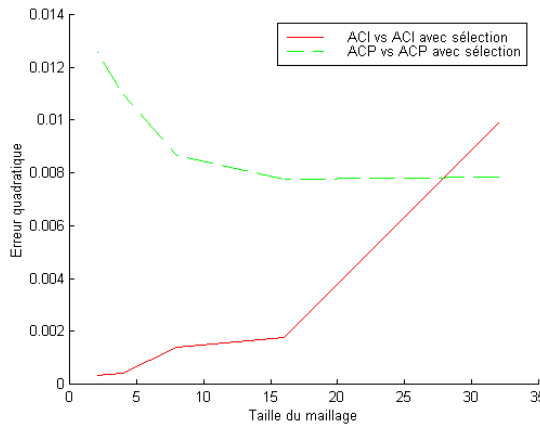


Figure 10 : Ecart quadratique moyen entre une reconstruction utilisant tous les détecteurs et une reconstruction utilisant seulement les détecteurs les plus actifs.

11.6. Propriété de catégorisation d'une base d'images

Pour analyser, les propriétés de catégorisation des filtres ACI, il faut préalablement définir la signature d'une image après son codage par filtre ACI, pour ensuite définir la « distance » entre 2 signatures. La tâche de catégorisation est simplement l'appartenance d'une image à l'une des catégories.

Les images de la base sont filtrées par les quatre populations Φ de primitives. Les distributions des énergies qui en résultent constituent la signature des images. Une densité de probabilité multivariée ($m=30$) est ainsi associée à chaque image. De ce fait, la différence entre images est implantée comme une différence entre distributions.

11.6.1 Signature d'une image par filtre ACI

Considérons les notations suivantes. On appelle $\Phi_v = \{\phi_{v_i}, i = 1..30\}$, $\Phi_o = \{\phi_{o_i}, i = 1..30\}$, $\Phi_i = \{\phi_{i_i}, i = 1..30\}$ et $\Phi_f = \{\phi_{f_i}, i = 1..30\}$, respectivement l'ensemble des primitives issues respectivement des villes, des paysages ouverts, des scènes d'intérieur et des paysages fermés. Explicitons la suite des notations en prenant comme exemple la famille des filtres issus des scènes fermées. Soit par exemple $I_{f_i}(x, y) = I(x, y) * \phi_{f_i}(x, y)$, avec * l'opérateur de convolution et $\phi_{f_i} \in \Phi_f$, l'image I filtrée par le $i^{\text{ème}}$ filtre de l'ensemble Φ_f des filtres (de même pour les autres catégories). Sur l'image $I_{f_i}(x, y)$, $e_{f_i}(x, y) = [I_{f_i}(x, y)]^2$ définit une variable aléatoire d'énergie. Soit e_f la variable aléatoire multivariée de dimension 30 générée à partir de la projection de l'image sur les 30 filtres de Φ_f ($e_f \in \mathfrak{R}^{30}$). Les composantes de e_f sont supposées indépendantes (de même que les composantes de e_v , e_o et de e_i), se traduisant par :

$$h_f(e_f) = h_f(e_{f_1}, \dots, e_{f_i}, \dots, e_{f_{30}}) = \prod_{i=1}^{30} h_{f_i}(e_{f_i}), \quad (4)$$

avec h_f la densité de probabilité de e_f et h_{f_i} la densité de probabilité marginale sur l'axe i .

VI.2 Différence entre images

Pour mesurer la différence entre images utilisant les propriétés d'indépendance des codes des images représentées par les filtres ACI, le choix s'est naturellement porté sur la divergence de Kullback, version symétrique de l'information de Kullback. Notons par ailleurs que cette mesure de dissimilarité est souvent utilisée (de même que la distance du Chi2) comme technique d'appariement dans les systèmes de recherche d'images [17,18].

Soient deux images, I_k et I_l . Par définition de l'information de Kullback entre deux distributions [1] et avec (4), on a :

$$\mathbf{K}(h_f^{(k)} \| h_f^{(l)}) = \int_{D_e} h_f^{(k)}(e) \cdot \log\left(\frac{h_f^{(k)}(e)}{h_f^{(l)}(e)}\right) \cdot de = \sum_{i=1}^{30} \int_{D_e} h_{f_i}^{(k)}(e) \log\left(\frac{h_{f_i}^{(k)}(e)}{h_{f_i}^{(l)}(e)}\right) de \quad (5)$$

La divergence de Kullback s'obtient donc par :

$$Div(h_f^{(k)}, h_f^{(l)}) = (K(h_f^{(k)} \| h_f^{(l)}) + K(h_f^{(l)} \| h_f^{(k)})). \quad (6)$$

