

# Extension du Modèle par Sac de Mots Visuels pour la Classification d'Images

## Mots clés :

- **Directeur de thèse** : Matthieu Cord
- **Co-encadrant(s)** :
- **Unité de recherche** : Laboratoire d'informatique de Paris 6
- **Ecole doctorale** : École Doctorale Informatique, Télécommunications, Électronique de Paris
- **Domaine scientifique principal**: Divers

## Résumé du projet de recherche (Langue 1)

L'information visuelle, représentée sous la forme d'images ou de vidéos numériques, est devenue si omniprésente dans le monde numérique d'aujourd'hui, qu'elle ne peut plus être considérée comme un "citoyen de seconde zone", par rapport à l'information textuelle. Néanmoins, contrairement aux documents textuels, les images sont constituées de pixels ne portant pas d'information sémantique directement accessible, ajoutant ainsi une difficulté à la tâche d'interprétation. Dans ce contexte, la classification d'images est devenue une tâche critique. En particulier, l'identification automatique d'objets complexes et de concepts sémantiques dans les images, a suscité de nombreux travaux récents, aussi bien en Recherche d'Information, Vision par Ordinateur, Traitement d'Image qu'en Intelligence Artificielle. Dans cette thèse, nous traitons le problème de la représentation des images. Notre objectif est la détection de concepts à partir d'une analyse du contenu visuel des images et des vidéos. Pour cela, nous introduisons une nouvelle représentation qui enrichit le modèle classique par sacs de mots visuels. S'appuyant sur la quantification de descripteurs locaux, et l'agrégation de ces descripteurs quantifiés en un vecteur de caractéristique unique, le modèle par sacs de mots visuels a émergé comme l'approche la plus efficace pour la classification d'images. Nous proposons BossaNova, une nouvelle représentation d'images permettant de conserver plus d'information lors de l'opération d'agrégation (pooling) en exploitant la distribution des distances entre les descripteurs locaux et les mots visuels. L'évaluation expérimentale sur plusieurs bases de données de classification d'images, telles que ImageCLEF Photo Annotation, MIRFLICKR, PASCAL VOC et 15-Scenes, a montré l'intérêt de Bossanova vis-à-vis des techniques traditionnelles, même sans utiliser de combinaisons complexes de multiples descripteurs locaux. Une extension de notre approche a également été étudiée. Elle concerne la combinaison de BossaNova avec une autre représentation basée sur des vecteurs de Fisher très compétitive. Les résultats obtenus sont systématiquement meilleurs atteignant l'état de l'art sur de nombreuses bases. Ils permettent ainsi de démontrer expérimentalement la complémentarité des deux approches. Cette étude nous a permis d'obtenir la seconde place lors de notre participation à la compétition ImageCLEF 2012 Flickr Photo Annotation Task parmi les 28 soumissions sur la partie visuelle. Enfin, nous avons appliqué notre stratégie de représentation BossaNova dans un contexte vidéo, en vue de faire de la détection de séquences à caractère pornographique. Les résultats ont permis de valider une nouvelle fois l'intérêt de notre approche par rapport à des détecteurs standards du marché sur une application réelle.