



## Rapport sur le mémoire de thèse de Mr Nicolas HERVE, intitulé « Vers une description efficace du contenu visuel pour l'annotation automatique d'images »

Le travail de N. Hervé porte sur l'annotation automatique de bases d'images, c'est-à-dire l'étiquetage d'images par des concepts représentatifs de leurs caractéristiques visuelles. N. Hervé se place d'emblée dans le cadre de bases d'images généralistes et cherche à développer des techniques d'annotation génériques, adaptées à tout type de base d'image. L'annotation est faite à partir d'un petit nombre fixé de concepts qui peuvent illustrer différentes caractéristiques de l'image. C'est le cadre habituel des campagnes ou des corpus d'évaluation dans ce domaine. Dans ce contexte, N. Hervé, développe une méthodologie d'annotation basée sur l'utilisation de classifieurs pour étiqueter les images par ces concepts. Le manuscrit comporte deux contributions principales. La première concerne l'analyse de descripteurs globaux des images et la seconde est consacrée à l'utilisation de caractéristiques locales. Le travail a été effectué dans l'équipe Imédia il fait suite à d'autres travaux sur la recherche d'image par le contenu sur lesquels N. Hervé s'est appuyé lors de sa thèse.

Le manuscrit suit cette organisation en deux grandes parties consacrées respectivement à chacun de ces problèmes. Chacune comprend un état de l'art et une description des contributions originales. Le manuscrit se lit facilement. Il reste des imprécisions à corriger pour la version finale.

Après une introduction générale à la problématique, la première partie donne les notions de base de la recherche d'image par le contenu et de l'annotation à partir de caractéristiques globales. N. Hervé passe ainsi en revue, l'extraction de caractéristiques images, les mesures de similarité, les mesures de performance. Il discute ensuite de l'intérêt de l'utilisation combinée d'informations texte et image et présente des algorithmes de base pour la classification. L'annotation globale consiste à étiqueter des images avec des concepts qui caractérisent globalement l'image. N. Hervé utilise pour cela des caractéristiques calculées sur l'image entière. Une première contribution introduit un nouveau descripteur image qui caractérise les orientations de contour présentes dans l'image. Ce descripteur sera utilisé par la suite en conjonction avec ceux déjà développés pour la recherche par contenu dans le moteur de recherche de Imédia. Ensuite, il décrit l'approche de classification à base de SVM qu'il a développée et testée dans le cadre de la campagne d'évaluation ImageVAL sur une tâche visant à caractériser la nature et le contexte des images. N. Hervé a effectué plusieurs soumissions en utilisant différentes caractéristiques et plusieurs variantes de l'algorithme de classification SVM. Il a obtenu les meilleures performances en 2006, sur 6 équipes, sur cette tâche de classification. Il fournit dans son manuscrit une étude détaillée de l'influence des différents paramètres (caractéristiques image et paramètres d'apprentissage) sur les performances et le comportement de l'algorithme. Cette étude est suivie d'une discussion sur les limitations des bases d'évaluation disponibles actuellement. Il termine ce chapitre par une étude des capacités de généralisation d'un classifieur entraîné sur une base et testé sur des bases de caractéristiques différentes. La conclusion est mitigée.

La deuxième partie est consacrée à l'annotation locale : comment étiqueter les images en fonction de caractéristiques locales comme par exemple des objets. N. Hervé fait le choix d'une représentation par sac de mots visuels qui est la plus en vogue actuellement. Il propose dans un premier temps une série d'expérimentations portant sur la création du dictionnaire de mots visuels utilisé pour cette représentation. Il compare différents algorithmes de clustering et propose une variante d'un algorithme existant qui, dans les tests effectués sur une base de référence, améliore légèrement les performances de classification par rapport à un algorithme k-moyenne de base ou à une sélection aléatoire des mots visuels. Différentes analyses des paramètres de l'algorithme complètent cette étude.

N. Hervé développe ensuite un parallèle intéressant entre les tâches de classification texte et image. Son idée est d'étudier les similarités entre les deux tâches et de voir s'il y existe des points communs. Il utilise un corpus de texte légèrement modifié afin que ses caractéristiques soient proches d'un corpus image. Les modifications reviennent à éliminer la notion de mot en enlevant tout ce qui est caractéristique des séparations inter-mots (espaces, ponctuations, majuscules). Les tests comparatifs sont effectués sur une part de la base Reuters 2004 pour le texte et une base d'ImageVAL pour l'image. Les principales conclusions sont la supériorité dans les deux cas des échantillonnages réguliers des mots des documents par rapport à des échantillonnages spécifiques (e.g. points d'intérêt ou mots rares) et une analyse de l'influence de la taille du vocabulaire. Un bémol dans cette comparaison, l'étude sur le texte revient principalement à comparer une approche n-grams à une approche par mots et les conclusions ne sont pas surprenantes. L'idée est donc intéressante, mais la portée des conclusions reste limitée. Deux autres problèmes sont étudiés dans cette partie. Le premier concerne l'utilisation de cooccurrence locale de termes images en complément à un codage sac de mots. Cette idée a souvent été déclinée dans le cas du texte. Elle apporte dans le cas présent des améliorations significatives. La deuxième idée concerne l'utilisation de méthodes de retour de pertinence pour l'étiquetage des données. Plusieurs méthodes sont comparées, dont une qui propose une stratégie originale pour exploiter simultanément des caractéristiques locales et globale des images. N. Hervé fait des essais à partir de plusieurs modèles d'utilisateurs types. La conclusion est qu'il peut avec 10 % des images sélectionnées et étiquetées par l'utilisateur atteindre les mêmes performances que celles qu'il a auparavant obtenues avec un apprentissage classique qui exploite une base d'images entièrement pré-annotée.

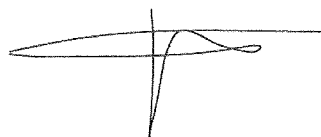
N. Hervé a fait un très gros travail expérimental pour proposer et étudier différentes stratégies d'annotation d'images. Sa démarche est assez générique et doit pouvoir être transposée à d'autres problèmes d'annotation d'images. Il a dégagé quelques conclusions claires dont la portée est certainement assez générale.

Je donne un avis favorable à la soutenance de ce travail pour l'obtention du grade de docteur de l'université Paris-Sud.

Paris le 24 - 05 - 2009

Patrick Gallinari

Professeur



University Paris VI, 104 Av. Président Kennedy, 75016 Paris, France

Tel : (33-1) 44 27 73 70 - Fax (33-1) 44 27 70 00 - E-mail: Patrick.Gallinari@lip6.fr