Grâce à la propriété d'indépendance (eq. 4), la divergence s'évalue comme une somme de divergences sur les densités marginales monovariées :

$$Div(h_f^{(k)}, h_f^{(l)}) = \sum_{i=1}^{30} Div(h_{f_i}^{(k)}, h_{f_i}^{(l)}). \quad (7)$$

Cette formulation montre l'avantage ici de la propriété d'indépendance où une estimation en 30 dimensions se calcule par la sommation de 30 estimations élémentaires en 1 dimension. On contourne ainsi le problème de « l'espace vide » pour les estimations en grande dimension. Toutes les formulations précisées ici avec la famille de filtres Φ_f se déclinent identiquement avec les trois familles de filtres Φ_v , Φ_o et Φ_i .

Pour une image I_k donnée (taille 128 x 128), les densités de probabilité constituent la signature de cette image. Elles sont calculées (équation 4) à l'aide d'une technique d'estimation par noyaux gaussiens [19] discrétisée sur 256 points. Afin de limiter le temps de calcul, tout en ayant une estimation fiable, les calculs de densité de probabilité sont effectués sur 50% des pixels de l'image (soit $(128-31)^2 / 2 = 4852$ observations).

Les distances entre images, donc entre distributions sont estimées suivant les équations 6 et 7. Nous avons ainsi généré 4 matrices de Kullback 200x200 en projetant les images sur chacune des quatre bases Φ . Notons D_f , D_v , D_i et D_o ces quatre matrices, calculées comme somme de matrices élémentaires D_{f_i} (équation 7).

11.6.3 Expérimentations en catégorisation

La discrimination entre images suivant le contexte est réalisée par apprentissage supervisé. L'information utile est synthétisée dans la matrice de dissimilarités entre les images, calculée par la divergence de Kullback. Ainsi il n'y a pas de représentation euclidienne des images dans un espace de caractéristiques multidimensionnelles, et l'outil le plus simple de discrimination disponible pour cela, est alors le classifieur aux K plus proches voisins. Le paramètre K et la matrice de confusion sont estimés sur l'ensemble de la base d'images par validation croisée.

Préalablement à la catégorisation en utilisant l'ensemble des filtres extraits, nous avons cherché à ordonner chaque filtre selon leur pouvoir de discrimination. La démarche suivie est issue de l'Analyse Discriminante Linéaire, où le pouvoir de discrimination est évalué par le rapport de l'inertie inter-classe sur l'inertie totale. Soit par exemple $D_{f_{i^*}}$, la matrice élémentaire de Kullback pour les images projetées uniquement sur le filtre i^* pris arbitrairement dans la population Φ_f . Notons $\tau_f(i^*)$ le pouvoir discriminant du filtre i^* . Ce critère classique en Analyse Discriminante Linéaire peut également s'évaluer ici par la connaissance de la matrice de dissimilarités. En toute rigueur, il s'agira en fait de « pseudo-inertie » puisque les mesures utilisées ne sont pas des distances euclidiennes. On appelle $IT_{f_{i^*}}$, la pseudo-inertie totale des images projetées sur le filtre i^* de Φ_f et $IW_{f_{i^*}}$, la pseudo-inertie intra-classe. Par analogie avec les distances euclidiennes et en développant le calcul de l'inertie totale, on a :

$$IT_{f_{i^*}} = \frac{1}{2.N^2} \sum_{k=1}^N \sum_{l=1}^N [D_{f_{i^*}}(k, l)]^2, \quad \text{et } IW_{f_{i^*}} = \sum_{c=1}^4 q_c . IT_{f_{i^*}}(c), \quad (8)$$

avec $IT_{f_{i^*}}(c)$, la pseudo-inertie totale restreinte à une catégorie c donnée d'images et q_c le poids de chaque catégorie (ici =1/4). Le pseudo-pouvoir discriminant s'évalue alors par :

$$\tau_f(i^*) = 1 - \frac{IW_{f_{i^*}}}{IT_{f_{i^*}}}. \quad (9)$$

Dans chaque population Φ , les filtres sont classés par ordre décroissant de leur pouvoir de discrimination (fig. 11a). Les performances en catégorisation sont analysées en cumulant progressivement les matrices élémentaires associées aux filtres ainsi triés. La figure 11b illustre l'évolution du taux de reconnaissance en fonction du nombre de filtres ainsi ordonnés pour le calcul cumulatif de la matrice de Kullback selon l'équation 7.

