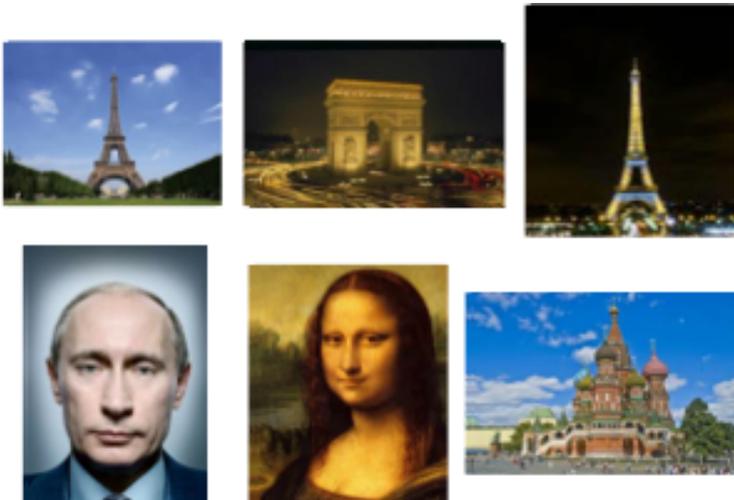


Sujet de stage :

Vers la conception interactive d'une hiérarchie sémantique de descripteurs d'images

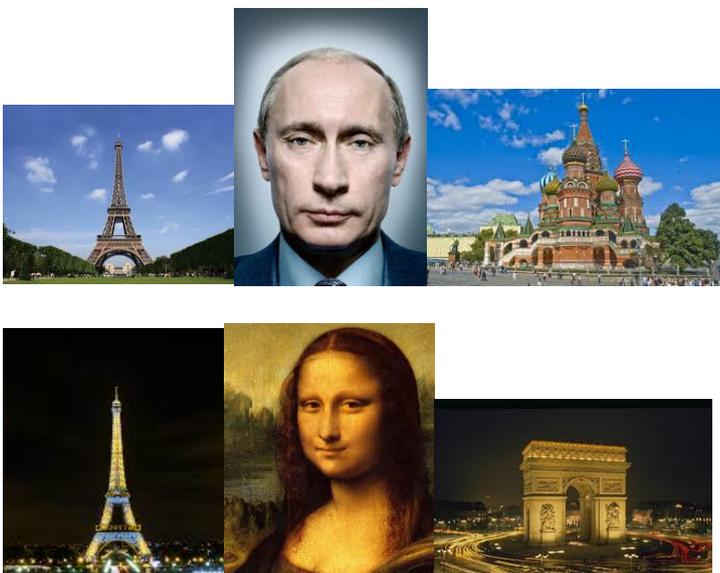
Résumé du travail proposé :

Ce travail s'intéresse à la différence entre la perception humaine des images, et la représentation que peut en faire la machine. Prenons l'exemple du jeu d'images ci-dessous :

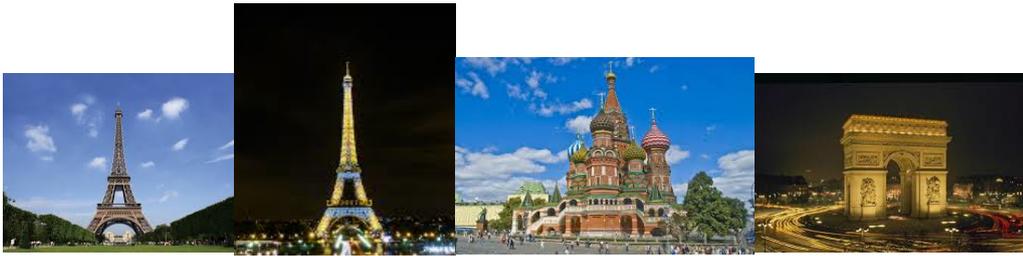


Admettons que nous voulions amener la machine à catégoriser ces images.

Si l'on confie la catégorisation à la machine selon des critères colorimétriques, il est possible que les catégories soient les suivantes : deux catégories, avec les images à prédominance jaunes et noires d'un côté, images à prédominance bleue de l'autre.



Si l'on donne à la machine des critères d'analyse un peu plus sémantiques au travers par exemple d'un apprentissage, il sera possible d'obtenir deux groupes, les bâtiments d'un côté, et les portraits de l'autre.



Ces deux catégories pouvant éventuellement être redécoupées en 3 catégories, pour différencier les bâtiments de vue de jour d'un côté et les bâtiments de vue de nuit de l'autre.

Tandis que la perception humaine, qui diffère bien sûr en fonction de la personnalité et de la culture de l'individu concerné, pourra résulter en 2 groupes de niveau sémantique bien plus élevé : Paris d'un côté, et Moscou de l'autre.



Il sera bien évidemment difficile d'amener la machine à constituer ce genre de groupes dès lors qu'elle se base uniquement sur le contenu pixellaire des images (sans méta-donnée textuelle ou GPS).

Le sujet de ce stage est de concevoir une méthode (et donc un programme informatique) permettant à la machine de se rapprocher de la perception humaine.

Pour cela, l'idée est de faire interagir l'utilisateur avec les résultats renvoyés par la machine, pour les faire converger vers ceux souhaités par l'utilisateur. Ce travail peut être fait à plusieurs niveaux. Au niveau de la catégorisation (*clustering*), nous avons proposé dans le contexte de la thèse de

Lai Hien Phuong une méthode interactive de *clustering* semi-supervisé qui vise à rapprocher incrémentalement les catégories découvertes par la machine de celles perçues par l'humain.

Ce stage s'intéresse à la possibilité de modifier directement les descripteurs d'images en fonction des retours donnés par l'utilisateur.

