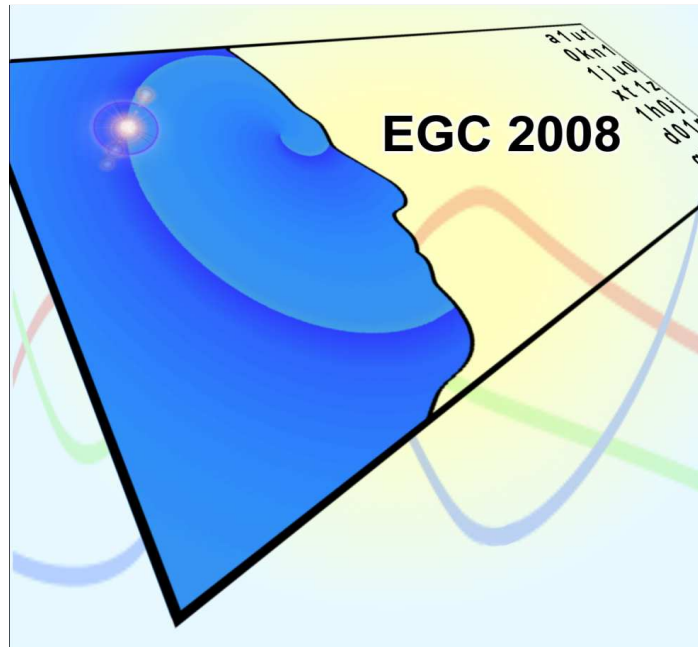


Atelier



ECOI : Extraction de COonnaissance et Images

Organisateurs :

- Nicolas Loménie (Univ. Paris V)
- Nicole Vincent (Univ. René Descartes)
- Rémy Mullot (L3i, La Rochelle)

Responsables des Ateliers EGC :

Alzenny Da Silva (INRIA, Rocquencourt)
Alice Marascu (INRIA, Sophia Antipolis)
Florent Masegla (INRIA, Sophia Antipolis)

<http://www-sop.inria.fr/axis/egc08>

EGC

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche SOPHIA ANTIPOLIS - MÉDITERRANÉE

Table des matières

Concepts visuels et reconnaissance d'objets et d'activités, <i>Monique Thonnat</i>	3
Ontologies pour l'analyse et l'interprétation d'images : état de l'art, <i>Céline Hudelot</i>	5
Un atelier d'ingénierie des connaissances en traitement d'images, <i>Régis Clouard, Arnaud Renouf, Marinette Revenu</i>	7
Vers la construction d'une ontologie appliquée à l'imagerie satellitaire, <i>Marine Campedel, Marie Lienou, Ivan Kyrgyzov, Henri Maitre</i>	19
Reconstruction 3D sémantique dans le contexte architectural, <i>Christophe Cruz, Franck Marzani, Frank Boochs</i>	31
Enrichissement sémantique d'annotations de documents multimédias à l'aide de connaissances explicites, <i>Nicolas James, Céline Hudelot</i>	43
Vers une ontologie modulaire multi-niveaux pour l'indexation sémantique multi-points de vue du compte-rendu d'imagerie médicale, <i>Sonia Mhiri, Sylvie Despres, Ezzedine Zagrouba</i>	47
Utilisation de structures sémantiques pour la recherche d'images sur Internet, <i>Adrian Popescu, Pierre-Alain Moellic, Ioannis Kanellos</i>	59
Extraction de détecteurs d'objets urbains à partir d'une ontologie, <i>Sébastien Deriveaux, Germain Forestier, Cédric Wemmert, Sébastien Lefevre</i>	71

Concepts visuels et reconnaissance d'objets et d'activités

Monique Thonnat

Equipe-projet PULSAR
INRIA Sophia Antipolis

Résumé. Dans cet exposé, j'aborderai deux aspects liés à l'interprétation sémantique de données visuelles (Maillot et Thonnat (2008); Bremond et al. (2004)). Tout d'abord je présenterai des travaux sur la reconnaissance d'objets dans des images statiques. Ces travaux effectués dans le cadre de la thèse de Nicolas Maillot s'appuient sur une ontologie de 101 concepts visuels. Un point important consiste en l'apprentissage automatique des détecteurs des concepts visuels permettant de décrire la couleur ou la texture des objets à reconnaître. La deuxième partie de mon exposé concerne l'analyse temps-réel de vidéos pour la reconnaissance d'activités humaines. L'interprétation sémantique de vidéos nécessite une bonne détection des objets mobiles mais aussi une modélisation de leurs activités. Un effort de standardisation récent sur la terminologie des événements vidéos a permis de définir une première ontologie. Je conclurai mon exposé par quelques perspectives de recherches sur ces thèmes.

Références

- Bremond, F., N. Maillot, T. M., et T. Vu (2004). Ontologies for video events. Research report, INRIA Sophia Antipolis. Equipe-projet PULSAR.
- Maillot, N. et M. Thonnat (2008). Ontology based complex object recognition. *Image and Vision Computing Journal* 26(1), 102–113.

Summary

Team-project PULSAR - INRIA Sophia Antipolis

Ontologies pour l'analyse et l'interprétation d'images : état de l'art

Céline Hudelot*

* Laboratoire MAS - Ecole Centrale Paris, Châtenay-Malabry, France
celine.hudelot@ecp.fr,
<http://www.mas.ecp.fr/>

Dans les domaines de l'analyse de scènes et de l'interprétation d'images, tout comme dans le domaine de l'indexation des images numériques, on assiste, grâce aux progrès récents en ingénierie des connaissances, à un regain d'intérêt pour les approches s'appuyant sur la modélisation de connaissances a priori sur le domaine étudié. Comme en témoignent les nombreux systèmes de vision à base de connaissances, dont on peut trouver un bon état de l'art dans Crevier et Lepage (1997), l'introduction de connaissances *a priori* dans le processus d'interprétation d'images n'est pas nouveau en soi. Cependant cette approche a souvent été critiquée notamment à cause de son manque de généralité (système *ad hoc*) mais aussi à cause du problème de la représentation et de l'acquisition des connaissances *a priori*. Des progrès récents en ingénierie des connaissances, en particulier les ontologies, permettent désormais de répondre à certaines de ces critiques (Hudelot (2005), Town (2006)). Celles-ci répondent en effet au problème de la formalisation, de manière cohérente et consensuelle, des connaissances d'un domaine donné.

Le problème de l'inter-opérabilité sémantique et plus particulièrement le problème du fossé sémantique défini par Smeulders et al. (2000) comme le problème *du manque de concordance entre les informations perceptuelles que l'on peut extraire des images et l'interprétation qu'ont ces images pour un utilisateur dans une situation déterminée* est une des principales raisons de l'importance prise par les ontologies dans le processus d'analyse et d'interprétation d'images.

Après une brève définition des ontologies, cette présentation présentera un panorama de l'apport des ontologies au domaine de l'interprétation d'images. Nous verrons que les ontologies peuvent intervenir à plusieurs niveaux dans le processus d'analyse. Nous distinguerons en particulier, trois types d'applications différentes des ontologies :

- *Les ontologies de description* en tant qu'outils pour la formalisation de la connaissance haut niveau sur le domaine d'application des images.
- *Les ontologies opérationnelles*, qui, ancrées à des informations perceptuelles permettent de guider l'analyse et l'interprétation des images.
- *Les ontologies de tâches* qui permettent de formaliser des connaissances sur les algorithmes de traitements et d'analyse d'images et de formuler les applications.

Nous nous intéresserons aussi aux raisonnements ontologiques et à leur application dans les différentes étapes d'une tâche d'interprétation d'images. Enfin, nous discuterons aussi des

Ontologies pour l'analyse et l'interprétation d'images

limites des approches ontologiques ainsi que des perspectives de recherche autour des ontologies et de l'interprétation d'images.

Références

- Crevier, D. et R. Lepage (1997). Knowledge-based image understanding systems : a survey. *Computer Vision and Image Understanding* 67(2), 160–185.
- Hudelot, C. (2005). *Towards a Cognitive Vision Platform for Semantic Image Interpretation ; Application to the Recognition of Biological Organisms*. Phd in computer science (in english), Université de Nice Sophia Antipolis.
- Smeulders, A., M. Worring, S. Santini, A. Gupta, et R. Jain (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(12), 1349–1380.
- Town, C. (2006). Ontological inference for image and video analysis. *Machine Vision and Applications* 17(2), 94–115.

Summary

The semantic interpretation of images and the domain of scene analysis can benefit from representations of useful concepts and the links between them as ontologies. In particular, ontologies are useful to tackle the famous *semantic gap* problem. In this communication, we briefly give a definition of ontologies. Then, we review the different applications of ontologies in scene analysis and image interpretation. At last, we conclude on the potential of knowledge engineering for the improvement of multimedia information systems.

Un atelier d'ingénierie des connaissances en traitement d'images

Régis Clouard, Arnaud Renouf, Marinette Revenu

GREYC, ENSICAEN & Université de Caen
6, boulevard Maréchal Juin, 104050 Caen cedex
Regis.Clouard@greyc.ensicaen.fr
<http://www.greyc.ensicaen.fr>

Résumé. Notre objectif de recherche à long terme est de développer un système à base de connaissances capable de produire automatiquement des logiciels de traitement d'images à partir d'une spécification faite par un utilisateur. Cet objectif a déjà été largement étudié et s'est heurté à un échec. Néanmoins, après avoir analysé les causes de cet échec, nous proposons de reprendre ces recherches en changeant de perspective pour adopter une approche constructiviste de la réalisation du système. Pour cela, il nous est nécessaire de construire un atelier d'ingénierie des connaissances en traitement d'images qui se veut le lieu de l'étude et du développement de ce logiciel. L'atelier se présente comme une ensemble d'outils logiciels partageant une même ontologie destinés à faire collaborer l'expert de traitement d'images et l'ingénieur de la connaissance afin d'aboutir à une théorie du développement d'application opérationnalisable dans un système.

1 Introduction

En dépit du nombre d'applications de traitement d'images réalisées grandeur nature variant autant les objectifs que les classes d'images, le développement d'une nouvelle application est encore décrit comme une tâche particulièrement complexe à la fois laborieuse et surtout à forte charge cognitive. Ceci fait que le développement d'applications de traitement d'images reste l'apanage de spécialistes et freine une réelle pénétration dans le milieu industriel alors que les besoins croissent très fortement à la mesure de l'essor des capteurs d'images.

C'est pourquoi très tôt, les chercheurs se sont intéressés à la conception de systèmes capables d'analyser automatiquement le contenu des images. Ces recherches ont débuté avec l'ambition de la généralité et du tout automatique. Les systèmes les plus emblématiques de ces travaux sont les systèmes d'interprétation d'images VISIONS/SCHEMA (Hanson, 1978) et SIGMA (Matsuyama, 1990). Mais très rapidement, le talon d'Achille de ces systèmes s'est révélé être la partie bas niveau, c'est à dire le traitement des images. L'effort de recherche avait été mis sur la partie haut niveau estimant que les erreurs et insuffisances de la partie bas niveau, réduite à de simples algorithmes généraux de segmentation, pouvaient être compensées par la partie haut niveau grâce aux connaissances modélisées sur le domaine d'application. Mais, l'analyse de ces systèmes montre au contraire que les impérities de la partie traitement d'images sont rédhibitoires pour le fonctionnement du système global (Draper, 1996). Dans un deuxième temps, les recherches se sont alors recentrées sur la partie bas niveau pour

tenter de la rendre plus adaptative en intégrant dès ce niveau les connaissances sur le domaine d'application. Pour réaliser cela, pratiquement toutes les architectures logicielles de résolution de problème ont été essayées (Crevier, 1997), (Clouard, 1999) :

- compétition d'algorithmes avec recherche de la meilleure solution ;
- systèmes experts ;
- réseaux de neurones ;
- modèles statistiques (réseaux bayésien, processus de décision markovien) ;
- systèmes à base de connaissances.

Malgré cette profusion de travaux qui a atteint son zénith au milieu des années 90, aucun système produit n'a permis une utilisation en routine. Cet échec a donné un violent coup d'arrêt aux travaux sur le développement de système automatique de traitement d'images à vocation générale.

1.1 Autopsie d'un échec

Les raisons de cet échec sont à rechercher d'une part dans la problématique même du traitement d'images et d'autre part dans l'approche de l'acquisition des connaissances injectées dans le système.

1.2 Les raisons liées au domaine du traitement des images

Le traitement d'images se caractérise principalement par la difficulté à poser le problème à résoudre. L'analyse des besoins est typiquement une activité de nature qualitative qui implique qu'il ne peut exister de spécification exhaustive ou exacte de l'application mais simplement une caractérisation approchée du comportement souhaité (Clouard, 2004).

La principale raison tient au fait que le problème à traiter n'est pas dans les images qui sont les seules informations objectives disponibles. En effet, d'une part les données image sont intrinsèquement incomplètes, dégradées et corrompues. La production d'image est responsable de la dénaturation de la scène (*i.e.*, le phénomène observé ou calculé) que ce soit la perte d'information réelle (*e.g.*, 3D, mouvement), le mélange de plusieurs sources dans la valeur d'un pixel dont la contribution relative est impossible à retrouver (*e.g.*, texture, illumination, géométrie), l'introduction de valeurs erronées (*e.g.*, bruit, aberration chromatique) ou la dégradation de l'information originelle (*e.g.*, distorsion géométrique, flou). D'autre part, le contenu de l'image n'a pas de sens en lui-même. Une image est ambiguë par nature et ne fournit pas d'information sur son contenu. Sans sujet de traitement, une image ne permet pas de distinguer les informations à considérer comme pertinentes de celles qui ne le sont pas. Par exemple, une information aussi simple que le contour d'objet est difficile à extraire avec précision sans connaissance sur le contenu de la scène. Les contours sont généralement modélisés comme une transition franche d'intensité, mais c'est aussi le cas du bruit, des ombres, de la texture, etc.

Pour espérer résoudre de tels problèmes, dits mal définis, il est nécessaire de disposer d'information sur le contexte de traitement pour contraindre l'espace des solutions. Ces informations ont pour but d'aider à corriger les données dégradées ou corrompues, à séparer les sources composant la valeur d'un pixel et à donner un sujet et un sens aux images. Elles sont disponibles pour une partie sous forme numérique et pour une autre sous forme symbolique. En conséquence pour réaliser un système de traitement d'images, il n'y a que deux solutions :

- faire un système dédié dans lequel les informations sur le domaine d'application sont implicitement codées ;

- faire un système général qui soit capable de collecter et d'organiser les informations numériques et symboliques sur le domaine d'application en connaissances utilisables pour la résolution.

1.3 Les raisons liées à l'acquisition des connaissances

Si on envisage la généralité, le choix de réalisation doit s'orienter vers un système à base de connaissances. En effet, si les traitements relèvent bien d'une approche numérique, le développement d'une solution logicielle complète ne se réduit pas à un problème calculatoire (Garbay-02) (Clouard-2004). La nature qualitative de l'analyse des besoins conditionne toute la chaîne de résolution de problèmes de traitement d'images. En particulier, la résolution ne peut pas s'envisager comme la recherche d'une solution optimale mais simplement comme la recherche d'une solution "satisfaisante" avec toute la subjectivité que cela comporte. Ceci fait du développement d'application de traitement d'images un domaine où l'approche algorithmique, type minimisation de fonctionnelle, apprentissage à partir d'exemples, modélisation statistique, est inapplicable pour concevoir des systèmes non dédiés.

Si les systèmes à base de connaissances sont légitimes, pourquoi ceux-ci n'ont-ils pas permis des systèmes opérationnels ? On peut constater à partir de la variété des approches proposées que ce n'est pas un problème d'architecture. La cause de l'échec est en fait à rechercher dans le goulot d'étranglement que représente l'acquisition des connaissances.

La première erreur, c'est d'avoir cherché à construire des modèles de connaissances qui soient le reflet du savoir des experts consultés, à partir d'une verbalisation de leurs connaissances et de leur démarche de résolution (Bachimont, 1996). Or, les connaissances de traitement d'image sont tacites et peu verbalisables et la démarche de résolution est essentiellement empirique en l'absence d'une théorie formelle du domaine. L'hypothèse de l'existence d'une expertise rationnelle s'est donc révélée fautive et la tentative d'acquisition des connaissances s'est avérée vaine.

La deuxième erreur, c'est d'avoir cru à l'hypothèse du monde fermé impliqué par la modélisation des connaissances a priori (Draper, 1996). L'essentiel de l'effort de modélisation s'est porté sur la représentation de différentes expertises de résolution de problèmes qui ont pu être identifiées dans le domaine du traitement d'images parce que l'on a fait l'hypothèse que le domaine se caractérisait par ses solutions. La formulation du problème se réduit alors à choisir une solution dont l'étiquette est libellée par le nom d'une tâche. L'utilisateur a l'impression de poser le problème mais en fait il choisit une solution. De plus pour rendre ces solutions effectivement opérationnelles, bon nombre de connaissances a priori y ont été implicitement codées, ce qui les a rendu peu réutilisables. Les informations sur le contexte (*e.g.*, caractéristiques des capteurs, des régions, des contours) servent simplement à paramétrer cette solution. Les systèmes produits se sont donc révélés peu adaptables parce que les solutions n'étaient pas construites spécifiquement pour le problème traité mais prédéfinies et simplement ajustées avec les informations du contexte.

1.4 Notre solution

Notre objectif de recherche à long terme consiste à concevoir un système à base de connaissances capable de produire automatiquement des logiciels de traitement d'images pour tout type d'application. Pour ne pas aboutir au même échec, nous avons choisi d'adopter une approche constructiviste de la construction du système à base de connaissances (Charlet, 2002). Il s'agit d'un changement de perspective pour passer d'une modélisation de l'expert et de son fonctionnement cognitif à une modélisation du système et de son fonctionnement ef-

fectif (Gaines, 1993).

C'est dans cette perspective que s'inscrit notre atelier d'ingénierie des connaissances. Il fournit l'environnement logiciel qui doit permettre d'élaborer les modèles originaux du développement d'application de traitement d'images qui seront ensuite opérationnalisés dans un système à base de connaissances.

La section 2 présente les objectifs, les hypothèses de base et l'architecture de notre atelier logiciel dans lequel les ontologies jouent un rôle central. La section 3 détaille la partie de l'atelier qui concerne la formulation d'application de traitement d'images. Nous l'avons souligné plus tôt, la formulation est un point critique dans la conception d'un système à base de connaissance pour le traitement d'images. En conclusion, nous faisons un bilan de l'état actuel de l'atelier puis des perspectives à court terme.

2 Un atelier d'ingénierie des connaissances

L'atelier a pour raison d'être la conception d'un système de génération automatique de logiciels de traitement d'images. Un tel système doit permettre à l'utilisateur de construire lui-même son logiciel dédié par interaction avec l'interface. L'utilisateur est considéré comme naïf du traitement d'images mais spécialiste d'un domaine d'application, c'est à dire qu'il a une réelle expertise dans le domaine des images et qu'il a un projet. Le logiciel produit est supposé capable de traiter toutes les images de la classe identifiée.

L'atelier se propose d'analyser toute la chaîne de développement depuis la formalisation des problèmes, la construction d'une solution logicielle jusqu'à son évaluation dans le but de construire des modèles qui forment les éléments d'une théorie du domaine du développement d'applications de traitement d'images du point de vue du système devant traiter ce problème. Pour cela, l'atelier doit être le lieu de l'étude de la connaissance en contexte. C'est l'interaction entre l'expert du traitement d'images et l'ingénieur de la connaissance par l'intermédiaire des outils logiciels fournis par l'atelier qui crée la connaissance (Linster, 1993).

L'atelier couvre tout le domaine du traitement d'images, c'est à dire qu'il est indépendant de tout domaine métier mais qu'il prend en compte toutes les tâches possibles en traitement d'images.

2.1 Le traitement d'images comme contexte

En l'absence de définition consensuelle du domaine du traitement d'images, nous proposons une définition établie selon un point de vue intentionnel. Le traitement d'images se définit à partir des six catégories de tâches atomiques résumées dans le tableau 1. Ces six catégories résultent de la combinaison des trois opérations de base - ajout, modification et suppression - appliquées à des fins de pure transformation ou à des fins épistémiques (augmentation de l'intelligibilité de l'information).

- La *restauration* a pour but de retrouver l'image d'origine à partir de sa version dégradée par ajout d'information au niveau signal connaissant la fonction de dégradation (ou au moins une estimation).
- L'*amélioration* a pour but d'adapter au mieux les données à une exploitation visuelle par modification de l'information au niveau signal en utilisant uniquement les valeurs de l'image d'entrée.
- La *compression* a pour but de réduire la quantité de données nécessaire pour stocker une image numérique en supprimant l'information identifiée comme redondante au niveau signal.

- La *reconstruction* consiste à créer une nouvelle image par ajout d'une information spatio-temporelle déduite d'une image ou de plusieurs (*e.g.*, la forme, le relief, le mouvement).
- La *segmentation* structure l'information signal à un niveau d'information symbolique : points d'intérêt, régions et contours.
- La *détection* fournit un masque positif des zones de l'image contenant les objets d'intérêt et supprime toutes les autres informations.

	Transformation	Augmentation
Ajout	<i>Restauration</i>	<i>Reconstruction</i>
Modification	<i>Amélioration</i>	<i>Segmentation</i>
Suppression	<i>Compression</i>	<i>Détection</i>

TAB. 1 – *Le traitement d'images recouvre six catégories de tâches.*

Une application complète de traitement d'images nécessite en généralement l'enchaînement de plusieurs de ces tâches élémentaires où une sortie considérée au niveau symbole peut se retrouver comme entrée au niveau signal dans la tâche suivante. Par exemple, une application de numérisation de livre ancien enchaîne une étape de reconstruction pour une remise à plat de la page par "shape from X" puis une amélioration du rendu visuel en homogénéisant la surface de chaque page avant la compression pour l'archivage numérique.

Cette hypothèse de l'existence d'un domaine du traitement d'images ne doit pas faire croire que nous allons retomber dans l'écueil du monde fermé. Notre thèse, c'est de considérer qu'il existe bien un domaine du traitement d'images indépendant des domaines d'application. Les connaissances de résolution ne se définissent pas par rapport à un domaine d'application mais bien par rapport au domaine de l'image numérique. Pour schématiser, il n'y a pas de technique de traitement d'images propre aux domaines biomédical ou astronomique ; ces techniques sont réutilisables en l'état dans d'autres domaines. Par contre, les informations sur le domaine d'application traité sont nécessaires à l'élaboration d'une solution. Il est donc indispensable d'acquérir ces informations auprès de l'utilisateur pour chaque application. Cela conduit à faire un vrai travail de modélisation de la formulation d'application qui doit distinguer ce qui doit être représenté dans le système de ce qui doit être acquis pour une application particulière.

2.2 Les hypothèses de base

L'hypothèse forte à la base de notre atelier, c'est de considérer que les connaissances de traitement d'images ne sont pas disponibles sous forme linguistique. Des travaux comme ceux de l'équipe LITIS de Rouen ont fait l'hypothèse inverse et ont basé l'élaboration de leur systèmes sur l'analyse de corpus linguistiques (interview, articles scientifiques et thèse) (Saidali-07) (Labiche-07). Notre point de vue, c'est de considérer que les corpus linguistiques (littérature ou interview) ne rendent pas compte d'une théorie de la résolution. S'ils détaillent bien les techniques du traitement d'images c'est uniquement par ce qu'elles sont, un peu par ce qu'elles font mais pas du tout par leurs conditions d'application. La raison est encore une fois que ces connaissances sont tacites et peu verbalisables. Ceci fait du traitement d'images un domaine très particulier pour lequel beaucoup de modèles et de méthodologies d'ingénierie des connaissances ne sont pas utilisables.

A défaut de corpus linguistique constitué, il nous est nécessaire de construire nos propres

modèles de capitalisation des connaissances afin de constituer les corpus. Ce qui est immédiatement disponible en traitement d'images, c'est le code du programme qui constitue l'application. Toutefois, nous proposons de capitaliser non pas les programmes directement mais une représentation sous la forme de graphes d'opérateurs. La notion de graphe d'opérateurs individualisés pour représenter les programmes de traitement d'images est maintenant communément admise dans la communauté du domaine du traitement d'images (Draper-1999). Il est seulement nécessaire d'ajouter dans la représentation des graphes tous les éléments d'un langage évolué : variables et structures de contrôle. Pour compléter le corpus, nous ajoutons une représentation de la formulation du problème faite par l'utilisateur ainsi qu'une description du plan de traitement élaboré par l'expert de traitement d'images. En effet, les travaux des années 90 ont mis en évidence l'intérêt d'une modélisation des applications sous la forme de plans de traitement pour modéliser l'expertise de conception (Thonnat, 1999).

L'atelier fournit donc les outils nécessaires pour constituer ces trois types de corpus au cours du développement d'applications réelles.

2.3 Vue structurelle de l'atelier

L'architecture de l'atelier est donnée figure 1. L'atelier présente macroscopiquement deux volets : l'un automatique et l'autre manuel. Le développement d'une application formulée par un utilisateur devrait pouvoir se faire indifféremment du côté expert ou du côté système à base de connaissances. Les deux volets partagent pour cela les mêmes représentations dont les ontologies sont notamment les garantes.

L'atelier se compose de l'interface de formulation d'application dédiée aux utilisateurs, d'un atelier de génie logiciel propre au traitement d'images, d'une interface de programmation visuelle qui permet de construire des programmes sous la forme de graphes d'opérateurs et

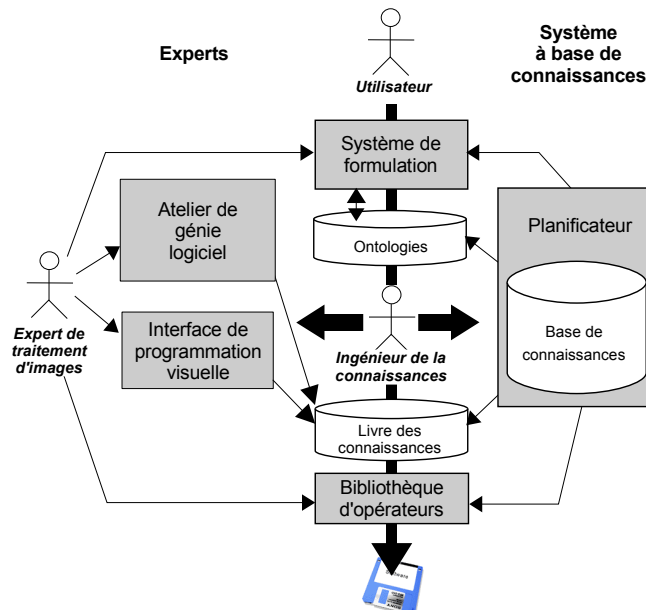


FIG. 1 – L'architecture globale de l'atelier.

d'une bibliothèque d'opérateurs exécutables.

Les corpus sont enregistrés sous la forme d'ontologies, de graphes d'opérateurs et d'un livre des connaissances. La formulation d'application faite par l'utilisateur est capitalisée sous la forme d'une ontologie d'application, le programme sous la forme d'un graphe d'opérateurs et le plan de traitement avec sa logique de conception dans un livre des connaissances (Ermine, 2000).

