

Exploitation de connaissances domaine pour l'interprétation d'images

Image Understanding using Domain Knowledge

Nicolas Zlatoff, Bruno Tellez, Atilla Baskurt

LIRIS FRE 2672
Université Claude Bernard
Bâtiment 710
43 bd du 11 novembre
69622 Villeurbanne cedex France
{nzlatoff, btellez, abaskurt}@liris.cnrs.fr

Résumé

Pour être réellement efficaces, les outils d'indexation d'images doivent être automatiques en terme d'analyse et de description du contenu. Le fossé sémantique existant entre les concepts haut-niveau et des descripteurs bas-niveau (couleur, texture), rend inévitable l'introduction dans le processus de connaissances externes. Ces connaissances sont celles du domaine que dénotent les images traitées. Nous présentons dans cet article un système d'interprétation d'images qui se voudra générique, c'est-à-dire capable de s'interfacer à différentes connaissances domaine sous forme d'ontologies. Il permet, à partir d'une image segmentée, d'identifier des objets par l'intermédiaire de deux processus successifs de fusion : le premier se base uniquement sur des critères bas-niveaux (indépendants du domaine), le second extrait des connaissances domaine l'information nécessaire à son fonctionnement. Nous montrerons également que ces connaissances domaine doivent en fait correspondre à un modèle de représentation, afin de pouvoir être correctement exploitées. Le contexte d'application décrit dans cet article est celui de l'archéologie, plus précisément, l'indexation d'images de stèles thessaliennes, fournies par la Maison de l'Orient et de la Méditerranée (MOM).

1. Introduction

Durant ces dix dernières années, l'informatique a connu une augmentation considérable des capacités de stockage, couplée à une interconnexion des réseaux sans cesse croissante. Un nombre gigantesque de documents numériques et plus particulièrement d'images est maintenant disponible en libre accès ou dans des collections privées.

Dès lors, pour pouvoir correctement exploiter ces banques d'images, il est indispensable de disposer d'outils qui permettront d'identifier automatiquement les images répondant aux requêtes des utilisateurs. Pour l'instant, de nombreux outils basés sur le contenu (*content-based*¹) existent : QBIC (Niblack, 1993), VisualSEEK (Smith, 1996), pour ne citer qu'eux. Ils permettent une indexation automatique d'images, sur des critères bas-niveau comme la couleur, la texture, voire la forme de l'objet principal de l'image. Néanmoins, ce type d'indexation n'est en général pas pertinent pour l'utilisateur qui, dans la plupart des cas, cherche une image d'après ce qu'elle représente (sa sémantique) : un objet, une personne, une scène particulière.

Chercher à étendre les principes des outils *content-based* pour essayer de déduire une sémantique à partir de primitives bas-niveaux est un objectif qui trouve rapidement ses limites. En effet, il est difficile d'associer de manière unique un concept haut-niveau à une série de descripteurs bas-niveau, justement parce qu'un concept est une abstraction visant à regrouper les différentes représentations de l'objet. Cette absence de lien direct entre haut-niveau et bas-niveau est connue sous le nom de fossé sémantique.

¹ Le terme « content-based » est mal adapté, puisqu'il renvoie à des notions bas-niveau comme la couleur et non pas à des *contenus* sémantiques.

Dès lors, pour tenter de combler ce fossé, il apparaît nécessaire d'introduire des connaissances extérieures qui permettront d'identifier, en fonction d'un contexte donné, l'interprétation la plus probable d'une partie de l'image.

L'étude présentée dans cet article s'intègre dans des travaux de recherche visant à l'élaboration d'une méthodologie d'interprétation d'image en exploitant des connaissances d'un domaine². En nous appuyant sur les trois niveaux classiquement admis en image (régions du bas-niveau puis du niveau intermédiaire, et objets sémantiques de haut-niveau), nous insisterons sur le rôle des connaissances du domaine dans le pilotage d'une partie du processus d'interprétation qui consiste à fusionner les objets du niveau intermédiaire en objets sémantiques, et dans la remise en cause des traitements de plus bas-niveau qui s'avèreraient erronés.

Un état de l'art complet sur les systèmes à base de connaissances en interprétation d'images (Crevier, 1997), insiste sur la nécessité de disposer d'outils génériques similaires aux noyaux des systèmes experts mais spécialisés en interprétation d'images. Aujourd'hui, de nombreux systèmes existent, comme *SIGMA* (Matsuyama, 1990), *Schema* (Draper, 1989) pour les systèmes géospatiaux ou d'autres pour des applications en biologie (Hudelot, 2003). Ces systèmes ont prouvé leur efficacité dans l'interprétation d'images au sein d'un domaine précis. Néanmoins, l'extension de ces mêmes systèmes à d'autres domaines est souvent problématique voire impossible car les connaissances liées au domaine sont trop fortement inscrites dans les procédures d'interprétation.

Nos recherches s'inscrivent dans la réalisation d'un système générique d'interprétation d'images capable de s'interfacer à différentes connaissances d'un domaine, définies selon les images à interpréter. Nous avons choisi de décrire les connaissances sous forme d'ontologies (une par domaine) dans lesquelles les différents mécanismes d'interprétation extraient les informations nécessaires à la réalisation de leurs tâches.

