

Recherche d'image par le contenu : requête partielle ou globale, apprentissage en ligne

Sylvie Philipp-Foliguet

ETIS, CNRS UMR8051 ,
6 avenue du Ponceau
95014 Cergy-Pontoise Cedex
philipp@ensea.fr
<http://www-etis.ensea.fr/~philipp/>

Résumé. Nous présentons dans cet article deux méthodes d'élaboration des signatures, une méthode globale à l'aide d'histogrammes et une méthode de description des régions et de leur disposition dans l'image. Nous exposons ensuite une méthode dédiée à la requête partielle qui est basée sur la mise en correspondance de graphes de régions et une méthode interactive basée sur l'apprentissage statistique.

1 Introduction

La recherche d'image dans des bases est un domaine de recherche en plein essor à cause des innombrables images acquises chaque jour dans des domaines aussi variés que la médecine, la télédétection ou avec les appareils photographiques personnels, pour n'en citer que quelques-uns. Cette nouvelle thématique emploie les techniques de la reconnaissance des formes dont elle est une application particulière. Ce qui la caractérise par rapport aux applications précédentes de l'interprétation d'images est de s'attacher davantage à la sémantique de l'image et de s'effectuer en interaction avec l'utilisateur. En effet les applications de la reconnaissance des formes qui fonctionnent sont par exemple la lecture des chèques ou la reconnaissance des empreintes digitales. La sémantique est ici quasiment inexistante, on recherche le "tout automatique". Autrement plus difficile est la recherche pour une agence de presse d'une photographie illustrant tel président devant tel édifice ou pire un concept aussi abstrait que la "misère".

Nous excluons dans ce papier l'emploi d'autre information que l'image elle-même, nous concentrant sur la recherche d'image par le contenu, à partir d'une image requête. Comme la sémantique est difficile à représenter numériquement, l'utilisateur est souvent mis à contribution pour guider la recherche dans une procédure dite de "bouclage de pertinence".

Un système de recherche d'images par le contenu comprend une phase de description de l'image qui a pour but de la représenter numériquement à l'aide d'une signature. Cette phase, parfois très longue s'effectue hors-ligne. La phase de recherche proprement dite s'effectue en ligne et consiste à comparer, via les signatures, les images entre elles ou avec une image fournie en entrée par l'utilisateur.

2 Description de l'image sous forme de signature

Les signatures d'images couramment utilisées font soit appel à des descripteurs globaux, caractérisant l'ensemble de l'image, soit à des descripteurs locaux calculés sur des régions ou des points caractéristiques.

2.1 Descripteurs globaux

Les descripteurs globaux sont les plus utilisés actuellement et sont en général constitués de distributions de couleur et de texture. On en trouvera un bon aperçu dans (Trémeau et al. 2004) ainsi que les similarités associées. Concernant la couleur, un problème est la réduction du nombre de couleurs en un nombre raisonnable de bins. La méthode que nous avons adoptée consiste à fabriquer un nombre fixé de classes couleur en appliquant un algorithme de type k -means sur l'ensemble des pixels de la base (ou sur un échantillon) (Cord et al. 2004). Nos essais ont montré que sur une base généraliste type Correl, 25 classes couleur constituaient un bon compromis entre compacité et représentativité de la signature (cf Fig 1).

Pour représenter la texture, nous avons opté comme de nombreux auteurs pour un ensemble de 12 filtres de Gabor codant 3 fréquences et 4 orientations. La distribution des textures est elle aussi codée en 25 classes obtenues par k -means à partir de l'ensemble des pixels de la base.

2.2 Descripteurs locaux

Il existe plusieurs méthodes d'extraction de points caractéristiques (souvent les angles des objets) et l'image est alors représentée par des descripteurs calculés au voisinage de ces points. On trouvera une synthèse comparative dans (K. Mikolajczyk et al. 2005).



(a) image originale



(b) image quantifiée en 25 classes couleur

FIG. 1- Exemple de quantification couleur ;

chaque pixel de l'image quantifiée a pour couleur le centre de la classe la plus proche

Les régions constituent des primitives que l'on peut extraire des images puis caractériser par des attributs de couleur, de texture, mais aussi de forme. Les méthodes de segmentation d'images ont été revisitées ces dernières années afin de les rendre robustes au réglage des paramètres. Les régions obtenues sont souvent moins précises (Carson et al. 2002) voire même floues (Philipp-Foliguet et al. 2002) (voir Fig 2). La composition de l'image est alors représentée par la position des primitives (points caractéristiques ou régions) et sous la forme d'un graphe d'adjacence pour les régions.