2.4 Vue fonctionnelle de l'atelier

A l'intérieur de l'atelier, le développement d'application est vu comme un processus itératif et incrémental. L'utilisateur et l'expert de traitement d'images collaborent pour développer une application par l'intermédiaire de l'interface de formulation de problème. L'utilisateur apporte sa connaissance du domaine et ses capacités d'évaluation des résultats produits si ceux-ci sont sous la forme d'images. L'expert apporte sa science du développement d'application. A partir d'une formulation partielle, l'expert ou le système à base de connaissances ou les deux développent une solution qui permet de créer des images de sorties sur lesquels l'utilisateur peut compléter, affiner ou remettre en cause sa formulation. L'important, c'est d'avoir un résultat très rapidement pour ne pas rompre le cycle de formulation. La programmation visuelle est ici un outil précieux : des prototypes exécutables sont créés immédiatement sans les contraintes de la programmation. Enfin l'ingénieur de la connaissance interagit avec l'expert du traitement d'images pour expliciter la logique de conception afin de la capitaliser dans le livre des connaissances.

Le bénéfique dans notre atelier, c'est qu'il n'augmente pas la charge cognitive de l'expert de traitement d'images. Au contraire, les outils permettent d'accroître la production par incitation à la réutilisation et de se recentrer sur la couche métier en se débarrassant des problèmes de programmation et de formalisation.

3 La formulation d'application

De cet atelier nous avons choisi de présenter plus spécifiquement la formulation d'application. La formulation consiste à décrire le problème de traitement d'images à résoudre.

Le rôle du système de formulation, c'est de faire franchir à l'utilisateur les 2 fossés sensoriel et sémantique (Smeulders, 2000) pour passer de sa perception du monde réel à une description à partir de caractéristiques visuelles (figure 2). Le fossé sensoriel résulte de l'écart existant entre le phénomène observé et sa représentation sous forme d'images. La perte d'information est due à l'acquisition d'image. Le fossé sémantique correspond à l'écart entre la représentation sous forme d'image et sa description sous forme d'un ensemble de caractéristiques qui peuvent être symboliques ou numériques. La perte d'information résulte de la représentation linguistique de la dénotation de l'image. Le système de formulation doit donc aider l'utilisateur à produire une « bonne » formulation, c'est à dire une formulation qui minimise la perte d'information pertinente et accroît l'abstraction des informations pour qu'elles décrivent au mieux la scène avec le moins d'information possible.

Parce que la formulation d'application de traitement n'a été que peu étudiée en tant que telle, nous avons choisi de construire notre propre modèle conceptuel de la formulation malgré le coût élevé que cela représente. Le modèle conceptuel de la formulation d'application développé par Arnaud Renouf (Renouf, 2007) distingue la spécification des objectifs de la définition de la classe d'images. Les figures 3 et 4 présentent successivement la représentation du modèle conceptuel de la formulation des objectifs et de la classe d'images.

Un atelier d'ingénierie des connaissances en traitement d'images

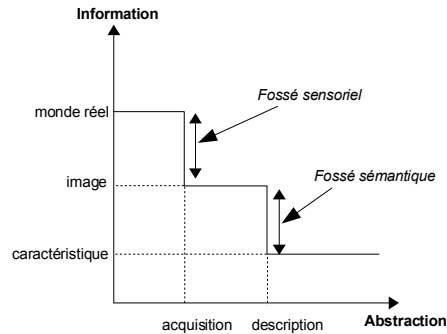


FIG. 2 – Le fossé sensoriel sépare le monde réel de l'image, obtenue par acquisition et le fossé sémantique sépare l'image de sa description linguistique.

Basée sur une analyse sémiotique¹, la classe d'image est décrite sur trois niveaux : le niveau physique porte sur la description des effets de la chaîne de production sur les images, le niveau perceptif décrit le contenu d'image par les caractéristiques des primitives visuelles qui le composent et le niveau sémantique s'intéresse à la description des objets présents dans l'image. Le niveau perceptif est utilisé à défaut ou en complément du niveau sémantique notamment quand les objets sont imprévisibles.

Basé sur une analyse systémique, l'objectif se formule en terme intentionnel sous la forme {tâche, contraintes}. La tâche s'écrit comme une action pouvant porter sur un objet du domaine, une primitive visuelle ou une propriété des images. Les contraintes spécifient les critères à optimiser et les niveaux de détail à atteindre. Chaque contrainte peut être précisée par une erreur acceptable qui permet ainsi de résoudre les compromis en cas de doute sur le respect de la contrainte.

Ce modèle est formalisé dans une ontologie du domaine. Le système de formulation utilise cette ontologie du domaine pour construire avec l'utilisateur une ontologie d'application.

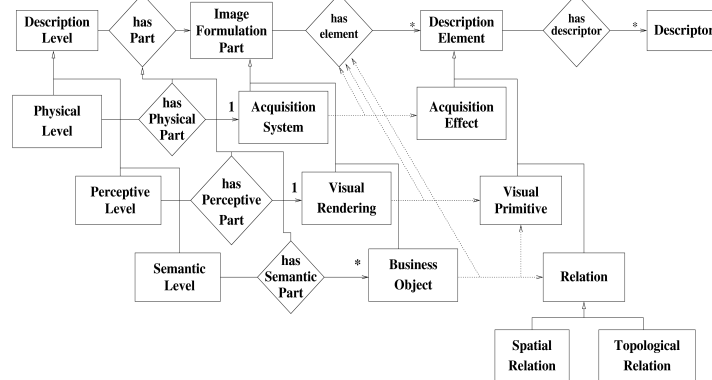


FIG. 3 – Le modèle conceptuel de la définition de la classe d'images décrit les éléments qui composent la description d'une classe d'images et les relations qu'ils entretiennent.

¹ La sémiotique est l'étude des signes et de leur signification.

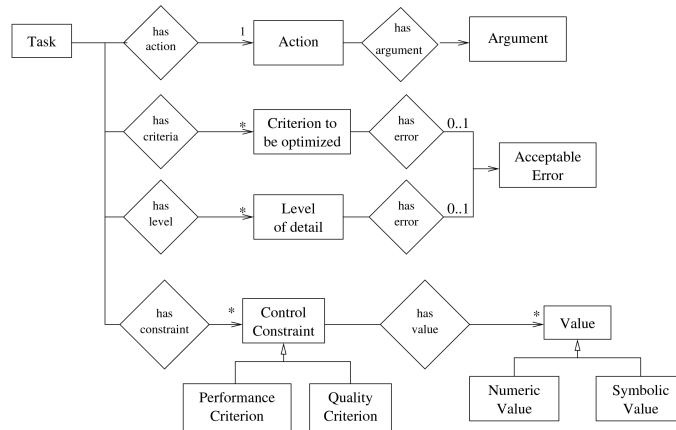


FIG. 4 – Le modèle conceptuel de la spécification des objectifs de traitement.

Pour tenter d'objectiver au maximum la formulation de l'utilisateur, le système de formulation met en œuvre un cycle de formulation. Pour illustrer ce cycle nous prenons un exemple d'une application en géographie. Le contexte de l'application est l'étude diachronique de l'utilisation des sols en milieu rural dont une image est présentée figure 5.

Étape 1. Le cycle est initié par la spécification des post-traitements. C'est un point d'entrée assez naturel. Les post-traitement sont décrits par la tâche à opérer et les mesures qui seront pratiquées sur les images de sortie du traitement d'images. Ces post-traitements ne font pas partie à proprement parlé de la formulation de traitement d'images mais ils constituent une source d'information permettant d'aider à la formulation (cf. étape 4).

Exemple : Le post-traitement concerne la comparaison d'images et les mesures pratiquées sont de type radiométrique, calcul de surface et localisation.



FIG. 5 – Les images proviennent de la base de données Google Map et visualise le paysage bocagier de la région de Bayeux (Calvados).

Étape 2. L'utilisateur renseigne ensuite les effets de la chaîne de production de l'image. Il est bien question des effets et non de la description des éléments de la chaîne. En effet seule une description phénoménologique des effets de la chaîne est utilisable pour orienter les traitements. De plus, ces informations sont relatives au domaine du traitement d'images et indépendante de l'évolution des capteurs.

Exemple : Les images sont altérées par l'effet de blocs du à la compression JPEG.

Étape 3. La définition du domaine est basée sur la construction de l'arbre des objets du domaine. Ce choix repose sur l'hypothèse que les spécialistes d'un domaine partagent une même taxonomie des objets de leur domaine. L'arbre fige les relations de composition entre objets (relation « partie-de »).

Exemple : la figure 6 donne l'arbre des objets présents dans les images.

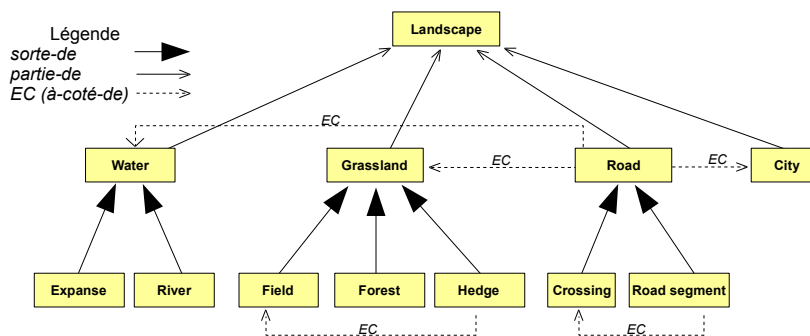


FIG. 6 – L'arbre des objets géographiques.

Étape 4. Il est temps ensuite de spécifier les tâches de traitement à opérer et les contraintes à y associer. Une liste des tâches et de contraintes est fournie par l'ontologie et reprend de façon plus fine les six catégories pointées par notre définition du domaine du traitement d'images. Les post-traitements donnés plus tôt peuvent être utilisés ici pour aider à la spécification des contraintes. Par exemple, si des mesures de radiométrie sont à opérer lors des post-traitements alors une contrainte sur un objet de segmentation devrait être de localiser avec précision la frontière des régions et en cas de doute préférer mettre la frontière à l'intérieur plutôt qu'à l'extérieur pour éviter des erreurs de mesures en prenant des pixels n'appartenant pas à la région.

Exemple : tâche : détection les villes.

Critère à optimiser : maximiser la détection (préférer des fausses alarmes aux oublis)

Niveau de détail : inclure tous les pixels dans la région (préférer plus de pixels que moins)

Étape 5. Sur la base de l'arbre des objets, l'utilisateur doit à présent décrire chacun des objets identifiés par la caractérisation de la primitive visuelle par laquelle il se manifeste (e.g., région, contour, point d'intérêt). Il est fait ici usage du paradigme différentiel qui veut que la description des objets se fasse par différence avec les autres objets. Si deux objets existent dans l'arbre c'est qu'ils présentent au moins une caractéristique différente. L'essentiel de la description d'un objet est fait par les caractéristiques qui le discriminent des autres. La liste des primitives et de leurs caractéristiques est fournie par l'ontologie.

Exemple : City : region
 texture : contrast
 topology : number of hole : 0

Forest : region
 texture : contrast
 color : hue = green

Étape 6. A défaut d'information disponible au niveau sémantique, par exemple dans le cas d'environnements imprédictibles, il est nécessaire de procéder à une description perceptive du contenu des images, c'est à dire une description « syntaxique » à partir de la caractérisation des primitives visuelles déconnectée de leur structuration objet. Par exemple, en robotique les images se décrivent par les caractéristiques des contours en espérant qu'ils correspondent à des bords d'objets réels.

Exemple : Les objets sont tous connus a priori et donc pas de description au niveau perceptif.

Étape 7. Enfin, l'utilisateur est amené à préciser les contraintes sur le système telles que la qualité du résultat, fiabilité *versus* robustesse et les limites de temps et de ressources.

Exemple : robustesse. Ce qui est important ici c'est la reproductibilité des résultats puisqu'il s'agit d'une comparaison d'images quitte à avoir de moins bons résultats sur chacune des images prise séparément.

4 Conclusion

L'atelier décrit a été défini au début des années 2000. Néanmoins, son utilisation réelle nécessite au préalable la réalisation effective des outils et l'élaboration des modèles de capitalisation des connaissances propres au traitement d'images. Aujourd'hui, la plupart des outils sont maintenant disponibles, que ce soit le système de formulation, l'atelier de génie logiciel, l'environnement de programmation visuelle et la bibliothèque d'opérateurs. Les modèles de capitalisation des connaissances sont en cours d'élaboration. Ce travail préalable extrêmement chronophage est combiné avec l'élaboration des modèles de capitalisation des connaissances. Tout ceci justifie la durée du projet.

Nos efforts de recherche actuels portent sur l'étude du système à base de connaissances qui doit opérationnaliser les modèles de connaissances que nous avons élaborés. Outre ses capacités de raisonnement pour développer de façon autonome des applications, le système doit disposer de capacité d'apprentissage pour lui permettre d'accroître ses capacités de résolution de problèmes. Dans l'idéal, nous souhaitons ne pas mettre de connaissances de traitement d'images dans le système mais des méta-connaissances lui permettant de créer lui-même sa connaissances. Pour cela, le système est basé sur une approche par raisonnement à partir de cas dont nous avons déjà fait l'expérience en traitement d'images (Ficet-Cauchard, 2000).

5 Références

- Bachimont B. (1996). Herméneutique matérielle et Artéfacture : des machines qui pensent aux machines qui donnent à penser. Thèse de doctorat, École Polytechnique, Paris, 1996.
- Clouard R., A. Elmoataz, C. Porquet, M. Revenu (1999). Borg: a knowledge-based system for automatic generation of image processing programs. *IEEE PAMI*, 21(2):128-144.
- Clouard R. (2004). Une méthode de développement d'applications de traitement d'images, *Traitement du signal*, 21(4):277-293.
- Crevier D., R. Lepage (1997). Knowledge-based image understanding systems: a survey,

- Computer Vision and Image Understanding*, 67(2):161-185.
- Charlet, J. *L'ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales*. Mémoire d'habilitation à diriger les recherches, Université Pierre et Marie Curie, Paris, 2002.
- Draper, B., A. Hanson, A. E. Riseman (1996), Knowledge-Directed Vision : Control, Learning, and Integration, *Proc. of IEEE*, 84(11):1625-1681.
- Ficet-Cauchard V. , M. Revenu , C. Porquet , R. Clouard (2000), Aide à la conception d'applications de traitement d'images : une approche fondée sur le raisonnement à partir de cas. *Ingénierie des connaissances: évolutions récentes et nouveaux défis*, Eyrolles, 93-108.
- Ermine, J-L., *Les systèmes de connaissances* (2e édition), Hermès, Paris, 2000.
- Gaines B.R., M.L.G. Dhaw, J.B. Woodward (1993). Modelling as framework for knowledge acquisition methodologies and tools. *Int. journal of Intelligent Systems*, 8(2):155-168.
- Garbay, C. (2002). Pour une conception distribuée des systèmes de vision, Revue L'Objet, Numéro Spécial « Systèmes distribués et connaissances », 8(4).
- Hanson, A. and E. Riseman (1978). VISIONS: A computer System for Interpreting Scenes. *Computer Vision Systems*, New York: Academic Press, 303-334
- Labiche, J., E. Trupin, Y. Saidali, D. Dionisi, M, Holzem, JL. Boudon (2007). De l'analyse à l'interprétation d'images de documents. *7ième journées francophones Extraction et Gestion des Connaissances, Actes de l'atelier ECOI*, 13-24.
- Matsuyama, T. et V. Hwang (1990). *SIGMA: A Knowledge-Based Aerial Image Understanding System*. Perseus Publishing.
- Linster M. (1993). Closing the gap between modeling to make sense and modeling to implement systems. *International Journal of Intelligent Systems*, 8(2):209-230.
- Renouf A. (2007). *Modélisation de la formulation d'applications de traitement d'images*, Thèse de Doctorat, Université de Caen.
- Smeulders, A. W. M., M. Worring, S. Santini, A. Gupta, R. Jain(2000), Content-Based Image Retrieval at the End of the Early Years, *IEEE PAMI*, 22:1349-1380.
- Saidali, Y., E. Trupin, M. Holzem, N. Baudoin (2007). Pour une aide à l'interprétation de connaissances traiteur d'images: une approche terminologique. *7ième journées francophones Extraction et Gestion des Connaissances, Actes de l'atelier ECOI*, 25-37.
- Thonnat M., S. Moisan et M. Crubézy, (1999). Experience in integrating image processing programs. *Proceedings of the 1st Int. Conference on Vision Systems*, 200-215.

6 Summary

Our goal long-term research is to develop a knowledge-based system that can automatically generate image processing software from a specification made by a user. This objective has already been widely studied and met with a setback. However, after analyzing the causes of this failure, we propose to take this research from a new perspective: the constructivist approach. For that we need to build an image processing engineering knowledge workbench that is the place of the study and development of such a software. The workbench is a set of software tools designed to encourage the collaboration between the expert image processing and engineering knowledge in order to define an application development theory from the point of view of the system.

Vers la construction d'une ontologie appliquée à l'imagerie satellitaire

Marine Campedel, Marie Lienou, Ivan Kyrgyzov, Henri Maître

* GET-Télécom Paris - TSI - 46 rue Barrault, 75013 Paris, France,
CNES-DLR-ENST Competence Center
marine.campedel@enst.fr

Résumé. Le nombre et la résolution des images satellitaires ne cessent d'augmenter. Il en résulte un accroissement phénoménal de la quantité d'information disponible mais non exploitable en totalité par des photo-interprètes. Il est donc nécessaire de développer des processus (semi-)automatiques facilitant l'indexation et la récupération du contenu de ces images. Cette indexation repose sur des processus d'analyse automatiques des images et des connaissances a priori sur leur domaine d'usage. Nous présentons nos résultats concernant les connaissances accessibles au travers de processus d'analyse automatique en termes de concepts et relations entre concepts ; nous proposons également quelques réflexions sur la façon d'organiser leur stockage et exploitation, en tenant compte du domaine d'application de ces images.

1 Introduction

L'analyse et l'interprétation des images satellitaires, généralement effectuées par des photo-interprètes, requièrent un effort humain considérable et coûteux. Ces experts ont appris à exploiter des résultats d'analyse des images (appariement, segmentation, classification, plus ou moins automatiques ...) afin d'extraire une information utile dans un but applicatif bien défini : par exemple, l'urbaniste s'intéresse à la morphologie des villes (distribution des types d'habitation, disposition des réseaux de transport, espaces verts, ...) alors que l'agronome étudie l'évolution des cultures (état des cultures, suivi des tailles des parcelles exploitées, ...). Les images doivent être bien choisies pour permettre la mise en évidence des objets d'intérêt de l'application choisie. Des dizaines de satellites parcourent notre ciel et fournissent régulièrement des images ; celles-ci sont issues de capteurs optiques ou radars. Dans le cas de l'optique, l'interprétation visuelle est plus immédiate mais très sensible aux conditions de luminosité, contrairement à l'imagerie radar.

Dans le cadre du centre de compétence dit CoC ¹, nous nous intéressons à l'étude du contenu informationnel des images satellitaires optiques (en France) et radar (en Allemagne) ainsi qu'à son exploitation de façon (semi-)automatique. En effet, les agences spatiales stockent de plus en plus d'images totalement inexploitées. En outre des satellites à très haute résolution

¹Competence Center on Information Extraction and Image Understanding : issu d'un accord tripartite entre les agences spatiales française (CNES) et allemande (DLR) et Télécom Paris. <http://www.coc.enst.fr>

ont été récemment lancés (TerraSar²) ou le seront prochainement (Pléiades³). Dans ce dernier programme, on prévoit jusqu'à 450 images journalières par satellite, à 70 cm de résolution : il s'agit d'une masse de données considérable, intraitable dans son ensemble par des photo-interprètes. Il est donc urgent de mettre au point des algorithmes pour indexer automatiquement ces images, i.e. pour caractériser leur contenu informationnel et permettre sa récupération.

Indexer des images satellitaires présente quelques particularités. Selon la résolution, le contenu informationnel est différent : spectral, textural au-dessus de 1m de résolution par pixel (zone urbaine, champs, forêt, ...) et plutôt géométrique (bâtiments, rond-points, parking ...) en dessous. En outre les objets d'intérêts sont plus ou moins étendus : un rond-point et un bâtiment sont localement définis alors qu'un réseau routier s'étend sur une grosse portion de l'image. Enfin une image seule est déjà très volumineuse (une image SPOT5 à 5m contient 12000×12000 pixels). Ainsi, il est nécessaire d'organiser l'ensemble des connaissances à disposition, qu'elles concernent les algorithmes de traitement d'images ou les applications d'usage des images satellitaires.

Ainsi cet article présente tout d'abord une analyse bibliographique et expérimentale de ce qui peut être obtenu à partir de l'image, vue comme un signal. Nous verrons notamment (section 2.2) comment, à partir de plusieurs "points de vue" non supervisés sur le contenu d'une (ou un ensemble d') image(s), nous pouvons mettre en évidence des concepts sémantiquement pertinents et des relations entre eux. Il s'agit d'un premier pas vers une ontologie des représentations visuelles telles que produites par des algorithmes automatiques, i.e. de la mise en évidence de concepts saillants et de relations entre ces concepts. Dans la section 3, nous nous intéressons à l'usage des images satellitaires, ce qui nous amène à considérer des moyens de stockage et d'organisation des connaissances disponibles. Nous présentons ainsi notre projet d'élaboration d'un système de fouille de données multimédia appelé PLATO, ainsi que des réflexions plus prospectives sur l'usage d'ontologies de domaine pour enrichir l'indexation des images. Enfin nous concluons et présentons des perspectives.

2 Analyse de l'image et conceptualisation

2.1 Etat de l'art

Dans cette section nous essayons de voir comment une analyse automatique de l'image peut nous donner accès à une forme de connaissance. Par connaissance nous désignons une information structurée et exploitée. Concrètement, nous limitons nos investigations actuelles à la production de concepts et à la caractérisation des relations entre ces concepts. Lorsque ces concepts peuvent être rattachés à un mot, nous les qualifions de *sémantiques*.

Concepts Dans la littérature, loin des problématiques ontologiques, les experts en traitement d'images tentent de produire un codage du contenu sémantique de ces images. La méthode la plus courante de codage est celle utilisée dans Barnard et al. (2003), Duygulu et al. (2002) et Jeon et al. (2003) : elle consiste à segmenter les images en régions (visual tokens), et à extraire pour chaque région, un vecteur de caractéristiques (spectrales, texturelles, formes, etc.) qui la représente. Ces représentations des régions sont ensuite quantifiées par un algorithme

²<http://www.infoterra.de/tsx/index.php>

³<http://www.cnes.fr/web/3227-pleiades.php>

de clustering (en général les k-means) ; les groupes de régions (clusters) obtenus sont appelés *blobs*. Ainsi, chaque blob a une étiquette qui permet de l'identifier ; l'ensemble des étiquettes produites est assimilé à un dictionnaire visuel pouvant être utilisé pour décrire le contenu de l'image. Une variante de la méthode de codage sus-présentée est de découper l'image en images en utilisant une grille régulière, plutôt que de la segmenter. Ainsi nous voyons émerger des concepts possédant une description numérique et un identifiant (l'étiquette).

Les relations Caractériser ensuite les relations entre blobs (*mots visuels*) n'est pas simple. Le plus souvent, cela se fait au niveau sémantique, i.e. après association à des mots. Par exemple dans Barnard et al. (2003) et Glotin et Tollari (2005), les auteurs utilisent cette caractérisation de l'image pour annoter les images naturelles. L'image codée et le texte de l'annotation sont groupés pour produire une représentation caractérisée par une distribution jointe reliant les images et les mots. Duygulu *et al.* Duygulu et al. (2002) utilisent un *modèle de traduction* pour associer un mot à une région individuelle de l'image. Ce modèle se comporte comme un *lexique* qui, sachant les mots dans une langue, les prédirait dans une autre langue.

Il est cependant intéressant de tenir compte des dépendances entre les mots visuels. L'idée est alors d'exploiter les relations syntaxiques entre les mots visuels pour déterminer la sémantique de l'image. Nous constatons en effet que les relations spatiales constituent un élément essentiel des descriptions d'agencement entre les régions d'une scène et sont donc utiles à un grand nombre de tâches liées à la reconnaissance des formes, la vision par ordinateur, les systèmes d'information géographique (SIG), et plus particulièrement l'interprétation des scènes. Freeman (1975) a étudié l'ensemble de ces relations pour la langue anglaise et a proposé une liste de treize relations spatiales de base pour décrire la position relative de deux objets dans un espace bidimensionnel. Dans la littérature, ces relations sont généralement classées en trois familles : relations topologiques, relations métriques (ou distances) et directions. On rajoute souvent à ces trois familles les relations morphologiques ou encore les relations de symétrie. Pour représenter ces relations spatiales, deux types de modèles sont utilisés : les modèles de représentation qualitative souvent liés à la logique formelle ou à la classification des intervalles temporels proposée par Allen (1983), et les modèles de représentation quantitative plutôt rattachés à la logique floue ou aux probabilités. Bloch (2005) propose une présentation détaillée de ces approches floues. Plus encore, Hudelot et al. (2006) décrit une ontologie des relations spatiales, exploitée en collaboration avec un formalisme flou. Ces approches n'ont pas encore été testées dans le cadre satellitaire mais une thèse va commencer dès janvier 2008, sous la direction d'Isabelle Bloch, à Télécom Paris, avec cet objectif.

2.2 Exploitation d'un consensus

L'idée de cette section est de démontrer qu'il est possible d'identifier des concepts sémantiquement pertinents, à partir de l'image et des attributs extraits. Le principe repose sur la production d'un consensus à partir de différents points-de-vue d'analyse de l'image. N'est-ce pas exactement notre façon humaine de procéder ? nous avons chacun une perception du monde différente et pourtant, par le biais d'une communication, nous confrontons nos vues et définissons les concepts, éléments stables, qui régissent notre monde. En pratique, ces *points-de-vue* sont produits par des classificateurs non supervisés (des algorithmes de clusterisation).

L' idée du consensus est d'identifier l'information commune portée par les différents résultats de classification pour ensuite en analyser la redondance et la complémentarité.

Combiner des classificateurs n'est pas nouveau, en particulier dans la communauté de la classification supervisée. Les objectifs sont généralement d'améliorer des performances de classification, d'identifier des résultats communs à plusieurs algorithmes, de classifier des données partiellement étiquetées ou des données distribuées, ... Cependant, lorsqu'il s'agit de classification non supervisée le problème n'est pas simple (pas de classes prédéfinies ni mesures d'erreurs bien établies) et la littérature est bien moins fournie. Nous en présentons une ci-dessous et démontrons son efficacité sur une image PELICAN très haute résolution de Toulouse.