Le plan de cet article sera le suivant : nous commencerons dans une première partie par présenter le domaine en question auquel nous nous sommes intéressés. Nous avons volontairement choisi de nous limiter à un domaine particulier et restreint : celui des stèles thessaliennes. Dans un deuxième temps, nous expliciterons le cadre global de notre méthodologie pour construire un système générique d'interprétation avant de détailler les traitements aboutissant au bas-niveau et au niveau intermédiaire (segmentation et fusion bas-niveau). Nous détaillerons ensuite la modélisation et l'exploitation des connaissances sous forme d'ontologies que nous avons effectuées dans le but d'aboutir au haut niveau. Enfin, une dernière partie présentera les différents résultats obtenus dans ce cadre restreint ainsi que les perspectives d'évolution pour un outil générique utilisable dans d'autres contextes.

2. Domaine

La Maison de l'Orient et de la Méditerranée (MOM), laboratoire de recherche en archéologie, dispose actuellement d'une base non indexée de plusieurs milliers d'images représentant des stèles thessaliennes. Ces images sont issues de la numérisation de photographies argentiques prises par l'Ecole Française d'Athènes il y a une dizaine d'années. Devant la taille imposante de cette base d'images, la MOM souhaiterait disposer d'outils permettant d'accéder automatiquement aux différents composants des stèles.

Les conditions de prises de vues globalement assez mauvaises freinent l'utilisation seule des méthodes traditionnelles de traitement d'images et posent rapidement de réels problèmes en terme d'automatisation des tâches d'interprétation.

La Figure 1 présente quelques-unes de ces images.

² Par « domaine » nous entendons « domaine dénoté par l'image »

Une stèle funéraire est un bloc de pierre, peint et sculpté, enfoncé dans le sol afin de marquer l'emplacement d'une sépulture. En première approximation, une stèle thessalienne comporte deux types d'objets: ceux liés à la forme générale (*corps de stèle*, *couronnement*, *geison*) et ceux liés à des travaux de finition comme les *ornements*, les *inscriptions*, etc...

La Figure 2 présente un exemple de photo de stèle ainsi que des éléments caractéristiques de cette stèle.

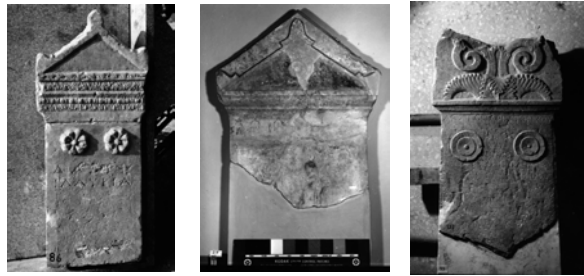


Figure 1- Différentes images de stèles

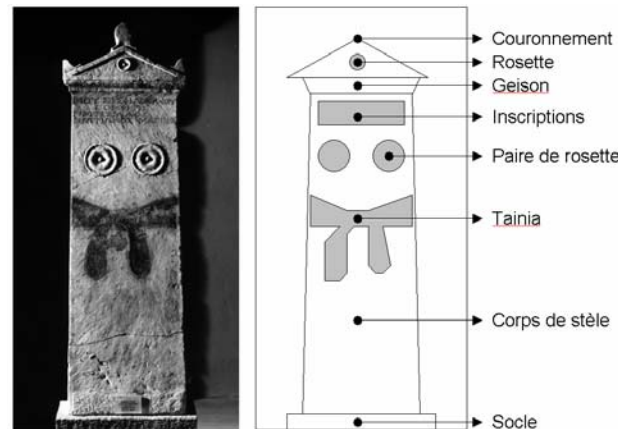


Figure 2- Différents composants d'une stèle

L'objectif est donc d'extraire aussi bien les éléments liés à la forme générale (liés à la typologie des stèles) que ceux liés à la finition.

3. Méthodologie

Notre approche pour traiter le processus d'interprétation s'appuie sur un moteur générique de traitement des images et sur des connaissances, décrites dans des ontologies. Le moteur reprend les trois niveaux de traitement d'une image (Figure 3).

