

Comparaison de la segmentation pixel et segmentation objet pour la détection d'objets multiples et variables dans des images

J. Pasquet^{2,3}, M. Chaumont^{1,2} et G. Subsol²

¹ Université de Nîmes, F-30021 Nîmes Cedex 1, France

² LIRMM, CNRS/Université Montpellier 2, France

³ Berger Levraut, 31676 Labège, France

{jerome.pasquet, gerard.subsol, marc.chaumont}@lirmm.fr

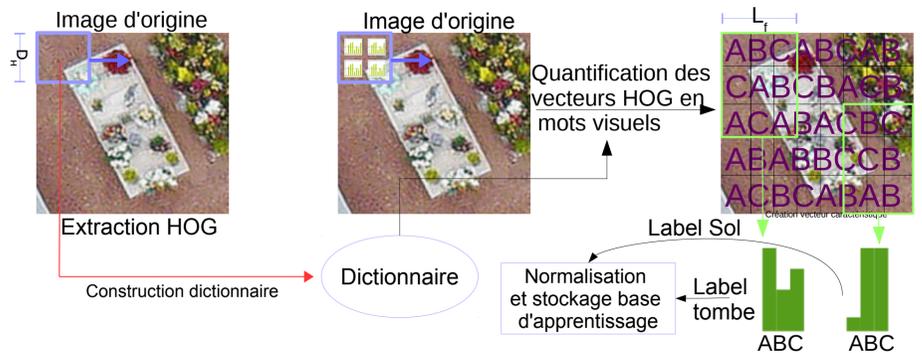
Segmentation Pixel

consiste à classifier directement chaque pixel de l'image [1].

Schéma classique :

- Extraction d'un vecteur caractéristique par pixel,
- Quantification des vecteurs en mots visuels [4],
- Histogramme des mots visuels du voisinage du pixel étudié,
- Labellisation et apprentissage des histogrammes de mots visuels.

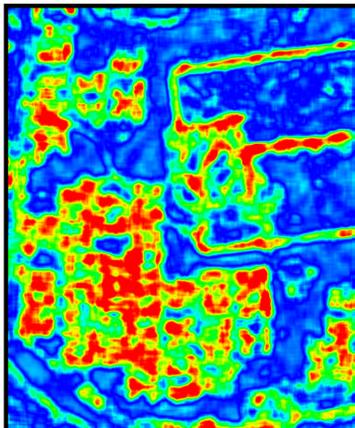
Construction base d'apprentissage



Phase d'apprentissage

Apprentissage par un SVM linéaire [5] sur des vecteurs contenant **655360 mots visuels**.

Phase de classification



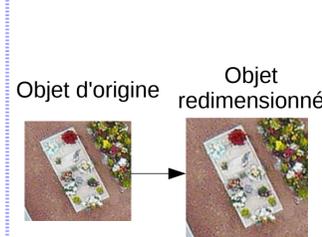
retourne une **carte de probabilité** de présence de l'objet.

Segmentation Objet

construit le modèle de l'objet [2].

Nous formons un seul vecteur caractéristique pour chaque fenêtre englobant un objet recherché.

Construction base d'apprentissage



Extraction des descripteurs HOG



f, vecteur contenant tous les descripteurs HOG pour une résolution :



g, vecteur contenant tous les descripteurs HOG pour une résolution :



h, vecteur contenant tous les descripteurs HOG pour une résolution :

ooo

Le vecteur caractéristique est formé de la concaténation de **f, g, h, ...**

Phase d'apprentissage

Apprentissage par un SVM linéaire [5] sur des vecteurs de dimension 160000.

