

# Un environnement efficace pour la classification d'images à grande échelle

Thanh-Nghi Doan\*, François Poulet\*,\*\*

\*\*Université de Rennes I, \*IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France,  
{thanh-nghi.doan | francois.poulet}@irisa.fr

## Résumé

La plupart des processus de classification d'images comportent trois principales étapes : l'extraction de descripteurs de bas niveaux, la création d'un vocabulaire visuel par quantification et l'apprentissage à l'aide d'un algorithme de classification (eg.SVM). De nombreux problèmes se posent pour le passage à l'échelle comme avec l'ensemble de données ImageNet contenant 14 millions d'images et 21,841 classes. La complexité concerne le temps d'exécution de chaque tâche et les besoins en mémoire et disque (eg. le stockage des SIFTs nécessite 11To). Nous présentons une version parallèle de LibSVM pour traiter de grands ensembles de données dans un temps raisonnable. De plus, il y a beaucoup de perte d'information lors de la phase de quantification et les mots visuels obtenus ne sont pas assez discriminants pour de grands ensembles d'images. Nous proposons d'utiliser plusieurs descripteurs simultanément pour améliorer la précision de la classification sur de grands ensembles d'images. Nous présentons nos premiers résultats sur les 10 plus grandes classes (24,817 images) d'ImageNet.

## Summary

The usual frameworks for image classification involve three steps : feature extraction, building codebook by feature quantization and training the classifier with a standard classification algorithm (eg. SVM). However, task complexity becomes very large when applying this approach on large scale datasets like the ImageNet dataset containing more than 14 million images and 21,000 classes. The complexity is both about the time needed to perform each task and the memory and disk usage (eg. 11TB are needed to store the SIFT descriptors computed on the full datasets). We have developed a parallel version of LIBSVM to deal with very large datasets in reasonable Temps. Furthermore a lot of information is lost when performing the quantization step and the obtained bag-of-words (or visual-words) are often not enough discriminative for large scale image classification. We present a novel approach using several local descriptors simultaneously to try to improve the classification accuracy on large scale image datasets. We show our first results on dataset made of the ten largest classes (24,817 images) from ImageNet.