Les descripteurs locaux sont plus adaptés à une requête partielle, c'est-à-dire à la recherche d'images contenant un objet spécifique, quel qu'en soit l'environnement.

3 Recherche d'images

Nous présentons deux approches de la recherche d'images, l'une basée sur des descripteurs locaux de type région et adaptée à la recherche d'objets et l'autre basée sur des descripteurs globaux, adaptée à la recherche de catégorie et comprenant un bouclage de pertinence.

3.1 Requête partielle : appariement de graphes

Dans ce premier système (S. Philipp-Foliguet et al. 2004), la signature d'une image est constituée du graphe d'adjacence des régions, et pour chaque région de sa distribution couleur en 25 classes ainsi que de la position de son centre de gravité. La requête est constituée d'un ensemble de régions d'une image, constituant l'objet recherché. Le problème est de retrouver dans la base des images comportant un ensemble de régions similaires en terme de couleur et de composition. Les deux ensembles de régions ne sont pas nécessairement de même cardinalité, il s'agit donc d'un problème d'appariement inexact de graphes.

Nous avons employé un arbre de recherche (Cordella et al. 1998) permettant d'éliminer rapidement les images trop dissemblables à la requête. Cette dissimilarité prend en compte la cohérence entre la composition des régions formant la requête et celle des régions de l'image cible. Elle permet de restreindre la recherche aux seuls ensembles de régions respectant l'adjacence et les positions relatives des régions requête. Mais elle autorise les appariements d'une région d'une des images avec plusieurs régions de l'autre image. La dissimilarité est finalement mesurée comme la moyenne des distances (de type L_1) entre les distributions couleur des régions appariées. Un exemple est donné dans la fig. 2, seules l'adjacence et les positions verticales relatives sont prises en compte.

3.2 Requête globale : apprentissage en ligne

Nous nous situons ici dans le cadre de la recherche de catégorie. Le problème est la grande variabilité des images au sein d'une catégorie. Les images pouvant être très dispersées dans l'espace de représentation, il est nécessaire d'explorer au mieux la base pour être sûr de ne pas oublier un des aspects de la catégorie. Pour cela, il faut présenter à l'utilisateur pour annotation une large variété d'images susceptibles d'appartenir à la catégorie.

Nous avons proposé dans RETIN (Cord et al. 2004) un système de recherche interactive qui explore la base lors des premières itérations, afin de détecter les différents modes de la catégorie. La recherche s'effectue à partir d'une requête initiale, qui est enrichie au fur et à mesure des itérations grâce aux annotations de l'utilisateur (qui signale si une image appartient ou n'appartient pas à la catégorie recherchée). Les images présentées à l'utilisateur sont tirées aléatoirement en fonction de leur probabilité d'être pertinente (calculée à partir de leur similarité à la requête) et du numéro d'itération. La recherche se resserre autour des images pertinentes au fil des itérations. La fonction de similarité est une somme pondérée des distances entre bins de chaque distribution, les poids étant mis à jour lors du bouclage par rétropropagation de l'erreur. Pour modéliser la multimodalité de la catégorie, les exemples annotés comme pertinents par l'utilisateur sont accumulés pour former le prototype de la catégorie. Le calcul de la similarité ne se fait plus par rapport à la requête initiale mais par rapport au prototype en employant une fusion des similarités (voir Fig 3).

Dans RETIN-AL (Cord et al. 2005), le problème est posé comme un problème de classification à deux classes (images de la catégorie et autres images), pour lequel on dispose de quelques exemples (les images annotées positives par l'utilisateur). Ce problème d'apprentissage actif se résout bien par la technique du SVM qui détermine au mieux la frontière entre les deux classes, pourvu qu'on présente à l'utilisateur les images les plus ambiguës, c'est-à-dire celles qui sont les plus proches de la frontière. A chaque itération, nous présentons à l'utilisateur pour annotation 5 images proches de la frontière, les plus

