

Compte-rendu de la journée
Passage à l'échelle dans recherche d'information
multimédia
(GDR ISIS et I3, 9 juin 2009)

Laurent Amsaleg (IRISA), Michel Crucianu (CNAM)

16 juin 2009

La journée, qui a regroupé 46 inscrits, s'est articulée autour de trois grandes familles de travaux concernant respectivement :

1. le passage à l'échelle de la recherche par similarité sur la base de descripteurs de bas niveau (Picard, Gouet, Joly, Douze, Gorisse, Tirilly, Hède),
2. le passage à l'échelle de méthodes d'apprentissage statistique (Poulet, Martinez, James),
3. les problèmes d'échelle liés à des données spécifiques (Campedel, Morand).

Bien naturellement, les lignes de coupe entre ces travaux ne sont pas franches. Par ailleurs, chacun focalise ses efforts pour contourner un ou plusieurs obstacles que la volonté de passer à l'échelle fait apparaître.

Certains travaux attaquent de front les problèmes d'échelle lorsque le volume des collections de données à indexer croît. Quelques travaux montrent que l'on peut accélérer les recherches exactes grâce à des structures d'index (Gouet) ou au parallélisme et à la distribution (Martinez, Hède). On distingue ensuite plusieurs approches de recherche très efficaces tant que les données tiennent en mémoire. Ces approches visent toutes l'obtention de performances grâce à des recherches approximatives avec ou sans contrôle de la qualité (Picard, Gouet, Joly, Douze). Certains travaux visent à éliminer certains faux positifs, dont l'apparition est la conséquence des approximations (Douze, Gorisse).

Une autre tendance est d'attaquer le problème de la taille des collections au travers de transformations des descriptions de bas niveau permettant l'utilisation de techniques issues du texte (Douze, Tirilly). En particulier, l'analyse des facteurs de pondérations locales et globales classiquement utilisées en texte montre qu'ils ont une très grande influence lorsque les bases d'images traitées sont plutôt homogènes (bases de bâtiments par ex.), le biais est alors fort, et que l'influence de ces pondérations est bien moins sensible lorsque la base d'image est plus grande et plus hétérogène.

L'apprentissage statistique (supervisé ou non supervisé) est souvent utilisé dans un contexte de structuration ou de recherche par le contenu multimédia, les problèmes de passage à l'échelle (données volumineuses, données distribuées) se posent également pour l'apprentissage. Une accélération des algorithmes de classification/apprentissage (SVM, *Random Forest*) traditionnels grâce aux GPU et aux grilles de calcul a été présentée (Poulet), rendant envisageables leur application à ces contextes.

Des problèmes d'échelle spécifiques se posent aussi au niveau des images mêmes, comme dans le cas d'images urbaines de très haute résolution où une perte en reconnaissance est constatée (Picard), avec la vidéo HD (Morand) ou les images satellitaires dont la taille, typiquement 24000 x 24000 pixels, soulève de nombreux problèmes (Campedel). La très haute résolution engendre des difficultés dans l'emploi d'outils de bas niveau comme la segmentation en régions d'intérêt ou la classification d'éléments de l'image.

De manière générale, les problèmes d'échelle sont considérés dans un environnement où les collections de données à traiter ne disposent d'aucune annotation de haut niveau, mais seulement de vérité terrain au niveau des descripteurs eux-même ou des images devant être mises en correspondance. Une des difficultés des annotations « sémantiques » est qu'elles contiennent peu ou pas de contexte, introduisant de nombreuses ambiguïtés sur le sens des mots (James).

Néanmoins, la constitution de benchmarks standardisés est une question centrale exprimée par tous les participants. Définir un corpus associé à une vérité terrain tant d'un point de vue strictement bas niveau avec la recherche à epsilon près ou la recherche de k -ppv, qu'à un niveau plus haut avec des classes ou même des annotations, est un besoin partagé.