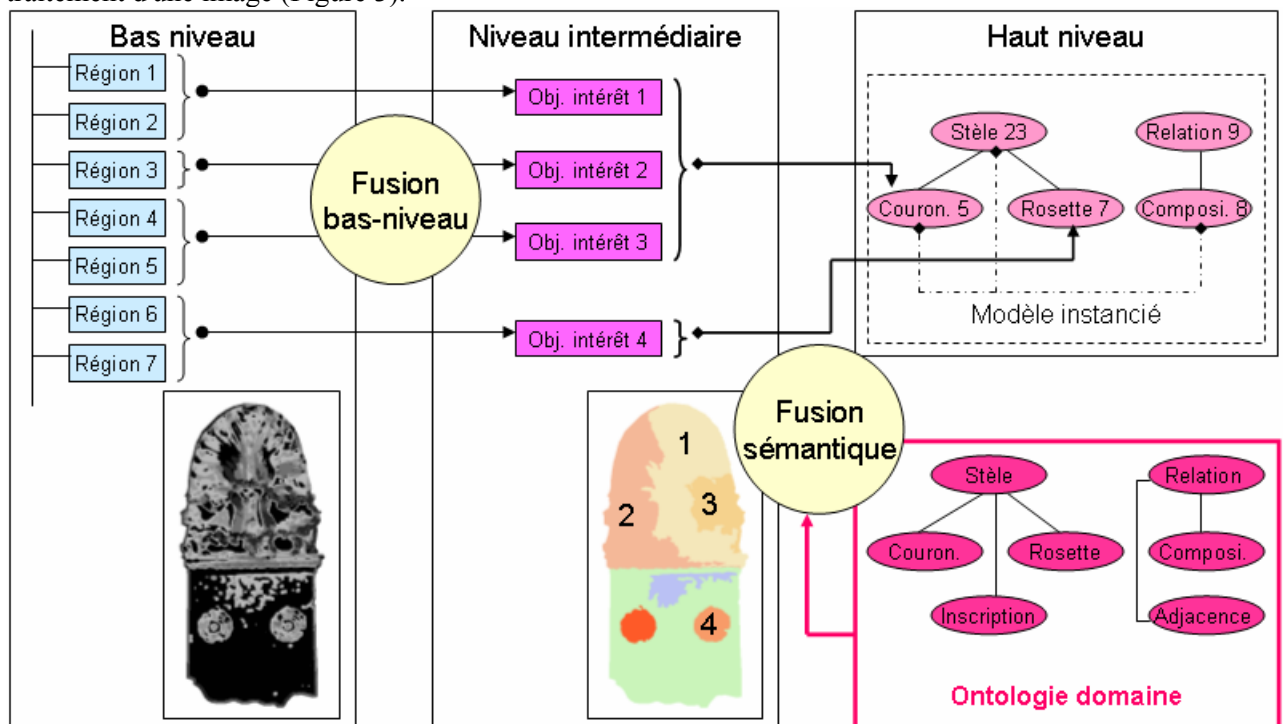


Figure 3 - Trois niveaux d'image

Le bas-niveau, issu du processus de segmentation, consiste en une tessellation de régions homogènes par rapport à un descripteur particulier (couleur, texture par exemple).

Le niveau intermédiaire est obtenu en regroupant plusieurs régions du bas-niveau, afin de former des objets présentant une certaine cohérence (sémantique ou géométrique). Dans le meilleur des cas, les régions issues de la segmentation ont été fusionnées pour former directement des objets sémantiques non encore nommés à ce stade (objet 4 par exemple). Néanmoins, nous considérerons que le niveau intermédiaire peut aussi être composé d'objets représentant seulement une partie d'un objet sémantique (objets 1, 2, 3). On parlera d'objets d'intérêt pour désigner les objets du niveau intermédiaire.

Enfin, le haut-niveau identifie les différents objets sémantiques présents dans l'image à partir des objets d'intérêt.

Le passage du bas-niveau au niveau intermédiaire s'effectue selon un procédé de fusion basé sur des critères bas-niveaux : différence des valeurs de descripteurs entre deux régions, imbrication mutuelle des régions. Il n'est donc en général pas possible d'aboutir directement à ce stade à la formation d'objets sémantiques, puisque de tels objets sont souvent constitués de plusieurs d'éléments, chacun ayant des descripteurs bas-niveaux associés très différents.

Nous proposons donc un deuxième procédé de fusion, appelé fusion sémantique, qui extrait dans les connaissances domaine les informations nécessaires à la reconnaissance des objets sémantiques présents dans l'image. Pour cela, le moteur générique sélectionne dans l'ontologie les connaissances domaine qui vont lui permettre de regrouper et de caractériser les objets d'intérêt. En outre, cette phase permettra également de remettre en cause des traitements de plus bas-niveau éventuellement erronés.

4. Segmentation et fusion en objets d'intérêt

Cette partie présente les traitements permettant d'aboutir au niveau intermédiaire.

4.1. Segmentation

La segmentation doit permettre d'obtenir, à partir d'une image originale, une tessellation de régions homogènes, qui servira de base pour tous les traitements futurs (fusion de bas-niveau et sémantique). Il ne s'agit donc pas de trouver la meilleure méthode de segmentation possible, mais de mettre en place une solution simple qui permette à coup sûr de séparer les différents objets sémantiques de l'image. Dans notre approche, une sur-segmentation est préférable à une sous-segmentation, puisqu'elle sera rectifiable directement par la suite via les différentes fusions.

Compte tenu du fait que les stèles présentent des zones très texturées (rosettes, inscriptions, cassures, etc...), des descripteurs de texture ont été choisis afin d'apporter une certaine valeur sémantique aux données extraites.

Parmi les différentes méthodes basées sur la texture proposées par la littérature, les filtres directionnels de Laws (Laws, 1980) se conformaient bien aux contraintes imposées : simplicité, rapidité et pertinence des résultats pour ce type d'images. Ces filtres sont au nombre de 25, issus des convolutions croisées de 5 filtres monodimensionnels (Voir Tableau 1).

L5	[	1	4	6	4	1	]	Level
E5	[	-1	-2	0	2	1	]	Edge
S5	[	-1	0	2	0	1	]	Spot
W5	[	-1	2	0	-2	1	]	Wave
R5	[	1	-4	6	4	1	]	Ripple

Tableau 1-Filtres monodimensionnels de Laws

Le principe consiste à convoluer chacun de ces filtres à l'image originale pour former 25 images dites de texture, puis, en calculant pour chaque pixel de chaque image la variance de ses valeurs de texture sur un voisinage, à obtenir 25 images d'énergie de texture. Après combinaison des images de directions complémentaires, on obtient 14 images descriptives, qui permettent donc de décrire l'image de départ comme un ensemble de N points P_i dans un espace à 14 dimensions.