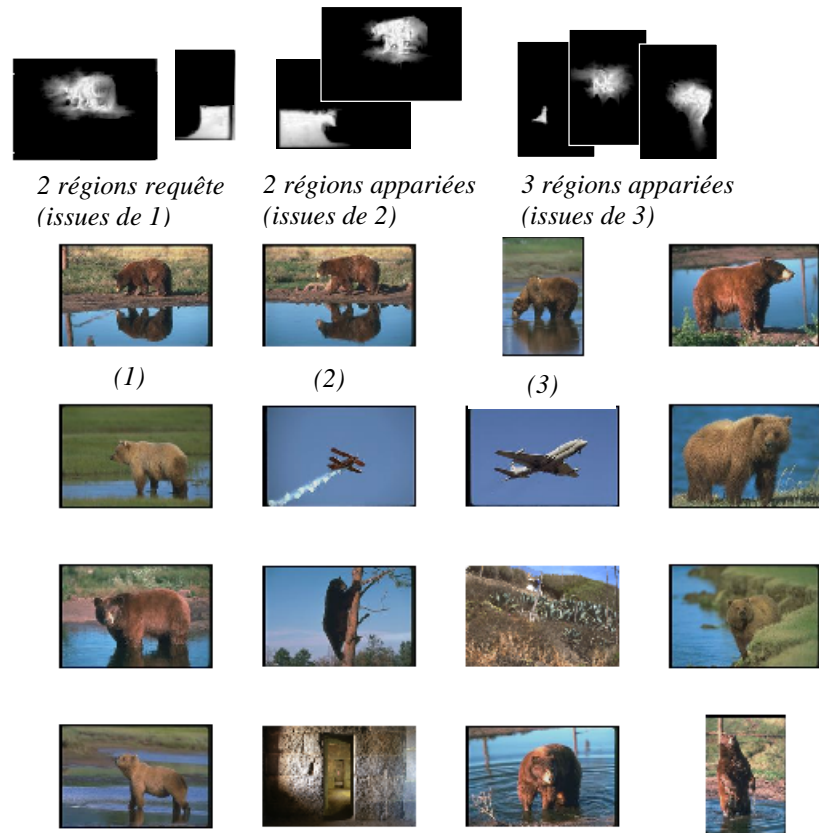


FIG. 2 - Les 16 images les plus proches pour la requête constituée par les deux régions floues tirées de l'image (1), triées par similarité décroissante. Pour les deux images les plus proches figurent les régions appariées aux régions requête (base de 1200 images)

différentes possibles ; la frontière est remise à jour en fonction de ces nouveaux exemples et contre-exemples et les similarités sont recalculées.

4 Conclusion

En ce qui concerne le calcul de signature, il existe de nombreux attributs soit globaux, soit calculés sur des primitives extraites de l'image. Le choix d'une signature adaptée aux images à traiter et au type de requête demeure un problème central. La robustesse face aux changements d'illumination, d'échelle ou d'angle de vue est un des éléments du choix. En étroite liaison avec les signatures, de nombreuses mesures de similarité ont été proposées.

Le gros effort doit porter sur les systèmes de recherche eux-mêmes. S'ils utilisent les outils classiques mais aussi récents de classification, ils ont des caractéristiques bien spécifiques, comme la grande dimension des espaces de représentation, le faible nombre

d'exemples, etc. En outre, grâce à l'interactivité des systèmes, on peut former des catégories sémantiques comprenant des images de contenus très divers. Ces catégories peuvent changer d'un utilisateur à l'autre et même d'une session à l'autre pour un même utilisateur. Un des challenges est de conserver les catégories ainsi formées, afin d'accélérer les fouilles ultérieures de la base.



Références

- C. Carson, S. Belongie, H. Greenspan, J. Malik (2002), Blobworld : image segmentation using Expectation-Maximisation and its application to image querying, *IEEE Trans. on PAMI*, 24 (8), 1026-1038
- M Cord, J Fournier, S Philipp-Foliguet (2004), Approche interactive de la recherche par le contenu, *TSI*, vol. 23, n°1
- M Cord, S Philipp-Foliguet, P-H. Gosselin, J Fournier (2005), Interactive exploration to image retrieval , *Journal of Applied Signal Processing*(à paraître)
- L. P. Cordella, P. Foggia, C. Sansone, M. Vento (1998), Subgraph Transformation for the inexact Matching of Attributed Relational Graphs, *Computing*, 12, 43-52
- K. Mikolajczyk, C. Schmid, (2004), Scale and affine invariant interest point detectors, *Int. J on computer vision*, vol 60, 63-86
- S. Philipp-Foliguet, M. B. Vieira, M. Sanfourche (2002), Fuzzy segmentation of colour images and indexing of fuzzy regions, *CGIV'02, Poitiers*, 507-512
- S. Philipp-Foliguet, M. Lekkatt (2004), Recherche d'images à partir d'une requête partielle utilisant la disposition des régions, *RFIA'04, Toulouse*
- A. Trémeau, C. Fernandez-Maloigne, P. Bonton (2004), Image numérique couleur ; de l'acquisition au traitement, Dunod

Summary

In this paper we present two methods for signature building, a global method based on histograms and a method describing regions and their disposition within the image. Two methods are then expounded : a method dedicated to partial request based on region graph matching and an interactive method based on statistical learning.