

Stage de recherche Master 2

CNAM Paris – Télécom ParisTech
2011

« Analyse d'images par descripteurs locaux et relations spatiales floues »

Contexte

Dans le cadre d'une collaboration de recherche, l'équipe Vertigo du laboratoire CEDRIC (CNAM, Paris) et l'équipe Traitement et Interprétation des Images du laboratoire LTCI (Télécom ParisTech, Paris) proposent un sujet de stage de recherche en master 2^{ème} année, sur le thème de l'analyse, l'interprétation et l'indexation des images.

Vertigo¹ est une équipe de recherche dont un des axes concerne l'indexation par contenu visuel des grandes collections d'images et de vidéos. Elle investigate d'une part les méthodes de description du contenu visuel par analyse d'images, notamment les approches de description locale [Gouet-Brunet and Lameyre 2008] [Law-To et al., 2009] et les relations spatiales strictes qui les caractérisent [Hoang et al., 2010], et d'autre part les méthodes d'accès rapides associées pour la gestion des grands volumes de contenus multimédia [Gouet-Brunet et al., 2009] [Barton et al., 2010].

Une partie des activités de recherche de l'équipe Traitement et Interprétation des Images² du LTCI est consacrée à l'indexation d'images et au raisonnement spatial flou [Bloch, 2005] [Bloch et al., 2006] [Hudelot et al., 2008] [Aldea, 2009] [Millet et al., 2010] [Vanegas et al., 2010]. Ces travaux ont été principalement appliqués à des contenus image spécifiques en imagerie médicale [Bloch et al., 2006] [Aldea, 2009], en imagerie satellitaire [Vanegas et al., 2010] et aussi aux contenus hétérogènes (images issues du Web [Millet et al., 2010]).

Présentation du sujet

Ce stage porte sur l'analyse du contenu des images fixes en vue de leur interprétation et indexation par contenu visuel dans un ensemble d'images. Parmi les nombreuses approches de description des contenus visuels [Gouet-Brunet, 2006] [Datta et al., 2008], nous nous intéressons aux méthodes de description locale à base de points d'intérêt [Tuytelaars and Mikolajczyk, 2008] [Li and Allinson, 2008], et notamment à la caractérisation des relations spatiales entre ces points selon le formalisme de la logique floue. Le choix d'étudier le thème des relations spatiales floues pour les points d'intérêt est motivé par les raisons suivantes :

En premier lieu, la description des relations spatiales entre points d'intérêt dans les images est une problématique de recherche d'actualité qui intervient dans de nombreux domaines applicatifs, surtout lorsque la représentation de ces points est en sacs de mots visuels [Sivic et al., 2003] où l'information spatiale est inexistante. Pour ces primitives image, il existe encore assez peu de solutions efficaces incluant une information spatiale. Classiquement, la comparaison d'images passe par la recherche des points les plus similaires visuellement, étape à laquelle on ajoute *a posteriori* une étape plutôt coûteuse de recalage, qui permet de vérifier la configuration spatiale des points appariés (selon des versions

¹ Equipe Vertigo : <http://cedric.cnam.fr/vertigo/>

² Equipe TII : <http://perso.telecom-paristech.fr/~bloch/tii/>

améliorées de la méthode RANSAC, par exemple [Philbin et al., 2007]). Plusieurs approches proposent d'intégrer *a priori* dans la description une information sur les relations spatiales, voir par exemple [Sivic et al., 2005], [Lazebnik et al., 2006] [Yuan et al., 2007] [Sahbi et al., 2010] [Hoang et al., 2010].

En second lieu, aucune de ces approches de description des relations spatiales rencontrées n'exploite le formalisme de la logique floue. Pourtant, il devrait s'avérer pertinent pour pallier l'imprécision de la représentation par description locale. En effet, les techniques de description locale reposent sur l'extraction de points d'intérêt dont la répétabilité d'une image à l'autre n'est pas très bonne. Le même point peut ne pas être détecté d'une image à l'autre à cause des différentes transformations que les images peuvent subir, rendant l'existence d'une relation spatiale avec ses voisins non certaine d'une image à l'autre. Son extraction peut aussi être imprécise (problème de délocalisation), à cause de plusieurs facteurs comme le bruit inhérent au signal ou la modification de l'image due à des étapes de filtrage. Enfin, certaines relations spatiales entre points peuvent être perturbées par la transformation qui existe entre deux images, comme le changement de point de vue par exemple. On comprend alors qu'une description fine des relations spatiales entre primitives de ce type sera sensible à toutes ces imprécisions et que l'utilisation du formalisme de la logique floue devrait permettre de définir des contraintes spatiales plus robustes.