2.2.1 Algorithme LSEC

Nous proposons ainsi d'exploiter une méthode de consensus, reposant sur une représentation par matrice de co-association des clusterisations individuelles. Ces matrices illustrent le fait que deux individus (parmi les N considérés) sont classifiés ou non dans le même groupe (cluster) ; elles ne dépendent pas d'un processus de numérotation des clusters et ne reflètent pas non plus les espaces de caractéristiques dans lequel sont représentées initialement les données. Différents critères objectifs de combinaison ont déjà été proposés, tels Fred et Jain (2005) et A. Topchy et Punch (2004) ; nous en proposons un, bien formalisé mathématiquement, pour lequel nous proposons une solution efficace s'affranchissant de tout paramètre d'initialisation. En effet, à partir de la matrice de co-association moyenne A , le problème est de trouver la matrice de partition consensuelle B^s (binaire) telle que $B^s B^{s'4}$ est la plus proche de A au sens des moindres carrés. L'exploration de l'ensemble des solutions n'est pas envisageable, du fait de la quantité de données qui nous intéresse. Nous proposons donc une heuristique, reposant sur l'algorithme agglomératif appelé "single-link". L'algorithme, dénommé LSEC (Least Square Error Combination), procède itérativement en réduisant progressivement l'erreur E ; il est présenté ci-après :

LSEC

Etape 1 Initialisation de B^s (avec la matrice identité), $i \leftarrow 1$ et calcul de l'erreur $E^{(i)} =$

$$\sum_{u=1}^N \sum_{v=1}^N \left(\sum_{r=1}^N (B_{ur}^s B_{rv}^{s'}) - A_{uv} \right)^2$$

Etape 2 Trouver les données appariées avec une probabilité maximale $(r, t) = \max\{A_{uv} : u, v = 1, \dots, N, u \neq v\}$

Etape 3 Si $A_{rt} = 0$, B^s est la partition recherchée, stop.

Etape 4 Mise à jour de A avec $A_{rt} \leftarrow 0$, $B^h \leftarrow B^s$.

Fusionner les deux clusters C_r et C_t en sommant les colonnes respectives de $B^h : B_{kr}^h \leftarrow (B_{kr}^h + B_{kt}^h)$ et supprimer la colonne B_{kt}^h (mise à 0), avec $k = 1, \dots, N$.

Mise à jour de l'erreur : $E^{(i+1)} \leftarrow \sum_{u=1}^N \sum_{v=1}^N \left(\sum_{r=1}^N (B_{ur}^h B_{rv}^h) - A_{uv} \right)^2$

Si $E^{(i+1)} \leq E^{(i)}$, alors $i \leftarrow i + 1$, $B^s \leftarrow B^h$, $A \leftarrow A * (\mathbf{I} - B^s B^{s'})$.

Aller en **Etape 2**.

^{4'} désigne l'opération de transposition.

'.*' représente la multiplication de matrice point par point.

Le nombre optimal de clusters est obtenu implicitement comme étant le nombre de colonnes non nulles de B^s . En pratique, il est possible de s'affranchir du stockage de la matrice A et d'initialiser efficacement B^s à l'aide de graphes de voisinage, afin de faciliter la convergence de l'algorithme. La présentation détaillée de ce nouvel algorithme est effectuée dans Kyrgyzov et al. (2007b).

2.2.2 Fouiller les données

A partir de la matrice de co-association A et du résultat du consensus B^s , différentes informations peuvent être obtenues. En particulier, nous appelons *stabilité* d'une donnée (région élémentaire dans notre cas particulier) la quantité $S_{d \in C} = 1/\#C \sum_{n \in C}^D A_{dn}$; ce critère illustre la force de la connection entre une donnée et son groupe. Nous pouvons également dériver un critère de stabilité pour un cluster C_k , à partir des probabilité de connections des éléments qui le composent (fournies par la matrice moyenne de coassociation A) :

$$S_{intra}^k = \frac{1}{\#C_k(\#C_k - 1)} \sum_{(i,j) \in C_k} A_{ij}$$

ainsi, qu'un critère inter clusters :

$$S_{inter}^{k,l} = \frac{1}{\#paires(k,l)} \sum_{i \in C_k} \sum_{j \in C_l} A_{ij}$$

Ces deux critères ne sont finalement que la variance interne des données d'appariement à leur cluster (matrice B concaténant les matrices d'appariement des données à chaque cluster produit par les différents algorithmes) et l'intercorrélation entre deux clusters. A l'aide de ces trois critères nous pouvons caractériser deux types de relations principales :

- entre une région et les clusters de régions produits par le consensus : ceci nous permet d'identifier (à la façon des machines à vecteur de support) des données bien classées (i.e. majoritairement clusterisées ensemble par les algorithmes de clustering) que nous qualifions de *stables*, des données isolées (qui produisent des clusters unitaires) et des données frontalières, i.e. dont l'appartenance à un cluster n'est pas bien définie.
- entre clusters : la force des connections $S_{inter}^{k,l}$ caractérisent la relation existant entre clusters. Si cette valeur est faible (resp. forte), les clusters sont faiblement (resp. fortement) connectés (i.e. qu'ils contiennent des données qui ont été minoritairement (resp. majoritairement) associées lors d'une opération de clusterisation élémentaire).

Nous présentons ci-après les résultats obtenus sur une image Pélican de Toulouse (cf Figure 1, en haut à gauche)

2.2.3 Expérimentations

L'expérimentation consiste tout d'abord à extraire des caractéristiques de bas niveaux sur l'image. Nous avons choisi ici de nous limiter à des caractéristiques spectrales et texturales (QMF : calculées sur une fenêtre recouvrante de taille 32x32 avec un pas de 8 pixels, uniquement sur la bande rouge). L'étape suivante applique différents algorithmes de clustering sur les vecteurs d'attributs obtenus. Il s'agit des algorithmes suivants :

- KMeans
- Kernel KMeans (Shawe-Taylor et Cristianini, 2004)
- Spectral KMeans (Ng et al., 2001)
- Algorithme hiérarchique de Ward (Jain et Dubes, 1988)

Chacun de ces algorithmes nécessite le nombre de clusters comme paramètre d'entrée. Ce dernier est estimé à l'aide d'une procédure dérivée du MDL (Minimum Description Length), issue de Kyrgyzov et al. (2007a) ; on dispose donc d'une réalisation de chaque algorithme de clustering, avec une valeur optimale (pour cet algorithme) de clusters. Les quatre résultats (cf Figure 1) vont être utilisés pour calculer la matrice de co-association moyenne puis produire la clusterisation consensuelle (cf Figure 1).

Mais les informations produites par le clustering consensuel permettent également de fouiller les images. La figure 2 nous permet de visualiser la force (mesurée à l'aide du score $S_{inter}^{k,l}$) des relations entre les clusters produits. L'ensemble des scores est présenté dans la table 1. Nous remarquons ainsi, que le cluster 5 est le plus stable (S_{intra}^5 très proche de 1) ; suivant le code de couleur présenté sur la figure 1, il nous semble pouvoir être étiqueté "grands bâtiments". Nous pourrions dire, qu'au sens du score S_{intra} , ce cluster est saillant. Nous remarquons ensuite qu'il existe une relation privilégiée entre les clusters 2 et 4 : effectivement, lorsque l'on regarde l'image, ces deux clusters semblent correspondre à de la végétation, avec une différence de densité. Ainsi, de façon complètement non supervisée, en s'appuyant sur plusieurs points de vue, nous parvenons à faire émerger des groupes de pixels correspondant à une sémantique et à définir des relations entre eux.

	1	2	3	4	5
1	0.59	0.26	0.25	0.12	0.16
2	0.26	0.57	0.12	0.30	0.26
3	0.25	0.13	0.75	0.20	0.25
4	0.12	0.30	0.20	0.89	0.19
5	0.16	0.26	0.25	0.19	0.99

TAB. 1 – Scores des clusters. les scores S_{intra}^k sont sur la diagonale et les $S_{inter}^{k,l}$ dans les autres positions.

2.3 Conclusion

Nous avons présenté une méthode de classification non supervisée par combinaison de résultats de clusterings. L'idée sous-jacente est de faire émerger des concepts sémantiquement valables à l'aide de points de vue non sémantiques issus d'algorithmes différents. Nous avons démontré que cette approche fonctionne, sur une image Pélican de Toulouse à 1m de résolution. Dans d'autres contextes, nous l'avons également appliquée à d'autres images (SPOT5 notamment, pour caractériser des paysages urbains, Kyrgyzov et al. (2007b)) ainsi qu'à de la sélection d'attributs Campedel et al. (2007). Dans tous les cas, les clusters produits étaient beaucoup plus proches d'une interprétation sémantique que les résultats des clusterisations individuelles ; d'autre part, les relations identifiées entre clusters étaient elles-aussi très informatives et sémantiquement pertinentes. Bien entendu, la qualité du résultat dépend des attributs

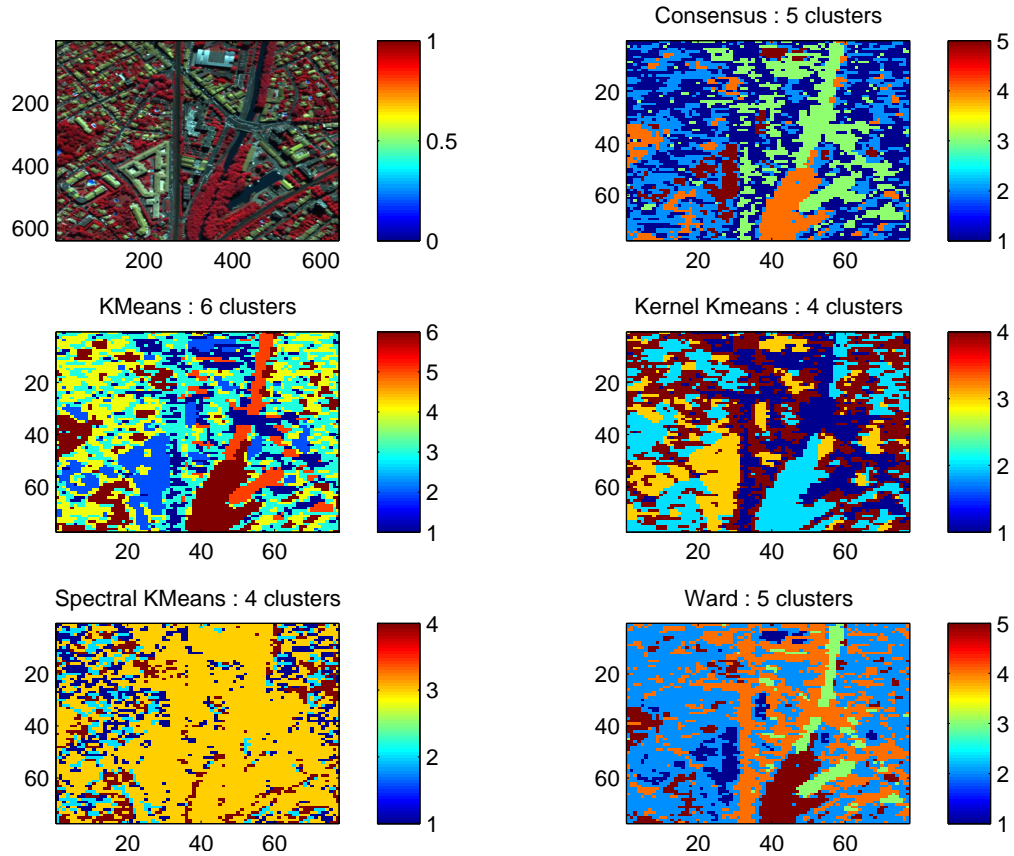


FIG. 1 – Image *Pélican de Toulouse*, à 1m de résolution (*fausses couleurs*), et résultats obtenus par les algorithmes de clusterisations consensuel et individuels. Les couleurs ne sont pas comparables : c’est d’ailleurs un avantage de la méthode de consensus, basée sur les matrices de co-association, de ne pas faire les mises en correspondances des clusters.

extraits de l’image. Dans le cadre de l’imagerie à très haute résolution (inférieure au mètre), les *bonnes caractéristiques* ne sont pas vraiment connues ; elles reposent a priori plus sur des descripteurs géométriques et des relations spatiales structurantes. Nous n’avons pas encore de résultats probants à ces résolutions.

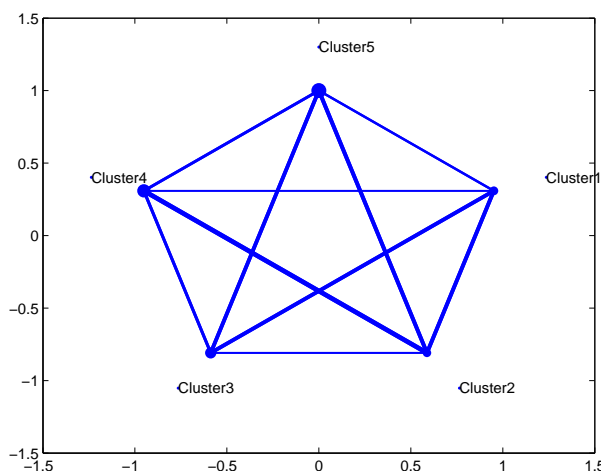


FIG. 2 – Graphe des relations entre les clusters produits par la méthode consensuelle. Plus les traits sont épais et plus les scores sont élevés.

3 Vers une organisation ontologique

3.1 PLATO

Partant de besoins urgents et concrets, Télécom Paris se dote actuellement d'une plateforme de stockage et fouille de données dénommée PLATO (PLATe-forme Online). Celle-ci, un outil de travail pour les chercheurs, tend à : i) gérer une quantité grandissante et hétérogène de données (sons, images, vidéos, textes), ii) faciliter la mise en place d'expérimentations sur des bases de données de taille importante, iii) aider la production et la diffusion des résultats de la recherche. Ainsi, Télécom Paris se trouve-t-elle au cœur des problématiques soulevées dans les projets auxquels elle participe activement tels Infom@gic du pôle CapDigital⁵, le réseau d'excellence K-Space⁶ et diverses ANR.

PLATO repose sur des outils technologiques de pointe, open-source et représentés par une communauté informatique active. Elle est composée d'un système de gestion de base de données (SGBD) PostgreSQL 7.4.7 (Douglas, 2005), d'un serveur web Apache 2.0, d'une interface web développée avec Python (Lutz, 2006), des machines de calcul. L'ensemble est connecté via le réseau local. L'utilisateur interagit au moyen d'une interface web utilisant les technologies : Python et Ajax (Asleson et Schutta, 2006). La technologie Ajax est utilisée pour exploiter les avantages des connexions client-serveur asynchrones. Le SGBD gère une base de données

⁵Infom@gic : <http://www.capdigital.com/xwiki/bin/view/Projet/Infomagic>

⁶K-Space : Knowledge Space of Semantic Inference for Automatic Annotation and Retrieval of Multimedia Content. Web : <http://www.kspace-noe.net/>

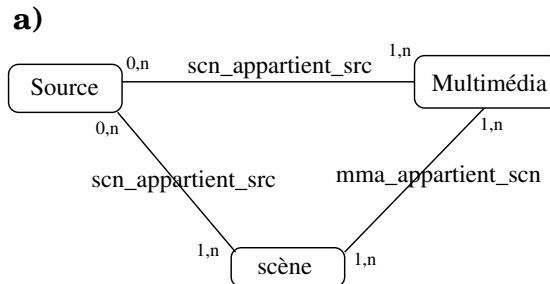


FIG. 3 – Relation entre les entités scène, source et multimédia.

relationnelle qui constitue le cœur de PLATO. Etant donnée leurs complexités, ni le modèle conceptuel de données (MCD), ni le modèle physique de données (MPD) ne peuvent être présentés dans leur globalité. La figure 3.1 présente trois entités majeures : i) l'objet multimedia (représente à la fois les médias de base et les objets multimédias qui sont des combinaisons d'au moins deux médias), ii) la source (modélise toutes les meta-données, i.e. des caractéristiques globales attachées aux documents, en particulier les informations sur la provenance des données et leurs conditions d'obtentions) et enfin iii) la scène (le contenu sémantique, notamment indépendant des conditions de prise de vue ou d'encodage de l'image). En imagerie satellitaire, cela revient à conserver des informations sur la provenance de l'image (type de capteur, satellite), sur la zone pointée et l'image elle-même (avec les traitements qui lui sont associés).

L'organisation à l'aide d'une base de données (BD) relationnelle nous permet un stockage efficace des informations à disposition ou extraites par des traitements. Qu'est-ce qui différencie une BD relationnelle d'une ontologie ? une ontologie pouvant elle-même être formalisée par une base de données, nous pensons que la façon de les employer les rend différentes. En effet, une base de données n'est manipulée qu'à travers des opérations simples de gestion (insertion/suppression/modification) alors qu'une ontologie est exploitée à l'aide d'un raisonneur qui peut apporter une information non directement stockée dans la BD.

3.2 Les ontologies du domaine satellitaire

Lorsque l'on considère la façon dont sont exploitées les images satellitaires, on observe un effort sémantique et ontologique important, en particulier cartographique. Mais les cartes ne sont qu'un type de représentation du contenu d'une image, valable à un temps donné.

Le projet Corine Land Cover⁷ a pour objectif de produire un inventaire biophysique de l'occupation des sols européens. Des cartes de couverture des sols sont générées à partir d'images satellitaires (Landsat) qui sont interprétées visuellement par des experts, s'aidant de données exogènes (photographies aériennes, cartes topographiques, etc). Des centaines de milliers de km^2 ont été décrits et les cartes sont disponibles gratuitement. Afin de rendre consistantes ces cartes, les experts disposent d'une nomenclature précise. Celle-ci est composée de 45 concepts

⁷Projet européen <http://reports.eea.europa.eu/COR0-landcover/en>

organisés hiérarchiquement sur trois niveaux. Chaque concept est précisément décrit, ainsi que ses liens et oppositions avec d'autres concepts (par exemple, le concept "tissu urbain discontinu" est défini en lui-même et par opposition au "tissu urbain continu"). Peut-on parler d'ontologie pour cette nomenclature détaillée ?

Dans différents domaines d'application, on retrouve cette envie de décrire les connaissances par un nombre fini de concepts décrits précisément et organisés. Parmi les projets marquants, on trouve une ANR FoDoMUST⁸ et un projet de recherche européen et suisse appelé Townontology⁹. Ces deux projets concernent la mise au point d'ontologies urbaines. Dans le premier cas, l'ontologie est conditionnée par les caractéristiques spectrales des images satellitaires et doit nourrir un SIG (Système d'Information Géographique) avec des résultats sur la localisation et l'identification des éléments du tissu urbain (surfaces minéralisées, végétation, eau). Dans le second cas, la description ontologique repose sur les objets physiques qui composent la matérialité de la ville ainsi que les événements historiques déterminant la composante temporelle des transformations du territoire ; il n'est pas fait référence aux images satellitaires. Combien d'ontologies différentes faut-il donc définir pour arriver à une conceptualisation acceptable par le plus grand nombre de la notion de "urbain" ?

3.3 Le projet DAFOE

Le projet DAFOE¹⁰ est un projet ANR débuté en janvier 2007, sous le pilotage de Jean Charlet (INSERM). La plupart des outils développés autour des ontologies permettent de les construire en précisant comment représenter les concepts et formaliser leur sémantique, ils ne précisent pas comment trouver les concepts ni comment expliciter leur signification. L'objectif du projet est de proposer une méthode complète associée à une plateforme technique pour concevoir des ontologies, de la modélisation à partir du domaine à leur évolution en passant par leur formalisation et exploitation. S'appuyant sur les acquis de travaux antérieurs, à la fois issus des partenaires et de la littérature du domaine, le projet a pour but de prendre en charge la modélisation sémantique des concepts ontologiques pour i) motiver et justifier les représentations formelles qui seront utilisées et ii) en faciliter la révision.

Nous participons à ce projet en tant qu'utilisateurs désireux de créer une ontologie avec pour objectif de faciliter/améliorer notre processus d'indexation des images satellitaires. Les résultats produits serviront à valider la plate-forme développée dans le cadre du projet DAFOE. Plus précisément, nous attendons de ce projet qu'il nous aide à i) enrichir nos annotations sémantiques, ii) raisonner à partir de concepts sémantiques et non sémantiques (issus d'une analyse de l'image) et iii) interagir avec l'utilisateur pour apprendre et affiner ses propres centres d'intérêt. L'enrichissement d'annotations sémantiques est une application classique des ontologies. Nous souhaitons dépasser cet aspect et exploiter au maximum les connaissances que nous pouvons acquérir à partir des images (même si elles ne peuvent être associées directement à une étiquette sémantique) et des usages faits de ces images. Il semble ainsi que nous ayons besoin de deux ontologies différentes qu'il faudra relier progressivement (des techniques d'apprentissage automatique sont envisagées). Il ne s'agit pour l'instant que d'un projet, encadré par des experts académiques et industriels, qui devrait donner ses premiers fruits fin 2008.

⁸<http://lsiit.u-strasbg.fr/afd/sites/fodomust/fr-accueil.php>

⁹<http://www.townontology.net>

¹⁰<http://www.dafoe4app.fr>

4 Conclusion

Nous avons proposé dans cet article notre vision (interne) actuelle des ontologies pour l'imagerie satellitaire. Nous sommes des traiteurs de signal et en tant que tels nous partons de l'image pour identifier les concepts et relations entre concepts qui émergent du signal. En particulier, nous avons montré qu'une méthodologie reposant sur la production d'un consensus de clusterisations (donc de façon non supervisée) met en évidence des concepts et relations sémantiquement pertinentes. D'autre part, nous participons à des projets nationaux et européens auxquels contribuent des géographes, des géologues, des photo-interprètes, etc. qui nous proposent d'autres points de vue sur nos images, en particulier au niveau de leur richesse sémantique. Organiser l'ensemble de ces connaissances n'est pas simple et l'outil "ontologie" nous paraît bien adapté. Afin de progresser dans l'élaboration d'une ontologie adaptée à nos besoins nous participons activement au projet DAFOE. Ce projet nous permettra de définir formellement deux ontologies : l'une issue de l'analyse automatique des images et l'autre issue des besoins applicatifs des images satellitaires. Ces deux ontologies devront, à l'aide de techniques d'apprentissage, être mises en relation et faire ainsi le lien entre les experts du traitement automatique des images et les experts de leur interprétation visuelle.

Références

- A. Topchy, A. J. et W. Punch (2004). A mixture model for clustering ensembles. In *in Proc. SIAM Conf. on Data Mining*, pp. 379–390.
- Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM* 26(11), 832 – 843.
- Asleson, R. et N. Schutta (2006). *Foundations of Ajax*. Apress.
- Barnard, K., P. Duygulu, D. Forsyth, N. de Freitas, D. Blei, et M. Jordan (2003). Matching words and pictures. *Journal of Machine Learning Research* 3, 1107 – 1135.
- Bloch, I. (2005). Fuzzy Spatial Relationships for Image Processing and Interpretation : A Review. *Image and Vision Computing* 23(2), 89 – 110.
- Campedel, M., I. Kyrgyzov, et H. Maître (2007). Sélection non supervisée d'attributs - application à l'indexation d'images satellitaires. In *SFC'07*, Paris, pp. 43–46.
- Douglas, K. (2005). *PostgreSQL : the comprehensive guide to building, programming and administering PostgreSQL databases*. Sams.
- Duygulu, P., K. Barnard, J. de Freitas, et D. Forsyth (2002). Object recognition as machine translation : learning a lexicon for a fixed image vocabulary. *Seventh European Conference on Computer Vision IV*, 97 – 112.
- Fred, A. L. et A. K. Jain (2005). Combining multiple clusterings using evidence accumulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(6), 835–850.
- Freeman, J. (1975). The modelling of spatial relations. *Computer Graphics and Image Processing* 4(2), 156 – 171.
- Glotin, H. et S. Tollari (2005). Fast image auto-annotation with visual vector approximation clusters. In *Proc. of Fourth International Workshop on Content-Based Multimedia Indexing*

(*CBMI2005*).

Hudelot, C., J. Atif, et I. Bloch (2006). Ontologie de relations spatiales floues pour l'interprétation d'images. In *Rencontres francophones sur la Logique Floue et ses Applications, LFA 2006*, Toulouse, France, pp. 363–370.

Jain, A. et R. C. Dubes (1988). *Algorithms for Clustering Data*. Prentice-Hall, Englewood Cliffs, NJ.

Jeon, J., V. Lavrenko, et R. Manmatha (2003). Automatic image annotation and retrieval using cross-media relevance models. *Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, 119 – 126.

Kyrgyzov, I., H. Maître, et M. Campedel (2007b). A method of clustering combination applied to satellite image analysis. In *IEEE - International Conference on Image Analysis and Processing ICIAP 2007*, Modena, Italy, pp. 81–86.

Kyrgyzov, I. O., O. O. Kyrgyzov, H. Maître, et M. Campedel (2007a). Kernel mdl to determine the number of clusters. In *5th International Conference on Machine Learning and Data Mining MLDM 2007*, (Perner Ed, *LNAI, 4571*), Leipzig, Germany, pp. 203–217.

Lutz, M. (2006). *Programming Python*. O'Reilly Media.

Ng, A., M. Jordan, et Y. Weiss (2001). On spectral clustering : Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, Number 14.

Shawe-Taylor, J. et N. Cristianini (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press.

Summary

The number and quality of satellite images is growing very rapidly, more rapidly than the photo-interpreter are able to use them. This results in a growing number of unexploited data: it is urgent to develop (semi-)automatic method to facilitate the indexing and access to their content. We usually index images while exploiting automatic image analysis process as well as knowledge about the usage domain. In this article, we present the knowledge we get using only a data driven approach in terms of concepts and relations; we also give our thoughts about how to store and organize all the knowledge we accumulate regarding both automatic process and domain application.

Reconstruction 3D sémantique dans le contexte architectural

Christophe Cruz*, Franck Marzani*
Frank Boochs**

* Laboratoire Le2i,
UMR CNRS 5158
Université de Bourgogne,
B.P. 47870, 21078 Dijon Cedex
prenom.nom@u-bourgogne.fr

**Institut i3mainz,
am Fachbereich 1 Geoinformatik und Vermessung,
Fachhochschule Mainz, Holzstrasse
3655116 Mainz,
boochs@geoinform.fh-mainz.de

Résumé. Cet article présente une méthode de reconstruction 3D de bâtiment à partir de nuages de points provenant de la mesure d'un scanner laser 3D. L'objectif est d'avoir, au final, une définition sémantique associée au modèle géométrique reconstruit afin d'être exploitable rapidement dans les logiciels CAO/DAO. Cette méthode se fonde sur l'idée qu'il est possible d'améliorer la reconstruction d'une scène 3D à partir d'un nuage de points de manière automatique, si le contexte sémantique de la scène est défini. Pour atteindre ces objectifs, les recherches entreprises depuis plusieurs années sont pluridisciplinaires. En effet, les techniques de reconstruction 3D font appel aux méthodes de vision artificielle de l'équipe M2D+ du laboratoire Le2i situé à Dijon et de l'institut de recherche appliquée i3mainz situé à Mayence en Allemagne. Concernant la partie gestion de la connaissance, celle-ci fait appel aux techniques des systèmes d'information et bases de données de l'équipe SISI du laboratoire Le2i, située également à Dijon.

1 Introduction

Dans le domaine de l'ingénierie civile, l'extension ou la mise aux normes d'un bâtiment existant est souvent problématique. La plupart du temps, les informations concernant la conception du bâtiment comme les plans ont tout simplement disparu après livraison. Ceci est dû à l'absence de politique de sauvegarde des données. Pourtant, ces dernières sont très utiles pour une estimation rapide du coût des travaux d'aménagement. Prenons l'exemple de la mise aux normes de sécurité : les lois évoluent rapidement en fonction des savoirs et des techniques. Les bâtiments doivent alors suivre cette évolution et être aménagés pour répondre aux nouvelles normes. Toutefois, lorsqu'il s'agit d'estimer le coût de cette évolution, un surcoût est automatiquement associé. En effet, avant toute phase d'étude, le bâtiment doit être mesuré en vue d'établir une maquette numérique exploitable par les logiciels de CAO/DAO. Seulement après cette phase, il est possible d'établir un modèle du bâtiment aux normes et ainsi d'estimer les coûts. Les mesures entreprises exclusivement par les ingénieurs

sont composées de plusieurs étapes comme l'établissement du géo-référencement des éléments géométriques et la capture des données par des méthodes de mesure d'ingénierie civile. Ces méthodes requièrent beaucoup de temps et par conséquent toute méthode automatique ou semi-automatique permettant un gain de temps est la bienvenue. En principe, les techniques photogrammétriques et de scanner laser sont potentiellement exploitables pour améliorer les techniques de reconstruction afin d'atteindre un degré d'automatisme plus élevé. Toutefois, lorsqu'elles sont mal employées, elles peuvent s'avérer chronophage. Cet article présente brièvement ces différentes techniques.