Un procédé de clustering non supervisé (Nuées dynamiques, ou K-Means (Gersho, 1991)) est alors employé pour permettre de regrouper les pixels jugés similaires dans un espace euclidien à 14 dimensions en K clusters R_k , caractérisés par leur centre G_k . Le seul paramètre de cette étape est le nombre K de clusters auquel on souhaite aboutir. Une valeur de 8 (déterminée par heuristique) conduit à une sur-segmentation efficace, c'est-à-dire permettant de bien séparer les différents objets sémantiques. L'influence du paramètre K sur les étapes suivantes est faible dès lors que les objets sémantiques ont bien été séparés les uns des autres. En effet, augmenter K ne ferait que morceler un peu plus les clusters, effet qui sera annulé par la fusion de bas-niveau qui va suivre.

La Figure 4 présente un exemple de traitement : l'image originale³ (a) est segmentée en texture puis classifiée avec 8 clusters (b). La figure 4.c montre l'extraction d'un cluster et des différentes régions qui le composent.

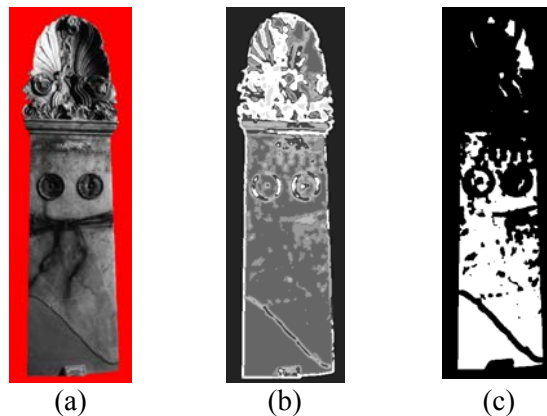


Figure 4 - Résultats de la segmentation

Afin de valider la pertinence de cette description, une série de test a été effectuée sur 15 images désignées par les experts de la MOM comme étant représentatives, avec pour but de montrer qu'il était possible de reconstruire les objets sémantiques en regroupant manuellement des régions segmentées. Par comparaison avec une extraction manuelle directement depuis l'image originale, nous obtenons à titre d'exemple sur les rosettes une précision de 80% (15% sont des faux positifs).

4.2. Fusion de bas-niveau

A l'issue de la segmentation, l'image se trouve ainsi partitionnée en un grand nombre de régions. Il est alors fréquent d'observer dans certaines zones une multitude de petites régions spatialement très proches les unes des autres et présentant par ailleurs des différences de texture relativement faibles. L'objectif consiste donc assez naturellement à regrouper ces régions entre elles. Nous nommons cette étape fusion bas-niveau, par opposition à la fusion sémantique qui impliquera des connaissances explicitement du domaine.

Cette fusion bas-niveau (Idrissi, 2003) mesure, pour deux régions voisines i et j , l'écart entre les valeurs de descripteurs de texture (ED_{ij}) qui est ensuite pondéré par deux coefficients, l'un privilégiant la fusion des petites régions (A_{ij}) et l'autre (C_{ij}) favorisant la fusion des régions fortement imbriquées l'une dans l'autre (Schettini, 1993).

A partir de l'image segmentée, le graphe d'adjacence des régions est calculé et réduit itérativement en calculant pour chaque frontière entre deux régions i et j leur différence pondérée D_{ij} et en fusionnant au final les deux régions dont la frontière présente la plus faible différence. L'algorithme possède deux critères d'arrêt : le nombre de régions minimales à obtenir et la différence D_{ij} maximale tolérée.

³ Du fait des mauvaises conditions de prises de vues (stèle en pierre sur fond de pierre), les stèles sont préalablement extraites du fond manuellement

$$D_{ij} = ED_{ij} \times C_{ij} \times A_{ij}$$

$$ED_{ij} = \sqrt{\sum_{k=0}^n (d_{j,k} - d_{i,k})^2}$$

$d_{j,k}$: $k^{ième}$ descripteur de la région numéro j
 n : nombre total de descripteurs

$$C_{ij} = \frac{\min(P_i, P_j)}{4P_{ij}}$$

P_i : périmètre de la région numéro i
 P_{ij} : périmètre commun entre la région i et la région j

$$A_{ij} = \begin{cases} \varepsilon \text{ si } (N_i < NbPixMin \text{ ou } N_j < NbPixMin) \\ 1 \text{ sinon} \end{cases}$$

N_i : nombre de pixels de la région i
 $NbPixMin$: nombre minimal de pixels pour une région (fixé en paramètre)
 ε : réel positif entre 0 et 1 (fixé en paramètre)

Cette phase de fusion permet de réduire significativement le nombre de régions (entre 400 et 600 suivant les images) en formant les objets d'intérêt du niveau intermédiaire (entre 10 et 30). La Figure 5 présente en (a) une image segmentée et en (b) l'image du niveau intermédiaire obtenue suite à la fusion bas-niveau. On constate que certains objets sémantiques sont déjà entièrement reconstruits (comme les *rosettes*), d'autres restent segmentés en plusieurs objets d'intérêt (comme le *couronnement*) et enfin certains ont été fusionnés dans un autre objet sémantique (ex : *geison* dans *couronnement*).

Une difficulté inhérente à cette étape consiste à fixer les différents paramètres de la fusion (ε , $NbPixMin$, seuil de fusion, critères d'arrêt...). A ce stade, ils sont fixés de manière heuristique. Néanmoins, nous avons observé qu'il est possible de "contrôler" et donc, de remettre en cause la fusion bas-niveau et de modifier la valeur de ces paramètres en fonction des résultats obtenus à l'issue de la seconde fusion.

