Analyse d'Images par Composantes Indépendantes : Application à l'Organisation Sémantique de Base d'Images

Hervé Le Borgne, Anne Guérin-Dugué
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. Ces détecteurs corticaux pourraient résulter de l'application d'un principe de réduction de redondance par indépendance statistique de leurs activités [1, 5, 10, 13].

Parmi les techniques proposées pour construire de tel code, l'Analyse en Composantes Indépendantes (ACI) [4, 6, 8] 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 [2,9,14].

Cette voie est expérimentée ici dans le contexte d'analyse de scènes complexes. Les applications visées sont les système 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, une analyse locale confirmant ou infirmant l'analyse globale fournira des cartes de traits caractéristiques saillants (stratégie « Coarse-to-Fine »). Seule l'étape d'analyse globale sera illustrée ici.

Les catégories sémantiques sont choisies à partir d'études psychophysiques [11] montrant la capacité des sujets humains à discerner des environnements de type « villes », « pièces d'intérieur », « paysages », …, à partir d'images en basse résolution.

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, elle est estimée en optimisant un critère d'indépendance sur les sorties. En appliquant un tel modèle aux images :

$X=A.S+\eta$, et $S= (A^t.A)^{-1}.A^t.X = B.X$,

on considère qu'une image est une combinaison linéaire d'un ensemble de fonctions de base

indépendantes. 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 ne peut 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).

Plus précisément l'algorithme utilisé est la « Fast ICA » [6] pour ses propriétés de convergence rapide et d'extraction séquentielle de sources. Les observations d'entrée sont des fenêtres de taille $n=l$ x l (l=32) extraites à des positions aléatoires dans les images (30 fenêtres apodisées -Hamming- par image). A la convergence, *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).

La méthodologie suivie se décompose en 3 étapes : (i) extraction et analyse de *m* (*m*=30) unités de codage « simples » obtenus par ACI séparément sur chacune des catégories, (ii) idem pour *m* (*m*=120) unités de codage « 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, …).

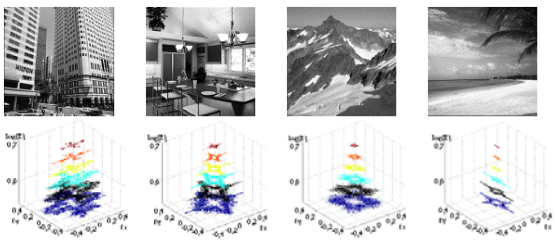

Figure 1 : Pour chacune des catégories, illustration des formes spectrales prototypiques sur une image et son spectre d'énergie.

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 [14]. (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 ($\sigma_x=\sigma_y$).

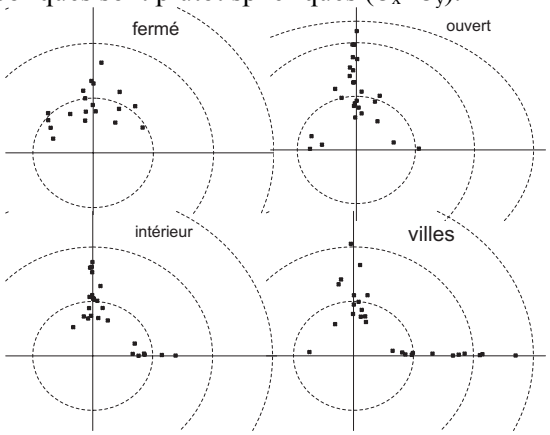

Figure 2 : Localisation des fréquences centrales (u_0,v_0) des filtres dans le demi plan spectral supérieur. Graduations tous les 4 cycles par image.

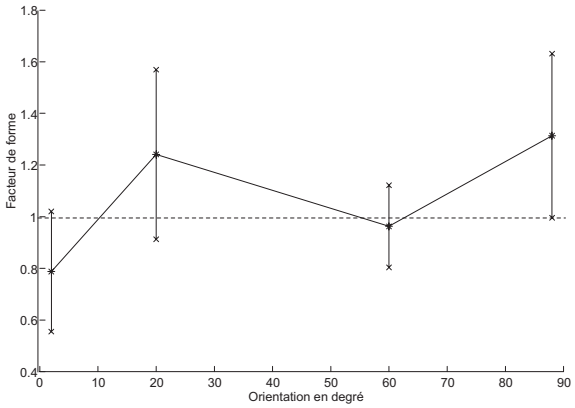

Figure 3 : Facteur de forme vs orientation (moyenne et +/- écart-type)