La plupart du temps, les maquettes numériques sont constituées d'éléments géométriques simples. Durant les différentes étapes de conception et d'échange de l'information, les objets constituant les bâtiments sont réduits à un ensemble de vecteurs utilisant généralement le format de fichier DXF ou DWG. Par conséquent, toute information sémantique manipulée par les logiciels CAO/DAO est définitivement perdue lors de l'échange des données. Une phase de qualification permettant d'identifier sémantiquement les éléments géométriques est alors nécessaire afin de réaliser des calculs spécialisés. Par exemple, un logiciel de métré pourra calculer la surface au sol et le coût de changement de revêtement, si le logiciel peut identifier les éléments géométriques qui définissent le sol. Le format IFC (Industrial Foundation Classes) défini par l'IAI (Agence Internationale pour l'Interopérabilité des logiciels CAD/DAO) commence à être utilisé pour éliminer ces phases de qualification. Ce standard associe une définition sémantique aux éléments géométriques dans le domaine de l'architecture. Ce format est actuellement reconnu par la majorité des logiciels CAD/DAO du marché. Par conséquent, l'extraction de la sémantique lors de la phase de reconstruction permet également un gain de temps conséquent lors de la phase d'exploitation des données.

Cet article présente une méthode basée sur la reconstruction à partir de données provenant de scanners laser 3D et fournit en sortie des objets géométriques auxquels sont associés de la sémantique. Les définitions sémantiques sont inspirées des IFC, permettant ainsi la conversion de ces données au format IFC, afin d'être exploitable directement dans les logiciels CAD/DAO. Nous avons défini également une ontologie architecturale qui est le résultat d'une spécification formelle des besoins du projet. Cette ontologie contient les définitions sémantiques des éléments géométriques du contexte architectural.

La section suivante présente un bref état de l'art sur les projets de reconstruction 3D existants. Dans certains projets, la sémantique qui décrit le contexte architectural tient une place importante. La section 3 présente notre approche inspirée par les projets décrits dans la section précédente. La section 4 décrit les points importants de notre méthode de reconstruction.

2 Etat de l'art

Aujourd'hui, le traitement des données géométriques provenant des appareils de mesure est limité par la complexité des objets à extraire. En quelque sorte, il est très difficile, et cela requiert beaucoup de temps, de formuler des règles de détection et d'extraire des objets géométriques corrects. La principale raison provient du fait que les objets sont divisés en de nombreuses sous-parties. Même si chaque élément peut être traité indépendamment, il est très difficile de les traiter tous en même temps. Par conséquent, l'utilisation de la connaissance et son introduction dans le processus de reconstruction est prometteuse. L'impact des informations sémantiques sur le processus de reconstruction dépend de la structure des données brutes. Les deux sous-sections suivantes présentent les principales méthodes de recons-

truction, ainsi que les structures associées. La première d'entre elles présente les méthodes de reconstruction basées sur la photogrammétrie tandis que la seconde présente les méthodes basées sur les nuages de points qui sont le résultat d'un relevé au scanner 3D. Il est évident que chacune des méthodes possède ses propres avantages dépendant des conditions d'utilisation. Toutefois, le choix des outils de mesure dépend bien souvent du matériel disponible, des objets à reconstruire, de la précision souhaitée, et du temps disponible pour réaliser la mesure. Grün et al. (2002), Bryan et al. (1999), Balletti et Mander (2004), Boehler et al. (2004).

2.1 Méthodes basées sur la photogrammétrie

Concernant les méthodes de reconstruction basées sur la photogrammétrie, celles-ci utilisent des photographies de l'objet à reconstruire prises sous différents angles. Il est possible de définir deux groupes de méthodes. Le premier groupe est constitué de méthodes semi-automatiques. Cela consiste à interagir avec l'utilisateur tout au long du processus de reconstruction. Le deuxième groupe de méthodes est constitué des méthodes automatiques où l'utilisateur n'intervient que pendant la phase d'initialisation du processus. Les méthodes manuelles ne sont pas référencées ici. Concernant les méthodes semi-automatiques, nous pouvons citer les projets TotalCalib - Robert (1995), Bougnoux et Robert (1997), Faugeras et al. (1997), Poulin et al. (1998), Rekon - Frasson (1999), Loscos et al. (1999), Marina - Huot et Colin (2002), Cantzler et al. (2002), Nüchter et al. (2003), Realise - (Zitova, 2003). Contrairement aux méthodes semi-automatiques, les méthodes automatiques sont beaucoup moins nombreuses. Nous dénombrons trois méthodes intéressantes que sont celle de Pollefeys et al. (2000) pour la première, celle de Zisserman et al. (2002) pour la seconde. Toutes les deux emploient la géométrie projective sur des images non calibrées. La méthode de Pollefeys combine différents algorithmes de vision artificielle comme la reconstruction projective, l'autocalibration et l'estimation des cartes de profondeur. La dernière méthode est celle qui a attiré notre attention. Elle est utilisée dans le projet Aida, Weik (1996) et est basée sur une définition sémantique des éléments géométriques à l'aide un réseau sémantique. Cette méthode ouvre une nouvelle voie en utilisant la sémantique pour aider la reconstruction. Malgré les années de recherche, la reconstruction automatique reste une tâche très difficile à réaliser, Marr et Poggio (1979), Backer et Binford (1981), Grimson (1981), Fleet et al. (1991), Jones et Malik. (1992), McMillan et Bishop (1995). Le problème majeur est l'impact du point de vue sur l'apparence de l'objet dans l'image. Ceci est dû aux différences entre les clichés qui sont caractérisés par la géométrie, la radiométrie, les occlusions et le manque de texture. Les grandes différences entre les clichés peuvent provoquer une destruction des relations d'adjacence entre les points et engendrent une confusion lors de la détermination des correspondances entre images, notamment quand la surface des objets possède de fortes variations géométriques. Dans le cas de faible texture, les algorithmes ne possèdent pas suffisamment d'informations pour résoudre les problèmes de correspondance, ce qui cause généralement l'échec de la reconstruction.

2.2 Méthodes basées sur les nuages de points

La reconstruction précise d'objets géométriques à partir de nuages de points provenant d'un scanner laser 3D est complexe et constitue un problème qui n'est toujours pas complètement résolu. Cette difficulté vient du fait que les points prélevés ne sont généralement pas

organisés, que le résultat est souvent bruité et qu'il ne reflète pas directement les caractéristiques de l'objet. Ces facteurs limitent l'efficacité des algorithmes de reconstruction. F. Remonido a réalisé un état de l'art très exhaustif de ces algorithmes, Remondino (2003). Une attention toute particulière est portée aux travaux de Cantzler et al. (2002) et aux travaux de Nüchter et al. (2003), car leurs projets utilisent des informations sémantiques. En effet, les plans en cours de reconstruction sont associés à une interprétation sémantique qui doit vérifier un modèle de réseau sémantique, Grau (1997). Un arbre de backtracking permet de trouver les meilleures correspondances entre l'interprétation de la scène et le modèle de réseau sémantique. Pour finir, l'étiquetage des surfaces est cohérent si toutes les surfaces possèdent une étiquette. Par comparaison avec la photogrammétrie, les problèmes semblent être moindres avec les scanners 3D, mais une reconstruction complètement automatique basée sur les nuages de points est aussi complexe qu'une reconstruction complètement automatique avec les méthodes photogrammétriques. La raison principale s'explique par la complexité des objets à reconstruire combinée avec la redondance, l'incomplétude et le bruit caractérisant les nuages de points. Comme le montrent les travaux de Cantzler et Nüchter, il est possible d'attendre des améliorations significatives en utilisant la connaissance des scènes et un modèle sémantique définissant les éléments composant cette scène.

3 Reconstruction guidée par la sémantique

Comme le montrent les projets présentés dans les sections précédentes, la sémantique supporte considérablement la reconstruction 3D. Dans le cas des nuages de points, elle permet de combiner les éléments déjà détectés dans une structure finale. Dans le cas de la photogrammétrie, la sémantique est utilisée pour grouper des points 2D dans les images ou pour former une structure spatiale entre plusieurs images. Dans les deux cas, la structure sémantique est la même, seuls son emploi et les interactions sont différents. Dans cette section, nous allons esquisser notre vision de l'utilisation de la sémantique dans la reconstruction 3D. L'idée principale se trouve dans la dualité entre « contexte » et « contrainte ». Le contexte sémantique de la scène va permettre d'établir des contraintes entre ces éléments. Afin d'utiliser le contexte lors du processus de reconstruction, nous définissons un modèle grossier géométrique et sémantique appelé « MG » lors de la phase d'initialisation de la reconstruction. Ce « MG » (Fig. 1) est une structure spatiale définissant les éléments du bâtiment. Il est à noter que le « MG » définit de manière grossière la géométrie du bâtiment et que les mesures ne sont pas réelles. Il forme en quelque sorte une esquisse définie rapidement par l'ingénieur. Une fois cette esquisse réalisée, elle est corrigée à l'aide des vraies mesures du bâtiment prélevées sur le nuage de points à l'aide de la connaissance définie par le contexte et la définition sémantique des éléments.

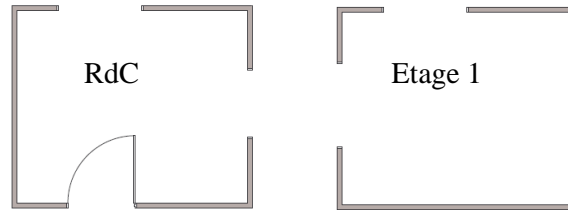


FIG. 1 – Exemple de modèle grossier. Ce « MG » est esquissé rapidement à l'aide d'une interface graphique spécialisée afin de définir les éléments recherchés dans le nuage de points, ainsi que la sémantique associée.

Pour atteindre cet objectif, la connaissance doit représenter le monde réel en reflétant les entités et les relations qui la composent. C'est pourquoi la connaissance constitue un modèle du monde et les processus informatiques utilisent ce modèle comme représentation du monde. De plus, la représentation de la connaissance est indépendante de son utilisation. Par conséquent, la représentation de la connaissance et les mécanismes d'inférence sont dissociés, Guarino et al. (1994). La conceptualisation d'un domaine ne peut être entreprise sans ambiguïté seulement si un contexte d'utilisation est donné. En d'autres termes, un mot ou un terme peut désigner deux concepts différents en fonction du contexte d'utilisation, Bachimont (2000). L'exemple du terme « coccinelle » est flagrant. La sémantique de la connaissance est fortement contrainte par la représentation symbolique des machines. Pour cette raison, N. Guarino (1994) a introduit un niveau ontologique entre le niveau conceptuel et le niveau épistémologique. Ce niveau ontologique forme une passerelle entre la sémantique interprétative avec laquelle l'utilisateur interprète le terme et la sémantique opérationnelle avec laquelle la machine manipule le symbole, Dechilly et Bachimont (2000). Pouvez-vous conduire une « coccinelle », si celle-ci est un insecte ? Pour définir la sémantique de la scène, nous allons utiliser une ontologie qui s'avère être un diagramme de métadonnées, Amann (2003). Son rôle est double. D'une part, il représente la connaissance partagée dans le domaine. D'autre part, il joue le rôle d'un schéma de base de données utilisé pour formuler des requêtes structurées sur les métadonnées. De plus, une ontologie permet de dissocier la représentation de la connaissance et les mécanismes d'inférence. Cette propriété nous a permis de définir une architecture dans laquelle il est possible d'ajouter de nouveaux éléments à l'ontologie sans avoir à modifier le code. Ces nouveaux éléments sont alors pris en compte dans les processus de stockage et d'inférences.

Finalement, lorsque le modèle grossier est corrigé, celui-ci est exportable au format IFC afin d'être exploitable directement dans les logiciels de CAO/DAO.

4 Présentation du processus de reconstruction

La méthode (Fig. 2) que nous avons développée a pour objectif la reconstruction sans intervention de l'utilisateur lors du processus de reconstruction. Par contre, lors de la phase d'initialisation, l'utilisateur intervient en mettant en correspondance un mur du « MG » avec un sous nuage de points représentant le même mur. Ensuite, l'algorithme de reconstruction se

Reconstruction 3D sémantique dans le contexte architectural

charge de rechercher les éléments dans le nuage de points définis dans le « MG » par propagation sur les éléments adjacents, c'est-à-dire les autres murs. L'étape finale de la reconstruction consiste en la détection des éléments contenus dans d'autres éléments, comme les portes et les fenêtres. Cette solution doit prendre en considération les trois interrogations suivantes. Comment définir un modèle grossier géométrique et sémantique ? Comment trouver les éléments géométriques dans les nuages de points ? Comment définir un algorithme de propagation pour rechercher et trouver tous les objets dans le nuage de points.

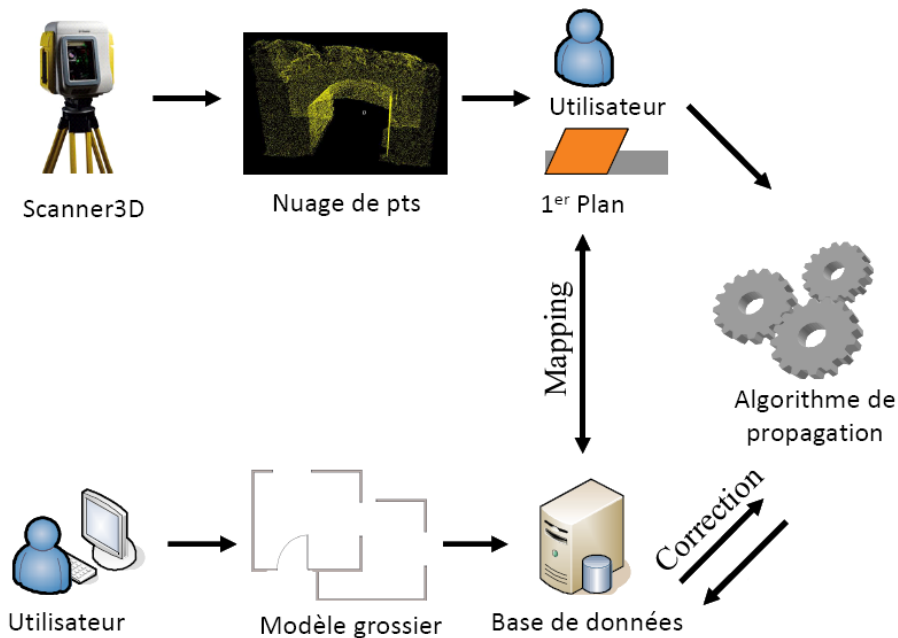


FIG. 2 – Vue globale de la méthode de reconstruction. En haut à gauche se trouve le scanner 3D qui fournit un nuage de points à l'utilisateur. À partir de ce nuage de points, l'utilisateur sélectionne un sous-nuage correspondant à un mur. Avant cette phase, l'utilisateur définit rapidement un « MG » qui est stocké dans la base de données. Une fois la mise en correspondance du sous-nuage de points et du mur dans le « MG », la reconstruction est démarrée et le « MG » est corrigé.

Les trois points suivants présentent brièvement les solutions mises en œuvre pour définir un « MG », la détection des objets dans les nuages de points et la recherche des éléments par propagation.

- Grâce à l'application développée (Fig. 3 droite), l'utilisateur définit la structure générale du bâtiment comme la position et la taille. Cette application permet également de définir automatiquement des contraintes entre les éléments du « MG » en fonction de leur nature. Ces contraintes sont décrites dans l'ontologie architecturale.

Par exemple, la fenêtre est définie comme un concept dans notre ontologie et possède la contrainte suivante : « une fenêtre doit être contenue dans un mur ». Pour calculer cette contrainte, nous nous basons sur la hiérarchie d'héritage défini dans l'ontologie. En effet, les éléments sémantiques sont avant tout des éléments géométriques horizontaux et verticaux auxquels une valeur sémantique a été attribuée. En toute logique, les murs et les fenêtres sont des éléments géométriques verticaux. Pour vérifier la contrainte définie précédemment, il est nécessaire que les plans des deux éléments géométriques soient coplanaires et que la distance entre ces deux plans soit inférieure à un seuil fixé. De plus, les quatre extrémités de la fenêtre doivent être incluses dans le rectangle défini par le mur. Pour vérifier ces contraintes, notre architecture est décomposée en deux niveaux. Le premier est le niveau sémantique où sont stockées les classes de l'ontologie définies à partir d'un fichier OWL (Web Ontology Language) et sur lequel nous baseront nos règles. Le deuxième niveau est le niveau des instances qui devront respecter les règles. Celui-ci permet de stocker les instances des classes d'objets de l'ontologie. Le processus de stockage et l'interface graphique sont suffisamment souples pour ne pas être modifié lors de l'ajout d'une nouvelle classe d'objets dans l'ontologie. De plus, nous avons défini des classes de contraintes associées aux classes d'objets de l'ontologie. Ces contraintes sont associables aux nouvelles classes d'objets. Par exemple, une porte double-battant possèdera les contraintes d'une porte simple par héritage.

- Les éléments géométriques que nous recherchons dans les nuages de points sont les plans. Ces primitives sont les plus faciles à retrouver, et surtout les plus rapides à retrouver, Remondino (2003), dans le cas de nuages dépassant le million de points. Durant le processus de recherche des plans, plusieurs étapes doivent être remplies. La première consiste à partitionner le nuage de points afin d'améliorer les performances du processus. Pour cela, l'espace est subdivisé en voxels. Si des voxels contiennent des points alors, ces voxels sont pris en compte pour le calcul de l'équation du plan. Une fois l'équation de plan calculée pour un voxel initial, les voxels voisins sont examinés afin de les intégrer, le cas échéant, dans le calcul de la nouvelle équation de plan initial. Si le plan formé par les pixels d'un voxel voisin n'est pas similaire à celui du voxel courant, alors il n'est pas utilisé pour le calcul du plan courant. Toutefois, il servira à calculer l'équation d'un autre plan. Pour prendre la décision de les intégrer, plusieurs solutions sont disponibles. La première consiste à utiliser la méthode des moindres carrés. Une meilleure solution consiste à sélectionner les voxels avec la meilleure erreur résiduelle dans le voisinage direct, et de tester la distance avec le plan. Si la somme des distances est plus basse qu'un seuil basé sur un angle d'orientation alors, le voxel est fusionné. Une fois le voxel fusionné, l'équation du plan doit être recalculée afin d'affiner le résultat.
- Le principe de ce projet est d'utiliser un nuage de points provenant du relevé d'un bâtiment pour corriger un modèle grossier de ce même bâtiment. La phase d'initialisation par mise en correspondance du « MG » avec le nuage de points est fondamentale pour l'amélioration du modèle. Tout d'abord, il permet de réajuster les dimensions initiales du mur dans le modèle grossier et de redresser le « MG » par rapport au nuage de points. De cette manière, le processus peut identifier les directions de l'espace, par exemple différencier le haut et le bas. Deuxièmement, par

Reconstruction 3D sémantique dans le contexte architectural

propagation et grâce aux contraintes établies entre éléments du modèle grossier, les autres éléments sont corrigés. Par exemple, si la hauteur du mur initial est modifiée, alors les murs adjacents qui avaient la même hauteur sont corrigés également. Une fois le premier élément géométrique corrigé, l'algorithme passe aux éléments adjacents dans le « MG » et recherche ces éléments dans le nuage de points. Ceci est possible grâce aux relations de voisinage établies durant la phase de définition du « MG ». L'algorithme s'arrête une fois tous les éléments passés en revue.

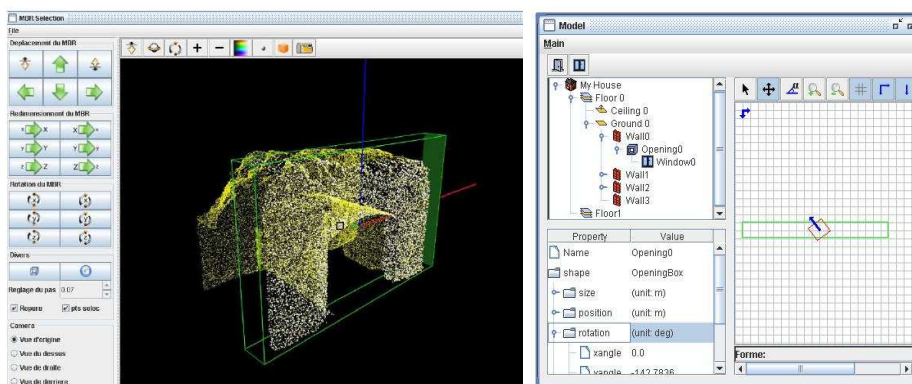


FIG. 3 – A gauche, capture d'écran de l'application permettant de sélectionner un sous-nuage de points correspondant à un mur. À droite, capture d'écran de l'application permettant de définir le « MG » ; sur la gauche de cette fenêtre, il est possible de voir la hiérarchie des éléments sémantiques, ainsi que les propriétés de l'élément sélectionné.

5 Définition et utilisation des relations spatiales

Après la réalisation d'une première version de l'ontologie, la question de l'extensibilité de l'ontologie est devenue rapidement problématique. Tous les concepts et relations (géométriques et sémantiques) de l'ontologie ainsi que les règles ont été définis en « dur » dans l'application. Il était alors très difficile de faire évoluer l'ensemble de l'application. Le problème majeur est l'interaction des éléments, et plus particulièrement leurs positionnements et leurs redimensionnements. Si un élément doit être mis à l'échelle et déplacé, alors un ensemble d'éléments associés doit également être mis à l'échelle et déplacé. Pour répondre à cette difficulté, un ensemble d'éléments sémantiques, de relations et de comportements sur ces éléments a été défini. Ces éléments sémantiques sont les éléments horizontaux, les sous-éléments horizontaux, les éléments verticaux et les sous-éléments verticaux. Les relations définies sont les relations d'appartenance (EtreSur Elément, EtreSur Elément Vertical, EtreSur Elément Horizontal), et les relations de contenance (Avoir Element, Avoir Elément Vertical, Avoir Elément Horizontal). A partir de ces éléments et relations, il est possible de définir des comportements qui correspondent aux règles devant être respectées par la correction du Modèle Grossier. Par exemple, si un élément horizontal est déplacé selon un vecteur de déplacement, alors tous les éléments liés par la relation « Avoir Elément Vertical » doivent être déplacés en respectant le même vecteur de déplacement. Cet exemple correspond à

l'exemple d'un mur et d'une fenêtre de la section précédente. Grâce à cette extensibilité de l'ontologie, il est possible de définir de nouveaux éléments sémantiques par le biais de l'héritage. Un « poteau » peut être défini comme un élément vertical et peut être lié à un élément horizontal « dalle » par la relation EtreSur Elément Horizontal. Ce nouvel élément « poteau » exploitera la règle « un poteau se trouve sur une dalle ». Le comportement exploitable par cette règle est que tout élément se trouvant sur un élément horizontal subira le même vecteur de placement que cet élément horizontal. Si cette « dalle » est la dalle du rez-de-chaussée, alors tous les éléments du bâtiment seront modifiés par propagation des règles.

6 Conclusion

Cet article a présenté brièvement une solution que nous proposons pour la reconstruction 3D sémantique guidée par une ontologie architecturale. Elle tente d'aller plus loin que les solutions existantes en intégrant des contraintes sur les éléments du bâtiment. Ces travaux entrepris depuis deux ans révèlent de nombreux verrous techniques, technologies et financiers que nous tâchons de résoudre un à un. Actuellement, plusieurs problèmes ont été résolus et le processus complet a été prototypé. Toutefois, de nombreux obstacles persistent, tels les cas particuliers qu'il est possible de rencontrer dans la réalité. Nous prospectons une nouvelle méthode basée sur la définition de règles afin d'éliminer la phase de définition de modèle grossier.

Références

- Amann, B. (2003). *Du Partage centralisé de ressources Web centralisées à l'échange de documents intensionnels*. Documents de Synthèse.
- Bachimont, B. (2000). *Engagement sémantique et engagement ontologique : conception et réalisation d'ontologie en ingénierie des connaissances*. In Charlet J., Zackland M., Kessel G. & Bourigault D., eds., *Ingénierie des connaissances : évolution récentes et nouveaux défis*, Eyrolles, pages 305-323.
- Backer, H. H. et Binford, T. O (1981). *Depth from edge and intensity based stereo*. In Proceedings of the seventh IJCAI, Vancouver, BC, pages 631-636.
- Balletti, C. et Mander, S. (2004). *Contemporary Master's Architecture: New Architectural Heritage, Approaches For Surveying and Representation*. Geo-Imagery Bridging Continents, XXth ISPRS Congress, 12-23 July, Istanbul, Turkey.
- Boehler, W. et al. (2004). *The potential of non-contact close range laser scanners for cultural heritage recording*. Actes du XVIII Symposium International CIPA, Postdam, Allemagne.
- Bougnoux, S. et Robert L. (1997). *TotalCalib: a fast and reliable system for off-line calibration of images sequences*. In Proceedings of International Conference on Computer Vision and Pattern Recognition, The Demo Session.
- Bryan, P.G., Corner, I. et Stevens, D. (1999). *Digital Rectification Techniques for Architectural and Archaeological*. Photogrammetric Record, 16(93): 399-415, April.

- Cantzler, H., Fisher, R. B. et Devy, M. (2002). *Quality enhancement of reconstructed 3D models using coplanarity and constraints*. Proc. Annual German Symposium for Pattern Recognition (DAGM02, Zurich), pp 34-41.
- Dechilly, T. et Bachimont, B. (2000). *Une ontologie pour éditer des schémas de description audiovisuels, extension pour l'inférence sur les descriptions*. In Actes des journées francophones d'Ingénierie des Connaissances (IC'2000).
- Faugeras, O., Laveau, S., Robert, L., Csurka, G., Zeller, C., Gauclin, C. et Zoghliami, I., (1997). *3-d reconstruction of urban scenes from image sequences*. CVGIP : Image Understanding.
- Fleet, D. J., Jepson, A. D. et Jenkin M. R. M. (1991). *Phase-Based Disparity measurement*. CVGIP : Image Understanding, 53(2):198-210.
- Frasson, M. (1999). *Reconstruction interactive de scènes tridimensionnelles à partir d'images*. M.Sc. Thesis, March.
- Grau, O. (1997). *A Scene Analysis System for the Generation of 3-D Models*. 3dim, p. 221, First.
- Grimson, W. E. L. (1981). *From Images to Surfaces*. MIT Press.
- Grün, A., Bär, S. et Beutner, S., (2002). *Signals in the Sand - 3D Recording and Visualization of the Nasca Geoglyphs*. PFG (Photogrammetrie, Fernerkundung, Geoinformation), No. 6/2000. pp. 385-398.
- Guarino, N., (1994). *The ontological level*. in R. Casati B. S. & White G., eds, Philosophy and the cognitive sciences, Hölder-Pichler-Tempsky.
- Guarino, N., Carrara C. et Giaretta P. (1994). *An ontologie of meta-level categories*. in J. Doyle F. S & Torano P., eds., Principles of Knowledge representation and Reasoning, Morgan-Kauffman, pages 270-280.
- Huot, S. et Colin C. (2002). *MARINA : reconstruction de bâtiments 3D à partir d'images*. Colloque Modélisation Multimodale appliquée à la reconstruction d'environnements architecturaux et urbains, Bordeaux, France.
- Jones, D. et Malik J., (1992). *Computational Framework for determining stereo correspondence from a set of linear spatial filters*. Image and Vision Computing, 10(10):699-708, December.
- Kuzo, P. M. (1999). *Des contraintes projectives en modélisation tridimensionnelle interactive*. Thèse de doctorat, Ecole des Mines de Nantes – Université de Nantes, novembre.
- Loscos, C., Frasson, M., Drettakis, G., Walter, B., Granier, X. et Poulin, P. (1999). *Interactive Virtual Relighting and Remodeling of Real Scenes*. Proc. Eurographics Workshop on Rendering 99, June.
- Marr, D. et Poggio, T. (1979). *A computational theory of human stereo vision*. Proceedings of the Royal Society of London, 204:301-328.
- McMillan, L. et Bishop, G. (1995). *Plenoptic modeling : An image-based rendering system*. In SIGGRAPH '95.