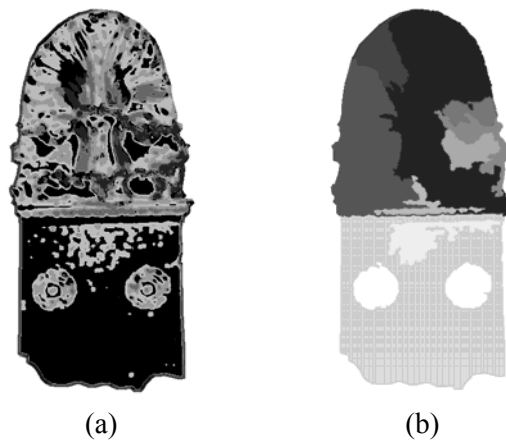


Figure 5- Segmentation et fusion bas-niveau

Il n'est toutefois pas possible de pousser plus loin la fusion bas-niveau pour tenter de reconstruire totalement les objets sémantiques. En effet, ceux-ci sont souvent constitués de plusieurs parties, de texture très différentes que l'algorithme ne peut donc pas fusionner. En revanche, nous proposons que l'utilisation des connaissances contribuent à la fin de la fusion et permettent du même coup de reconnaître les différents objets sémantiques.

5. Modéliser les connaissances du domaine

Les connaissances du domaine une fois modélisées constituent une ontologie. En effet, une ontologie est une « spécification explicite d'une conceptualisation admise » (Gruber, 1993). Une ontologie regroupe des connaissances (c'est-à-dire les concepts et les relations dans le monde étudié) sous un formalisme donné.

S'il n'existe pas une unique ontologie pour un problème donné (Noy, 2001), il est néanmoins possible de considérer une méthodologie de construction en quatre étapes (Uschold, 1995). La première, fondamentale, consiste à définir l'usage futur qui sera fait de l'ontologie. Dans notre contexte, l'ontologie doit permettre d'exploiter des connaissances d'un domaine (valables pour tous les objets du domaine) pour fusionner des objets d'intérêt en objets sémantiques. Il s'agit donc d'une double problématique de classification: étant donnés différents objets d'intérêt, comment les assembler pour former des objets sémantiques ? Étant donnés ces objets sémantiques, dans quelle catégorie du modèle les placer ? La définition de l'usage de l'ontologie est fondamentale car elle conditionne toute la modélisation.

La deuxième étape consiste à construire l'ontologie : identifier les concepts et les relations du domaine et transcrire ces notions dans un formalisme. Enfin, l'ontologie produite doit être évaluée et documentée.

5.1. Construire l'ontologie

Construire une ontologie consiste à identifier les concepts et les relations du domaine étudié, à les regrouper par leur sémantique, à donner à tous ces composants une définition et enfin à les nommer. Concernant l'ontologie domaine des stèles thessaliennes, nous avons choisi de modéliser en tant que concepts les différents objets sémantiques constituant la stèle (*corps*, *geison*, *rosette*...). Comme tous ces objets ne peuvent pas être considérés au même niveau de spécialisation dans la description, il est nécessaire de les organiser dans une hiérarchie. Par exemple, un concept abstrait *ornement* a été créé pour regrouper tous les objets de type ornement (*rosette* et autres).

Dans un deuxième temps, afin de pouvoir modéliser toutes les connaissances du domaine, il est nécessaire de créer d'autres relations que celle de hiérarchie entre les objets sémantiques (relation d'adjacence par exemple). D'après les experts de la MOM, les relations de composition, d'adjacence et de voisinage sont suffisantes pour décrire le domaine : *adjacence* indique nécessairement une frontière commune entre deux objets alors que *voisinage* traduit une notion de proximité sans contact direct (distance de 2 dans le graphe d'adjacence). Enfin, pour ces deux relations, nous avons précisé la direction selon les situations *haut* et *bas*.

Considérant alors l'utilisation future de l'ontologie, une restriction importante s'impose sur la modélisation. En effet, le processus de contrôle de la fusion sémantique des objets d'intérêt en objets sémantiques s'appuiera sur une mise en correspondance des connaissances génériques du domaine (valable de manière générale, dans le domaine) avec les connaissances relatives aux objets d'intérêt en cours de traitement. Or, ces dernières se réduisent à des relations topologiques, directionnelles ainsi qu'à des descripteurs bas-niveau. Ainsi, la modélisation des connaissances domaine doit-elle se restreindre à des connaissances correspondant au moins approximativement à la description des objets intermédiaires (relations topologiques et directionnelles). Les autres connaissances, comme la relation de composition, relèvent de notions plus abstraites et ne trouvent pas directement leur correspondance dans la description intermédiaire : par exemple, comment traduire en relations topologiques que *A compose B* ? Il est possible d'affirmer que *si A est inclus dans B alors A compose B*, mais la réciproque est fausse.

Ainsi, nous prenons le parti de ne conserver dans l'ontologie que les connaissances *adjacence*, *voisinage* (non modifiées) et *composition* restreinte à *inclusion géométrique*. Ceci a pour effet de transformer les connaissances domaine (concepts et relation du domaine) en connaissances de scène (modèles d'apparition dans l'image des concepts et des relations du domaine).

