Propriétés des détecteurs corticaux extraits des Scènes Naturelles par Analyse en Composantes Indépendantes.

Hervé Le Borgne[1], Anne Guérin-Dugué[2]
L.I.S – I.N.P.G.
46 avenue Félix Viallet
F-38000 Grenoble Cedex
Mél : {hleborgn, guerin}@lis-viallet.inpg.fr

Résumé

L'Analyse en Composantes Indépendantes appliquée à une collection d'images permet d'obtenir des détecteurs qui ont des similarités avec les cellules du cortex visuel de type passe-bande orienté pour les cellules « simples », et réparties en fréquence et en orientation. Les détecteurs obtenus, en relation avec la cohérence spectrale des images dans la collection d'images considérée, permettent d'organiser sémantiquement les bases d'images. Cette étape est incontournable dans un système d'analyse de scènes complexes suivant une approche de perception visuelle « Coarse-to-Fine ». Les applications visées concernent l'indexation d'images par le contenu.

I. 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 [1, 6, 11, 14] 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) [5, 7, 9] 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 [2, 10, 15].

Cette voie est expérimentée ici dans le contexte de l'analyse de scènes complexes. Les applications visées sont en autres, les systèmes d'indexation de bases d'images par le contenu.

II. Objectif, Méthodologie

L'objectif est de construire automatiquement une base de détecteurs corticaux permettant d'extraire le contexte sémantique de la scène à partir d'une première étape d'analyse globale. Ensuite, en s'inspirant de la stratégie « Coarse-to-Fine » du système visuel humain, une analyse locale confirmant ou infirmant l'analyse globale fournira des cartes de traits caractéristiques saillants. Cette étude ne présente ici que la phase d'analyse globale par Composantes Indépendantes.

Les catégories sémantiques sont choisies à partir d'études psychophysiques [13] montrant la capacité des sujets humains à discerner des environnements de type « villes », « pièces d'intérieur », « paysages », …, à partir

[1] Hervé Le Borgne est financé par la région Rhône-Alpes dans le cadre du projet EMERGENCE 2000 « ASCII : Architecture Sémanti-Cognitique d'Indexation d'Images ».
[2] Anne Guérin-Dugué est actuellement Chargé de Recherche, détachée auprès de l'INRIA Rhône-Alpes.

d'images en basse résolution. Cette discrimination est basée sur les propriétés spectrales des images en basses fréquences spatiales.

Le codage des images est obtenu par Analyse en Composantes Indépendantes (ACI). L'ACI ou séparation de sources permet à partir de la connaissance d'observations de *n* signaux X, d'estimer les *m* sources primitives statistiquement indépendantes S, sous l'hypothèse que les observations sont issues d'un mélange inconnu additif des sources primitives. Si A(n,m) est une telle transformation linéaire, nous pouvons écrire :

$$X=A.S+\eta$$

où η est généralement un bruit gaussien souvent considéré comme une source indépendante ajoutéé à S. Le problème revient alors à inverser ce système d'équations et à calculer les sorties S par :

$$S= (A^t.A)^{-1}.A^t.X = B.X.$$

De nombreux critères ont été proposés pour estimer la matrice A, par exemple (i) la minimisation de l'information mutuelle entre les sorties (s_1, s_2, …, s_m) (c'est une mesure naturelle de dépendance entre les variables) [14], (ii) l'optimisation d'une fonction de contraste ou d'entropie utilisant les statistiques d'ordre 4 [3, 7]. Dans notre cas, l'algorithme utilisé est celui appelé « Fast ICA » [7] pour ses propriétés de convergence rapide et d'extraction séquentielle de sources, il est basé sur l'optimisation d'une fonction d'entropie (appelée entropie négative ou « negentropy ») appliquée sur les sorties.

En appliquant un tel modèle aux images, on considère donc qu'une image est une combinaison linéaire d'un ensemble de fonctions de base. Les coefficients de pondération de cette combinaison linéaire forment le code S de l'image, et les fonctions de base sont les colonnes de la matrice A. Les observations d'entrée X du système sont extraites des images sous forme d'imagettes (ou patches) de taille $n=l$ x l (l=32) extraites à des positions aléatoires dans les images. 30 imagettes sont extraites par image, elles sont préalablement apodisées par une fenêtre de Hamming (réduction de l'effet de Gibbs augmentant artificiellement les énergies des fréquences spatiales verticales et horizontales et effet de focalisation au centre de l'imagette). A la convergence, les *m* fonctions de base sont extraites à partir des *m* colonnes de A, formant les unités de codage ou détecteurs à utiliser comme des filtres 2D (taille l x l =n).

