

Noyaux avec information spatiale pour la classification d'images et de vidéos

Stage Master 2 Recherche 2010 - UPMC LIP6

1 Informations pratiques

Mots-clés : Sciences pour l'ingénieur, Traitement des images, Apprentissage, Indexation

Durée : 5 à 6 mois

Localisation : LIP6, 104 ave du President Kennedy, 75016 Paris

Encadrement : David Picard, Matthieu Cord, Nicolas Thome

Contact : david.picard@lip6.fr, Matthieu.Cord@lip6.fr, nicolas.thome@lip6.fr

Rémunération LIP6 : 2k euros pour le stage

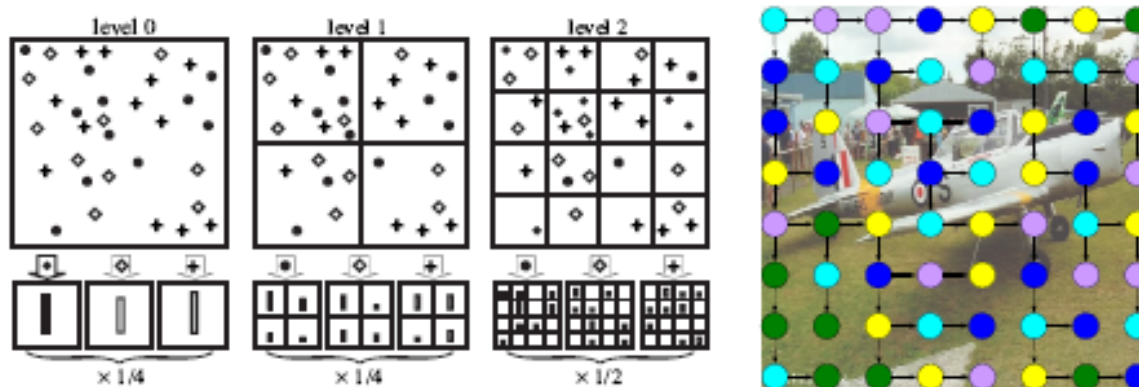


FIG. 1 – Représentation par mot visuels : approche par pyramide spatiale (gauche), et approche par chaîne de Markov (droite).

2 Contexte

Pour opérer une classification d'images basée contenu, les méthodes proposées dans l'état de l'art repose aujourd'hui presque toutes sur l'extraction d'un ensemble descripteurs locaux (typiquement des descripteurs SIFT[4]), l'image étant alors représentée par un sac contenant ses descripteurs. Afin de comparer des images l'espace des descripteurs est quantifié de manière à produire un dictionnaire de mots visuels. Chaque descripteur du sac de l'image est alors projeté sur le(s) mot(s) le(s) plus proche(s) afin de constituer l'histogramme des occurrences des mots visuels dans l'image. La représentation de l'image de vient alors vectorielle (vecteur de fréquence d'appartition des mots visuels dans l'image), facilement exploitable par les outils de classification [6] (figure 1). Afin de rajouter de l'information spatiale, plusieurs histogrammes peuvent être construits sur des sous-parties de l'image (par un découpage pyramidal [3]). Cette méthode à l'avantage de fournir directement une mesure exploitable par les outils d'apprentissage statistique largement utilisés aujourd'hui, tels que les SVM[1], cependant, dans le cas de recherche d'objets, l'information spatiale reste pauvre (découpage statique, pas d'invariance, etc).

3 Objectifs

Les approches par dictionnaire de mots visuels avec découpage pyramidal de l'image montrent aujourd'hui leurs limites en terme de pouvoir de représentation spatiale. **L'objectif de ce stage est d'introduire de nouveaux modèles de représentation alliant caractéristiques visuelles et structures spatiales.** On s'appuiera sur les travaux de [5], qui propose de représenter les images par une chaîne de Markov des transitions entre les points d'intérêts. Ce type de modèle semble bien plus pertinent car non seulement il présente une certaine robustesse au transformation (invariances), mais de plus, il prend en compte l'information visuelle et la structure spatiale dans une représentation unifiée.

Dans un premier temps, les expériences seront conduites sur la base PASCAL [2], afin de déterminer la justesse de telles approches dans le cadre de la catégorisation d'objets.

4 Compétences requises

Des connaissances minimales en traitement d'images et en apprentissage statistique sont nécessaires. Un bon niveau de programmation est demandé (le développement se fera en Java). Ce sujet comporte un fort côté expérimental, aussi, de la rigueur dans la conduite du plan expérimental sera-t-elle exigée.

Références

- [1] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3) :273–297, 1995.
- [2] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2009 (VOC2009) Results. <http://www.pascal-network.org/challenges/VOC/voc2009/workshop/index.html>.
- [3] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features : Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2169–2178, 2006.
- [4] D. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 20, pages 91–110, 2003.
- [5] G.J. Qi, X.S. Hua, Y. Rui, J.H. Tang, Z.J. Zha, and H.J. Zhang. A joint appearance-spatial distance for kernel-based image categorization. pages 1–8, 2008.
- [6] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (in press), 2010.