La Figure 6 montre quelques concepts de l'ontologie, organisés suivant les relations de spécialisation, composition restreinte et adjacence.

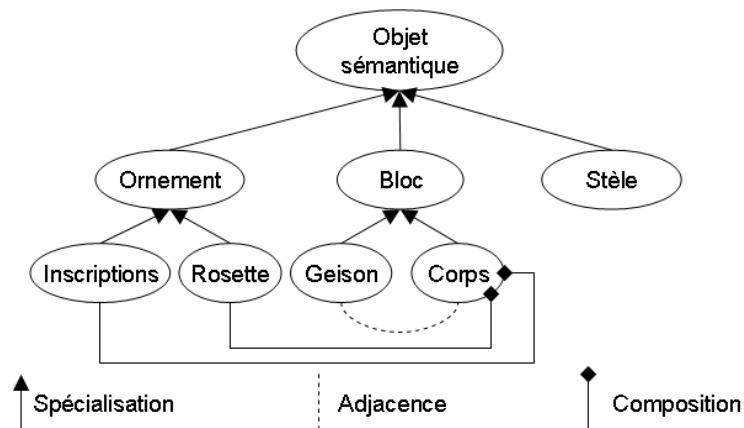


Figure 6- Concepts de l'ontologie

5.2. Formaliser l'ontologie

Une fois que l'ontologie est construite conceptuellement, il s'agit d'adopter un formalisme pour la stocker. Parmi les différentes solutions existantes présentées dans (Roussey, 2001), les systèmes à base de *frames* semblent les plus adaptés pour les problèmes de classification (Marino Drews, 1993).

Dans ces systèmes (Minsky, 1975), la connaissance est regroupée en « paquets ». Un *frame* est un prototype, c'est-à-dire un objet représentatif d'une famille. Il comporte d'une part des *slots* ou attributs qui décrivent les propriétés du prototype, d'autre part, des facettes qui permettent d'imposer des contraintes aux *slots* (type de données par exemple). Une évolution de la description de base des *frames* a conduit à distinguer deux types de *frames* : les *frames* classes qui représentent les catégories d'objets du monde et les *frames* instances qui représentent les individus.

Dans le cadre de cette formalisation, tous les concepts sont modélisés sous la forme de *frames* classes car ils correspondent à des connaissances génériques. Lorsque le modèle sera appliqué à une image particulière, son instanciation permettra de décrire l'image en cours de traitement. Si la relation de spécialisation trouve son équivalent exact dans la modélisation par *frame*, grâce à la notion d'héritage, la modélisation des autres relations entre concepts ne va pas de soit. En effet, deux méthodes sont possibles : gérer la relation en tant que *frame* ou bien l'intégrer comme *slot* à un *frame* déjà existant (donc à un concept). Néanmoins, une relation modélisée par un *slot* perd sa sémantique de relation. En effet, rien ne la distinguerait d'un autre type de *slot* (propriétés intrinsèques taille ou forme par exemple). C'est pourquoi nous préférons modéliser les relations entre objets sémantiques du domaine par des *frames*.

Ainsi, notre modélisation inclut trois types de connaissances liées directement au domaine. La première consiste en une hiérarchie d'objets sémantiques selon la relation de spécialisation. La seconde consiste en une liste exhaustive des différentes relations pouvant exister entre ces objets. Enfin, la troisième consiste en l'introduction de connaissances « par défaut » par l'intermédiaire de facettes. Celles-ci introduisent des contraintes sur certains *slots* des *frames* concepts ou relations. Par exemple, la Figure 7 illustre la modélisation de la connaissance « une rosette est ronde ».

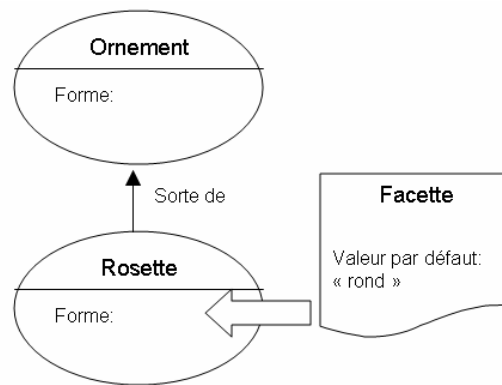


Figure 7- Introduction de connaissances par défaut via une facette

5.3. Implantation

L'ontologie a été réalisée avec l'outil *Protégé2000* proposé par l'université de Stanford aux États-unis. Développé en Java, *Protégé2000* (Noy, 2000) est un outil de modélisation et d'acquisition de connaissances. Il est compatible avec le protocole OKBC et permet ainsi une bonne interopérabilité avec d'autres outils orientés connaissances. En outre, *Protégé2000* est spécifiquement orienté pour la modélisation sous formes de *frames* puisqu'il reprend les structures de classes, *slots* et facettes. Il est fréquemment utilisé aujourd'hui pour la modélisation de connaissances d'un domaine, par exemple, pour des applications d'indexation d'images par annotations (Schreiber, 2001) (Hyvonen, 2002).

De plus, il est possible d'ajouter facilement des fonctionnalités à Protégé sous forme de *plug-ins*. Ces fonctionnalités supplémentaires relèvent aussi bien de la visualisation que de la manipulation des connaissances.

La modélisation des connaissances sous *Protégé2000* nous a conduit à créer 262 *frames* classes, dont environ 70 spécifiquement dédiées à la connaissance du domaine.

