

Comparaison de la composition de deux images, et application à la recherche automatique

Yann GOUSSEAU¹,

¹Laboratoire Traitement du Signal et des Images
Ecole Nationale Supérieure des Télécommunications
46 rue Barrault, 75643 Paris Cedex 13
gousseau@tsi.enst.fr

Résumé – Nous présentons une méthode nouvelle permettant de tester la proximité de deux images, en rejetant une hypothèse d’indépendance. Cette méthode repose sur une représentation géométrique de l’image : ses ensembles de niveaux sont vus comme des fermés aléatoires et sont comparés grâce à des résultats élémentaires de géométrie stochastique. Le critère de ressemblance résultant compare les grandes lignes de l’organisation globale des images. Enfin, nous appliquons ce test à la recherche automatique dans une base d’images à partir d’un exemple, qui peut être une image ou un schéma de composition. Nous indiquons également comment décider si la mise en correspondance de deux images est pertinente, en fixant les seuils de comparaison grâce à des réalisations d’un modèle aléatoire d’image, le modèle feuilles mortes.

Abstract – This communication introduces a new method to test the similarity between two images, by rejecting a test of independence. This approach relies on a geometric representation of the image : its level sets are modeled as random closed sets, and their comparison is achieved thanks to elementary results from stochastic geometry. The resulting similarity criterium enables to compare the global organization of images. Eventually, we apply our test to automatic query by example in an image database (examples can be provided either by an image or a sketch). An important point we address is the one of the meaningfulness of the result of the query. We fix the thresholds of our method using a random model of images, the so called dead leaves model from mathematical morphology.

1 Introduction

La recherche automatique d’images dans une base est un domaine de recherche particulièrement actif ces dernières années. Les critères de recherche les plus utilisés reposent sur des attributs de couleur et/ou de texture, par l’intermédiaire de réponses à des bancs de filtres. La cohérence de ces attributs est prise en compte la plupart du temps par des histogrammes locaux ou par une segmentation de l’image, [3], [4] (les indications bibliographiques ne prétendent en aucun cas à l’exhaustivité). Ces méthodes prennent peu en compte l’organisation spatiale de ces régions. Plusieurs travaux touchent à l’organisation globale de l’image, mais reposent sur des procédures non automatiques, ou réduisent l’organisation de l’image à un schéma très simplifié dans lequel le nombre de régions est un paramètre ([6]). De nombreux auteurs se sont intéressés au problème de la comparaison d’organisations spatiales par structures symboliques (voir par exemple [1]), mais s’intéressent aux positions relatives d’objets, sans aborder le problème de la composition de l’image. Dans [11], il est proposé de comparer la composition de l’image, après segmentation par couleurs, en représentant vectoriellement l’organisation spatiale des régions (après un parcours séquentiel de l’image). Les résultats sont intéressants, mais ne permettent de distinguer que des relations verticales et occultent la notion de région.

Dans cette communication, nous introduisons une méthode simple pour comparer la proximité des régions d’une image, reposant sur la modélisation de ces régions comme ensembles aléatoires. Nous représentons une image I par ses ensembles

de niveau inférieurs, c’est à dire par les

$$\chi_\lambda(I) = \{x/I(x) \leq \lambda\},$$

pour $\lambda \in \mathbb{R}$. Nous supposons que les χ_λ sont les réalisations d’ensembles fermés aléatoires (FA).

2 Une mesure d’indépendance de deux fermés aléatoires

Un fermé aléatoire F est une application mesurable d’un espace de probabilité dans l’espace \mathcal{F} des ensembles fermés de \mathbb{R}^2 (fermés au sens de la topologie usuelle), muni de la tribu des boréliens pour une topologie adéquate ([8]). Pour l’application qui nous intéresse, il suffit de savoir que deux FA sont indépendants si, pour tout ensembles compacts K_1 et K_2 ,

$$\begin{aligned} Pr(K_1 \cap F_1 \neq \emptyset \text{ et } K_2 \cap F_2 \neq \emptyset) \\ = Pr(K_1 \cap F_1 \neq \emptyset) Pr(K_2 \cap F_2 \neq \emptyset). \end{aligned} \quad (1)$$

Vérifier cette relation pour tout compacts K_1 et K_2 (même si l’on peut se restreindre à des sous-classes) est bien entendu irréalisable, et nous nous restreignons à $K_1 = K_2 = O$, l’origine. La quantité que nous considérons pour tester l’indépendance de deux FA, F_1 et F_2 , est

$$T(F_1, F_2) = \frac{1 - Pr(O \in F_1 \cap F_2)}{1 - Pr(O \in F_1)Pr(O \in F_2)}. \quad (2)$$

On a ainsi $T = 1$ dans le cas de FA indépendants, et $T = (1 + Pr(O \in F_1))^{-1}$ si $F_1 = F_2$.