(iv) On observe également une relation décroissante entre la bande passante et la résolution du filtre (conséquence du spectre des images naturelles, globalement en « 1/f »).
Enfin, la localisation des filtres (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. Ici, les détecteurs extraits sont supposés être complètement indépendants. Or il existe une dépendance résiduelle que l'on met en évidence par une simple mesure de corrélation. La figure 4 illustre l'organisation topologique 2D des filtres obtenue en respectant les similitudes par corrélation par les distances euclidiennes dans le plan. Cette auto-organisation a été obtenue par Analyse en Composantes Curvilignes (ACC) [3]. Cette illustration 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 [7].

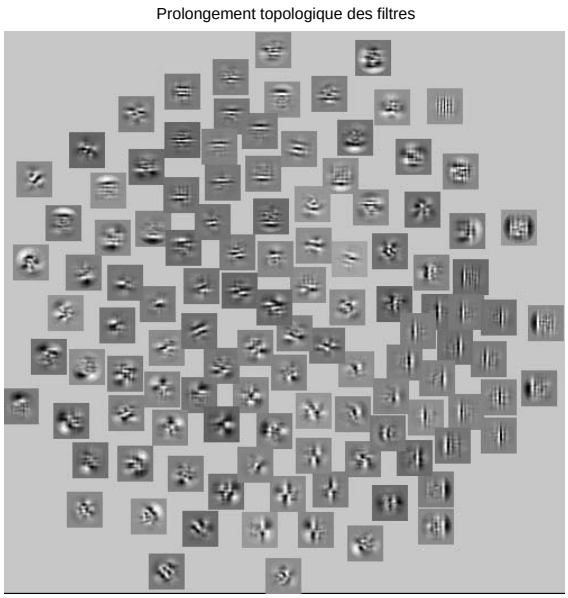

Figure 4 : Organisation topologique des filtres « simples » par Analyse en Composantes Curvilignes

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, 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 : tableau d'énergie (200 images x 120 filtres). Cette recherche s'effectue en maximisant le rapport d'inertie inter-classe sur l'inertie intra-classe [9]. Pour chaque filtre i, le pouvoir discriminant pour chaque classe 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 : 29 filtres différents sont ainsi retenus.

V.2 Organisation sémantique des scènes

Cette sélection étant effectuée, la dissimilitude entre 2 images est modélisée par la distance euclidienne entre les deux profils d'énergie sur ces filtres. La représentation euclidienne en 2D des images est obtenue par ACC [3] réalisant la réduction non linéaire de dimension de 29D à 2D (cf. fig 5). Cette représentation plane met en évidence l'auto-organisation des images globalement cohérente avec les classes sémantiques et pourra être le principe original d'un système de navigation dans une base d'images, respectant notre perception visuelle.

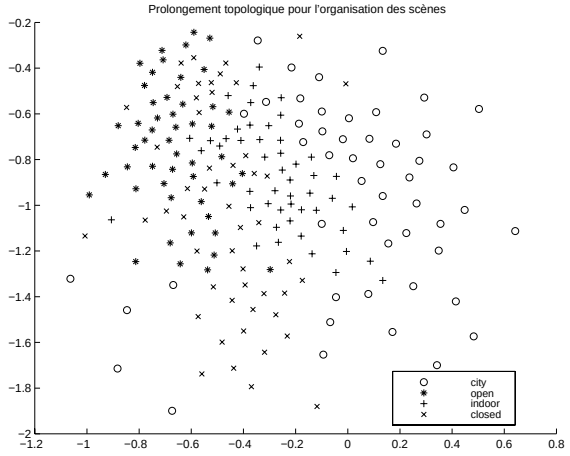

Figure 5 : Organisation topologique des scènes

VII. Conclusions

Les unités de codage obtenues par application du principe d'indépendance statistique de leurs activités sont cohérente s avec les études précédentes [2, 9, 13, 14]. 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. 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] 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

[4] 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.

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

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

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

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

[9] 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.

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

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

[12] 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.

[13] 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.

[14] 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.