5.4. Vérification de la faisabilité de l'exploitation de la connaissance

Afin de vérifier que notre modélisation des connaissances pouvait être utilisée efficacement, une première série de test a été effectuée : celle-ci consistait à faire reconnaître par le système des objets sémantiques extraits manuellement d'une image segmentée (Zlatoff, 2003). Cette extraction manuelle simplifie la procédure car elle considère que la fusion des objets d'intérêt a déjà été effectuée par l'opérateur.

Pour exploiter les connaissances, il est nécessaire de disposer d'un mécanisme de raisonnement. Nous avons choisi d'utiliser un système expert car son moteur de raisonnement est indépendant de la base de connaissances. Un système expert dispose d'une base de faits qu'il cherche à unifier à une série de règles.

Une règle est du type

SI <condition> ALORS <action>

<action> pouvant conduire à l'affirmation de nouveaux faits et permettre ainsi d'inférer de la connaissance.

Jess est un système expert qui peut s'intégrer à *Protégé* sous la forme d'un *plug-in*, réalisant ainsi une correspondance (*mapping*) en temps réel entre les instances et classes de la base de connaissance et les faits.

Pour notre application, une vingtaine de règles ont été développées. Ces dernières sont des règles génériques de reconnaissance d'objets sémantiques, dans la mesure où elles ne dépendent pas du domaine modélisé. A titre d'exemple, une des règles s'articule de la façon suivante :

```

Soit
  OI_1, instance de ObjetIntérêt,
  OI_2, instance de ObjetIntérêt,
  OI_1 adjacente supérieure à OI_2.
Si
  OI_1 représente OS_1 de la classe C_OS_a
  Il existe Adj_1, instance de RelationAdjacence,
{Adj_1 :
  élément supérieur : classe C_OS_a
  élément inférieur : classe C_OS_b
}
Alors OI_2 représente une instance OS_2 d'objet sémantique de la classe C_OS_b.

```

Dans la plupart des cas, il n'est pas possible de conclure directement qu'un objet d'intérêt désigne un objet sémantique. L'objet d'intérêt est alors considéré comme candidat pour représenter un ou plusieurs objets sémantiques. Dans un deuxième temps, des règles cherchent s'il existe des paramètres discriminants pour les candidats et les appliquent éventuellement.

Les différentes propriétés des objets d'intérêt utilisées pour ces tests sont les suivantes : taille (en pourcentage de la surface totale de la stèle), forme, autres objets adjacents (haut et bas), autre objet dans lequel il est directement inclus, autres objets dont il est voisin (nord et sud).

Des tests ont été menés sur 15 images représentatives désignées par les experts de la MOM et ont permis, pour chacune d'elles, d'identifier entre 60% et 100% des objets sémantiques.

6. Vers un pilotage de la fusion sémantique

Nous disposons ainsi à ce stade d'un module de traitements bas-niveau (segmentation en régions et fusion bas-niveau) qui permet d'aboutir à une tessellation d'objets d'intérêt. En outre, les connaissances du domaine sont formalisées et ont déjà servi à résoudre des problèmes d'interprétation simplifiés.

Nous présentons maintenant la structure que nous proposons pour piloter la fusion sémantique et les premiers tests que nous avons effectués.

6.1. Extraction de la description intermédiaire

Cette partie détaille la description que nous cherchons à extraire à partir du niveau intermédiaire. Puisqu'il va s'agir de mettre en correspondance les connaissances haut-niveau avec cette description, cette dernière doit comprendre les relations topologiques et directionnelles entre les objets d'intérêt, leur taille et leur forme.

Classiquement, la topologie se définit comme la géométrie qui s'intéresse, sans aucune métrique, à des notions comme la connectivité et la contiguïté entre objets. Le modèle des 4-intersections (Egenhofer, 1990) fait référence pour décrire ces relations. Il aboutit, si on ne considère que des objets régions à 6 cas: *disjoint*, *dans*, *touche*, *égal*, *couvre* et *recouvre*. Néanmoins, le niveau intermédiaire étant composé d'une tessellation d'objets, seuls trois cas subsistent : *disjoint*, *dans*, *touche*. Cette topologie est décrite par un graphe d'adjacence.

Dans un deuxième temps, nous recherchons la position relative des objets en contact (haut/bas). Puisque les objets d'intérêt sont déjà issus d'une fusion bas-niveau de régions, ils présentent une certaine convexité et peuvent raisonnablement être approximés par leur rectangle englobant. Afin de lever les fréquentes ambiguïtés se produisant lorsque les rectangles englobants se chevauchent, Goyal (1997) propose une matrice de relation-direction, qui repose sur la partition de l'espace autour du rectangle englobant, définissant ainsi 9 quadrants. En remplissant chacun de ceux-ci, par l'aire interceptée du deuxième objet, on caractérise la relation. Le Tableau 2 donne la matrice de direction pour le cas de la Figure 8 (pour des raisons de clarté, les objets A et B ne sont pas adjacents). A partir de cette matrice, en considérant les quadrants qui interceptent l'aire principale de l'objet B, on caractérise la relation en *haut* ou *bas*.