Par ailleurs, si un FA F est stationnaire, alors

$$Pr(O \in F) = E\nu(F \cap [0, 1]^2),$$

formule où E représente l'espérance mathématique, et ν la mesure de Lebesgue sur \mathbb{R}^2 . Si de plus on suppose que F est ergodique, alors

$$Pr(O \in F) = \lim_{R \rightarrow \infty} \frac{\nu(F \cap B(0, R))}{\nu(B(0, R))},$$

où $B(0, R)$ est une boule de rayon R . Ce résultat permet l'estimation de la quantité (2) à partir d'une seule réalisation des FA. Dans ce cas, le test consiste simplement en une comparaison de la proportion d'espace occupée dans une fenêtre par l'intersection de deux ensembles au produit des proportions d'espace occupées par chacun des ensembles.

3 Comparaison de deux images et recherche automatique

Pour comparer deux images I_1 et I_2 , nous nous intéressons aux ensembles $\chi_\lambda(I_1)$ et $\chi_\lambda(I_2)$, pour $\lambda = \lambda_1, \dots, \lambda_n$. En vertu des résultats rappelés au paragraphe précédent, et notant Ω le domaine de l'image, nous utiliserons la quantité

$$C_\lambda(I_1, I_2) = \frac{1 - \nu(\chi_\lambda(I_1) \cap \chi_\lambda(I_2)) \cdot \nu(\Omega)^{-1}}{1 - \nu(\chi_\lambda(I_1)) \nu(\Omega)^{-1} \nu(\chi_\lambda(I_2)) \nu(\Omega)^{-1}},$$

qui constitue une approximation de l'expression (2) appliquée aux ensembles de niveau λ de I_1 et I_2 . Nous supposons donc que les ensembles de niveau sont des FA stationnaires et ergodiques, et que la taille de l'image est suffisante pour remplacer les probabilités de la formule (2) par une moyenne spatiale. Enfin, nous comparerons deux images grâce à

$$C(I_1, I_2) = \sum_{i=1}^n C_{\lambda_i}(I_1, I_2).$$

Précisons que nous n'utilisons pas les ensembles de niveau supérieurs car le test correspondant mène à une comparaison similaire sur les ensembles inférieurs, ce qui se voit aisément en passant au complémentaire dans (2). Remarquons que les ensembles $\chi_{\lambda_i}(I)$ et $\chi_{\lambda_j}(I)$ ne sont clairement pas indépendants (l'un est inclus dans l'autre), mais que si I_1 et I_2 sont indépendantes, on peut montrer que $C(I_1, I_2) = n$ (chacun des facteurs du produit vaut un). Remarquons également que la fonction C n'est pas une distance, en particulier $C(I, I) \neq 0$.

Si nous voulons tester l'indépendance de deux images I_1 et I_2 , nous devons décider de seuils S_1 et S_2 , puis comparer $C(I_1, I_2)$ à $1 - S_1$ et $1 + S_2$. Remarquons qu'il y a deux sortes d'écart à l'indépendance. Soit $C(I_1, I_2) < 1 - S_1$, soit $C(I_1, I_2) > 1 + S_2$. Ces deux cas de figure peuvent s'interpréter comme suit : le premier cas correspond à deux images dont les ensembles de niveau sont plus proches que dans le cas d'indépendance, le deuxième à deux images dont les ensembles sont plus distants que dans le cas d'indépendance. Nous utilisons ce test pour quantifier la ressemblance entre images, et nous restreignons donc au cas $C(I_1, I_2) < 1 - S_1$: les deux images seront alors déclarées semblables.

Pour fixer S_1 , nous choisissons d'utiliser un modèle aléatoire d'images, le modèle « feuilles mortes » de la morphologie mathématique ([7], [10]), qui consiste en la superposition d'objets aléatoires. Ce modèle a la particularité de rendre compte

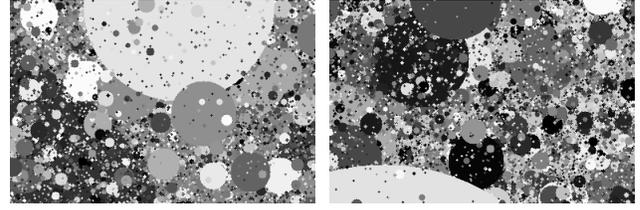


FIG. 1 – Deux réalisations du modèle feuilles mortes, avec des disques dont les rayons sont distribués selon la loi r^{-3} .