Nous proposons ainsi de tenter de constituer une première ébauche de corpus composé de :

1. Un ensemble d'images du Web, par exemple issues du corpus COPHIR. Sur cet ensemble, on se propose de calculer des descriptions SIFT (dans un premier temps) et de distribuer celles-ci (et non les images). À partir de ces descriptions, on constituera une vérité terrain au niveau descripteur à la manière de TREC, c'est à dire bâtie en fonction des résultats de ce que retournent tous les partenaires intéressés. On ne peut pas envisager une vérité terrain calculée grâce à une approche séquentielle, sauf si on utilise une grille... Ici, on se limite à des tâches de recherche à epsilon près et k -ppv.
2. Un corpus vidéo de plusieurs milliers d'heures, à partir de contenus libres de droits, issus de *European Web Archive* et *Internet Web Archive*. Des transformations générées avec l'outil INRIA (également employé pour Trecvid 2008 CBVCD), classées par type et valeurs des paramètres, permettront de constituer la vérité terrain.
3. Il est également envisagé de faire une vérité terrain sur la base d'images mentionnée, en générant un ensemble de transformations type STIRMARK.

Accès aux présentations en PDF :

Etude des descripteurs SIFT pour le matching de scènes dans un contexte urbain de type Google StreetView, David Picard, Eduardo Valle, Matthieu Cord, LIP6
<http://cedric.cnam.fr/~crucianm/src/echelle/DPicard.pdf>

HiPeR : Un modèle hiérarchique pour la recherche exacte, progressive et approximative par similarité, Nouha Bouteldja, Valérie Gouet-Brunet, Michel Scholl, Vertigo - CEDRIC & Wisdom, CNAM Paris

<http://cedric.cnam.fr/~crucianm/src/echelle/VGouet.pdf>

A posteriori multi-probe LSH for large scale image retrieval, Alexis Joly (INRIA-IMEDIA), Olivier Buisson (INA)

<http://cedric.cnam.fr/~crucianm/src/echelle/AJoly.pdf>

Recherche d'images rapide pour des similarités basées noyaux, David Gorisse (ETIS), Frédéric Précioso (ETIS), Matthieu Cord (LIP6), Sylvie Philipp-Foliguet (ETIS)

<http://cedric.cnam.fr/~crucianm/src/echelle/DGorisse.pdf>

Propriétés et limites des représentations en sac de mots visuels, Pierre Tirilly (IRISA-CNRS), Vincent Claveau (IRISA-CNRS), Patrick Gros (INRIA Rennes-IRISA)

<http://cedric.cnam.fr/~crucianm/src/echelle/PTirilly.pdf>

Web-scale image search, Matthijs Douze (INRIA-LEAR)

<http://cedric.cnam.fr/~crucianm/src/echelle/MDouze.pdf>

Indexation hiérarchiques des contenus vidéo HD par objets dans le cadres du standard scalable JPEG2000, C. Morand, J. Benois-Pineau, J.-Ph. Domenger (LABRI)

<http://cedric.cnam.fr/~crucianm/src/echelle/CMorand.pdf>

Passage a l'échelle d'algorithmes de classification, François Poulet, Université de Rennes I – IRISA

<http://cedric.cnam.fr/~crucianm/src/echelle/FPoulet.pdf>

Accès à l'information multimédia classifiée, José Martinez (INRIA-ATLAS)

<http://cedric.cnam.fr/~crucianm/src/echelle/JMartinez.pdf>

Extraction d'information dans les images satellitaires, Marine Campedel, Telecom ParisTech, CoC

<http://cedric.cnam.fr/~crucianm/src/echelle/MCampedel.pdf>

De 2 à 22 millions d'images, création indexation et recherche par le contenu avec PiRiA, Patrick Hède, Commissariat à l'Énergie Atomique (CEA) LIST

<http://cedric.cnam.fr/~crucianm/src/echelle/PHede.pdf>

Vers une base d'apprentissage pour l'annotation automatique d'images à grande échelle sémantisée, Nicolas James, Laboratoire MAS, École Centrale Paris

<http://cedric.cnam.fr/~crucianm/src/echelle/NJames.pdf>