- Nüchter, A., Surmann, H. et Hertzberg, J. (2003). *Automatic Model Refinement for 3D Reconstruction with Mobile Robots*. Fraunhofer Institute for Autonomous Intelligent Systems (AIS) Schloss Birlinghoven, D-53754 Sankt Augustin, Germany.
- Pollefeys, M., Koch, R., Vergauwen, M. et Van Gool, L. 2000. *Automated reconstruction of 3D scenes from sequences of images*. ISPRS Journal Of Photogrammetry And Remote Sensing (55)4, pp. 251-267.
- Poulin, P., Ouimet, M. et M. Frasson, 1998. *Interactively Modeling with Photogrammetry*. Proc. Eurographics Workshop on Rendering 98, June.
- Remondino, F. (2003). *From point cloud to surface: the modeling and visualization problem*. Proc. Int. Worksh. Visualization and Animation of Reality-Based 3D Models, Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIV-5/W10, Feb.
- Robert, L. (1995). *Camera calibration without feature extraction*, Computer Vision, Graphics, and Image Processing, 63(2) :314–325, March also INRIA Technical Report 2204.
- Weik, S. et Grau, O. (1996). *Recovering 3-D Object Geometry using a Generic Constraint Description*. In ISPRS96 - 18th Congress of the International Society for Photogrammetry and Remote Sensing, July, Vienne.
- Werner, T. et Zisserman, A. (2002). *New Techniques for Automated Architecture Reconstruction from Photographs*. Proc. 7th European Conference on Computer Vision, Copenhagen, Denmark.
- Zitova, B. et Flusser, J. (2003). *Image registration methods: A survey*, Image and Vision. Computing 21, 977–1000.

Summary

This paper presents an ontology-driven 3D architectural reconstruction approach based on the survey with a 3D scanner. This solution is powerful in the field of civil engineering projects to save time during the cost process estimation. This time is saved using efficient scanning instruments and a fast reconstruction of a digital mock-up that can be used in specific software. The reconstruction approach considers the three following issues. How to define an ontology to drive the reconstruction process? How to find semantic objects in a cloud of points? How to control an algorithm in order to find all objects in the cloud of points? This paper underlines the solutions found for these questions.

Enrichissement sémantique d'annotations de documents multimédias à l'aide de connaissances explicites

Nicolas James*, Céline Hudelot*

*Laboratoire MAS - Ecole Centrale Paris, Châtenay-Malabry, France
nicolas.james@ecp.fr,
<http://www.mas.ecp.fr/>

Résumé. Cette communication traite du problème de l'annotation automatique de grandes collections de documents multimédia et de l'enrichissement sémantique de ces annotations. L'annotation sémantique d'images est un axe de recherche très actif à l'heure actuelle qui permet notamment d'adresser le problème du *fossé sémantique* et qui a donc un impact important sur l'amélioration des performances des systèmes de recherche d'information. Selon le type du système de recherche d'information multimédia (basé sur le contenu, basé sur le texte, par retour de pertinence ou hybride), on peut distinguer deux approches principales pour l'annotation de documents multimédia : (1) l'annotation basée sur le contenu qui se base sur des techniques de propagation sémantique ou d'inférence statistique; (2) l'annotation basée sur les informations linguistiques et contextuelles accompagnant les documents multimédia. Un autre constat est l'importance prise par les ontologies, au sens large, dans le processus d'annotations de documents multimédia. Dans cette communication, nous proposons une approche hybride pour l'annotation sémantique de documents multimédia qui se base d'une part sur des techniques d'analyse de liens entre documents comme présentées dans Pan et al. (2004) et d'autre part des techniques d'enrichissement sémantique se basant sur des ontologies et du *concept mining*.

1 Description

L'essor du Web et la démocratisation des technologies de l'information ont générés une véritable explosion du nombre de documents multimédias nécessitant le développement de nouvelles méthodes, plus efficaces, pour la gestion et la recherche d'information dans ces documents multimédia. Les systèmes opérationnels actuels, comme par exemple le leader Google, se basent toujours principalement sur une recherche textuelle de mots clés dans le texte accompagnant les documents alors qu'un effort important de recherche a porté sur les techniques d'analyse, d'indexation et de recherche par le contenu des documents multimédias. Une des principales limitations des systèmes et des méthodes de recherche par le contenu est le problème du fossé sémantique défini par Smeulders et al. (2000) comme le problème *du manque de concordance entre les informations perceptuelles que l'on peut extraire des données multimédia et l'interprétation qu'ont ces données pour un utilisateur dans une situation déterminée*.

Dans un tel contexte, l'annotation (automatique) sémantique d'images, c'est à dire associer de manière automatique aux documents un ensemble de concepts décrivant leur contenu, permet en partie de combler le fossé sémantique. Cette problématique est donc largement étudiée depuis quelques années et a donné lieu à une littérature abondante sur le sujet (Duygulu et al. (2002); Carneiro et al. (2007); Jeon et Manmatha (2004); Monay et Gatica-Perez (2003); Pan et al. (2004); Wang et al. (2006); Zhang et al. (2006); Bloehdorn et al. (2005); Hollink et al. (2003)). Quelques prototypes pour l'annotation d'images sont d'ailleurs disponibles sur Internet (ALIPR¹, Behold²). Cependant, dans la plupart de ces travaux, le résultat de l'annotation est souvent limité à un ensemble de mots clés représentatifs du contenu du document multimédia. Pour améliorer cette annotation, il est nécessaire d'une part d'évaluer la pertinence et de mettre à jour l'annotation et d'autre part d'enrichir ces annotations pour permettre des requêtes à granularité multiples et sémantiquement riches. Dans cette communication, nous proposons une approche hybride pour l'annotation sémantique de documents multimédia. Dans un premier temps, un ensemble de mots clés est associé au document multimédia par une technique d'analyse de liens proposée par Pan et al. (2004) et améliorée dans Bartolini et Ciaccia (2007). Ensuite, une phase d'enrichissement sémantique des annotations basée sur des ontologies et des techniques de concept mining permet de préciser la sémantique des annotations et de prendre en compte des informations contextuelles et textuelles associées au document.

Références

- Bartolini, I. et P. Ciaccia (2007). Imagination : Accurate image annotation using link-analysis techniques. In *AMR 07 : Proceedings of the International Workshop on Adaptive Multimedia Retrieval*, Paris, France.
- Bloehdorn, S., K. Petridis, C. Saathoff, N. Simou, V. Tzouvaras, Y. Avrithis, S. Handschuh, I. Kompatsiaris, S. Staab, et M. Strintzis (2005). Semantic Annotation of Images and Videos for Multimedia Analysis. *Proceedings of the 2nd European Semantic Web Conference (ESWC 2005)*.
- Carneiro, G., A. B. Chan, P. J. Moreno, et N. Vasconcelos (2007). Supervised learning of semantic classes for image annotation and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(3), 394–410.
- Duygulu, P., K. Barnard, J. F. G. de Freitas, et D. A. Forsyth (2002). Object recognition as machine translation : Learning a lexicon for a fixed image vocabulary. In *ECCV '02 : Proceedings of the 7th European Conference on Computer Vision-Part IV*, London, UK, pp. 97–112. Springer-Verlag.
- Hollink, L., A. Schreiber, J. Wielemaker, et B. Wielinga (2003). Semantic annotation of image collections. *Knowledge Capture*, 41–48.
- Jeon, J. et R. Manmatha (2004). Using Maximum Entropy for Automatic Image Annotation. *Image and Video Retrieval : Third International Conference, CIVR 2004, Dublin, Ireland, July 21-23, 2004 : Proceedings*.

¹<http://www.alipr.com/>

²<http://photo.beholdsearch.com/>

- Monay, F. et D. Gatica-Perez (2003). On image auto-annotation with latent space models. In *MULTIMEDIA '03 : Proceedings of the eleventh ACM international conference on Multimedia*, New York, NY, USA, pp. 275–278. ACM.
- Pan, J.-Y., H.-J. Yang, C. Faloutsos, et P. Duygulu (2004). Gcap : Graph-based automatic image captioning. *cvprw 09*, 146.
- Smeulders, A., M. Worring, S. Santini, A. Gupta, et R. Jain (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(12), 1349–1380.
- Wang, X., L. Zhang, F. Jing, et W. Ma (2006). AnnoSearch : Image Auto-Annotation by Search. *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2*.
- Zhang, R., Z. Zhang, M. Li, W. Ma, et H. Zhang (2006). A probabilistic semantic model for image annotation and multi-modal image retrieval. *Multimedia Systems* 12(1), 27–33.

Summary

This communication deals with the problem of the automatic annotation of very large multimedia databases and with the semantic enrichment of these annotations. Semantic image annotation is a very active research domain which tackles the *semantic gap* problem and has therefore an important impact on the improvement of information retrieval systems. Depending on the type of multimedia information retrieval systems (content-based, text-based, by relevance feedback or hybrid), two main approaches can be distinguished for multimedia document annotation : (1) content-based annotation, based on semantic propagation techniques or statistical inference; (2) annotation based on linguistic or contextual information retrievable from or included in multimedia documents. Another fact is the importance taken by ontologies, in its broad meaning, in the multimedia document annotation process. We propose in this communication an hybrid approach for multimedia semantic annotation. Our approach is based, on one hand, on link analysis techniques between documents such as those defined in Pan et al. (2004), and on another hand, on semantic enrichment techniques based on ontologies and *concept mining*.

Vers une ontologie modulaire multi-niveaux pour l'indexation sémantique multi-points de vue du compte-rendu d'imagerie médicale

Sonia Mhiri* ***, Sylvie Despres**, Ezzeddine Zagrouba***

*Université René Descartes/Paris V, CRIP5/Equipe IAA,
UFR Mathématiques et Informatique,
45 rue des Saints-Pères, 75006 Paris, France
sonia.mhiri@math-info.univ-paris5.fr

**Université Paris-Nord/Paris XIII, LIPN, Institut Galilée,
99 avenue Jean-Baptiste Clément, 93430 Villetaneuse, France
sylvie.despres@lipn.univ-paris13.fr

***Université de Tunis El Manar, Institut Supérieur d'Informatique, URSIIVA,
2 rue Abou Rayhane El Bayrouni, 2080 Ariana, Tunisie
ezzeddine.zagrouba@fsm.rnu.tn

Résumé. Notre objectif est de construire une ontologie dans le but d'assister les spécialistes dans leur processus d'indexation du compte-rendu d'imagerie médicale (informations textuelles et images). Comme ces comptes rendus peuvent être objectivement comme subjectivement décrits, nos premiers travaux se sont tout d'abord orientés vers la proposition d'un modèle de représentation des points de vue possibles et privilégiés lors de l'indexation. Conformément à ce modèle et dans le cadre de la construction de l'ontologie, le recours à une approche modulaire et à la réutilisation d'ontologies est suggéré. Pour organiser les différents modules de l'ontologie, trois niveaux sont élaborés : le niveau commun, le niveau d'ancrage et le niveau domaine. Quant aux ontologies existantes, elles sont répertoriées en vue de déterminer d'une part les possibilités et les techniques appropriées de réutilisation et d'entamer d'autre part, l'expérimentation de notre approche méthodologique de construction dite mixte et fondée sur un niveau d'ancrage.

1 Introduction

En imagerie médicale, les médecins spécialistes bénéficient aujourd'hui d'une variété de systèmes d'imagerie (imagerie par résonance magnétique, imagerie nucléaire, imagerie à rayons X, échographie...) et par conséquent d'une masse importante d'informations multimodales et numérisées. Face à cette croissance rapide, la mise en œuvre de systèmes d'indexation et de recherche adéquats se révèle indispensable. Dans un but d'aide au diagnostic, nos travaux de recherche visent à exploiter l'ingénierie des connaissances et plus particulièrement les ontologies au service d'un futur système d'indexation et de recherche

sémantique de comptes-rendus d'imagerie médicale conformes aux standards existants (DICOM SR¹, HL7/CDA²...). Au sein du service d'imagerie, les requêtes formulées par les médecins spécialistes devront conduire à la recherche de comptes-rendus similaires au compte-rendu soumis (Mhiri et Després, 2007). Le compte-rendu numérisé du patient comprend l'ensemble des examens médicaux réalisés (séries d'images) ainsi que les observations établies (compte-rendu textuel). Dans cet article, nous allons nous intéresser à l'étape d'indexation et porter un intérêt plus spécifique aux techniques relevant de l'indexation sémantique (Liu et al., 2007). Ce procédé permet d'associer à un document (texte, image statique ou dynamique, séquence sonore...) un ensemble de descripteurs relatifs à l'interprétation de son contenu. Mais, cette sémantique n'est pas toujours explicitement présente dans le document parce qu'elle dépend d'une part des connaissances a priori sur le domaine et d'autre part du contexte de l'interprétation (Bouchet et al., 2005). Autant de difficultés qui ont amené la communauté de chercheurs à être de plus en plus séduits par la représentation possible de cette sémantique sous la forme d'ontologies.

C'est dans ce contexte que s'inscrivent les travaux de recherche présentés dans ce papier. L'objectif est de construire une ontologie dans le but d'assister les médecins spécialistes en imagerie dans leur processus d'indexation du compte-rendu. Question délicate dès lors que les praticiens disposent d'une large liberté d'expression lors de la formulation de leur requête de recherche. Dans un premier temps, un état de l'art sur les travaux abordant cette problématique est évoqué. Par la suite et pour répondre à ce problème, nous nous sommes tout d'abord orientés, dans la section 3, vers l'élaboration d'un modèle de représentation des points de vue possibles et privilégiés pour toute recherche effectuée. Puis conformément à ce modèle et sachant que les activités de recherche et d'indexation sont indissociables, nous suggérons dans la section 4 de construire une ontologie pour une indexation multi-points de vue. Le recours d'une part à la modularisation et d'autre part à la réutilisation est justifié. Trois niveaux structurent les modules de l'ontologie : le niveau commun, le niveau d'ancrage et le niveau domaine. Quant aux ontologies existantes, elles sont répertoriées en plusieurs classes en vue d'étudier les possibilités de réutilisation de certaines d'entre elles. L'approche méthodologique de construction mixte que nous décrivons par la suite s'articule autour du niveau d'ancrage. Diverses techniques propres à la réutilisation comme l'extraction et l'enrichissement ainsi que des procédés classiques de modularisation telles que le partitionnement et le couplage sont mises en œuvre. Avant de conclure, nos premières expérimentations sont présentées.

2 Travaux existants

Dans le domaine de l'imagerie médicale, peu de travaux font référence à l'utilisation d'une ontologie à la fois modulaire et structurée en niveaux pour l'indexation du compte rendu d'imagerie. En particulier, la problématique de l'indexation selon des points de vue multiples est rarement prise en compte. Nous pouvons néanmoins mentionner les recherches menées en neuro-imagerie dans le projet NeuroBase et la construction de l'ontologie OntoNeuroBase (Temal et al., 2006). Pour servir de support à une recherche sémantique

¹ Digital Imaging and COmmunications in Medecine Structured Reporting

² Health Level 7/ Clinical Document Architecture

multi-critères d'images, une ontologie modulaire est structurée en trois niveaux au moyen d'une ontologie de haut niveau, de cinq ontologies noyaux et de deux ontologies de domaine. De nombreuses recherches sont également menées dans d'autres domaines mais la majorité d'entre elles explorent les approches de construction de type modulaire (Ribeiro et al., 2006, Thomas et al., 2006, Grau et al., 2006...). Les ontologies à la fois modulaires et structurées en couches sont moins nombreuses. Nous pouvons souligner les expérimentations menées dans le cadre d'une application d'aide à la prise de décision pour l'aménagement du territoire urbain (Benslimane et al., 2000). Une ontologie modulaire est organisée selon trois niveaux : le haut niveau, le niveau générique et le niveau domaine. D'autres travaux ont également été réalisés dans le cadre des systèmes distribués hypermédias (Alexakos et al., 2006). Une ontologie pour l'indexation des contenus hypermédias est structurée en un haut niveau, un niveau domaine et un niveau spécifique.

3 Compte-rendu d'imagerie médicale

Les médecins spécialistes en imagerie médicale sont quotidiennement conduits à réaliser les activités suivantes : la réception du patient muni de la prescription de son médecin traitant, la réalisation d'un ou de plusieurs examens, l'archivage approprié des séries d'examens produits avec une possibilité d'impression, l'interprétation des examens et finalement la réalisation du compte-rendu. Dans ce circuit classique (hôpital, clinique ou cabinet), comment les professionnels de l'imagerie peuvent-ils réexploiter les images produites et les comptes-rendus textuels établis afin de conforter leur pré-diagnostic ?

3.1 Vers un compte-rendu multimédia standardisé

Dans le domaine de la santé, il existe actuellement plusieurs standards internationaux pour l'échange des données. Les médecins sont convaincus que ces standards doivent absolument prendre en considération les images (Dossier Snitem Info, 2006). Pour l'imagerie médicale par exemple, les comptes-rendus effectués après chaque examen ne peuvent remplacer les images. Elles interviennent dans le processus de diagnostic et constituent une aide précieuse dans le suivi thérapeutique du patient. En attendant l'adoption des normes définitives pour assurer l'interopérabilité entre les systèmes d'information existants, trois standards de représentation de la structure du document/dossier médical émergent : HL7/CDA (document), CEN TC 251/EHRCOM (dossier) et DICOM/SR (document).

Health Level 7/Clinical Document Architecture³. La norme CDA organise les informations du document en trois niveaux :

- le *Coded Header* de documents repérés de façon unique grâce aux informations enregistrées dans le *header* (nom, identifiant, etc.) sans contrainte sur le contenu textuel ;
- le *Coded Context* adopte un niveau de description plus fin tout en respectant les mêmes *headers* que le niveau précédent avec en plus un contenu textuel organisé logiquement ;
- le *Coded Content*, niveau de découpage permettant de structurer de façon détaillée les informations médicales grâce au Reference Information Model (RIM). Le RIM

³ <http://www.hl7.org/>

Ontologie modulaire multi-niveaux pour l'indexation du compte-rendu d'imagerie médicale

présente une vue des données à échanger et permet de construire les messages qui pourront être échangés entre des systèmes hétérogènes

CEN TC 251/Electronic Healthcare Record COMMunication⁴. EHRCOM est une norme européenne conçue de façon à mettre à la disposition des utilisateurs des composants permettant de structurer le dossier. Ce sont les composants et non leur assemblage qui sont standardisés. Les choix pour la structuration du dossier sont laissés libres. Des indications sont uniquement données pour l'utilisation des différents composants. Un EHRCOM est composé de « dossiers », composés de « compositions », composées de « sections », composées « d'entrées », etc.

Comité DICOM/Structured Reporting⁵. DICOM SR est une évolution du standard mondial d'imagerie médicale DICOM (communication, archivage et gestion des images) vers la représentation du compte-rendu d'imagerie structurée. SR répertorie toutes les données utilisables dans le compte-rendu structuré. Les domaines couverts sont pour l'instant l'imagerie cardio-vasculaire, la mammographie, l'imagerie thoracique, les actes obstétriques et gynécologiques. Cette norme est représentée sous la forme d'une arborescence de nœuds appelés « Contents Items » représenté par un couple (nom /text /image /date /time..., valeur) reliés par des relations spécifiques (contains, has properties...). Avec une telle structure, tous les types de comptes-rendus peuvent être envisagés (simple, complexe et multimédia).

3.2 Modèle de représentation des points de vue d'indexation

Comment les médecins spécialistes en imagerie peuvent-ils interroger une base de comptes-rendus multimédias standardisés ? En réalité, plusieurs formes de requêtes de recherche sont possibles (texte, image...) mais notre attention s'est plus particulièrement portée sur les différents points de vue privilégiés par les praticiens lors de la formulation de leur requête.

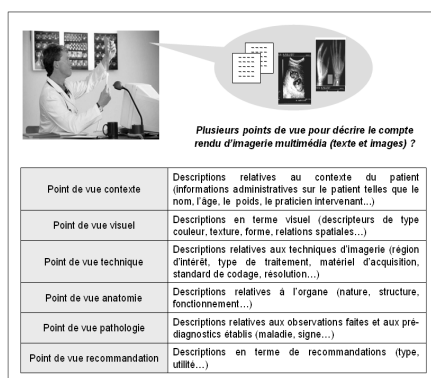


FIG. 1 – *Modèle des points de vue d'indexation du compte-rendu d'imagerie médicale*

⁴ <http://www.centc251.org/>

⁵ <http://medical.nema.org/dicom/>

Quel que soit le praticien, l'accent mis sera différent et diverses formulations sont alors possibles : les images contenant une tumeur (un accent sur la pathologie), les images contenant une couleur donnée à telle position (un accent sur l'aspect visuel), les comptes-rendus établis par Mr X en 2007 (un accent sur le contexte), etc. De plus, la médecine contrairement à d'autres disciplines scientifiques, ne possède pas de vocabulaire unique standardisé et le nombre élevé de terminologies en est un parfait exemple (Zweigenbaum, 1999). Il est opportun de mettre l'accent sur la diversité des termes médicaux dont la définition est rarement rigoureuse parfois même ambiguë. Une telle imprécision est un problème en soi. Face à cette situation et sachant que les activités d'indexation et de recherche sont indissociables, nous avons défini un modèle de représentation des points de vue possibles et privilégiés par les praticiens. En nous appuyant sur les standards de représentation du contenu du compte-rendu précédemment évoqués, six points de vue d'indexation sont suggérés et illustrés par la figure 1 : contexte, visuel, technique, anatomie, pathologie et recommandation.

4 Vers une ontologie modulaire multi-niveaux

Dans ce paragraphe, nous traitons les premiers éléments de réponse à la question suivante : comment concevoir une ontologie pour l'aide à l'indexation multipoints de vue d'un compte-rendu d'imagerie médicale comprenant à la fois des informations textuelles et une ou plusieurs images ?

4.1 Modèle de structuration en niveaux des modules

Pour représenter la connaissance relative aux six points de vue proposés, nous sommes confrontés à une alternative pour la construction de l'ontologie. Soit une construction *ex nihilo* qui s'appuie sur l'expertise humaine du domaine et sur la documentation existante, soit une réutilisation d'ontologies existantes. Le premier choix exige du temps et un effort considérable de conceptualisation des terminologies médicales déjà existantes (Mesh⁶, Snomed⁷, Umls⁸, Radlex⁹...). Pour remédier à cette situation, nous suggérons d'explorer la réutilisation d'ontologies. Ce second choix est prometteur en raison du nombre particulièrement élevé d'ontologies du domaine médical disponible sur Internet. Même si aucune ontologie existante ne répond exactement à notre besoin, parce qu'une ontologie par principe est conçue pour couvrir une tâche particulière d'une application dans un domaine donné, parmi celles qui existent certaines peuvent servir de base à la construction. Cette réutilisation est une voie à la fois encourageante mais encore difficile et de nombreux travaux en témoignent (Klein, 2001). De ce fait, nous avons également opté pour une ontologie modulaire afin de faciliter d'une part la maintenance (évolution vers un contenu de type images animées, séquences sonores...) et d'autre part les futures réutilisations de l'ontologie par la communauté biomédicale en général et celle de l'imagerie médicale en particulier. L'ontologie des points de vue d'indexation du compte-rendu d'imagerie s'apparente donc à l'assemblage cohérent de différents modules construits de manière

⁶ <http://www.nlm.nih.gov/mesh/>

⁷ <http://www.snomed.org>

⁸ <http://www.nlm.nih.gov/research/umls/>

⁹ <http://www.rsna.org/radlex/>

Ontologie modulaire multi-niveaux pour l'indexation du compte-rendu d'imagerie médicale

indépendante, couvrant des domaines distincts et structurés par trois niveaux d'abstraction illustrés par la figure 2. Le niveau domaine correspond à l'ontologie modulaire. Le haut niveau, plus générique, mentionne l'ontologie de haut niveau nécessaire à l'intégration sémantique ultérieure des ontologies médicales existantes. Quant au niveau intermédiaire dit d'ancrage, il devra garantir un accrochage spécifique pour l'aide à la structuration de l'ontologie de domaine avec l'ontologie de haut niveau.

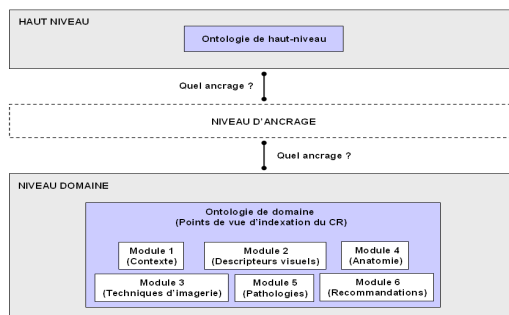


FIG. 2 – *Modèle de structuration en 3 niveaux des modules*

4.2 Classes d'ontologies OWL existantes

Les ontologies existantes ont été réparties en deux catégories (haut niveau et domaine) afin de déterminer les possibilités de réutilisation. Notre choix s'est porté sur le *Web Ontology Language* (OWL¹⁰). Recommandé par le W3C¹¹, ce langage est actuellement considéré comme le standard de représentation des ontologies. De plus et particulièrement en biomédecine, plusieurs travaux sont entrepris pour la traduction de terminologies médicales telles que Radlex, UMLS... en ontologies OWL.

4.2.1 Ontologies de haut niveau

Une ontologie de haut niveau décrit des concepts très généraux comme l'espace, le temps, la matière, les objets, les événements, les actions, etc. Ces concepts sont en principe partagés par une large communauté d'utilisateurs, universels en ne faisant référence à aucun domaine spécifique et multi-usage pour n'avoir été conçus pour une aucune tâche particulière. Dans l'état de l'art (Masolo et al., 2003), plusieurs ontologies existent dont l'ontologie des particuliers DOLCE¹², l'ontologie SUMO¹³ et l'ontologie BFO¹⁴ illustrées par la figure 3. SUMO introduit deux concepts importants : *Physical* (entités du monde réel) et *Abstract* (représentations abstraites que l'on se fait d'une réalité). L'ontologie des particuliers DOLCE représente deux mondes : des *Particuliers* avec des *Perdurant* (entités

¹⁰ <http://www.w3.org/TR/owl-features/>

¹¹ World Wide Consortium (<http://www.w3.org/>)

¹² Descriptive Ontology for Linguistic and Cognitive Engineering

¹³ Suggested Upper Merged Ontology

¹⁴ Basic Formal Ontology

du monde réel partiellement présentes) et des *Endurant* (entités du monde réel toujours présentes) et des *Universels* comprenant les concepts nécessaires à la description des *Particuliers*. Quant à l'ontologie BFO, plus récente et conçue spécifiquement pour le domaine biomédical (Grenon et al., 2004), elle distingue comme DOLCE les SPAN entités *Enduring* (entités du monde réel toujours présentes) des SNAP entités *Concrete* (entités du monde réel partiellement présentes). En revanche, le reste de l'arborescence est adaptée au monde biomédical.