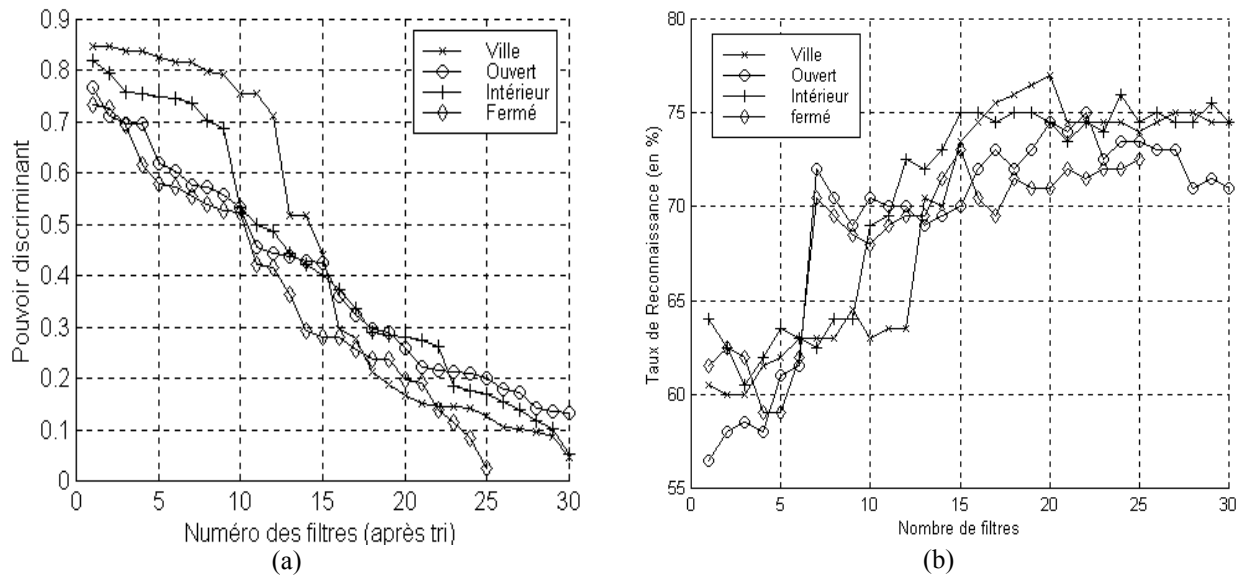


Figure 11 : (a) Evolution du pouvoir discriminant des filtres ACI ordonnés, (b) Evolution du taux de reconnaissance suivant le nombre de filtres ordonnés et utilisés pour la discrimination

Ces deux courbes montrent que les populations de filtres issues des villes et scènes d'intérieur sont les plus aptes à catégoriser la base d'images. Ces deux populations de filtres sont en fait assez voisines, car elles sont issues de deux classes d'images ayant des propriétés spectrales proches, comme nous l'avons signalé au paragraphe 11.5.1. De plus, on remarque qu'à partir de vingt filtres les informations détectées n'ont plus d'apport significatif dans la discrimination. Le taux de reconnaissance maximal est de 77 % avec la matrice de confusion obtenue avec les vingt filtres de « ville » les plus discriminants:

%	Ville	Ouvert	Intérieur	Fermé
Ville	84	0	12	4
Ouvert	0	74	14	12
Intérieur	10	4	78	8
Fermé	0	20	8	72

11.7. Conclusions

Les unités de codage obtenues par application du principe d'indépendance statistique de leurs activités sont cohérentes avec les études précédentes [3, 13, 18, 20]. Des données statistiques connues sur la population des champs récepteurs simples dans le cortex visuel se retrouvent ici dans cette population de détecteurs.

Nous avons montré comment des filtres obtenus par Analyse en Composantes Indépendantes pouvaient simplement catégoriser une base d'images. La principale originalité de cette étude par rapport aux approches de type « caractérisation par banc de filtres » est que les filtres utilisés ici sont extraits directement des catégories que l'on cherche à discriminer. Ils ne sont pas imposés a priori pour couvrir tout l'espace de Fourier. Il y a une relation forte entre les caractéristiques fréquentielles des détecteurs ainsi extraits et celles des images dont ils sont issus. Il s'agit donc d'une approche flexible pouvant se généraliser naturellement pour s'adapter à d'autres classes d'images.

Suivant cette approche, l'utilisation de l'ACI est fondamentale par la propriété d'indépendance qui rend plus robuste le processus aval d'estimation des signatures par les densités de probabilité marginales. Le problème de « l'espace vide » en grande dimension est ainsi contourné en se ramenant à des estimations mono-dimensionnelles. Une autre conséquence de cette propriété d'indépendance est la justification naturelle d'un modèle simple additif pour l'établissement d'une mesure de dissimilarité entre images. Si ce modèle est souvent utilisé de par sa simplicité dans les systèmes de reconnaissance, il est ici justifié théoriquement par la propriété d'indépendance statistique des codes.