Deux difficultés fondamentales se posent alors. (i) Comment estimer ici le nombre *m* de sources, alors qu'aucun a priori lié à la physique de l'application guide ce choix ? (ii) Le modèle de mélange A peut ne pas être unique considérant l'extrême variabilité des images perçues. Notre démarche sera ici pragmatique. (i) Le nombre de sources est fixé empiriquement pour limiter la complexité tout en garantissant une exploitation satisfaisante des résultats. (ii) Un nombre a priori de modèles (4) est fixé par le nombre de classes sémantiques choisies pour lesquelles les images ont une cohérence spectrale (cf §III).

La méthodologie suivie se décompose en 3 étapes : (i) extraction et analyse de *m* (*m*=30) unités de codage appelées « simples » obtenus par ACI séparément sur chacune des catégories, (ii) idem pour *m* (*m*=120) unités de codage appelées « complexes » obtenus à partir de l'ensemble de la base d'images, et enfin (iii) illustration de l'auto-organisation de la base d'images grâce aux détecteurs « complexes ».

III. Description de la base d'images

La base d'images utilisée est constituée de 200 images de scènes naturelles (extraction de la base COREL) se répartissant en 4 catégories sémantiques de 50 images chacune : villes, pièces d'intérieur, paysages « fermés » (forêts, montagnes, vallées, …), paysages « ouverts » (plages, déserts, champs, …).

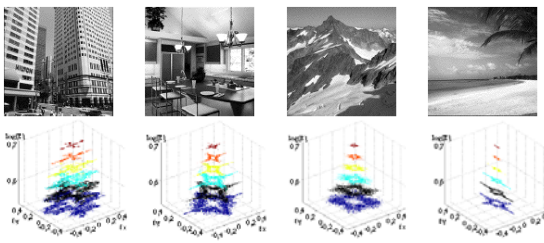

Figure 1 : Pour chacune des catégories, illustration des formes spectrales prototypiques sur une image et son spectre d'énergie.

Les images ont été sélectionnées pour former une base d'images ayant une grande variabilité intra-classe. Les images dans

chacune des classes ont une cohérence spectrale (cf. fig. 1) se modélisant par un gabarit spectral suivant l'anisotropie/isotropie en orientation du spectre d'énergie dans l'espace de Fourier [12].

IV. Extraction des filtres par catégorie

IV.1 Analyse et modélisation par filtrage de Gabor

Pour chaque catégorie, on extrait m=30 détecteurs simples à partir de 30 x 50 fenêtres (cf. annexe 1). 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 :

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

où F(u,v) est un détecteur 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). Via ces paramètres, l'analyse statistique de ces 4 ensembles de filtres montre des similitudes avec les populations de cellules simples du cortex visuel [15]. (i) Il y a un plus grand nombre de filtres sensibles aux orientations de référence (0°, 90°) (cf. fig. 2) et (ii) ces filtres sont plutôt anisotropes suivant leur orientation privilégiée (cf. annexe 1). La figure 3 met en évidence cette dépendance entre le facteur de forme (σ_x/σ_y) et l'orientation du filtre $(tg\theta=v_0/u_0)$. (iii) Les filtres sur les orientations obliques sont plutôt sphériques $(\sigma_x=\sigma_y)$. (iv) On observe également une relation décroissante entre la bande passante et la résolution du filtre (conséquence du spectre d'énergie des images naturelles qui est globalement en « 1/f »).