4.2.2 Ontologies de domaine relatives aux descripteurs visuels

Dans cette section, nous évoquons les ontologies développées dans le cadre de la représentation des descripteurs visuels d'une image. Nous citons en ordre chronologique les travaux évoqués : la *Photo Annotation Ontology* (Schreiber et al., 2001), la *Visual Descriptor Ontology* – VDO (Simou et al., 2005), l'ontologie des concepts visuels – VCO (Maillot et al., 2005), l'*Object Ontology* (Mezaris et al., 2004) et l'ontologie des relations spatiales (Hudelot et al., 2006). La figure 3 illustre la structure générale de chacune d'entre elles.

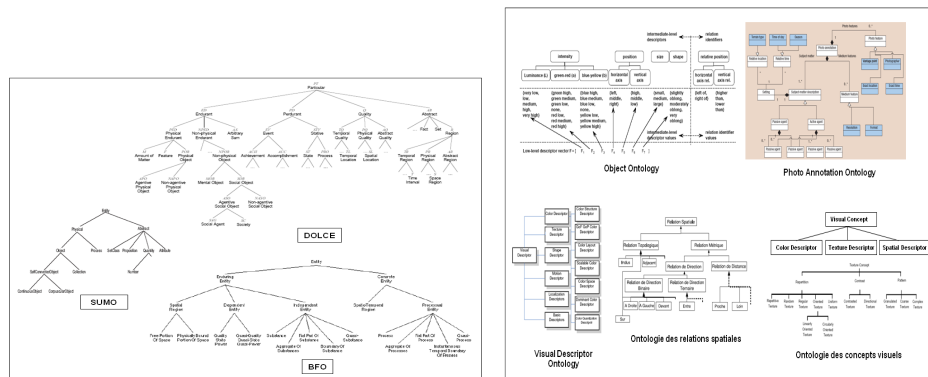


FIG. 3 – Exemples d'ontologies de haut niveau et d'ontologies des descripteurs visuels

4.2.3 Autres ontologies de domaine

De nombreuses ontologies de domaine proposent un modèle du domaine biomédical dans son ensemble¹⁵ et de l'imagerie médicale en particulier. Une ontologie de domaine décrit les concepts d'un domaine général (les pathologies, l'anatomie humaine, ...) ou plus spécifique (l'anatomie du cerveau, le cancer du sein...) et les relations qui les lient. Seules les ontologies réutilisables dans le cadre de notre problématique seront présentées. Nous pouvons mentionner, tout d'abord, l'ontologie de référence de toute l'anatomie humaine FMA¹⁶ (Golbreich et al., 2006) ainsi que l'ontologie de la réalité biomédicale OBR¹⁷.

¹⁵ Open Biomedical Ontologies (<http://obo.sourceforge.net/>)

¹⁶ Foundation Model of Anatomy

¹⁷ Ontology of Biomedical Reality

D'autres travaux plus récents proposent une ontologie des signes pathologiques (Bertaud et al., 2005). Cette ontologie modélise la connaissance relative aux signes élémentaires (ou observables) constituant de signes composés. Les signes composés ont un niveau de croyance et peuvent susciter une recommandation d'action. Dans le domaine du traitement d'image, deux travaux ont retenu notre attention, ceux de l'équipe Visages (Rennes) qui s'orientent vers une ontologie de l'image médicale et du traitement d'images pour la neuro-imagerie (Temal et al., 2006) ainsi que l'ontologie de l'équipe Greyc (Caen) pour la formulation d'applications générales du traitement de l'image (Renouf et al., 2007).

4.3 Approche méthodologique de construction

Lorsqu'une ontologie couvre une grande partie du domaine, et cela est souvent le cas pour le domaine médical, son processus de construction doit être considéré comme un projet à part entière. Dans le cadre de l'ingénierie des ontologies, il n'existe pas de standard exprimé en terme de cycle de vie. Mais certains travaux proposent de multiples méthodologies inspirées du génie logiciel (Pinto et Martins, 2004). Pour une large majorité, ce processus comprend essentiellement cinq étapes : la spécification des besoins, l'acquisition des connaissances, la conceptualisation, la formalisation, l'évaluation et la maintenance. Les premières interrogations qui se sont posées dans ce papier ont porté sur les étapes d'acquisition et de conceptualisation.

4.3.1 Acquisition des connaissances par réutilisation

Pour acquérir la connaissance, plusieurs scénarios sont possibles. Dans notre cas, nous commencerons par :

- nous approprier, dans un premier temps, les ontologies existantes afin de mettre en évidence les hétérogénéités constatées en terme de langage (versions de OWL : OWL-Lite, OWL-Full ou OWL-DL), en terme de nomination des entités, en terme conceptuel (granularité ou niveau de détail, couverture...) et pragmatique (l'engagement ontologique) ;
- puis par sélectionner parmi les ontologies existantes, les ontologies candidates à la réutilisation. Pour cela, certaines techniques pour la mise en œuvre de la réutilisation seront expérimentées telles que l'extraction, l'enrichissement et l'alignement (Euzenat et Shvaiko, 2005) ;
- et enfin par établir des entretiens informels avec les experts du domaine afin de raffiner certaines parties d'ontologies réutilisées et d'analyser le contenu des comptes-rendus textuels existants. On pourra alors amorcer une première représentation des connaissances en exploitant les méthodes appropriées d'extraction à partir de textes (Aussenac-Gilles, 2003).

4.3.2 Conceptualisation modulaire

Une fois la connaissance acquise l'étape de conceptualisation est mise en œuvre. Les concepts et leurs relations sémantiques seront progressivement classifiés selon une approche middle-out (combinée ou centrifuge). Cette approche permet l'émergence de thématique et améliore la modularité. Les concepts centraux seront repérés puis généraliser et spécialiser selon une arborescence.

Modularisation. Plusieurs aspects devront être pris en considération notamment le nombre, la couverture, le couplage et la taille de chaque module (Cuenca et al., 2006). Pour garantir une meilleure réutilisabilité, le couplage inter-modules doit être minimisé c'est-à-dire que les modules de l'ontologie doivent être aussi autonomes que possible. Les couplages représentés dans la figure 4 sont préconisés : anatomie/descripteur visuel, technique/descripteur visuel, anatomie/pathologie, pathologie/recommandation, anatomie/recommandation. La question de la couverture se pose également. Comment peut-on minimiser les redondances possibles entre les modules ? Quant au partitionnement de l'ontologie, il doit conduire à un nombre réduit de modules de taille raisonnable afin d'optimiser leur réutilisation. Dans notre cas, six modules sont proposés pour refléter les différents points de vue d'indexation.

Niveau d'ancrage. Evoqué à la section 4.1, ce niveau devra permettre l'ancrage des différents modules de notre ontologie à l'ontologie de haut niveau. D'après la figure 5, il ne s'agit pas d'une ontologie mais d'un ensemble de concepts moins génériques que ceux de l'ontologie de haut niveau mais plus génériques que ceux des modules de notre ontologie. Se pose alors la question suivante : est-ce que le nombre de concepts d'ancrage sera supérieur, inférieur ou identique à celui des modules ?

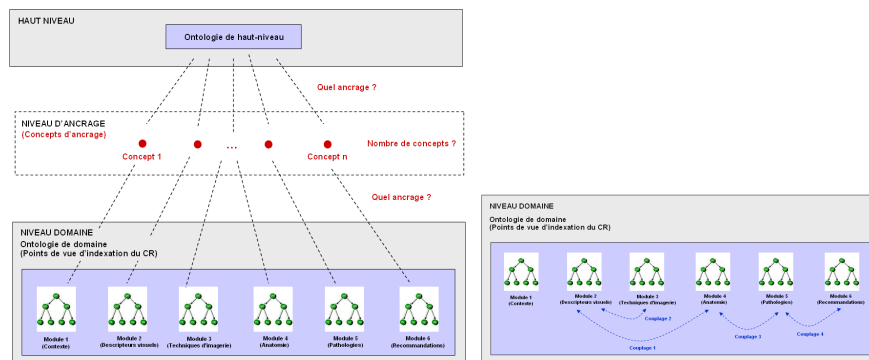


Fig. 4 – Concepts d'ancrage et couplage inter-modules

4.4 Premières expérimentations

Nos premières expérimentations, en matière d'acquisition des connaissances, ont porté sur la récupération de certaines ontologies existantes. Nous disposons actuellement des versions OWL de l'ontologie FMA, de la *Visual Descriptor Ontology* (VDO) et de l'ontologie des signes. Dans une deuxième phase, nous avons édité les ontologies à l'aide de l'outil Protégé¹⁸, environnement graphique de développement d'ontologies développé à Stanford. Basé sur une architecture ouverte, Protégé peut être complété par de nouvelles fonctionnalités grâce à l'ajout de nouveaux composants (*plug-ins*) dont certains sont développés par les utilisateurs. Actuellement, 69 *plug-ins* sont disponibles dont l'outil

¹⁸ <http://protege.stanford.edu/>

Ontologie modulaire multi-niveaux pour l'indexation du compte-rendu d'imagerie médicale

d'extraction Prompt (Noy et al., 2003) testé sur l'ontologie des signes. La figure 5 montre l'ontologie des signes éditée à l'aide de Protégé.

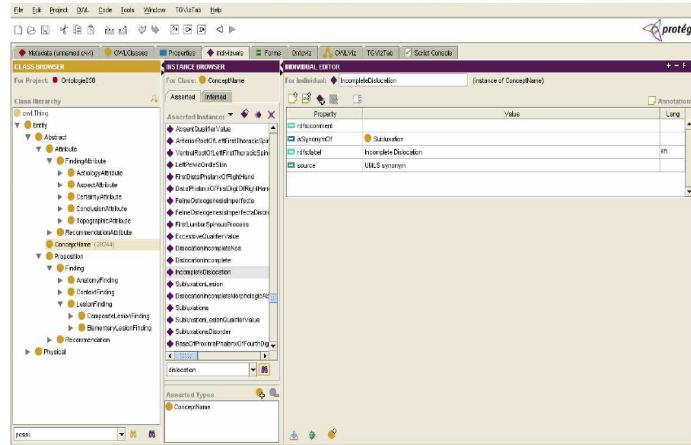


FIG. 5 – L'éditeur Protégé et l'ontologie des signes

5 Perspectives de recherche

Ce papier est une contribution à la mise en place future d'une ontologie modulaire multi-niveaux au sein d'un système d'indexation et de recherche par le contenu sémantique de comptes rendus patients standardisés. Les modalités de l'imagerie médicale ainsi que les pathologies mises en jeu sont nombreuses mais à l'heure actuelle, nous envisageons de réaliser nos expérimentations en imagerie ostéo-articulaire. En matière d'ingénierie des ontologies, une approche de construction mixte (modularité et réutilisation) fondée sur un niveau d'ancrage est proposée et requiert bien évidemment une certaine consolidation. Nos premières perspectives de recherche seront tout d'abord orientées vers l'appropriation des ontologies existantes afin de proposer, en premier lieu, des fiches descriptives et dans une deuxième phase, une étude comparative pour chaque classe d'ontologies.

Références

- Alexakos, C., K. Votis, B. Vassiliadis, et S. Likothanassis (2006). *An Ontology-based Integration Mechanism for Web Searching in Distributed Digital Content Repositories*, Knowledge Systems and Coalition Operations (KSCO). IEEE Workshop on Distributed Intelligent Systems (DIS).
- Aussenac-Gilles, N., B. Biebow et S. Szulman (2003). *D'une méthode à un guide pratique de modélisation des connaissances à partir de textes*. In Actes des 5es journées Terminologie et Intelligence Artificielle. 41-53.

- Benslimane, D., E. Leclercq, M. Savonnet, M.N. Terrasse, et K. Yetongnon (2000). *On the Definition of Generic Multi-Layered Ontologies for Urban Applications*. Computers, Environment, and Urban Systems (CEUS). 24:191-214.
- Bertaud, V., J. Lasbleiz, F. Mougin, F. Marin, A. Burgun, et R. Duvauferrier (2005). *Toward a Unified Representation of Findings in Clinical Radiology*. Stud Health Technol Inform. 116:671-6.
- Boucher, A. et T-L. Le (2005). *Comment extraire la sémantique d'une image ?* 3rd International Conference: Sciences of Electronic, Technologies of Information and Telecommunications (SETIT).
- Euzenat, J. et P. Shvaiko (2005). *Tutorial on Schema and Ontology Matching*. Knowledge Web.
- Golbreich, C., S. Zhang, et O. Bodenreider (2006). *The Foundational Model of Anatomy in OWL: Experience and Perspectives*. Journal of Web Semantics. 4:181-195.
- Grau, B. C., B. Parsia, E. Sirin, et A. Kalyanpur. *Modularity and Web Ontologies*. In Proc. 10th International Conference on Principles of Knowledge Representation and Reasoning (KR). AAAI Press. 198-209.
- Grenon, P., B. Smith, et L. Goldberg (2004). *Biodynamic Ontology: Applying BFO in the Biomedical Domain*. In P.M. Pisannelli Editor. Ontologies in Medicine: Studies in Health technology and Informatics. IOS Press. Volume 102:20-38.
- Hudelot, C., N. Maillot, et M. Thonnat (2005). *Symbol Grounding for Semantic Image Interpretation: From Image Data to Semantics*. 10th IEEE International Conference on Computer Vision.
- Hudelot, C., J. Atif, et I. Bloch (2006). *Ontologie de relations spatiales floues pour l'interprétation d'images*. In Rencontres francophones sur la Logique Floue et ses Applications (LFA). 363-370.
- Klein, M. (2001). *Combining and relating ontologies: An analysis of problems and solutions*. In IJCAI-2001 Workshop on Ontologies and Information Sharing. 53-62.
- Kompatsiaris, I., V. Mezaris, et M. G. Strintzis (2004). *Multimedia Content Indexing and Retrieval using an Object Ontology*. Multimedia Content and Semantic Web - Methods, Standards and Tools. Editor G.Stamou. Wiley. New York.
- Liu, Y., D. Zhang, G. Lu, et W-Y. Ma. (2007). *A Survey of Content-Based Image Retrieval with High-Level Semantics*. Pattern Recogn., 40(1):262-282.
- Masolo, C., S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, et L. Schneider (2003). *The Wonderweb Library of Foundational Ontologies and the Dolce Ontology*. Wonderweb Deliverable D18. Final report Vr. 1.0.
- Mhiri, S. et S. Despres (2007). *Ontology Usability via a Visualization Tool for the Semantic Indexing of Medical Reports (DICOM SR)*. 3rd Symposium USAB 07-HCI&UE for Medicine and Health Care. LNCS. 4799: 412-417.
- Noy, N. F. et M. A. Musen (2003). *The PROMPT Suite: Interactive Tools for Ontology Merging and Mapping*. International Journal of Human-Computer Studies.

Ontologie modulaire multi-niveaux pour l'indexation du compte-rendu d'imagerie médicale

- Pinto, H.S. et J.P. Martins (2004). *Ontologies : How can they be built ?* In Knowledge Information Systems. 6(4):441-464.
- Renouf, A., R. Clouard, et M. Revenu (2007). *A Platform Dedicated to Knowledge Engineering for the Development of Image Processing Applications*. ICEIS. Vol. AIDSS, 271-276.
- Ribeiro, R., F. Batista, J. P. Pardal, N. J. Mamede, et H. S. Pinto (2006). *Cooking an Ontology*. AIMSA. 213-221.
- Rosse, C., A. Kumar, J. L. V. Mejino, D. L. Cook, L. T. Detwiler, et S. Barry (2005). *A Strategy for Improving and Integrating Biomedical Ontologies*. In Proc. American Medical Informatics Association Fall Symposium (AMIA).
- Schreiber, A. T., B. Dubbeldam, J. Wielemaker, et B. J. Wielinga (2001). *Ontology-Based Photo Annotation*. IEEE Intelligent Systems. 66-74.
- Simou, N., V. Tzouvaras, Y. Avrithis, G. Stamou, et S. Kollias (2005). *A Visual Descriptor Ontology for Multimedia Reasoning*. In Proc. of Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS). 13-15.
- Temal, L., P. Lando, B. Gibaud, M. Dojat, G. Kassel, et A. Lapujade (2006). *OntoNeuro-Base: A Multi-Layered Application Ontology in Neuroimaging*. In Proceedings of the 2nd Workshop: Formal Ontologies Meet Industry (FOMI). Trento (Italy).
- Thomas, C., W.S. York, et A. Sheth (2006). *Modular Ontology Design Using Canonical Building Blocks in the Biochemistry Domain*. International Conference on Formal Ontology in Information Systems. 9-11.
- Zweigenbaum, P. (1999). *Encoder l'information médicale: des terminologies aux systèmes de représentation des connaissances*. Innovation Stratégique en Information de Santé (ISIS). 27-47.
- Imagerie Médicale : Une Vision Transatlantique* (2006). Dossier Snitem Info Hiver n°165.

Summary

This paper is a contribution to the semantic indexing process. We intend to build an ontology in order to index imagery reports (textual information and images) from different viewpoints. Due to the increasing number of various sources, specialists may establish different descriptions of their reports, objective or subjective ones. For these reasons, we propose firstly a model which represents six indexing viewpoints of a standardized report. Then, according to this model we suggest to build our ontology by using a modular approach and by reusing existing ontologies. To organize the various modules, three levels are planned: an upper level, an anchorage level and a domain level. Existing ontologies are listed in order to determine reuse possibilities and appropriate techniques to be used in a methodological approach called mixed and based on a level of anchorage. First experiments are showed.

Utilisation de structures sémantiques pour la recherche d'images sur Internet

Adrian Popescu*, **, Pierre-Alain Moëllic*
Ioannis Kanellos**

*CEA LIST, 18 route du Panorama, 92260 Fontenay aux Roses
{adrian.popescu, pierre-alain.moellic}@cea.fr
<http://www-list.cea.fr/>

**ENST de Bretagne, Technopôle Brest-Iroise, 29238 Brest
ioannis.kanellos@enst-bretagne.fr
<http://www.enst-bretagne.fr/>

Résumé. La recherche d'images (RIM¹) constitue une partie importante de la recherche d'information sur Internet. Malgré leur succès, les systèmes de RIM ont une série de limitations dont une partie découlent de leur insensibilité à la sémantique. Nous analysons ici l'introduction de structures sémantiques à grande échelle dans la recherche d'images sur Internet. Plus précisément, nous présentons quelques ressources utilisables, ainsi qu'un modèle d'agrégation de l'accès aux images par le texte et des techniques de traitement d'images. Enfin, nous illustrons nos propos avec quelques applications déjà implémentées ou en cours de développement. Une attention particulière est accordée au couplage des solutions proposées avec des usages existants ou potentiels. Les principaux avantages et difficultés liés à l'introduction d'une couche sémantique dans les systèmes de recherche d'images sont systématiquement mis en avant.

1 Introduction

La recherche de documents multimédia sur Internet est un domaine apparenté à la recherche de textes, mais elle se différencie sur un nombre de points importants. Ici, nous abordons le cas de la recherche d'images photographiques et articulons notre analyse autour de deux problèmes principaux. Premièrement, la recherche d'images à grande échelle, comme celle de données textuelles, est basée sur une exploitation du texte brut associé aux images. Cela entraîne une opacité totale des systèmes par rapport à la signification des requêtes et des réponses proposées par l'application. Deuxièmement, le paradigme de recherche multimédia actuel ne rend pas suffisamment compte des différences entre les types de documents traités impliquant une sous-exploitation de certaines techniques spécifiques aux types de données, qui pourraient améliorer le processus. Même si l'objet de la recherche est un ensemble d'images, les techniques de traitement d'images ne sont que très peu utilisées.

¹ RIM est un acronyme des auteurs pour Recherche d'IMages

Le premier des deux problèmes signalés peut se résoudre par l'introduction de structures sémantiques² dans l'architecture d'un système RIM. Les mêmes structures interviennent dans l'articulation de la recherche d'images par mots clé et la recherche basée sur le contenu visuel afin d'améliorer l'efficacité du dernier type d'accès aux données. Les bases de données textuelles structurées peuvent remplir plusieurs rôles dans les systèmes de recherche d'images. Dans ce papier nous illustrons leur utilité pour:

- une reformulation automatique des requêtes
- une présentation structurée des réponses
- une amélioration de l'interactivité des applications
- une meilleure précision des résultats
- une recherche par contenu visuel dans des régions conceptuellement cohérentes du corpus d'images

Les fonctions citées ci-dessus s'obtiennent au prix d'une complexification de l'architecture d'un système RIM. Il est cependant crucial d'intégrer ces structures sémantiques de façon à ne pas compromettre la facilité d'utilisation des applications de recherche. De plus, des ressources textuelles à très large échelle sont nécessaires afin de couvrir une partie significative de l'espace de requêtes dont Jansen et al. (2000) ont noté la grande complexité (requêtes variées en contenu et nombre de termes). Il est aussi nécessaire d'assurer une bonne qualité des connaissances incluses dans les bases de données structurées.

La suite de ce papier est structurée de la façon suivante: nous discutons tout d'abord de quelques travaux apparentés, puis la section 3 est dédiée à la présentation des structures sémantiques exploitées et à l'analyse de la relation entre les concepts inclus et leur représentation picturale. Les deux types d'accès aux données photographiques, par mots clef et par contenu visuel, ainsi qu'une possible agrégation, sont discutés dans la section 4. Avant de conclure, nous introduisons et évaluons quelques applications employant des ressources textuelles pour la RIM.

2 Travaux apparentés

Notre travail s'apparente à plusieurs domaines dont nous donnons ici un bref aperçu. En recherche d'images, Yang et al. (2001) incluent WordNet Miller (1990) dans un système afin de proposer des requêtes proches par rapport au concept courant. Wang et al. (2006) extraient la base taxinomique d'une ontologie des animaux de WordNet et la complètent avec des propriétés importées d'une autre ontologie. Leur structure comprend, à part les relations textuelles, des propriétés de bas niveau (comme la couleur ou la texture) caractérisant les classes dans la hiérarchie conceptuelle. Ruiz-Casado et al. (2007) quant à eux abordent la construction de ressources sémantiques à partir de Wikipedia.

O'Hare et al. (2005) et Ahern et al. (2007) s'intéressent à la recherche d'images possédant des coordonnées géographiques. Wilkins et al. (2005) et Quack et al. (2004) discutent le problème du passage à l'échelle dans les systèmes de recherche d'images par le contenu. Une discussion complète sur la RIM par le contenu augmentée par la sémantique se trouve dans Liu et al. (2007).

² Structure sémantique et ressource sémantique sont des termes désignant plusieurs types de bases de données textuelles structurées, comme: les thésauri, les hiérarchies conceptuelles ou les ontologies formelles.

3 Caractérisation des structures sémantiques

Les ressources linguistiques que l'on peut intégrer dans les systèmes de recherche d'images peuvent être caractérisées selon plusieurs critères. Nous retenons ici leur disponibilité, le degré de structuration et le domaine conceptuel des données incorporées.

3.1 Disponibilité des ressources

Une première approche pour l'introduction de ressources sémantiques et RIM vise à exploiter des données préexistantes et réutilisables (éventuellement après adaptation). Ces ressources assurent une couverture parfois insuffisante et il est nécessaire d'en créer de nouvelles. Etant donnée la diversité de l'espace des requêtes sur le Web, il est nécessaire de construire des structures à très large échelle afin de couvrir une partie significative des requêtes. La création manuelle de ressources étant trop coûteuse, une construction automatique doit être envisagée. Une première solution serait l'exploitation de documents textuels non structurés afin d'en extraire des connaissances. Cette approche s'avère inadaptée car les volumes de données qu'il est possible d'extraire, de qualité raisonnable, sont trop réduits comme l'indiquent Cimiano et al. (2004), Clough et al. (2005).

Alternativement, Ruiz-Casado et al. (2007) proposent l'utilisation de données semi structurées, comme les articles Wikipedia, pour construire des bases de connaissances. Cette approche est plus adaptée pour la RIM sur Internet et nous discutons dans la suite de cet article d'un cas d'application. D'autres bases de données structurées et spécifiques à des domaines existent et pourraient être exploitées, mais elles sont rarement libres de droits. IMDB³, une base de données cinématographique très détaillée, en est un parfait exemple.

3.2 Structuration des données dans les ressources

Les ressources linguistiques ont des degrés d'organisation variable qui influence grandement les possibilités d'exploitation des données. Un maximum de structuration logique est atteint avec les ontologies formelles Guarino (1998). Dans ces structures, il est possible de proposer une représentation compacte des connaissances et de raisonner automatiquement sur leur contenu. Malheureusement, deux problèmes subsistent : premièrement, les algorithmes de raisonnement automatique ne sont pas actuellement adaptés pour gérer des grands volumes de données ; deuxièmement, les contraintes imposées par le formalisme logique soutenant les ontologies sont parfois trop strictes pour modéliser la recherche d'information pour le grand public. Un bon exemple est l'impossibilité de définir un héritage multiple pour un concept. Les thésaurus ont une structure moins stricte que les ontologies formelles mais les possibilités de raisonnements sont plus réduites. La connaissance incluse doit être explicite afin de pouvoir en profiter, aussi, en contrepartie, on observe une plus grande flexibilité pour la définition des relations inter-conceptuelles. L'utilisation de thésaurus nous semble donc plus adaptée à la recherche d'images.

³ <http://imdb.com>

3.3 Données dans les structures sémantiques

L'espace des requêtes en RIM impose l'utilisation de ressources linguistiques comprenant des concepts appartenant à des domaines variés. Les ressources généralistes comme WordNet ou Cyc sont les seuls à remplir cette contrainte. Leur couverture est toutefois limitée et il est nécessaire de les compléter. Parmi les domaines pertinents pour la recherche d'images sur le Web, nous trouvons les noms de personnes célèbres et les entités géographiques Jansen (2000) qui sont insuffisamment renseignées. Cependant, il existe des bases de données utilisables permettant d'étendre WordNet à ces domaines.

3.4 Ressources exploitées

Nous utilisons plusieurs bases de connaissances pour améliorer le fonctionnement des applications de recherche d'images sur le Web que nous passons en revue dans les sous-sections suivantes.

3.4.1 WordNet

WordNet (Miller (1990)) est un réseau lexical construit manuellement par des experts comprenant plus de 100000 concepts nominaux structurés sous forme d'arbre. Les deux relations fondamentales dans WordNet sont la relation d'héritage conceptuel, déterminant la structure en profondeur et la synonymie, permettant de créer les atomes de la hiérarchie : les *synsets*. Parmi les propriétés importantes du réseau lexical, nous citons la séparation des sens pour les termes polysémiques et la possibilité d'inclure un héritage multiple pour un concept. Une adaptation de WordNet est nécessaire pour l'utiliser en RIM et nous décrivons ce processus en détail dans Popescu et Grefenstette (2008). Par exemple, il est nécessaire d'ordonner les concepts selon un critère de pertinence pour la recherche d'images. Nous avons choisi une mesure simple : la fréquence des termes sur Internet. Afin d'éviter les biais dus à la polysémie, les requêtes correspondant à chaque terme ont été reformulées utilisant le parent immédiat du concept.