du phénomène d'occultation (les objets opaques se cachent les uns les autres), et de rendre compte de nombreuses statistiques des images naturelles lorsque les objets le constituant ont certaines distributions en taille ([2], [5]). Suivant ces études, nous choisissons d'utiliser le modèle correspondant à des disques (dans un souci de simplicité et d'isotropie) de rayon suivant la loi $r^{-\alpha}$, voir la figure 1. Ne sachant pas, dans le cadre du modèle feuille morte, calculer le risque du test proposé, nous l'évaluons empiriquement. Nous fixons deux nombres N et p , et retenons pour $1 - S_1$ le quantile d'ordre p/N évalué sur les $C(I_0, I_i)$, pour $1 \leq i \leq N$, où les I_i sont des réalisations d'un modèle feuilles mortes isotrope, et I_0 l'image de référence. Nous fixons donc ainsi empiriquement un taux de fausses alarmes inférieur à p/N . L'idée est que ce modèle reproduit l'essentiel de la structure des images tout en restant indépendant d'une image donnée. Nous avons ainsi la possibilité de générer des échantillons d'images exempts des relations de dépendance que peuvent par exemple faire apparaître la présence du ciel, le fait que la plupart des objets reposent sur le sol, etc. Il est à signaler qu'il s'agit ici d'un choix, et donc que notre algorithme sera sensible à ces relations de dépendance lorsqu'on l'appliquera aux images. En particulier, comme nous le verrons lors des expériences, nous trouverons de nombreuses correspondances dans le cas d'images avec un horizon. En quelque sorte, nous sommes surpris de trouver un horizon dans les images.

La méthode développée ci-dessus permet donc de chercher les images les plus proches d'une image requête I_R dans une base $B = \{I_1, \dots, I_M\}$. Nous cherchons simplement les images pour lesquelles $C(I_R, I_i)$ est le plus petit. Signalons que l'on peut montrer que $C(I_R, I_i) > C(I_R, I_R)$ pour toute image I_i . De plus, le seuil S_1 calculé ci-dessus nous permet de décider quelles images de B sont gardées comme suffisamment proches de I_R . Il s'agit d'une particularité de cette méthode (voir aussi [9]), la plupart des algorithmes de recherche automatique se contentant de retourner un score pour chaque image de la base, sans chercher à évaluer la pertinence de la mise en correspondance.

Précisons enfin que la méthode est rapide : de l'ordre de la minute pour tester une image contre une base de 5000 images 300x200, sans indexation. Il serait cependant nécessaire d'utiliser une représentation économique des régions pour utiliser la méthode en temps réel sur une base importante. Une première solution consiste à utiliser des images à basse résolution. Nous avons observé une relative stabilité des résultats jusqu'à une taille de 60x40 pixels. Une approche multi-échelle est ici envisageable. Cependant, signalons que le seuil S_1 (qui dépend de la requête) est coûteux à évaluer, le paramètre N devant augmenter avec le nombre d'images de la base (si l'on veut gar-

der le rapport N/p de l'ordre du nombre d'images de la base). La méthode est par ailleurs robuste au bruit : la surface des ensembles de niveau varie peu, dans des proportions que l'on peut généralement calculer en fonction du type de bruit.

Parmi les limitations, il faut signaler la non-invariance par transformations géométriques simples. Cependant, nous rappelons que nous nous intéressons à la composition de l'image, et que seule une invariance à une classe très restreinte de transformations sera utile dans ce contexte. Une autre limitation est la nécessité d'uniformiser le support des images. Dans les expériences du paragraphe suivant, nous avons distingué entre images horizontales et verticales.

Plusieurs prolongements de cette méthode nous semblent intéressants. Tout d'abord, l'extension aux images couleurs peut se faire par l'intermédiaire d'une quantification dans un espace adéquat de type HSV (voir par exemple [11]), et les premiers essais dans cette direction sont satisfaisants. Plus généralement, la méthode peut s'appliquer à tous types de représentation de l'image par régions. Une autre direction est bien entendu l'utilisation de compacts K_1 et K_2 moins triviaux que l'origine O dans la formule (2), en particulier dans un cadre d'analyse de textures.