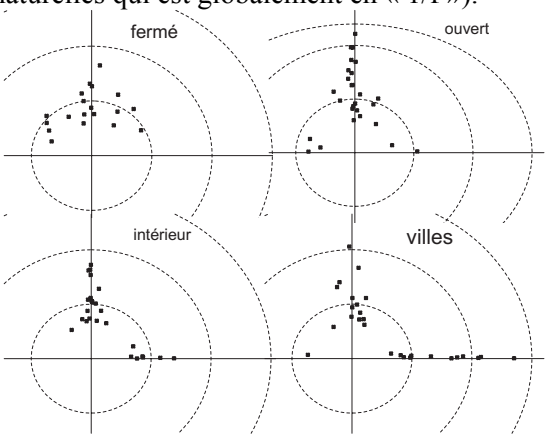

Figure 2 : Localisation des fréquences centrales (u_0,v_0) des filtres dans le demi plan spectral supérieur. Graduations tous les 4 cycles par image.

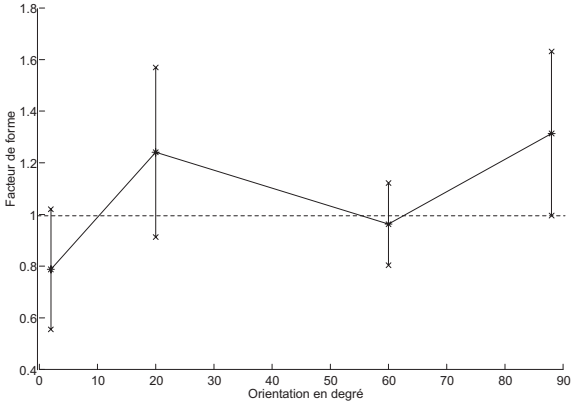

Figure 3 : Facteur de forme vs orientation (moyenne et +/- écart-type).

Enfin, la localisation des filtres dans l'espace de Fourier (cf. fig. 2), est en adéquation avec les caractéristiques spectrales de la catégorie d'images dont le filtre a été extrait . Les filtres provenant des catégories « villes » et « pièces d'intérieur » mettent essentiellement en valeur les orientations locales verticales et horizontales. Au contraire, dans les cas des images « fermées » ne présentant pas d'orientation privilégiées, les filtres extraits se distribuent sur toutes les orientations. Les images « ouvertes » conduisent à des filtres orientés verticalement, donc détectant la ligne d'horizon caractéristique de ces images.

IV.2 Organisation topologique

Dans les aires visuelles, les champs récepteurs sont organisés rétinotopiquement et par continuité en fréquence et en orientation. Pour une région spatiale quelconque, cette famille de détecteurs s'auto-organise en 2D à la surface corticale. La figure 4 illustre l'organisation topologique 2D obtenue en respectant au mieux les similitudes par corrélation des filtres : deux filtres ainsi similaires sont placés dans le plan tels qu'ils ont entre eux une faible distance euclidienne. L'auto-organisation par prolongement euclidien de la matrice des corrélations entre filtres a été obtenue par Analyse en Composantes Curvilignes (ACC) [4]. C'est un algorithme de type « MultiDimensionnal Scaling ». L' illustration de la figure 4 montre une organisation cohérente des filtres, au niveau des orientations et des résolutions permettant a posteriori

l'émergence de groupes cohérents de filtres actifs pour un stimulus donné [8].

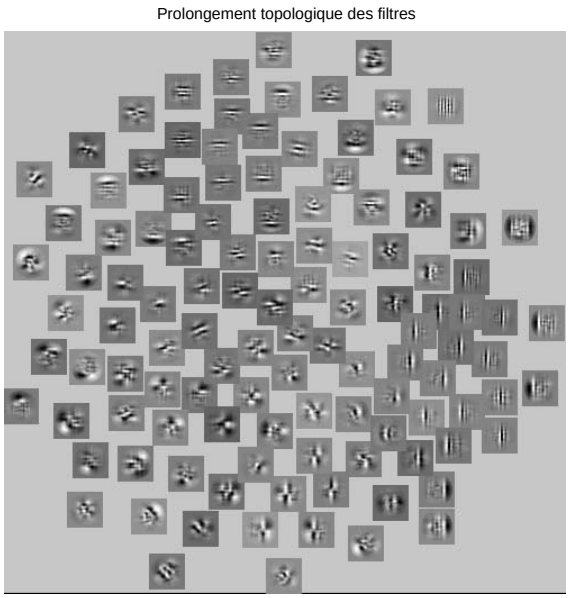

Figure 4 : Organisation topologique des filtres « simples » par Analyse en Composantes Curvilignes.