Mots clés :

Content-Based Image Retrieval, Image descriptors, human interaction.

Informations complémentaires :

Encadrant(s) : Muriel Visani (Université de La Rochelle), Thierry Urruty (Université de Poitiers)

Thématiques :

- Ingénierie des connaissances
- Analyse et gestions de contenus
- Interactivité et dynamique des systèmes

Domaine d'application :

- Pertinence – contenu – interactions
- Environnement

Cadre de coopération : projet CINEDI, Communauté d'Universités Léonard de Vinci, Fédération de Recherche CNRS MIRES

Date de début du stage : Janvier 2016 (ou plus tard, selon l'étudiant)

Durée du stage : 5 à 6 mois.

Financement : projet CINEDI

Contexte de l'étude:

Ce stage s'inscrit dans le contexte du projet CINEDI, financé par le CNRS au travers du Groupe de Recherche ISIS (Information, Signal, Image et ViSion). Ce projet traite de l'analyse d'images par le contenu, en vue de l'utilisation d'outils de navigation dans des bases d'images, ou de recherche par exemple.

Il sera co-encadré par Muriel Visani (université de La Rochelle) et Thierry Urruty (Université de Poitiers).

Description du sujet :

L'objectif principal du stage est de concevoir une hiérarchie sémantique de descripteurs d'images locaux afin d'améliorer la pertinence des moteurs de recherche image.

Beaucoup de systèmes de recherche d'images existants reposent sur une description des images basées sur des « mots visuels » [Sivic 2003, Chatfield 2011, Avila 2013]. Dans la pratique, on s'aperçoit souvent que les lexiques de sacs de mots visuels construits « à plat » contiennent naturellement de nombreux synonymes et/ou polysèmes (en fonction notamment de la taille du lexique considéré), ce qui est source d'ambiguïtés et d'incertitude dans la description.

Quelques travaux se sont intéressés à générer un lexique de niveau plus global dans l'image, en agrégeant non plus simplement des mots, mais des chaînes de mots [Ros 2009] ou des phrases,

constituées de groupes de mots visuels significatifs que l'on retrouve fréquemment ensemble [Yuan 2007], voire dans une même configuration spatiale [Zhang 2011] dans les images de la collection.

Dans le contexte du projet CINÉDI, en plus de ces informations spatiales, nous souhaitons in fine nous aider également d'une information de plus haut niveau sémantique concernant les images, obtenue soit par une annotation préalable de celles-ci, soit incrémentalement dans une phase d'interaction avec l'utilisateur, pour créer des phrases visuelles qui soient plus représentatives des concepts perçus par un humain dans les images.

Ce stage s'intéressera prioritairement aux deux questions suivantes :

- La proposition d'une approche de structurer les mots visuels selon une hiérarchie de phrases visuelles qui tienne compte à la fois de leur agencement spatial dans l'image, et des classes ou attributs associées aux images par l'interaction humaine. Il faudra, dans la conception de cette approche prototype, prendre en compte sur les contraintes liées au contexte interactif pour nos développements futurs (faible complexité calculatoire, convivialité des modes d'interaction avec l'utilisateur, etc.).
- L'implémentation d'un premier prototype de cette approche.

Prérequis et contraintes particulières :

Le candidat doit :

- être actuellement en master 2 d'informatique ou justifier de compétences équivalentes
- de préférence, avoir suivi quelques cours de traitement et d'analyse d'images
- avoir un certain goût pour la recherche (étude bibliographique à prévoir)
- avoir un bon niveau de programmation
- avoir un bon niveau d'anglais lu et écrit

Merci d'envoyer une lettre de motivation, ainsi qu'un CV détaillé, aux deux encadrants du stage, avant le 30 Octobre 2015. Des auditions seront organisées dans le courant du mois de novembre. Le stage pourra se dérouler, au choix de l'étudiant, soit à La Rochelle, soit à Poitiers.

Références bibliographiques :

[Sivic 2003] J. Sivic et A. Zisserman. Video Google : a text retrieval approach to object matching in videos. Proceeding of the IEEE International Conference on Computer Vision, pp. 1470–1477, 2003.

[Chatfield 2011] K. Chatfield, V. Lempitsky, A. Vedaldi et A. Zisserman. The devil is in the details : an evaluation of recent feature encoding methods. In proceedings of the British Machine Vision Conference (BMVC), pages 1–12, 2011.

[Avila 2013] S. Avila, N. Thome, M. Cord, E. Valle et A.A. Araújo. Pooling in image representation : The visual codeword point of view. Computer Vision and Image Understanding, vol. 117, no. 5, pages 453–465, 2013.

[Ros 2009] J. Ros, C. Laurent et J.M. Jolion. A bag of strings representation for image categorization.

Journal of Mathematical Imaging and Vision, vol. 35, no. 1, pages 51–67, 2009.

[Yuan 2007] J. Yuan, Y. Wu et M. Yang. Discovery of collocation patterns : from visual words

to visual phrases. In proceedings of the IEEE international conference on Computer Vision and Pattern Recognition (CVPR), pages 1–8, 2007.

[Zhang 2011] Y. Zhang, Z. Jia et T. Chen. Image retrieval with geometry-preserving visual phrases. In proceedings of the IEEE international conference on Computer Vision and Pattern Recognition (CVPR), pages 809–816, 2011.

[Philbin 2008] J. Philbin, O. Chum, M. Isard, J. Sivic et A. Zisserman. Lost in quantization : Improving particular object retrieval in large scale image databases. In proceedings of the IEEE international conference on Computer Vision and Pattern Recognition (CVPR), pp 1–8, 2008.

Contacts – liens :

Email : mvisani@univ-lr.fr, thierry.urruty@univ-poitiers.fr

Lien vers le fichier de description : (PDF) (si nécessaire)