Plan de travail

Etudier un modèle décrivant les relations spatiales floues entre points d'intérêt requiert de traiter plusieurs problèmes, dont ceux-ci :

- Quels sont les types de relations spatiales pertinents pour cette catégorie de primitive image, parmi celles existantes (voir par exemple l'ontologie de relations spatiales floues définie dans [Hudelot et al., 2008] et [Steiniger and Weibel, 2007] pour les relations structurelles) ? De part la nature de la primitive point, les relations spatiales topologiques ne sont pas adaptées, mais d'autres types semblent éligibles parmi les relations métriques et structurelles : nous pensons notamment aux relations binaires directionnelles (absolues ou relatives), de distances [Bloch, 2005], à certaines relations ternaires comme la relation *entre* [Bloch et al., 2006], à des relations *n*-aires comme les alignements [Vanegas et al., 2010] ou les regroupements. Cette partie de l'étude conduira aussi à mettre en place des mécanismes de raisonnement spatial, de manière à pouvoir combiner plusieurs types de relations spatiales, comme en inférer de nouvelles à partir de celles déjà établies. Dans le cas de relations spatiales floues, il est possible d'utiliser des opérateurs de fusion issus de la logique floue [Bloch, 1996].
- Quel modèle de représentation choisir et quelle mesure de similarité lui associer ? Appariement deux ensembles de points d'intérêt peut s'apparenter à de l'appariement de graphes. Le formalisme de la logique floue existe pour les graphes, voir par exemple [Perchant and Bloch, 2002] pour les graphes flous et [Cesar et al., 2005] pour les graphes d'attributs flous. Certaines des relations spatiales étant *n*-aires, les hypergraphes pourraient aussi être pertinents [Bunke et al., 2008] [Rital, 2009]. Cependant, de part le nombre de points d'intérêt mis en jeu dans chaque image (plusieurs centaines ou même milliers dans le cas d'images satellite), ces modèles pourront se révéler trop complexes pour traiter un grand nombre d'images et il faudra alors proposer un modèle de représentation *ad hoc*, voir par exemple [Gouet and Boujemaa, 2001] [Hoang et al., 2010] pour les relations spatiales strictes.

Organisation du stage

Il s'agit d'un stage de 6 mois à partir de mars 2011, rémunéré (750 euros net/mois) et basé à Paris entre le CNAM et Télécom ParisTech.

Compétences requises

De bonnes connaissances et une pratique en indexation d'images par contenu visuel, analyse d'images ou vision par ordinateur sont requises, ainsi que la maîtrise de la programmation C/C++ ou Java.

Modalités de candidature

Envoyer par email aux deux encadrants :

- CV
- Lettre de motivation ciblée sur le sujet
- Plusieurs lettres de recommandation
- Relevé de notes des deux dernières années d'étude
- Liste des enseignements suivis et validés durant les deux dernières années d'étude

Encadrement et contacts

Encadrants :

- Valérie Gouet-Brunet, Maîtres de Conférences HDR au CNAM – valerie.gouet@cnam.fr
- Isabelle Bloch, Professeur à Télécom ParisTech – isabelle.bloch@telecom-paristech.fr

Bibliographie

[Aldea, 2009] Aldea, E., Apprentissage de données structurées pour l'interprétation d'images. *Thèse de doctorat*, Télécom ParisTech, 2009.

[Barton et al., 2010] S. Barton, V. Gouet-Brunet, M. Rukoz, C. Charbuillet and G. Peeters, Estimating the indexability of multimedia descriptors for similarity searching, *International Conference RIAO'10 on Adaptivity, Personalization and Fusion of Heterogeneous Information*, 12 pages, Paris, April 2010.

[Bloch, 1996] Bloch, I., Information combination operators for data fusion : A comparative review with classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 26(1) :52–67, 1996.

[Bloch, 2005] Bloch, I., Fuzzy spatial relationships for image processing and interpretation : a review. *Image Vision Computing*, 23 :89–110, 2005.

[Bloch et al., 2006] Bloch, I., Colliot, O., and Cesar, R., On the ternary spatial relation between. *IEEE Transactions on Systems, Man, and Cybernetics SMC-B*, 36 :312–327, 2006.

[Bunke et al., 2008] Bunke, H., Dickinson, P., Neuhaus, M., and Stettler, M., Matching of hypergraphs - algorithms, applications, and experiments. *Applied Pattern Recognition*, ed. by Bunke, H. and Kandel, A. and Last, M., pages 131–154, 2008.

[Cesar et al., 2005] Cesar, Jr., R. M., Bengoetxea, E., Bloch, I., and Larrañaga, P., Inexact graph matching for model-based recognition : Evaluation and comparison of optimization algorithms. *Pattern Recognition*, 38 :2099–2113, 2005.