V. Extraction des filtres sur toute la base

Les détecteurs « complexes » (*m*=120) sont extraits par ACI appliquée sur toute la base (30x200 fenêtres) (cf. annexe 2). Le gabarit fréquentiel est beaucoup plus complexe et ne s'exprime pas par un simple modèle de Gabor. Par contre, il suggère des interactions pouvant se modéliser par combinaison de filtres « simples ».

V.1 Caractérisation par le pouvoir discriminant

En considérant ces détecteurs comme une deuxième couche de filtres, nous avons cherché les filtres les plus aptes à organiser sémantiquement une base d'images selon les 4 catégories à partir de l'énergie globale fournie par les détecteurs. L'information traitée est le tableau d'énergie pour les 200 images filtrées par les 120 détecteurs. A partir de cette base de donnée, la recherche des filtres les plus discriminants s'effectue en maximisant le rapport d'inertie inter-classe sur l'inertie intra-classe [10]. Pour chaque filtre i et chaque classe k, le pouvoir discriminant est :

$$W(i,k) = \frac{\left|M_k(i) - M_{\bar{k}}(i)\right|}{S_k(i) + S_{\bar{k}}(i)} \text{ où k=1…4,}$$

avec $M_k(i)$ la moyenne des réponses sur la classe k du filtre i ($\bar{k}$ = classe « non-k ») et $S_k(i)$ l'écart-type des réponses.
On sélectionne ainsi les 10 filtres les plus discriminants dans chaque catégorie ; certains filtres étant très discriminants pour plusieurs catégories, 29 filtres différents sont ainsi retenus.

V.2 Organisation sémantique des scènes

Cette sélection étant effectuée, la dissimilarité entre 2 images est modélisée par la distance euclidienne entre les deux profils d'énergie en sortie des filtres. La représentation euclidienne en 2D des images est obtenue par ACC [4] réalisant ainsi une réduction non linéaire de dimension de 29D à 2D (cf. fig. 5).

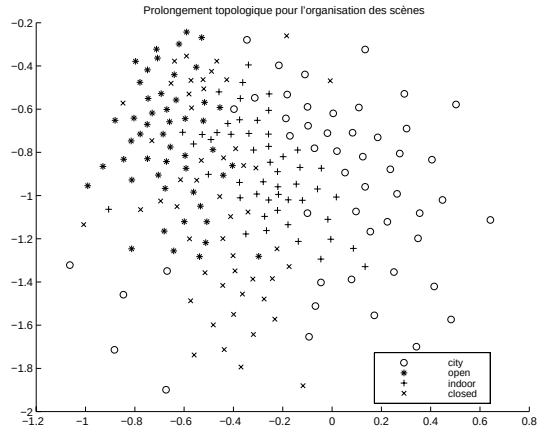

Figure 5 : Organisation topologique des scènes.

Cette représentation plane met en évidence l'auto-organisation des images globalement cohérente avec les classes sémantiques et pourra être le principe initial d'un système de navigation dans une base d'images, respectant notre perception visuelle.

VI. Caractère épars et reconstruction

VI.1 Caractère épars des codes

Les études [5, 6, 14] 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 [14, 15]. 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. Pour une collection d'images sémantiquement semblables, une majorité de détecteurs resterait inactive. De tels codes sont qualifiés de « codes épars ».

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 épars, chaque composante est essentiellement inactive, si bien que sa densité de probabilité présente un pic important autour de zéro.

Afin de valider cette hypothèse, nous avons comparé le caractère épars 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.

Soit $S=(S_1,\ldots, S_{30})$, le code pour une observation (imagette l x l extraite d'une image) représentée dans sa base de détecteurs « simples ». Le caractère épars 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 :

$c = \sum_{i=1}^{30} F_1(S_i)$, avec F_1 la fonction à seuil ainsi définie, $F_1(S) = 0$ si $\mu-\sigma < S < \mu+\sigma$, et $F_1(S) = 1$, 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 6a, respectivement 6b 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.