4 Expériences

Nous considérons une base d'environ 5000 images (extraites de la base Corel), 3000 horizontales et 2000 verticales, préalablement réduites à une taille de 300x200. Pour chaque image requête I_R , nous calculons $C(I_R, I_i)$ pour $I_i \in B$. Nous utilisons 5 ensembles de niveau dans cette expérience. Puis nous ordonnons les I_i selon les valeurs croissantes de C . De plus, comme expliqué au paragraphe précédent, nous pouvons décider quelles sont les images qui sont significativement proches de I_R . Dans cette expérience, nous avons choisi $N = 1000$, de manière à fixer empiriquement l'ordre de grandeur des fausses alarmes sur la base à un. Sur la figure 2, nous montrons trois exemples de requête parmi les images verticales (les deux types d'images ont été traitées séparément). Deuxième ligne, nous sommes dans un cas de figure où il n'y a que deux images significativement proches de la requête. La deuxième image est juste au seuil de significativité. La troisième ligne présente un exemple de requête par schéma de composition. Sur la figure 2, nous donnons des exemples de requête dans la base d'images horizontales. Généralement, l'ordre dans lequel sont retournées les réponses est en accord avec la perception que nous avons de la composition des scènes du strict point de vue des niveaux de gris, et bien sûr indépendamment de la sémantique des images. On remarquera cependant qu'à la troisième ligne, le seuil S_1 ne semble pas assez contraignant, et qu'à la cinquième ligne, la quatrième image a été préférée à la suivante malgré l'absence de bande noire centrale (rappelons que cette bande n'influe pas plus sur la mise en correspondance que la présence du ciel). Sur la figure 4, nous illustrons l'effet de S_1 dans deux cas.

Références

[1] I. Ahmad and W. I. Grosky. Spatial similarity based retrieval in image databases. *J. of Comp. Sci. and Inf. Manag.*, 2(2) :1–10, 1999.

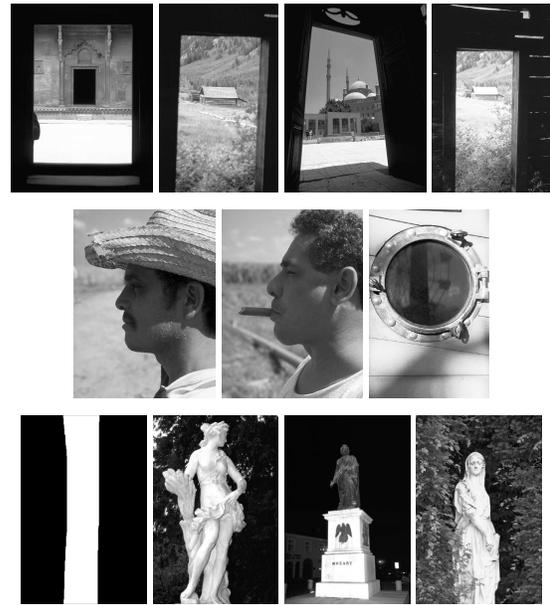


FIG. 2 – **Première ligne** : à gauche la requête, puis les trois images les plus proches. Pour cet exemple, il y a 26 images significativement proches de la requête. **Deuxième ligne** : à gauche la requête, puis les deux seules images significativement proches. La troisième image est juste au seuil S_1 . **Troisième ligne** : exemple de requête schématique.

[2] L. Alvarez, Y. Gousseau, and J.-M. Morel. The size of objects in natural and artificial images. *Advances in Imaging and Electron Physics, Academic Press*, 111 :167–242, 1999.

[3] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld : image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. on PAMI*, 24 :10026–1038, 2002.

[4] J. Fauqueur and N. Boujemaa. Image retrieval by range query composition of region categories. Technical report, INRIA, 2002.

[5] A. Lee, D. Mumford, and J. Huang. Occlusion models for natural images : A statistical study of a scale invariant dead leaves model. *Int'l J. of Computer Vision*, 41 :35–59, 2001.

[6] P. Lipson, E. Grimson, and P. Sinha. Configuration based scene classification and image indexing. In *IEEE Proc. Comp. Vis. Patt. Rec.*, pages 1007–1013, 1997.

[7] G. Matheron. Modèle séquentiel de partition aléatoire. Technical report, CMM, 1968.

[8] G. Matheron. *Random Sets and Integral Geometry*. John Wiley & Sons, New York, 1975.

[9] P. Musé, F. Sur, F. Cao, and Y. Gousseau. Unsupervised thresholds for shape matching. A paraître, ICIP 2003.

[10] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, London, 1982.

[11] J. R. Smith and C.-S. Li. Image classification and querying using composite region templates. *Computer Vision and Image Understanding*, 75(1/2) :165–174, 1999.