[Datta et al., 2008] R. Datta, D. Joshi, J. Li and J. Z. Wang, Image Retrieval: Ideas, Influences and Trends of the New Age, *ACM Computing Surveys*, 40(2): 1-60, 2008.

[Hudelot et al., 2008] Hudelot, C., Atif, J., and Bloch, I., Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets and Systems*, 159 :1929–1951, 2008.

[Gouet and Boujemaa, 2001] Gouet, V. and Boujemaa, Object-based queries using color points of interest. *IEEE Workshop on Content-Based Access of Image and Video Libraries (CVPR/CBAIVL 2001)*, pages 30–36, Kauai, Hawaii, USA, 2001.

[Gouet-Brunet, 2006] V. Gouet-Brunet, Chapitre de livre « Recherche par contenu visuel dans les grandes collections d'images », *Encyclopédie de l'Informatique et des systèmes d'information*, pages 564–576, J. Akoka and I. Commy-Wattiau (eds.), Vuibert, 2006.

[Gouet-Brunet et al., 2009] V. Gouet-Brunet, N. Bouteldja and M. Scholl, HiPeR : a hierarchical model for exact, approximate and progressive retrieval in multi-dimensional spaces, *International Journal on Data Engineering and Management*, 1(1) :14–33, 2009.

- [Gouet-Brunet and Lameyre, 2008] V. Gouet-Brunet and B. Lameyre, Object recognition and segmentation in videos by connecting heterogeneous visual features, *Computer Vision and Image Understanding (CVIU)*, 111(1) :86–109, February 2008.
- [Hoang et al., 2010] N. Vu Hoang, V. Gouet-Brunet, M. Rukoz and M. Manouvrier, Embedding spatial information into image content description for scene retrieval, *Pattern Recognition Journal*, 43(9) :3013–3024, 2010.
- [Law-To et al., 2009] J. Law-To, V. Gouet-Brunet, O. Buisson and N. Boujemaa, ViCopT : a robust system for content based video copy detection in large databases, *ACM Multimedia Systems Journal*, 15(6) :337–353, December 2009.
- [Lazebnik et al., 2006] Lazebnik, S., Schmid, C., and Ponce, J., Beyond Bags of Features : Spatial Pyramid Matching for Recognizing Natural Scene Categories, *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2169–2178, Washington, DC, USA, 2006.
- [Li and Allinson, 2008] Jing Li and Nigel M. Allinson, A comprehensive review of current local features for computer vision, *Neurocomputing Journal*, Volume 71, Issue 10-12, June, 2008.
- [Millet et al., 2010] Millet, C., Bloch, I., Hède, P., and Moëllic, P.-A., Automatic cleaning and segmentation of web images based on colors to build learning databases. *Image Vision Computing*, 28 :317–328, 2010.
- [Perchant and Bloch, 2002] Perchant, A. and Bloch, I., Fuzzy morphisms between graphs. *Fuzzy Sets and Systems*, 128 :149–168, 2002.
- [Philbin et al., 2007] Philbin, J., Chum, O., Isard, M., Sivic, J. and Zisserman, Object retrieval with large vocabularies and fast spatial matching, *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [Rital, 2009] Rital, S., Hypergraph cuts & unsupervised representation for image segmentation. *Fundam. Inform. Journal*, 96(1-2) :153–179, 2009.
- [Sahbi et al., 2010] Sahbi, H., Audibert, J.-Y., and Keriven, R., Context-dependent kernels for object classification, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.
- [Sivic et al., 2005] Sivic, J., Russell, B., Efros, A., Zisserman, A., and Freeman, W., Discovering objects and their location in images, *IEEE International Conference on Computer Vision*, pages 370–377, 2005.
- [Sivic et al., 2003] Josef Sivic and Andrew Zisserman, Video Google: A Text Retrieval Approach to Object Matching in Videos, *International Conference on Computer Vision*, Vol. 2, pages 1470 – 1477, 2003.
- [Steiniger and Weibel, 2007] Steiniger, S. and Weibel, R., Relations among map objects in cartographic generalization. *Cartography and Geographic Information Science*, 34(3) :175 -197, 2007.
- [Tuytelaars and Mikolajczyk, 2008] T. Tuytelaars and K. Mikolajczyk, Local invariant feature detectors: a Survey, *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2008.
- [Vanegas et al., 2010] Vanegas, M.-C., Bloch, I., and Inglada, J., Detection of aligned objects for high resolution image understanding. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2010)*, pages 464–467, Honolulu, Hawaii, USA, 2010.
- [Yuan et al., 2007] Yuan, J., Wu, Y., and Yang, M., Discovery of Collocation Patterns : from Visual Words to Visual Phrases, *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.