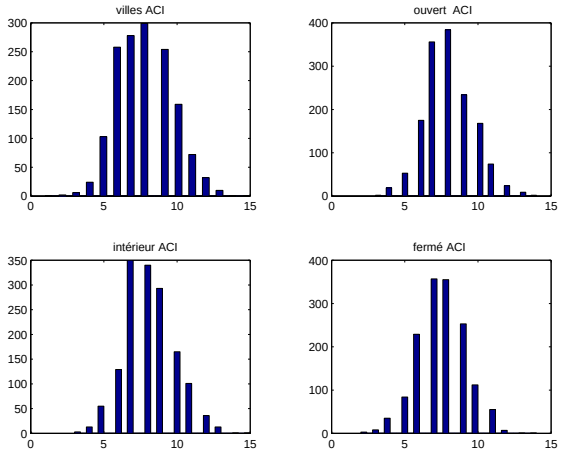

Figure 6a : Histogramme de c illustrant la structure des codes ACI.

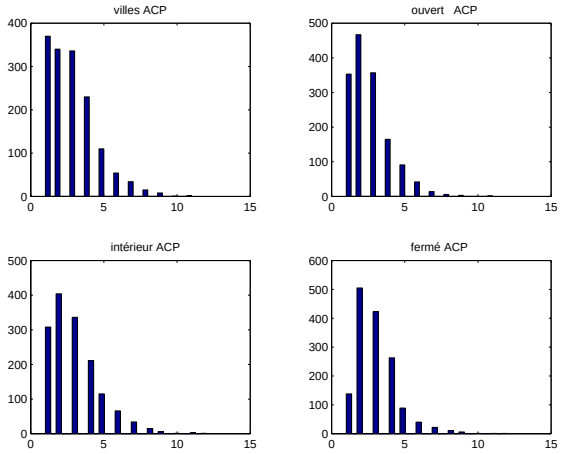

Figure 6b : Histogramme de c illustrant la structure des codes ACP.

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 plus épars pour les codes ACI que pour les codes ACP.

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

VI.2 Reconstruction et Information détectée par les filtres ACI

Le caractère épars du code obtenu par projection d'une imagette sur une collection de filtres ACI, permet d'espérer la reconstruction de cette image avec un nombre restreint de primitives. Ceci dénoterait un propriétés de compression de données avec des pertes minimes.

La figure 8a, respectivement 9a illustre la reconstruction d'une images 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. On obtient donc un rendu visuel meilleur avec cette technique qu'avec l'ACI. Deuxièmement, les primitives obtenues par Analyse en Composantes Indépendantes étant des filtres passe-bande, la valeur moyenne de l'image ne peut pas être restituée et on ne pourra reconstruire que la part des hautes fréquences spatiale de l'image (d'où le rendu visuel des figures 8a-8b).

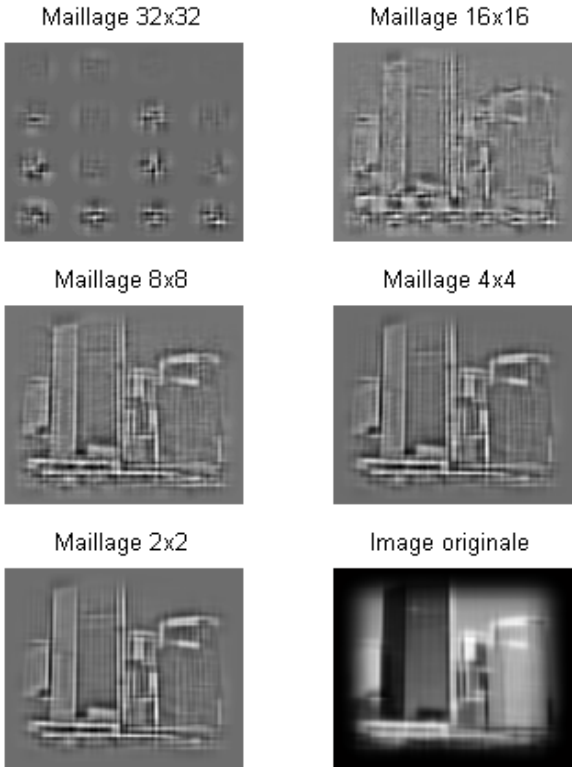

Figure 8a : Reconstruction d'une image à l'aide de primitives ACI.

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.

