

# Stage M2 CNAM/CEDRIC

## Synergies entre descripteurs visuels de bas-niveau

2010

**Lieu :** CNAM, laboratoire CEDRIC, Paris

**Encadrement :** Valérie Gouet-Brunet (CNAM), Marcelo Hashimoto (CNAM et Université de São Paulo, Brésil), Ricardo Torres (Université UNICAMP, Brésil)

**Rémunération :** 500 euros net/mois

**Contact :** valerie.gouet@cnam.fr

## 1 Cadre de travail

Ce stage sera réalisé dans le cadre d'une collaboration entre le laboratoire CEDRIC du CNAM, l'Université de São Paulo et l'Université UNICAMP au Brésil, portée par le projet CAPES-COFECUB AMIB (2010-2013). L'étudiant(e) travaillera en étroite collaboration avec Valérie Gouet-Brunet (Mdc CNAM) et avec Marcelo Hashimoto (doctorant de l'Université de São Paulo, en séjour de recherche invité au CNAM pendant la période du stage) et sera aussi en relation avec Ricardo Torres (Mdc Université UNICAMP).

Ce stage débouchera sur une thèse dans la continuité du travail déjà commencé. Dans le cadre du projet CAPES-COFECUB AMIB, le doctorant sera amené à faire un ou plusieurs séjours au Brésil et à collaborer avec des doctorants et chercheurs brésiliens.

## 2 Sujet

Le sujet de ce stage s'inscrit dans la problématique de l'indexation et de la recherche par contenu visuel dans les collections d'images et de vidéos. Plus particulièrement, le travail concernera les méthodes de description dites bas-niveau du contenu visuel, telles que l'extraction de points d'intérêt, de contours ou encore la segmentation en régions, les approches de description par couleur, texture et forme [DJLW08] [DKN08]. Nous nous focaliserons sur les méthodes de description locale à base de points d'intérêt, qui sont particulièrement efficaces pour la reconnaissance d'objets ou la détection de copies, applications considérées dans ce travail.

Depuis plusieurs années, la littérature relative à ce type de description s'est beaucoup enrichie, voir par exemple la synthèse [TM08]. Les propositions faites diffèrent principalement par le type de détecteur de points, par la méthode de description locale (impliquant différents niveaux de robustesse ou d'invariance) et par le type de classifieur utilisé. Avec la récente proposition de nouveaux types de descripteurs locaux impliquant différentes natures de supports de points (par exemple les patches de textures, les régions homogènes MSER, les formes locales et les points de symétrie), plusieurs travaux [SZ03] [OSP05] [OPFA06] [LGBB06] ont proposé d'améliorer la description des images en exploitant une combinaison de ces descripteurs. Avec le même objectif, d'autres approches, comme dans la thèse de B. Lameyre [GL08] (thèse menée au CEDRIC et en voie d'être soutenue), combinent une description locale avec une description globale des contenus visuels, de manière à bénéficier des avantages de ces deux classes d'approches, tout en minimisant leurs faiblesses.

Dans la continuité de ces travaux, le travail portera sur l'étude de la combinaison de différents descripteurs de bas-niveau, dont des points d'intérêt, et sera appliqué à la reconnaissance d'objets ou la détection de copies dans les images fixes et animées. En particulier, on choisira d'investiguer une piste parmi plusieurs pistes complémentaires déjà répertoriées (les pistes non traitées durant le stage feront l'objet de la thèse) :

1. **Combinaison de descripteurs visuels hétérogènes.** Les approches existantes qui combinent plusieurs natures de descripteurs locaux mettent en avant l'homogénéité ou au contraire les discontinuités existant sur les types de support adressés. Cependant, ces travaux se contentent la plupart du temps de combiner deux types de support, sans aucune justification de la qualité et de l'optimalité des associations réalisées. Le premier objectif sera l'évaluation de la complémentarité des détecteurs et des descripteurs existant dans la littérature, en proposant des critères, notamment de répartition dans l'image, dans le temps ou dans l'espace de description. Cette étude devra permettre de mettre en lumière une synergie entre certains descripteurs, tout en maintenant une compacité optimale de la description complète.
2. **Mise en place d'une mesure de similarité fine.** Alors que le choix de ces associations de descripteurs ainsi que leurs mesures de similarité est souvent déterminé *a priori*, Ricardo Torres (Université UNICAMP, Brésil) a proposé un modèle reposant sur la programmation génétique, permettant la détermination automatique de la mesure de similarité optimale pour un ensemble de descripteurs visuels [TFG09]. Ce modèle ayant été évalué sur des descripteurs d'images globaux simples, une piste de travail pertinente sera d'évaluer une mesure de similarité optimale pour des descripteurs bas-niveau plus complexes tels que ceux étudiés dans l'étape 1.
3. **Sélectivité de la description.** Les approches de description usuellement rencontrées sont par définition invariantes (ou robustes) à une ou plusieurs transformations géométriques et/ou photométriques, et sont donc naturellement moins sélectives. Une troisième piste de travail consistera au contraire à proposer une hiérarchie de descripteurs spécifiques à chaque transformation. Chaque descripteur sera robuste à une ou plusieurs transformations et sera placé dans la hiérarchie en fonction de son degré d'invariance. La recherche consistera alors à parcourir cette hiérarchie par raffinements jusqu'à localiser le meilleur descripteur pour l'image évaluée. En plus d'être particulièrement sélective, une telle description devrait permettre de diagnostiquer les traitements éventuellement appliqués aux contenus retrouvés et de donner ainsi la possibilité de formuler des requêtes ciblées par rapport au degré d'invariance souhaité.

### 3 Publications en relation avec le sujet

[DJLW08] 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), 2008.

[DKN08] T. Deselaers, D. Keysers, and H. Ney. Features for image retrieval: an experimental comparison. *Information Retrieval.*, 11(2):77-107, 2008.

[GL08] V. Gouet-Brunet et B. Lameyre. Object recognition and segmentation in videos by connecting heterogeneous visual features . *Computer Vision and Image Understanding*, 111(1): 86-109, Elsevier, 2008.

[LGBB06] J. Law-To and V. Gouet-Brunet and O. Buisson and N. Boujemaa, Labeling Complementary Local Descriptors Behavior for Video Copy Detection, in *Proceedings of the International Workshop on Multimedia Content Representation, Classification and Security*, Istanbul, 2006.

[OPFA06] A. Opelt and A. Pinz and M. Fussenegger and P. Auer, Generic object recognition with boosting, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol 28 (3), pages 416-431, 2006.

[OSP05] A. Opelt and J. Sivic and A. Pinz, Generic object recognition from video data, *1st Cognitive Vision Workshop*, 2005.

[SZ03] J. Sivic and A. Zisserman, Video Google: A Text Retrieval Approach to Object Matching in Videos, *IEEE International Conference on Computer Vision*, 2003.

[TFG09] R. da S. Torres, A. X. Falcão, M. A. Goncalves, J. P. Papa, B. Zhang, W. Fan, and E. A. Fox. A Genetic Programming Framework for Content-based Image Retrieval. *Pattern Recognition*, 42:283-292, 2009.

[TM08] Tinne Tuytelaars and Krystian Mikolajczyk. Local invariant feature detectors: a survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177-280, January 2008.