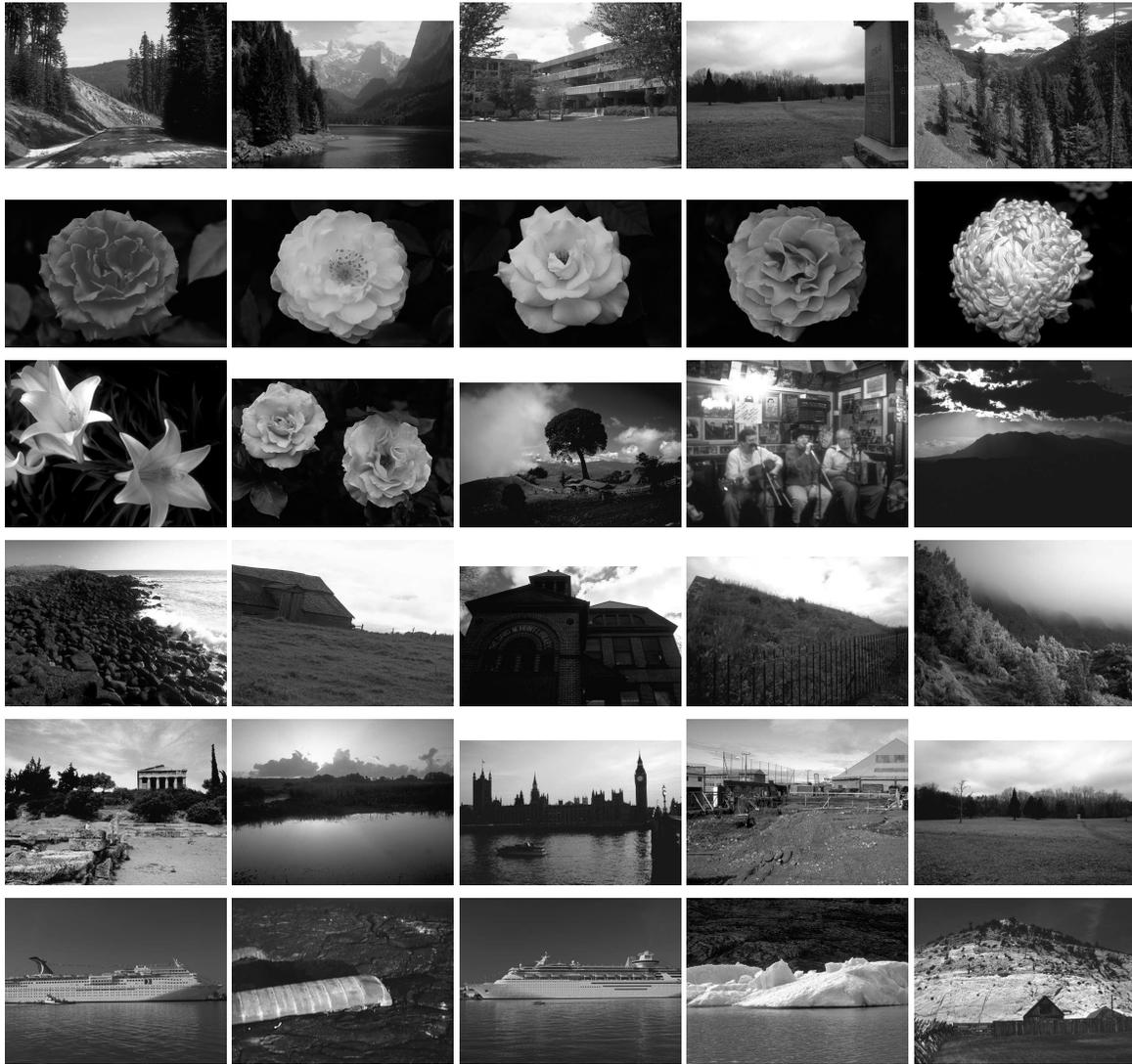


FIG. 3 – Pour chaque ligne : à gauche, image requête, puis les images les plus proches (toutes au dessus du seuil S_1 , voir le texte). L’algorithme compare les grandes lignes de l’organisation des images : présence d’une zone claire en haut au centre à la première ligne, fond noir autour d’une forme centrale claire à la deuxième ligne, etc. A la ligne 3, le seuil S_1 semble trop petit : trop d’images sont gardées comme significativement proches de la requête. Ligne 5, les réponses contiennent une bande sombre centrale, sauf la troisième : la présence d’une bande sombre n’est pas jugée plus surprenante que celle du ciel ou du sol.



FIG. 4 – **Première ligne** (de gauche a droite) : requête, puis les trois images les plus proches. Seules les deux premières vérifient $C < 1 - S_1$. **Deuxième ligne** : requête, puis l’image la plus proche, pour la quelle $C > 1 - S_1$. Aucune réponse n’est retenue.