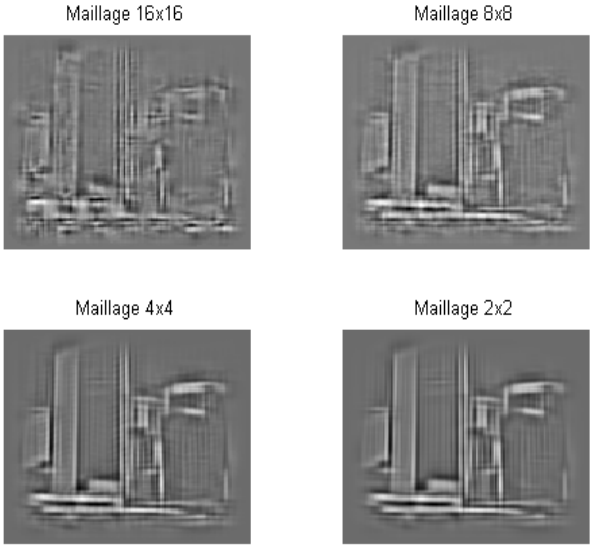

Figure 8b : Idem figure 8a, mais avec une reconstruction avec seulement les détecteurs les plus actifs.

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 9a 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.

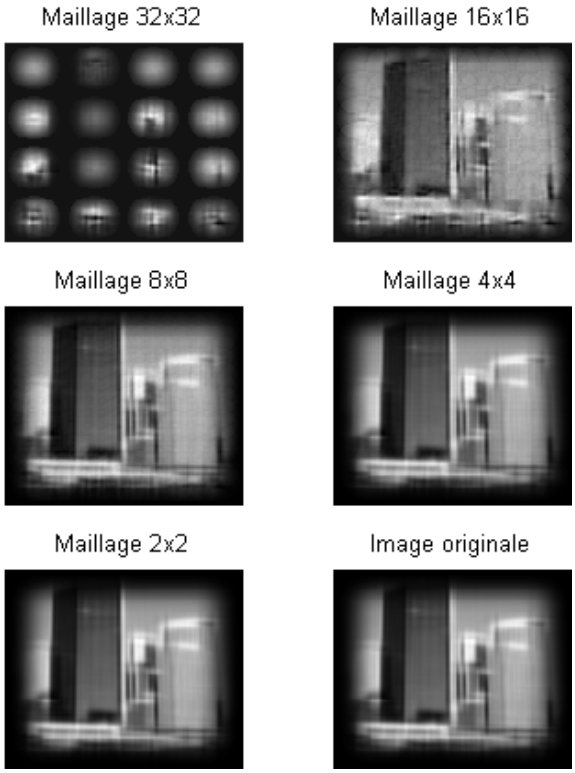

Figure 9a : Reconstruction d'une image à l'aide de primitives ACP.

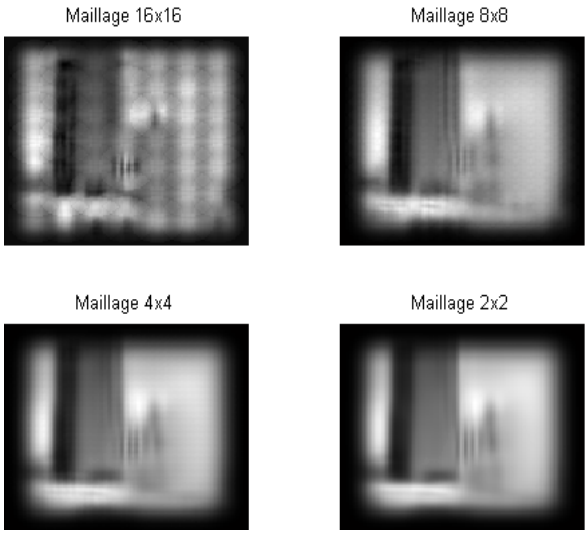

Figure 9b : Idem figure 9a, mais avec une reconstruction avec seulement les détecteurs les plus actifs.

Les figures 8b et 9b illustre le résultats de cette procédure de reconstruction en ne prenant que les détecteurs les plus actifs au sens défini au paragraphe VI.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.