Enfin, cette approche se met en perspective avec les modèles de perception visuelle par la similitude entre les détecteurs extraits par ACI et les cellules du cortex visuel. Ce travail se poursuivra (i) en étendant cette approche à des bases d'images plus complexes, (ii) en intégrant l'information de chrominance et (iii) en tenant compte de l'organisation spatiale dans les images par des statistiques locales, pour mettre en place des stratégies de type « Coarse to Fine ».

11.8 Références

- [1] Basseville M., (1989) Distance measures for signal processing and pattern recognition, *Signal Processing*, vol. 18, pp. 349-369.
- [2] Barlow, H.B. (1989) Unsupervised Learning. *Neural Computation*, vol. 1, pp. 295-315.
- [3] Bell A.J, Sejnowski T.J. (1997) The Independent Component of Natural Scenes are Edge Filter , *Vision Research*, vol 37, n° 23, pp 3327-3338.
- [4] Comon (1994) Independent Component Analysis – a new concept ?, *Signal processing*, vol 36, pp 287-314.
- [5] De Valois, R.L, Yund, E.W., Hepler,N (1982) The orientation and direction selectivity of cells in macaque cortex. *Vision Research*, vol 22, pages 531-544. 1982
- [6] De Valois, R.L, Albrecht, D.G., Thorell, L.G. (1982) Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, vol 22, pages 545-559. 1982
- [7] Donoho D.L. (2000) Nature vs. Math : Interpreting Independent Component Analysis in light of computational harmonic analysis, ICA'2000, Helsinki, june 2000 ;, pp. 459-470.
- [8] Field, D.J. (1994). What is the Goal of Sensory Coding ?, *Neural Computation*, vol. 6, pp. 559-601.
- [9] Guérin-Dugué A., Oliva A., (2000) Classification of Scene Photographs from Local Orientations Features, *Pattern Recognition Letters*,. vol 11, pp. 1135-1140.
- [10] Hyvärinen A., Oja.E. (1997) A Fast fixed-point algorithm for Independent Component Analysis, *Neural Computation*, vol 9, no 7, pp. 1483-1492.
- [11] Jutten C., Héroult J. (1991) Blind separation of sources : an adaptive algorithm based on neuromimetic architecture, *Signal processing*, vol. 24, pp. 1-10.
- [12] Labbi A., Bosch H., Pellegrini, Ch. (1999). Image Categorization using Independent Component Analysis, Workshop on Biologically Inspired Machine Learning, BIML'99, July 14 (invited talk), Crete, Greece.
- [13] Le Borgne H., Guérin-Dugué A.(2001) Propriétés des détecteurs corticaux extraits des Scènes Naturelles par Analyse en Composantes Principales, *Revue Valgo*, http://www.supelec-rennes.fr/acth/valgo/Valgo_Numero-01-01.html.
- [14] Li, Z., Attick, J.J. (1994) . Toward a Theory of Striate Cortex, *Neural Computation*, vol. 6, pp. 127-146.
- [15] Oliva A., Schyns P.G. (1997) Coarse blobs or fine edges ?, *Cognitive Psychology*, vol. 34, pp. 72-102.
- [16] Olshausen B.A, Field D.J (1997) Sparse coding with an overcomplete basis set : a strategy employed by V1 ?, *Vision Research*, vol 37, n° 23, pp. 3311-3325.
- [17] Pusicha J., Bushmann J.M., Rubner Y., and Tomasi C. (1999) Empirical Evaluation of Dissimilarity Measures for Color and Texture. *ICCV'99*, pages 1165-1172, Kerkyra-Greece, 1999.
- [18] Rubner Y., Tomasi C., Guibas L.F., (1998) A metric for distribution with application to images databases, *ICCV*, Bombay, India, pp. 59-66, jan. 1998.
- [19] Silverman B.W, (1986) *Density Estimation for Statistics and Data Analysis*, New York: Chapman and Hall.
- [20] Van Hateren J.H, Van der Schaaf A.(1998) Independent component filters of natural images compared with simple cells in visual cortex, *Proc. R. Soc. London*, B265, pp. 359-366.
- [21] Willmore B., Watters P.A., Tolhurst J., (2000), A comparison of natural-image-based models of simple-cell coding, *Perception*, vol. 29, pp. 1017-1040.