3.4.2 Geonames

Geonames⁴ est une base de données géographiques comprenant plus de six millions d'entrées structurées sous la forme d'une hiérarchie à trois niveaux. Le plus général contient huit catégories (ex. : *parcs, constructions, montagnes, eaux* etc.). Le niveau intermédiaire en inclut 645 (sous concepts des catégories principales) comme : *monument, plage, ville, palais* ou *lac*. Les instances des 645 concepts intermédiaires forment le niveau le plus détaillé de Geonames (environ 6 millions de concepts). Pour chaque instance, il existe plusieurs informations disponibles : noms alternatifs, coordonnées géographiques, pays d'appartenance etc. L'introduction de l'information sur le placement de chaque entité crée une nouvelle dimension hiérarchique, caractérisée par la relation de méronymie entre les concepts, qu'il est possible d'exploiter. L'ordonnement des concepts dans Geonames est obtenu de manière similaire à celui de WordNet (fréquence du terme). La désambiguïsation est réalisée ici en employant dans la requête le terme et le nom de l'entité géographique englobante.

⁴ <http://geonames.org>

3.4.3 Base de données structurée extraite de Wikipedia

Les études des usages en recherche d'images sur Internet montrent qu'il y a un grand intérêt du public pour les photos de personnes célèbres. Avec moins de 20000 entrées, Word-Net est clairement insuffisant dans ce domaine. Wikipedia constitue une source de données riche que nous avons utilisé pour extraire des informations concernant *les chanteurs* (plus de 6000 instances), *les acteurs* (plus de 13000 instances) et *les footballeurs* (plus de 26000 instances). L'intérêt principal des pages Wikipedia par rapport au texte libre est qu'elles incluent de l'information semi structurée facile à extraire. La description pour chaque type d'entité inclut :

- pour toutes les catégories : date de naissance (et, selon le cas, du décès), lieu de naissance, nationalité, célébrités associées
- *chanteurs* : groupes associés, genres abordés, nom réel, pseudonymes
- *acteurs* : films, rôles, prix obtenus
- *footballeurs* : club(s), poste(s)

De plus, chaque instance est rattachée à une des trois grandes catégories de célébrités (et, dans certains cas à des catégories plus précises comme *chanteur de jazz* ou *défenseur*), la base de données créée a une dimension hiérarchique. Il est ainsi possible de présenter des classes d'images structurées, de manière similaire à celle décrite et 3.4.1.

L'ordonnement des différentes instances de relations est réalisé en utilisant la fréquence jointe d'apparition de deux concepts sur le Web. Pour un acteur, on trouvera les acteurs associés, les rôles et les films les plus connus. Notons que la méthode d'extraction d'information mise en place pour les trois catégories citées est facilement adaptable à d'autres types de célébrités.

3.5 Concepts et représentation visuelle

Une partition possible des catégories donne les grandes classes suivantes : *concepts nominaux*, *concepts naturels* et *artefacts* (Keil (1992)). La plupart des termes de la première classe sont difficiles à représenter directement dans les images. Le lecteur peut vérifier cette affirmation en essayant de s'imaginer quel serait le contenu des classes d'images pour des concepts comme : *vérité*, *association* ou *processus*. Parmi les concepts nominaux, les événements constituent une exception car il est possible d'en trouver des représentations cohérentes (il est facile d'avoir une image mentale d'*un match de football* ou d'*une grève*). Pour une autre partie des *concepts nominaux*, il y a des représentations culturellement encodées (ex. *justice*). La représentation des concepts nominaux dans les images passe par une interprétation. Il serait intéressant d'approcher leur description imagée mais ici on se concentre sur les termes directement représentables dans des images photographiques.

Les concepts naturels et les artefacts sont, pour la plupart, représentables dans des photographies. S'ils ont une sous hiérarchie conceptuelle associée, il est possible de les représenter par l'intermédiaire de leurs sous-catégories. Si l'on considère la relation d'héritage conceptuel, les races comme *berger allemand*, *Norfolk terrier* ou *husky* forment une partie de la classe *chien*. La relation de méronymie est plus adaptée pour décrire d'autres concepts. Par exemple, *la France* peut être visualisée à travers des images *des villes* ou *des parcs naturels* contenus. A leur tour, *les villes* sont bien décrites par des concepts plus spécialisés (ex. *monuments*, images de *rues*). De même, les endroits les plus marquants représentent bien *les*

parcs naturels. Les ressources sémantiques décrites en section 3.4 incluent, selon le cas, des relations type – sous-type, de méronymie (ou autres) entre les concepts inclus.

3.6 Limites

L'introduction des ressources linguistiques en RIM n'est pas immédiate. Parmi les difficultés majeures rencontrées nous citons:

- la couverture conceptuelle des bases de connaissances : détermine la couverture de l'espace de requêtes. Premièrement, pour les requêtes portant sur un seul concept (ex. *Nice*), ce terme doit être renseigné dans la ressource. Deuxièmement, nous n'avons pas abordé le cas des requêtes contenant plusieurs termes (ex. *Nice EGC 2008*), majoritaires sur le Web (Jansen et al. (2000)). La définition d'une règle (ou de règles) de décomposition de ce type de requêtes utilisant l'héritage conceptuel est difficile.
- la qualité des ensembles d'images attachées aux concepts précis dans les hiérarchies conceptuelles. Nous avons montré dans Popescu et al. (2007b) que le bruit dans les réponses images est réduit quand on lance des requêtes avec *berger allemand* et *setter irlandais* pour définir la classe *chien* par rapport à une requête directe avec ce terme. Toutefois, il reste un certain nombre d'images non représentatives même parmi les images rattachées aux deux races de chiens citées.
- la qualité des connaissances incluses dans les ressources. Le principal désavantage de la constitution automatique de structures sémantiques est l'imprécision d'un nombre important des relations inter-conceptuelles découvertes (Ruiz-Casado et al. (2007)). Il est par exemple possible de trouver que *chien berger* hérite de *berger allemand* et non pas l'inverse. Un autre désavantage est l'incomplétude de la description des termes. Dans Wikipedia, les pages de certains acteurs sont plus complètes que d'autres. Et, habituellement, on obtient le plus d'informations pour les acteurs les plus connus.
- la présentation des requêtes proches les plus adaptées. A ce stade de nos travaux nous n'avons pas encore abordé la notion de personnalisation des résultats. L'adaptation des résultats de la recherche compte tenu des centres d'intérêt de l'utilisateur est importante pour améliorer la qualité des résultats. Nous illustrons cela par deux exemples : un utilisateur passionné par les *chiens bergers* formule une requête avec *chien*. Dans ce cas on devrait favoriser la présentation des sous-types intéressants pour l'utilisateur même s'ils sont moins fréquents que d'autres. Un utilisateur italien demande de voir des images de *Zinedine Zidane* : il serait intéressant de lui proposer, en priorité parmi les requêtes proches, des noms de joueurs italiens associés à *Zidane*.

Malgré ces limites, l'utilisation des bases de connaissances présente un certain nombre d'avantages. De plus, une partie des limites ne sont pas insurmontables mais plutôt un effet de l'état actuel de notre travail. Nous pensons notamment à la personnalisation des résultats et au traitement des requêtes complexes. Dans la section 4 on regarde de plus près l'articulation entre les modalités d'accès aux images du Web et les structures sémantiques utilisées.

4 Types d'accès aux images sur Internet

Vu l'énorme volume de données sur le Web, il est nécessaire de proposer des modalités de sélection rapide de l'information intéressante pour l'utilisateur. Il existe deux principaux types d'accès aux données photographique : la recherche par mots clef et la recherche par le contenu visuel. Deux dimensions principales les différencient :

- la formulation de la requête: elle consiste respectivement en une chaîne de caractères et en une image en recherche par le contenu
- les réponses sont obtenues exploitant des descriptions différentes des données : les index textuels et les propriétés perceptuelles des images

Les moteurs de RIM existants implémentent la recherche par mots clef, tandis que la recherche par le contenu reste l'apanage des travaux de recherche et des prototypes issus de ces travaux. Dans cette section, nous présentons les deux types d'accès, ainsi qu'une possible agrégation. Nous présentons, dans les deux cas, l'utilité des ressources sémantiques.

4.1 Recherche par mots clef

Ce processus peut être décrit de la façon suivante: un utilisateur formule une requête composée d'un ou plusieurs termes et le système lui présente les images considérées les plus proches par rapport aux termes introduits. La liaison entre l'espace des requêtes et celui des réponses se réalise en exploitant la similarité entre les chaînes de caractères introduites par les utilisateurs et les termes associés aux images dans les index. Les systèmes actuels sont opaques par rapport à la signification de la requête (aucune analyse de son contenu n'est réalisée). Avec l'exploitation des ressources sémantiques, il devient possible de traiter la requête à un niveau plus symbolique et de rendre des résultats améliorés. Nous discutons ici quelques avantages de l'introduction des structures sémantiques en RIM, ainsi qu'une série de limites de cette approche.

4.1.1 Détection de l'ambiguïté

Une bonne partie des termes de la langue sont polysémiques et, par conséquent, les images associées à ces concepts sont différentes. Spärck-Jones (2007) souligne que les systèmes actuels gèrent mal l'ambiguïté des requêtes et propose des moyens pour un meilleur traitement de demandes d'information polysémiques. L'introduction d'une couche sensible à la polysémie, rend possible la présentation des sens d'un terme dans des classes différentes de réponses.

4.1.2 Reformulation automatique des requêtes

Rosch et al. (1976) distinguent trois niveaux de généralité pour les catégories linguistiques. WordNet recense jusqu'à 20 niveaux conceptuels entre les catégories les plus générales et les plus spécialisées. Geonames propose une structure à trois niveaux. L'utilisation des structures hiérarchiques permet la représentation des concepts sur les niveaux intermédiaires de la hiérarchie utilisant leurs sous-catégories. Par exemple quand un utilisateur demande des images pour *chien*, cette requête sera remplacée ou complétée automatiquement par ses sous-types (ex. *chien+berger allemand* ; *chien+setter irlandais*). Pour les requêtes localisées, l'utilisation de la reformulation permet de distinguer les différents noms géographiques ho-

monymes. Si l'utilisateur veut voir des images de *Montréal*, et plus précisément de l'église *Notre Dame* située dans cette ville, la reformulation automatique de la requête *Notre Dame* en *Notre Dame+Montréal* évitera la présentation d'images d'églises du même nom d'autres villes. Par expansion des requêtes, on élimine une partie des résultats non représentatifs et, en même temps, on propose une désambiguïsation des différents sens pour les termes polysémiques. De plus, les ensembles d'images sont conceptuellement et visuellement cohérents pour les concepts spécialisés dans les hiérarchies, permettant une agrégation efficace entre la recherche par mots-clé et celle par le contenu.

4.1.3 Structuration des résultats

Les moteurs de RIM existants proposent les résultats sous la forme d'une liste ordonnée par des critères d'adéquation des réponses à la requête et de popularité des images. La description des concepts par leurs sous-catégories facilite une présentation structurée des réponses, préservant en même temps les avantages offerts par les méthodes de classement actuelles.

4.1.4 Précision des réponses

Wang et al. (2006), Popescu et al. (2007b) montrent que la précision des résultats pour une requête dans Google Images se situe autour de 60% - 70%. Les stratégies d'expansion de requêtes proposées dans les deux papiers permettent une amélioration sensible de la précision dans les pages de réponses. Wang et al. (2006) combine les propriétés conceptuelles et les caractéristiques visuelles des objets afin d'obtenir l'amélioration, tandis que dans Popescu et al. (2007), seul l'héritage conceptuel est utilisé.

4.2 Recherche par contenu visuel

L'accès aux données à travers des images exemples est une technique introduite dans beaucoup de travaux de recherche (Liu et al. (2007)), mais qui tarde à s'imposer dans les systèmes de recherche photographique sur Internet comme une alternative ou un complément à l'interrogation par mots-clé. Cette situation est causée principalement par l'inadéquation des critères de similarité mis en oeuvre dans l'interrogation par contenu visuel et la similarité entre les images telle qu'elle est perçue par les humains. Le premier type de similarité se base sur des caractéristiques perceptuelles de l'image (couleur, texture), tandis que le deuxième est principalement conceptuel. Les deux types de similarité ne sont pas de la même nature et il serait difficile de les mettre sur le même plan. Nous proposons une approche simple, faisant de la recherche par contenu visuel un complément de l'interrogation par mots-clé. L'accès aux données par images exemples est proposé seulement dans des régions de la base de données ayant une cohérence conceptuelle (images qui ont en commun les mots clé de la requête initiale). De plus, avec la reformulation automatique, l'utilisateur est guidé vers des régions du corpus photographique définies par des concepts spécialisés (ayant une certaine cohérence visuelle). On contourne également un problème non négligeable de la recherche par le contenu visuel: la réduction de l'espace de recherche dans des bases de données à large échelle (Wilkins et al. (2005)). Dans le papier cité, cette réduction se basait sur une analyse des images, tandis qu'ici on tient compte de l'information textuelle déjà introduite par l'utilisateur pour limiter la région potentiellement intéressante.

L'agrégation entre les deux types d'interrogation présentée ici repose sur une supposition: une fois sélectionnée une image illustrant une certaine requête, l'utilisateur ne voudra pas voir des photographies avec un contenu conceptuellement différent. Si on pose comme requête une image de *Renault Twingo*, il est évident qu'on n'attend pas des images de *chevaux* ou de *couteaux*. Cela arrive souvent dans les systèmes qui exploitent principalement les caractéristiques visuelles⁵.

Nous avons présenté les modalités d'accès aux images, ainsi qu'une agrégation possible. Dans la section suivante, nous discutons quelques exemples de systèmes de recherche d'images reposant sur ce type d'approche.

5 Applications

Nous présentons ici quatre exemples d'applications, achevées ou en cours de développement, qui combinent l'utilisation de ressource sémantiques et la RIM par le contenu. Pour les systèmes fonctionnels, des évaluations et des références à des publications pertinentes sont proposés.

5.1 Recherche dans une base d'images structurée

Nous avons sélectionné une partie de WordNet (la sous hiérarchie des *animaux* contenant environ 1000 noeuds), et formé des requêtes avec les feuilles de cette hiérarchie et récupéré un maximum de 100 images d'Internet pour chaque concept feuille. Ensuite, la base a été filtrée pour éliminer les dessins et les images contenant des visages, et indexée en utilisant un descripteur de texture et de couleur. L'utilisateur peut introduire un mot-clef du domaine retenu et, s'il s'agit d'un concept ayant des héritiers, des images de ces derniers sont proposées. L'utilisateur peut cliquer sur les images pour réaliser une recherche par le contenu. On présente d'abord les images les plus proches dans la même classe feuille, mais il est possible de choisir des régions conceptuelles plus larges pour l'interrogation par le contenu. Une évaluation sur 30 concepts montre que la précision décroît quand on recherche des images similaires dans des régions de plus en plus larges. Cette application illustre bien la différence de performances entre une recherche par le contenu guidée par la sémantique des concepts et une interrogation exclusivement basée sur le contenu visuel. Une vidéo⁶ illustrant le comportement du système est également disponible. Pour plus de détails, nous renvoyons le lecteur à Popescu et al (2007).

5.2 Recherche conceptuelle d'images sur Internet

Inspirés par l'application de 5.1, nous avons utilisé l'intégralité de WordNet pour proposer un accès aux images d'Internet basé sur l'utilisation de la hiérarchie conceptuelle. Si on compare l'application, nommée Olive, à Google Image, on ajoute : une désambiguïsation des requêtes polysémiques; une reformulation automatique et une présentation de requêtes proches utilisant les connaissances dans WordNet. WordNet est exploité de la même manière que dans 5.1. Des différences notables sont toutefois à souligner : le nombre de requêtes trai-

⁵ Voir par exemple Alipr – <http://alipr.com>

⁶ <http://moromete.net/semretriev.wmv>

tées par les systèmes, 1000 dans 5.1, plus de 100000 ici et le fait que, pour Olive, la base de données image se remplit au fur et à mesure de l'utilisation du système. Il n'est donc plus possible de proposer un accès par le contenu pour tout le corpus, mais seulement aux niveaux des feuilles.

Des évaluations sur 40 concepts comparant notre système à Google Images montrent qu'on obtient une amélioration de la précision d'environ 15%, passant de 70% à 85%. En ce qui concerne la recherche par le contenu, nous avons comparé ce système à Cortina (Quack et al. (2004)), l'application de RIM par contenu visant le plus large volume de données (11 millions d'images) . 80 images représentant des concepts spécialisés ont été utilisées et la précision moyenne des résultats passe de 10% dans Cortina à 35% dans Olive. Un deuxième paramètre important qui est amélioré est le nombre de requêtes images pour lesquelles l'utilisateur considère qu'au moins une réponse est similaire à la question : moins de la moitié des requêtes pour Cortina, 95% des cas pour Olive. De plus, un test utilisateur réalisé sur un échantillon de 10 personnes confirme l'utilité de la proposition des requêtes proches et la préférence pour la présentation des réponses structurées conceptuellement contre la présentation de Google. Une vidéo⁷ illustrant le comportement du système est disponible. Pour plus de détails, nous renvoyons le lecteur à Popescu et Grefenstette (2008).

5.3 Recherche d'entités géographiques

Avec l'introduction de systèmes de géolocalisation (GPS) directement intégrés dans les appareils photo numériques, le volume d'images géo localisées devient important et il est intéressant d'exploiter l'information de localisation dans des applications de recherche d'images. Il y a un nombre croissant de travaux de recherche s'intéressant aux images géo référencées (Ahern et al. 2007) mais les résultats sont encore loin de témoigner d'une maturité des approches. Nous avons combiné Geonames, un service de visualisation de cartes de Yahoo!⁸, et la recherche par le contenu visuel pour construire ThemExplorer, un système de recherche d'images géographiques thématique.

Par rapport à l'application de référence (Ahern et al (2007)), il est notamment proposé d'explorer la carte choisissant un ou plusieurs sujets (on peut par exemple demander de voir que *les églises*, *les montagnes* ou *les églises et les monuments* dans une certaine région de la planète). De plus, une fois sélectionnée une entité précise, on propose une exploration de ces images utilisant le contenu visuel. L'application est encore en développement et il n'est pas possible de présenter des évaluations de ses performances. Toutefois, une vidéo de présentation⁹ est disponible.

5.4 Recherche de célébrités

La base de données structurée extraite de Wikipedia ciblant les noms de célébrités (section 3.4.3) est utilisable en recherche d'images sur Internet. Une présentation adaptée à chaque type de personne est rendue possible par l'exploitation des informations dans la ressource constituée au préalable. Ainsi, pour un acteur, il est possible de présenter des images relati-

⁷ <http://moromete.net/olive.wmv>

⁸ <http://tagmaps.research.yahoo.com/>

⁹ <http://moromete.net/themexplorer.avi>

ves à ses rôles les plus connus ou, pour un footballeur, des photos dans les clubs pour lesquels il a évolué. Une proposition de requêtes proches (noms de personnes, clubs, films, groupes musicaux) est aussi possible par l'utilisation de la ressource sémantique. Par rapport aux systèmes actuels, comme Ask¹⁰, on obtient ici une meilleure contextualisation et structuration des requêtes.

6 Conclusions

Notre travail s'intéresse à l'introduction de structures sémantiques et de techniques de traitement d'images dans la recherche d'images sur Internet. Cette approche vise à donner plus d'importance à l'utilisateur, en lui proposant une interaction enrichie, mais intuitive avec le système de RIM. L'exploitation de bases de connaissances aide l'utilisateur à s'orienter dans un espace conceptuel qu'il expliciterait difficilement autrement. L'agrégation de la recherche par mots-clef et de celle par le contenu proposée dans ce papier allie deux types de similarité, celle propre aux humains (conceptuelle) et celle propre aux machines (perceptuelle), dans le but de proposer une nouvelle dimension de navigation pour les requêtes ayant un grand nombre de réponses. En même temps, cette méthode simple d'agrégation contourne le problème du passage à l'échelle propre aux applications de recherche par le contenu.

Les résultats présentés dans ce papier nous encouragent de continuer dans la même ligne dans le futur. Un premier axe de travail est constitué par l'implémentation et l'évaluation des applications actuellement en développement. Deuxièmement, nous allons continuer à exploiter le contenu de Wikipedia pour d'autres domaines conceptuels, comme l'extraction de bases de données géographiques. Troisièmement, il est nécessaire d'analyser plus finement la relation entre la recherche par le contenu visuel et la recherche par mots-clef. Cela se traduit par une adaptation de la description de bas niveau de l'image au type de concept demandé. Par exemple, pour les objets manufacturés, l'utilisation de descripteurs locaux (comme les points d'intérêt) peut s'avérer utile.

Références

- Ahern, S., M. Naaman, R. Nair, J. Yang (2007) World Explorer: Visualizing Aggregate Data from Unstructured Text in Geo-Referenced Collections. In Proc of. JCDL 2007.
- Cimiano, P., A. Pivk, L. Schmidt-Thieme, and S. Staab, S. (2004) Learning taxonomic relations from heterogeneous evidence. In Proc. of ECAI 2004, OLP Workshop.
- Clough, P., H. Joho, and M. Sanderson (2005), *Automatically Organising Images using Concept Hierarchies*, In Proc of ACM SIGIR -MIR Workshop.
- Guarino, N. (1998) *Formal Ontologies in Information Systems*. IOS Press.
- Jansen, J., A. Goodrum and A. Spink (2000) *Searching for multimedia: analysis of audio, video and image Web queries*. In World Wide Web Journal 3(4).
- Keil F. C. (1992) *Concepts, Kinds, and Conceptual Development*, Bradford Books, 1992.

¹⁰ <http://ask.com>

- Liu, Y., D. Zhang, G. Lu and W.-Y. Ma (2007) *A survey of content-based image retrieval with high-level semantics*, Pattern Recognition 40(1).
- Miller, G. A. (1990) *WordNet: An on-line lexical database*. Int. Journal of Lexicography 3, 4 (Winter 1990), 235-312.
- O'Hare, N., C. Gurrin, A. F. Smeaton, G.J.F Jones (2005) *Combination of content analysis and context features for digital photograph retrieval*. In Proc. of EWIMT 2005.
- Popescu, A., C. Millet, and P.-A. Moëllic (2007a) *Ontology Driven Content Based Image Retrieval*, In Proc. of ACM CIVR 2007.
- Popescu, A., G. Grefenstette, P.-A. Moëllic (2007b) *Improving Image Retrieval Using Semantic Resources*, Springer SCI series.
- Popescu, A., G. Grefenstette (2008) *Concept-based Internet Image Retrieval*, Article soumis pour LREC 2008.
- Quack, T., U. Monich, L. Thiele, and B. S. Manjunath (2004) *Cortina: A System for Large-scale, Content-based Web Image Retrieval*. In Proc. of ACM Multimedia 2004.
- Rosch, E., C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem (1976) *Basic objects in natural categories*, Cognitive Psychology, 8.
- Ruiz-Casado, M. , E. Alfonseca and P. Castells (2007) *Automatising the learning of lexical patterns: An application to the enrichment of WordNet by extracting semantic relationships from Wikipedia*, Data and Knowledge Engineering, 61(3).
- Spärck-Jones, K., S. E. Robertson, M. Sanderson (2007) *Ambiguous Requests: Implications for Retrieval Tests, Systems and Theories*, ACM SIGIR Forum, December 2007.
- Wang, H., Liu, S. and Chia L.T. (2006) *Does ontology help in image retrieval?: a comparison between keyword, text ontology and multi-modality ontology approaches*, In Proc. of ACM Multimedia.
- Wilkins, P., P. Ferguson, A.F. Smeaton, C. Gurrin (2005) *Text Based Approaches for Content-Based Image Retrieval on Large Image Collections*. In Proc. of EWIMT.
- Yang, J., Wenyin, L., Zhang, H., Zhuang, Y. (2001) *Thesaurus-aided Approach for Image Browsing and Retrieval*, In Proc. of IEEE – ICME.

Extraction de détecteurs d'objets urbains à partir d'une ontologie

Sébastien Derivaux, Germain Forestier,
Cédric Wemmert et Sébastien Lefèvre

LSIIT, UMR 7005 CNRS - Université Louis Pasteur
Boulevard Brant, BP 10413, 67412 Illkirch Cedex
{derivaux,forestier,wemmert,lefevre}@lsiit.u-strasbg.fr

Résumé. Afin de parvenir à une méthode d'interprétation automatique d'images de télédétection à très haute résolution spatiale, il est nécessaire d'exploiter autant que possible les connaissances du domaine. Pour détecter différents types d'objet comme la route ou le bâti, des méthodes très spécifiques ont été développées pour obtenir de très bons résultats. Ces méthodes utilisent des connaissances du domaine sans les formaliser. Dans cet article, nous proposons tout d'abord de modéliser la connaissance du domaine de manière explicite au sein d'une ontologie. Ensuite, nous introduisons un algorithme pour construire des détecteurs spécifiques utilisant les connaissances de cette ontologie. La séparation nette entre modélisation des connaissances et construction des détecteurs rend plus lisible le processus d'interprétation. Ce découplage permet également d'utiliser l'algorithme de construction de détecteurs dans un autre domaine d'application, ou de modifier l'algorithme de construction de détecteurs sans modifier l'ontologie.

1 Introduction

L'interprétation automatique d'images de télédétection à très haute résolution (résolution spatiale de l'ordre du mètre) est un domaine de recherche très actif ces dernières années. La volumétrie et la complexité des données rend en effet coûteux l'interprétation manuelle. Ces dernières années, deux approches se sont imposées. La première, nommée interprétation basée objets (Yongxue et al., 2006), effectue en premier lieu une segmentation de l'image afin de créer des régions. Ces régions sont ensuite classées en utilisant des caractéristiques de régions (i.e. taille, forme) plus discriminantes. La seconde approche est de concevoir des détecteurs indépendants pour chaque type d'objets. Dans cet article, nous nous intéressons au second paradigme.

De nombreuses méthodes spécifiques existent pour détecter des objets particuliers dans une image tels que la route (Zhao et al., 2002; Peteri et al., 2003; Yager et Sowmya, 2004) ou le bâti (Jin et Davis, 2005; Lefèvre et al., 2007). Ces méthodes intègrent les connaissances implicites qu'ont leurs concepteurs sur les types d'objets recherchés. Ces connaissances sont ensuite difficiles à décrire si l'on souhaite améliorer ou adapter l'algorithme.

Afin de modéliser ces connaissances implicites, l'utilisation d'ontologie devient de plus en plus courante. Une ontologie (Gruber, 1995) est une spécification abstraite, une vue simplifiée du monde représentée dans un but précis. Une ontologie définit un ensemble de concepts, leurs caractéristiques et les relations entre ces concepts. L'utilisation d'ontologies est devenue de plus en plus courante dans les systèmes d'informations géographiques (Fonseca et al., 2002). De plus, un intérêt croissant a été porté aux ontologies spécialisées pour l'analyse d'image (Bittner et Winter, 1999) depuis ces dix dernières années. La plupart des méthodes (Mezaris et al., 2004; Maillot et Thonnat, 2008; Panagi et al., 2006; Durand et al., 2007) formalisent les concepts pouvant être présents dans une image puis proposent une analyse sémantique de celle-ci en cherchant à identifier des représentants de ces concepts dans l'image.