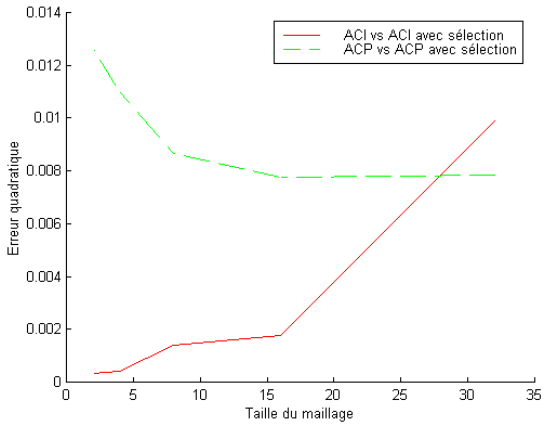

Figure 10 : Ecart quadratique moyen entre une reconstruction utilisant tous les détecteurs et une utilisant seulement les détecteurs les plus actifs.

On met en évidence ici, une conséquence du caractère épars 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.

VII. Conclusions

Les unités de codage obtenues par application du principe d'indépendance statistique de leurs activités sont cohérentes avec les études précédentes [2, 10, 14, 15]. Des données statistiques connues sur la population des champs récepteurs simples dans le cortex visuel se retrouvent ici dans cette population de détecteurs. Il est intéressant de noter l'apparition de détecteurs plus « complexes » quand la variabilité des stimuli visuels augmente.

L'organisation topologique 2D des détecteurs à partir des dépendances résiduelles fait émerger un placement continu suivant leur sélectivité en orientation et en résolution, devant mettre en évidence des motifs connexes d'activation, facilitant l'interprétation des stimuli de scènes.

Le codage par les détecteurs ACI peut être qualifié de creux ou d'épars, contrairement à celui obtenu par les détecteurs ACP. Cette propriété indique que dans la population des détecteurs, pour une observation donnée, certains ont une valeur d'activité importante rendant négligeable la contribution des détecteurs moins actifs. Ainsi, une observation pourra être représentée par un plus petit nombre de primitives. Avec les détecteurs ACP, le codage est plus distribué sur toutes les unités et il faut donc quasiment toutes les conserver pour bien représenter l'observation.

Notre objectif applicatif est de concevoir une architecture d'indexation d'images, à partir de la navigation dans une base en élaborant des similarités entre images en accord avec notre propre perception visuelle. Ces travaux se poursuivront en construisant une représentation locale de traits détectés, guidée par cette analyse globale.

Références

[1] Barlow, H.B. (1989) Unsupervised Learning. Neural Computation, vol. 1, pp. 295-315.

[2] Bell A.J, Sejnowsky T.J. (1997) The « Independant Component » of Natural Scenes are Edge Filter », Vision Research, vol 37, n° 23, pp 3327-3338.

[3] Comon (1994) Independant Component Analysis – a new concept ?, Signal processing, vol 36, pp 287-314.

[4] Demartines P., Hérault J. (1997) Curvilinear Component Analysis : A Self-Organising Neural Network for Non Linear Mapping of Data Sets, IEEE Transactions on Neural Networks, vol. 8, n° 1, pp. 148-154

[5] 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.

[6] Field, D.J. (1994). What is the Goal of Sensory Coding ?, Neural Computation, vol. 6, pp. 559-601.

[7] Hyvärinen A., Oja.E. (1997) A Fast fixed-point algorithm for Independent Component Analysis, Neural Computation, vol 9, no 7, pp. 1483-1492.

[8] Hyvärinen A, Hoyer P.O., Inki M., Topographic Inpedendent Component Analysis : Visualizing the dependence structure, ICA'2000, Helsinki, june 2000, pp. 591-596.

[9] Jutten C., Hérault J. (1991) Blind separation of sources : an adaptative algorithm based on neuromimetic architecture, Signal processing, vol. 24, pp. 1-10.

[10] 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.

[11] Li, Z., Attick, J.J. (1994) . Toward a Theory of Striate Cortes, Neural Computation, vol. 6, pp. 127-146.

[12] Oliva A., Schyns P.G. (1997) Coarse blobs or fine edges ?, Cognitive Psychology, vol. 34, pp. 72-102.

[13] Oliva A., Torralba A.B., Guérin-Dugué A., Hérault J., (1999) Super-Ordinate representation of scenes from power spectrum shapes, CIR-99, The challenge of image retrieval, Newcastle, march 1999.

[14] 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.

[15] 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.