Phase de classification



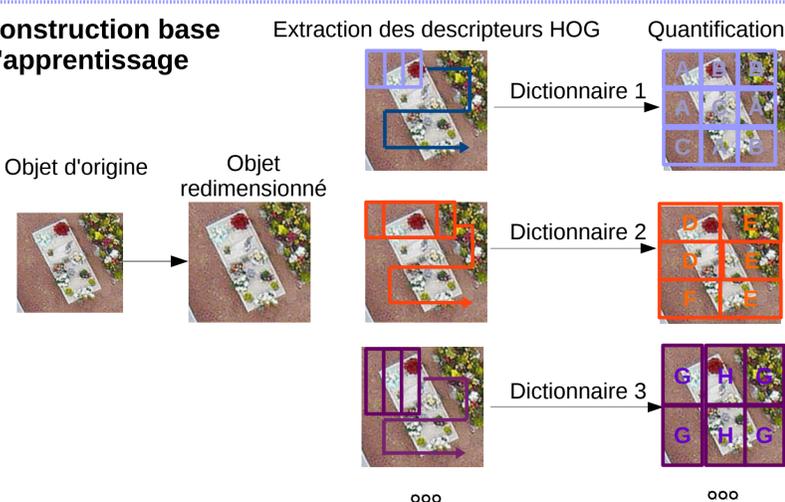
Une fenêtre glissante va extraire le vecteur caractéristique en toutes les positions et échelles de l'image de test.

Pour chaque **fenêtre** le **modèle détermine** si l'objet est présent ou pas.

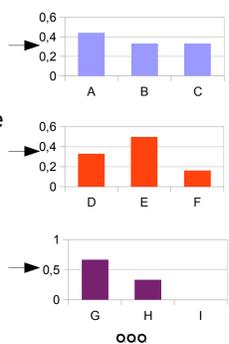
Proposition

Intégration de l'aspect "sac de mots visuels" de la segmentation pixel dans une approche par segmentation objet sans perdre l'aspect multi-résolution !

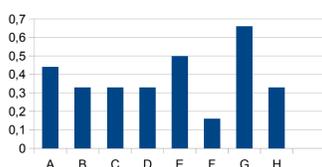
Construction base d'apprentissage



Calcul des fréquences de mots visuels pour chaque résolution



Concaténation des différents histogrammes de mots visuels



L'utilisation de l'approche pyramide [3] permet de garder une partie de l'information spatiale des mots visuels.

Résultats et conclusion

Comparaison de la segmentation pixel et Objet

Nécessité de segmenter la carte de probabilité :

- transformée de Hough pour détecter les rectangles,
- assignation à chaque rectangle d'une probabilité.

Détermination du seuil minimal pour qu'une probabilité corresponde à l'objet.

Pour un rappel fixé (à 41%) nous comparons le taux de détection des deux approches.

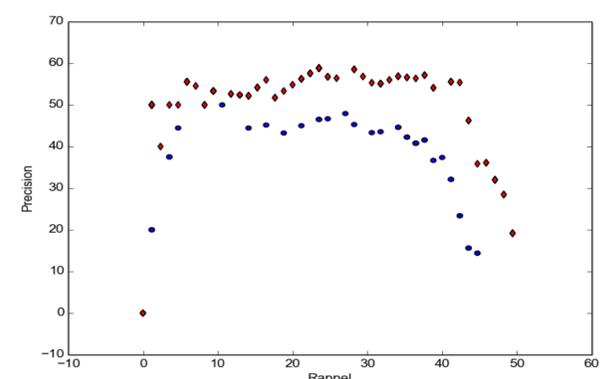
	Segmentation Pixel	Segmentation Objet
Précision	23 %	30 %

Évaluation de notre proposition

Courbe Précision/Rappel en fonction du seuil minimal.

● : Segmentation objet

◆ : Notre proposition



Conclusion

Dans notre cas d'étude, la **segmentation pixel est moins précise que la segmentation objet** (de 7% en précision). Ceci s'explique par la proximité des objets avec des contours partiellement visibles.

Notre approche combinant des mots visuels de multi-résolution permet de quantifier et concentrer l'information, ce qui **augmente la précision moyenne de 9%** par rapport à la segmentation objet.

¹ D. Aldavert, A. Ramisa, R. Toledo, R. López de Mántaras, *Fast and Robust Object Segmentation with the Integral Linear Classifier*, CVPR, 2010.

² Q. Zhu, S. Avidan, M. Yeh and K. Cheng, *Fast Human Detection Using a Cascade of Histograms of Oriented Gradients*, CVPR, 2006.

³ S. Lazebnik, C. Schmid and J. Ponce, *Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories*, CVPR, 2006.

⁴ D. Nister, H. Stewenius, *Scalable Recognition with a Vocabulary Tree*, CVPR, 2010.

⁵ R.E.Fan, K.W. Chang, C.J. Hsieh, X.R. Wang and C.J. Lin, *LIBLINEAR: A Library for Large Linear Classification*, Journal of Machine Learning Research, 2006.