L'objectif de cet article est de proposer une méthode d'extraction de détecteurs pour des objets spécifiques à partir d'une ontologie. L'originalité de cette approche est d'utiliser le *design pattern* de séparation des préoccupations (Hürsch et Lopes, 1995). Alors que dans les solutions existantes la connaissance est mêlée à l'algorithmique, notre approche permet de séparer la partie connaissance de la partie algorithmique. L'objectif ici est de permettre à l'expert du domaine de pouvoir facilement exprimer ses connaissances sans être un expert en traitement d'images ou en extraction de connaissances.

De plus, afin de faciliter le transfert de la connaissance de l'expert vers l'ontologie et d'assurer la population de l'ontologie, nous présentons également une méthode pour enrichir l'ontologie à partir d'exemples.

Le plan de cet article est le suivant. Dans la première section, nous étudierons le modèle de l'ontologie utilisé. Ensuite, la méthodologie d'enrichissement de l'ontologie à partir d'exemples sera présentée suivie de l'algorithme de construction de détecteurs. Enfin, la méthode proposée sera évaluée avant de conclure et de proposer des pistes de recherches.

2 Ontologie

Une ontologie permet de regrouper des connaissances relatives à un domaine donné. Dans le cadre de cet article, nous étudions le domaine des images et plus particulièrement des images de télédétection à très haute résolution spatiale. L'ontologie utilisée est composée d'une hiérarchie de concepts potentiellement présents (i.e. visibles) dans ces images et ayant un sens pour l'expert du domaine. Elle est stockée sous forme d'un ensemble de fichiers texte extensible afin de permettre à l'expert de rajouter ses propres informations. Chaque concept peut avoir des nœuds de niveau supérieur et inférieur. La hiérarchie utilisée dans cet article est illustrée par la figure 1. Un premier niveau de séparation est effectué entre les objets de type minéral et non minéral. Même si cette distinction n'a pas d'influence sur la description des concepts, elle permet de fournir une sémantique utile à l'expert du domaine.

Une fois la taxonomie établie, il convient de renseigner chaque concept par des contraintes, c'est-à-dire par des propriétés qui permettent d'identifier une région d'une image comme étant une instance d'un concept. Ces contraintes sont pour la plupart modélisées sous forme d'un intervalle de valeurs admises pour une propriété de niveau pixel (réponse spectrale) ou niveau région (surface). Comme cela est souvent le cas en gestion des connaissances, les entretiens effectués avec des géomaticiens experts du domaine montrent que la formalisation de telles connaissances n'est pas aisée. Le fossé sémantique qui réside entre les descripteurs de haut niveau (abstrait) et de bas niveau (réalité) implique des difficultés de modélisation.

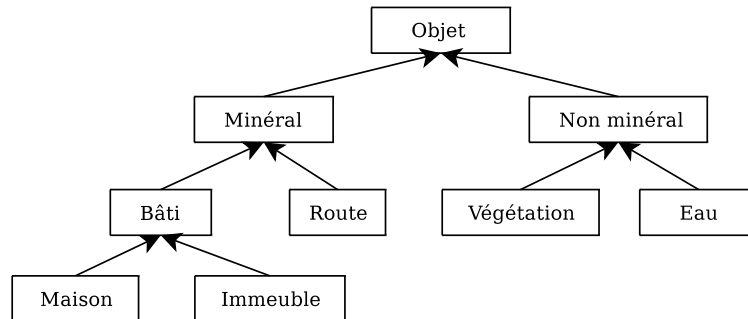


FIG. 1 – Hiérarchie des concepts de l'ontologie.

Prenons par exemple le cas de l'affectation d'une notion de couleur aux concepts. Pour un géomaticien, le toit d'une maison en tuile céramique dans la région Alsace est orange (ce cas couvrant la majorité des maisons). Néanmoins, affirmer que les toits sont oranges est une abstraction de haut niveau par rapport à une réalité beaucoup plus contrastée. Un toit est soumis à différentes conditions d'éclairage, notamment, il peut être très sombre sur sa partie non exposée au soleil et très brillant sur sa partie exposée. Sa couleur peut aussi varier du fait de son vieillissement (formation de mousse sur certaines parties). Enfin, le toit peut contenir des fenêtres qui ne seront pas oranges. Dans la littérature, ce type de problème est souvent résolu par l'utilisation d'intervalles larges et flous pour les différentes réponses spectrales. En effet, les objets recherchés étant fortement hétérogènes (même pour une seule instance d'objet), de tels intervalles se révéleraient trop larges car couvrant une trop grande partie de l'image pour être exploitables. Des expérimentations ont révélé que cette approche montrait ses limites (pour notre domaine d'application) en l'utilisant au niveau des pixels. Une autre approche consiste à raisonner au niveau des régions. Dans ce cas, l'intervalle des moyennes des valeurs des pixels pour chaque région est bien plus réduit qu'un intervalle calculé au niveau des pixels. Les résultats obtenus sont alors meilleurs.

3 Enrichissement et validation à partir d'exemples

Afin de résoudre les problèmes recensés dans la section précédente, nous avons mis en oeuvre une méthode indirecte d'expression des connaissances pour l'expert par l'utilisation d'exemples. A chaque concept de l'ontologie, l'expert peut affecter un ensemble d'exemples. Ces exemples sont composés d'une image et d'un masque binaire définissant les parties représentatives du concept donné. Un exemple pour le concept maison est donné en figure 2

A partir de tels exemples, il est possible d'inférer les propriétés géométriques de l'objet. L'objectif est double :

- **validation** : à partir des exemples, il est possible de vérifier les intervalles donnés dans l'ontologie par l'expert.
- **enrichissement** : pour les concepts où des données n'ont pu être renseignées, les exemples peuvent être utilisés pour définir les valeurs minimales et maximales pour les différentes contraintes.

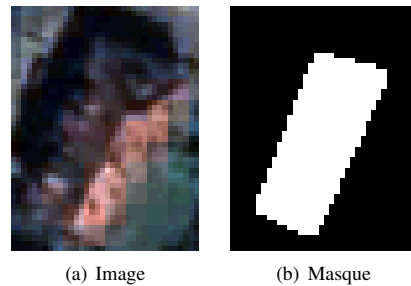


FIG. 2 – Exemple pour le concept maison.

L'utilisation d'exemples reste complémentaire au savoir direct de l'expert. En effet, pour le concept de végétation, l'intervalle associé à la contrainte de surface extrait à partir d'exemples ne peut être considéré comme pertinent. Un espace végétal peut avoir une surface arbitraire et donc supérieure à la surface maximale des exemples. L'expert joue donc ici un rôle primordial pour fournir des méta-informations sur les informations extraites. Il peut par exemple noter dans l'ontologie qu'un type de contrainte n'est pas pertinent pour un concept donné. Nous proposons d'utiliser la complémentarité de ces approches, notre ontologie représentant donc une base de connaissances.

4 Extraction de détecteurs

Le patron des algorithmes de détection est présenté dans la figure 3. Avant l'étape de détection, tous les pixels sont considérés comme appartenant au concept du détecteur. Le détecteur agit donc comme un filtre qui va retirer les pixels qui ne satisfont pas les contraintes associées au concept.

La première étape opère un filtrage sur les valeurs spectrales de chaque pixel pris individuellement. Si le concept possède des exemples associés et que l'expert a jugé opportun d'en extraire des informations spectrales, le détecteur procède à une classification basée pixel en utilisant les exemples donnés comme base d'apprentissage. Dans notre implémentation, nous avons choisi un classifieur 5 plus proches voisins qui utilise aussi les exemples des autres classes comme exemples négatifs. Il est néanmoins possible d'utiliser d'autres classifieurs, notamment des classifieurs à une classe. Si le concept n'a pas d'exemple, un filtrage est effectué en utilisant les contraintes sur les intervalles spectraux définis par l'expert dans l'ontologie.

Une fois cette première étape effectuée, un traitement spécifique pour les objets de forme rectangulaire est appliqué. Si un concept est défini comme étant de forme rectangulaire, on calcule le côté du carré le plus grand pouvant être contenu dans les instances de ce concept. Cette valeur peut être fournie par le côté minimal s'il est renseigné ou bien par l'analyse conjointe de la surface minimale et l'élongation maximale. Une ouverture morphologique (consistant en une érosion puis une dilatation) est appliquée en considérant comme élément structurant ce carré (sans et avec une rotation de 45 degrés). Ainsi, les pixels seront considérés comme

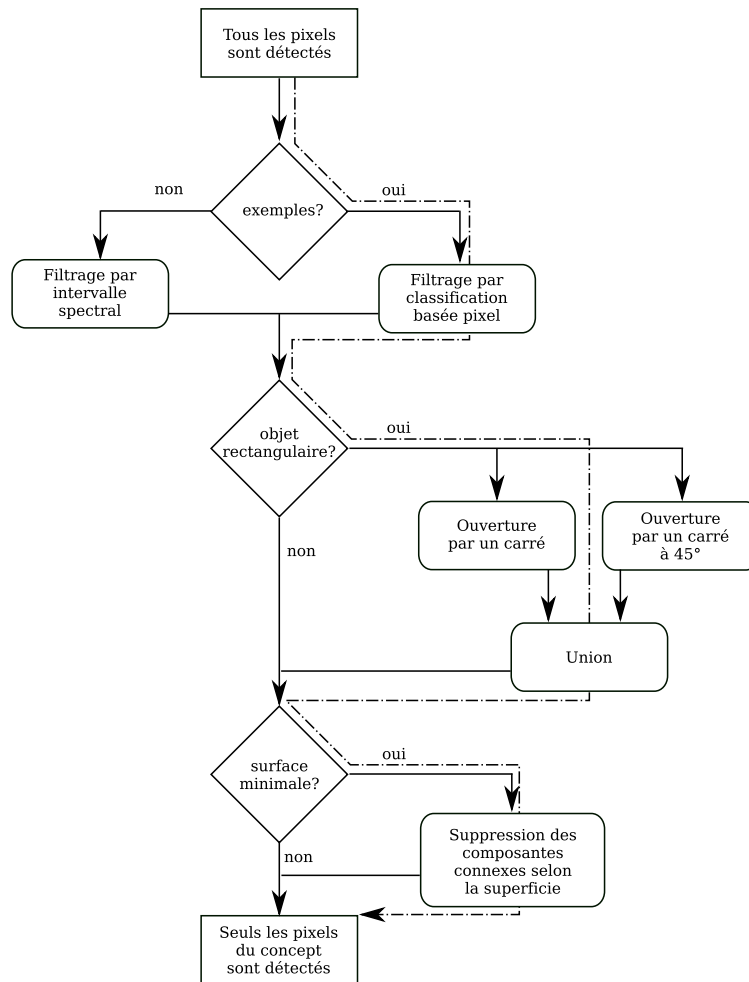


FIG. 3 – *Patron des algorithmes de détection d'un concept. En pointillé le détecteur extrait pour le concept maison.*

pouvant appartenir au concept si et seulement si ils appartiennent à des zones de pixels bien classés pouvant contenir le carré.

Enfin, une dernière étape consiste à créer des régions à partir des composantes connexes de l'image des pixels encore candidats pour le concept. Les régions dont la superficie n'est pas comprise dans l'intervalle défini par l'expert pour ce concept seront supprimées de la liste des régions candidates.

5 Évaluation

La méthode proposée a été évaluée sur une image de Strasbourg prise par le satellite Quickbird. Les capteurs Quickbird renvoient une bande panchromatique d'une résolution de 70 cm et 3 bandes couleur d'une résolution de 2,8 mètres. Les bandes couleurs ont été fusionnées avec la bande panchromatique pour obtenir une image à 4 bandes d'une résolution de 70 cm. La taille de l'image est de 900×900 pixels et la résolution spectrale est de 8 bits par pixels. L'image et les exemples utilisés pour chaque concept sont donnés dans la figure 4.

Nous allons à présent appliquer l'algorithme pour chaque concept recherché. L'évaluation sera faite par inspection visuelle. L'objectif ici n'est pas tant de mesurer la performance du détecteur mais d'évaluer :

- la qualité de l'ontologie : s'il y a un manque de connaissance dans l'ontologie ;
- la généralité de l'algorithme : s'il prend en compte toutes les connaissances de l'ontologie.

5.1 Évaluation pour le concept maison

Le concept *maison* est bien défini. Étant subsumé par le concept *bâti*, il hérite des propriétés de ce dernier (qui se limitent à la contrainte d'une forme rectangulaire). De plus, l'expert nous a fourni un intervalle de surface (entre 100 et 1000 m²) et indiqué que les bandes spectrales des exemples sont pertinentes et pourront donc être utilisées pour l'étape de classification des pixels.

Le résultat de l'application de ce détecteur est donné en figure 5. L'étape de filtrage par classification basée pixels conserve une grande partie des pixels, notamment ceux représentant des routes ou immeubles et des pixels bruités. L'utilisation de la connaissance sur la forme rectangulaire du bâti permet néanmoins d'améliorer considérablement le filtrage. Les maisons non détectées se classent en deux catégories :

- erreur de détection : la classification basée pixel a détecté trop peu de pixel de la maison sous-jacente, elle est donc éliminée du fait de sa surface trop faible ;
- maisons adjacentes : dans certains cas, plusieurs maisons proches sont regroupées, ces régions sont donc également supprimées du fait de leur surface (ici trop importante).

5.2 Évaluation pour le concept immeuble

Dans la figure 6, on remarque notamment que les immeubles qui n'ont pas été fournis comme exemples ont toutefois été détectés. L'utilisation des connaissances de l'expert en complément des informations fournies par les exemples montre ici son intérêt. Si la contrainte de superficie avait été définie à partir des exemples, les immeubles massifs auraient été supprimés

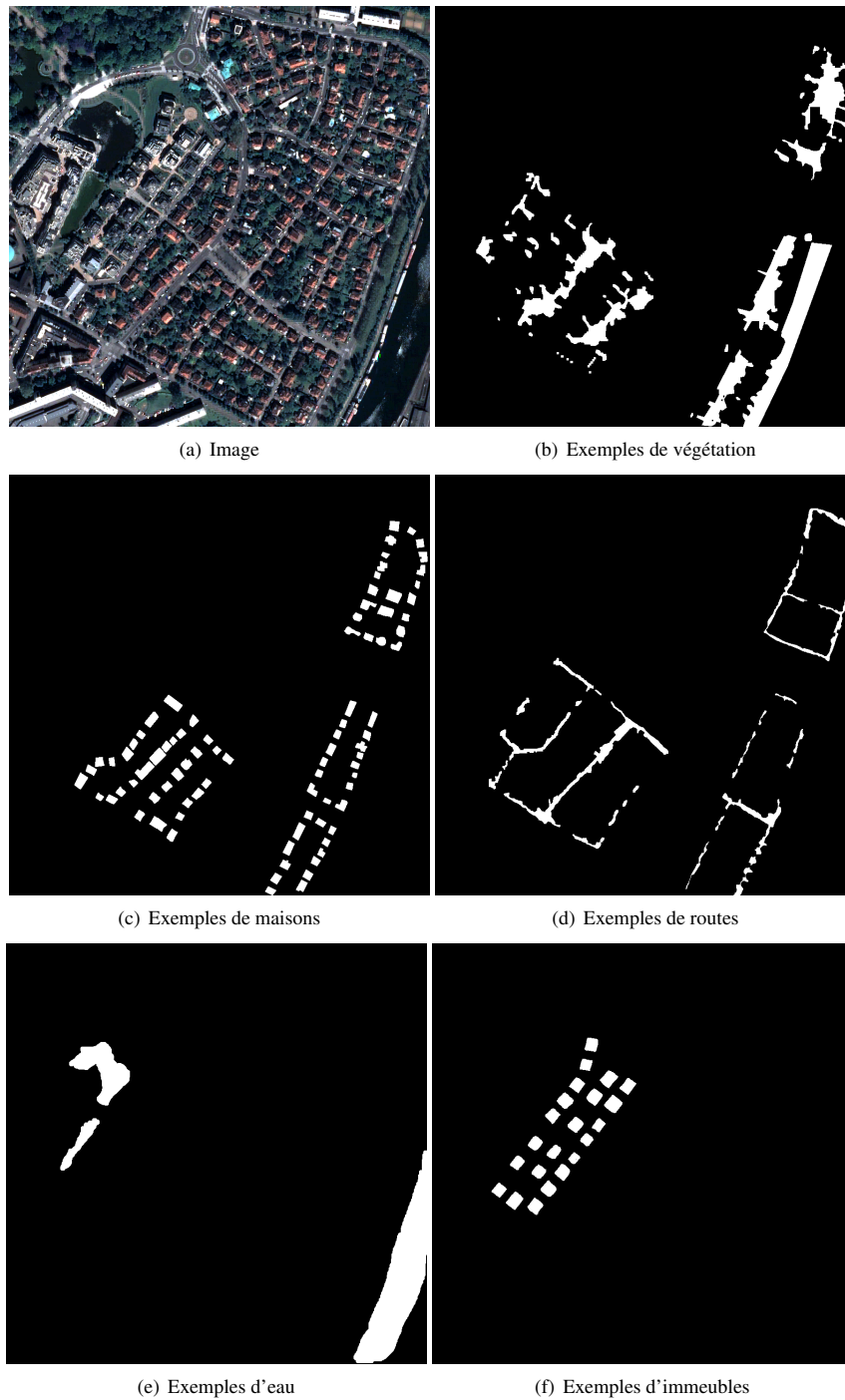


FIG. 4 – Image d'évaluation de la méthode et exemples de différentes classes d'objets considérés.

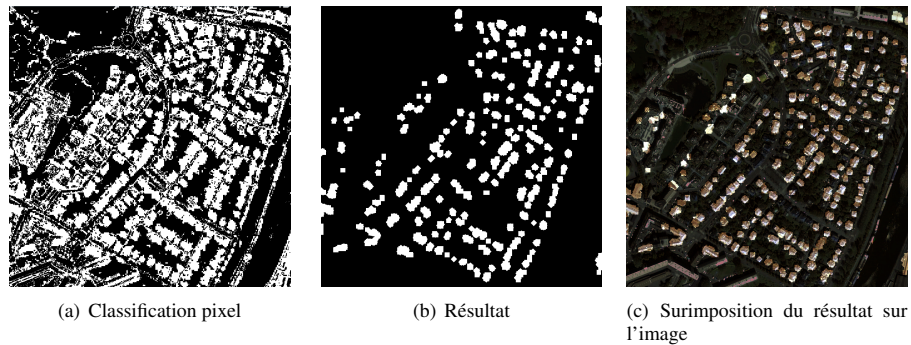


FIG. 5 – Application du détecteur pour le concept maison.

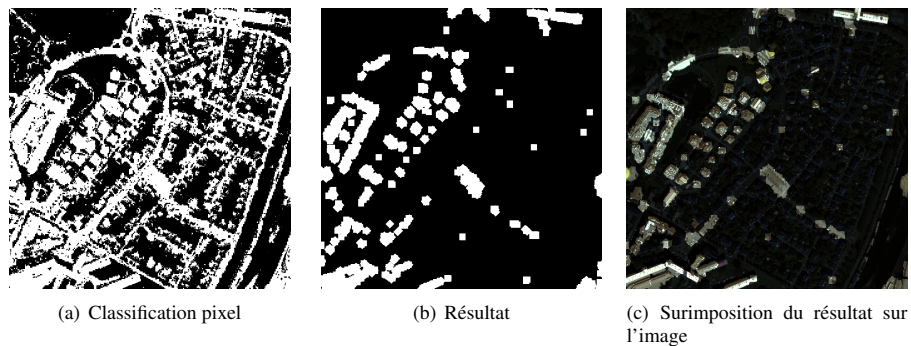


FIG. 6 – Application du détecteur pour le concept immeuble.

du fait de leur taille. Néanmoins, l'expert ayant renseigné un critère de taille plus large que les exemples donnés, ces immeubles ont été conservés.

Les erreurs portent ici sur des portions de routes, la réponse spectrale de la route étant semblable à celle des immeubles.

5.3 Évaluation pour le concept eau

L'eau présente la particularité d'avoir une réponse spectrale confondue avec l'ombre, comme on peut le voir sur la classification pixel de la figure 7. Ce problème est partiellement résolu en utilisant la connaissance de l'expert relative à la superficie des plans d'eau. Néanmoins, certaines grandes zones d'ombres issues d'immeubles ne sont pas éliminées.

5.4 Évaluation pour le concept végétation

La végétation est un concept très simple à extraire des images. Elle peut en effet être détectée en utilisant directement l'indice pixel *NDVI* (Normalized Difference Vegetation Index)

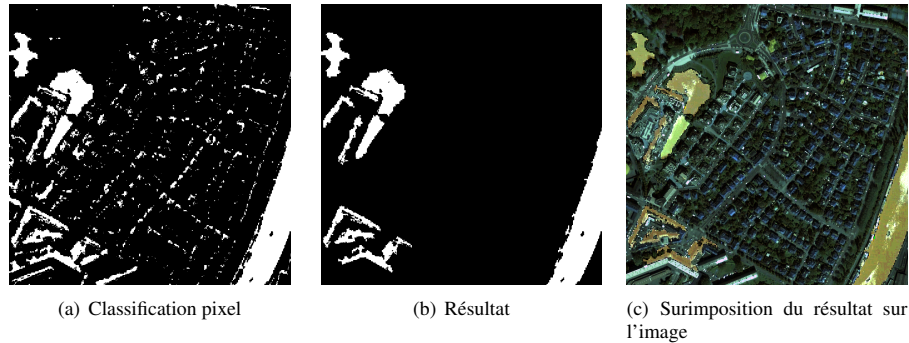


FIG. 7 – Application du détecteur pour le concept eau.



FIG. 8 – Application du détecteur pour le concept végétation.

dont la formule est :

$$NDVI = \frac{IR - R}{IR + R} \in [-1; 1]$$

avec IR étant la valeur du pixel dans la bande spectral du proche infra-rouge et R la valeur du pixel dans la bande spectrale de la couleur rouge. Pour cet indice, l'intervalle reconnu de la végétation est $[0, 3; 0, 8]$. L'expert ne produisant aucune information à partir des exemples, le détecteur utilisera un simple filtrage sur la valeur $NDVI$. De même, il n'y aura pas de filtrage ensuite puisqu'aucune autre contrainte n'est définie pour ce concept (ni directement, ni extraite des exemples).

Le résultat de la figure 8 montre que le résultat obtenu est amplement satisfaisant.

5.5 Évaluation pour le concept route

Le dernier concept à détecter est le concept *route*. Malheureusement l'algorithme proposé ne permet pas encore de le détecter correctement. En effet, le concept route est fort complexe.

Comme on le voit sur les exemples, une route peut être une région sans trou et allongée ou bien une région avec de grands trous. Ces deux types de représentants du concept n'ont pas du tout les mêmes propriétés.

Nous pensons que la solution réside dans la séparation de ces deux types d'instances en deux sous-concepts. De plus, notre algorithme n'est pour l'instant pas en mesure d'extraire l'information suffisante pour détecter des graphes routiers, et, de façon plus générale, des objets à trous.

6 Conclusion

Dans cet article, nous avons proposé une méthode de construction d'ontologie associant connaissances directes de l'expert et connaissances indirectes par l'utilisation d'exemples. Nous avons également introduit un algorithme de construction de détecteurs d'objets dans des images en utilisant l'ontologie. Les premières évaluations montrent l'intérêt d'une telle approche et sa généralisation à tous les concepts urbains considérés (sauf celui de route).

Par la suite, nous envisageons d'accroître les connaissances disponibles dans l'ontologie en intégrant notamment des connaissances sur le voisinage. Par exemple, un immeuble projeté une ombre et toute région candidate au concept immeuble qui ne serait pas à côté d'une zone d'ombre ne saurait être un immeuble. De même, les maisons sont proches d'autres maisons : une région candidate au concept de maison mais isolée ne saurait être une maison. Enfin, comme déjà indiqué, nous souhaitons gérer les concepts de forme ou topologie complexe (à trous par exemple).

Références

- Bittner, T. et S. Winter (1999). On ontology in image analysis. *International Workshop On Integrated Spatial Databases ISD'99 1737*, 168–191.
- Durand, N., S. Derivaux, G. Forestier, C. Wemmert, P. Gancarski, O. Boussaid D, et A. Puissant (2007). Ontology-based object recognition for remote sensing image interpretation. In *IEEE International Conference on Tools with Artificial Intelligence*, Patras, Greece, pp. 472–479.
- Fonseca, F., M. Egenhofer, P. Agouris, et G. Camara (2002). Using ontologies for integrated geographic information systems. *Transactions in Geographic Information Systems* 6(3), 231–257.
- Gruber, T. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human Computer Studies* 43(5/6), 907–928.
- Hürsch, W. et C. Lopes (1995). Separation of concerns. In *Technical report by the College of Computer Science*, Northeastern University.
- Jin, X. et C. H. Davis (2005). Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *EURASIP Journal on Applied Signal Processing* (14), 2196–2206.
- Lefèvre, S., J. Weber, et D. Sheeren (2007). Automatic building extraction in vhr images using advanced morphological operators. In *IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas (URBAN)*, Paris.

- Maillot, N. E. et M. Thonnat (2008). Ontology based complex object recognition. *Image Vision Computing* 26(1), 102–113.
- Mezaris, V., I. Kompatsiaris, et M. G. Strintzis (2004). Region-based image retrieval using an object ontology and relevance feedback. *EURASIP Journal on Advances in Signal Processing* 2004(1), 886–901.
- Panagi, P., S. Dasiopoulou, G. Papadopoulos, I. Kompatsiaris, et M. Strintzis (2006). A genetic algorithm approach to ontology-driven semantic image analysis. *IEEE International Conference of Visual Information Engineering (VIE) 2006*(CP522), 132–137.
- Peteri, R., J. Celle, et T. Ranchin (2003). Detection and extraction of road networks from high resolution satellite mages. In *IEEE International Conference on Image Processing*, pp. 301–304.
- Yager, N. et A. Sowmya (2004). Support vector machines for road extraction from remotely sensed images. In *Computer Analysis of Images and Patterns*, pp. 285–292.
- Yongxue, L., L. Manchun, M. Liang, X. Feifei, et H. Shuo (2006). Review of remotely sensed imagery classification patterns based on object-oriented image analysis. *Chinese Geographical Science* 16(3), 282–288.
- Zhao, H., J. Kumagai, M. Nakagawa, et R. Shibasaki (2002). Semi-automatic road extraction from high-resolution satellite image. In *ISPRS Symposium on Photogrammetry and Computer Vision*, pp. A : 406.

Summary

Building an efficient interpretation method for very high resolution images requires to rely on domain knowledge. Many specific methods have been developed to provide good results for detection of different kinds of objects like roads or buildings. These approaches involve domain knowledge about the sought object, but they do not express it explicitly. In this article, we propose first to model the knowledge available on the domain under consideration within an ontology. Thus, specific detectors are automatically built from the ontology to extract the interesting objects from the image. This enables a clear separation between the knowledge representation and the extraction methods. This separation makes it possible to apply the detection algorithm on various application domains and to modify the extraction algorithm without modifying the ontology.