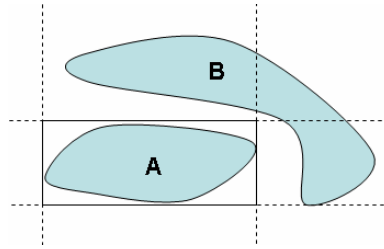


Figure 8- exemple de 2 objets

0	X%	Y%
0	0	Z%
0	0	0

Tableau 2- Matrice de direction associée à A

Enfin, un descripteur de forme est calculé pour chacun des objets d'intérêt. Dans le domaine modélisé, seuls cinq cas sont exploités : *rond*, *rectangle*, *triangle*, *ogive*, *autres*.

Comme les cas *rond* et *triangle* sont très caractéristiques (leur forme d'apparition dans l'image varie peu), des outils simples liés à la convexité ou comme ART (Kim, 1999) sont utilisés pour les reconnaître.

6.2. Fusion sémantique

Afin de valider la pertinence de la modélisation des connaissances, un premier test a été effectué, toujours avec le moteur *Jess*. Il consiste à lancer une fusion sémantique sur une image composée d'objets intermédiaires. Cette fusion se déroule en trois étapes successives : initiation, propagation des connaissances puis relaxation.

La phase **d'initiation** consiste à fournir des points de départ au système. Elle extrait de l'ontologie une série d'objets sémantiques nécessairement présents dans chaque image de stèle et recherche, parmi eux, l'élément supérieur *OS_sup* et inférieur *OS_inf*. Dès lors, l'objet d'intérêt supérieur de l'image est désigné candidat pour représenter une partie de *OS_sup* et l'objet d'intérêt inférieur est désigné candidat pour représenter une partie de *OS_inf*.

La phase de **propagation** consiste à exploiter au maximum les connaissances domaine pour pouvoir assigner à chaque objet d'intérêt une liste d'objets sémantiques qu'il sera susceptible de représenter (tout ou partie). Cette phase utilise les relations d'adjacence et d'inclusion sur le même modèle que décrit au paragraphe 5.4. Néanmoins, il est nécessaire de bloquer à un certain stade cette propagation des connaissances, si l'on ne veut pas, à l'issue de la propagation, que n'importe quel objet d'intérêt soit candidat pour représenter n'importe quel objet sémantique. Ainsi, les règles de propagation sont classées en catégories (types de relations utilisées, types d'hypothèses) et une seule application par objet d'intérêt d'un type donné de règle est autorisée.

Enfin, la phase de **relaxation** élimine les incohérences éventuelles obtenues. Elle utilise également les relations d'adjacence et d'inclusion, mais permet aussi, parmi une liste de candidats, de rechercher des paramètres discriminants (taille minimale ou maximale, forme) et de les appliquer.

Les règles utilisées (2 pour l'initiation, 8 pour la propagation, 10 pour la relaxation) s'appuient en outre sur 15 fonctions de manipulation de la connaissance. Ces règles et fonctions sont totalement indépendantes du domaine considéré : elles ont pour caractéristiques d'extraire de l'ontologie la connaissance nécessaire à leur tâche.

Cette fusion sémantique ne peut toutefois pas aboutir à une interprétation complète des objets d'intérêt. En effet, la description intermédiaire est relativement pauvre (une relation d'adjacence est réduite à deux cas : haut et bas, sans prendre en compte la taille et la forme de la frontière par exemple). Ainsi, ce type d'application ne peut servir qu'à une étape préliminaire qui donnerait, par exemple, pour chaque objet d'intérêt une liste d'objets sémantiques qu'il est susceptible de représenter (en tout ou partie).

La Figure 5 donne un exemple de résultats obtenus en présentant en (a) la stèle initiale qui comporte trois objets blocs (*couronnement ogival*, *geison*, *corps*) et trois objets de finition (deux *rosettes*, une *inscription*). L'image (b) est le résultat de la fusion bas-niveau. On constate que le geison a été fusionné dans le couronnement.

Le Tableau 3 présente en ligne supérieure, les objets sémantiques extraits par les experts et sur la ligne inférieure, le résultat de l'interprétation automatique des objets d'intérêt. Tous les objets que le processus automatique a été capable d'identifier ont été jugés valides par les experts.

Certains objets d'intérêt sont déclarés ambigus, c'est-à-dire qu'ils peuvent représenter une liste d'objets sémantiques. Néanmoins, pour chacun, l'objet sémantique qu'il désigne se trouve dans la liste de candidats.

Segmentation des experts					
	Couronnement	Geison	Corps	Rosettes	Inscriptions
Interprétation automatique		Fusionné au couronnement			Ambigu

Tableau 3- résultats de la fusion sémantique

7. Conclusion et perspectives

Nous avons proposé une méthodologie globale d'interprétation d'images, qui s'appuie sur des connaissances domaine. Nous avons également montré comment ces connaissances domaines devaient être traduites en des connaissances de scène afin de pouvoir être mises en correspondance avec la description du niveau intermédiaire. Nous avons implanté et validé des outils bas-niveau de segmentation et de fusion qui nous fournissent une tessellation d'objets d'intérêt. Enfin, nous avons réalisé les premières exploitations des connaissances pour réaliser des interprétations « simples » et une fusion sémantique.

Pour l'instant, la fusion sémantique n'est pas complète et ne peut constituer qu'une étape préliminaire, soit à une intervention humaine (expert) soit à des traitements complémentaires. En effet, le contrôle de l'application des règles se révèle rapidement complexe et il sera peut-être plus judicieux d'utiliser, par la suite, une méthode de fusion plus procédurale : fusion itérative d'objets d'intérêt jusqu'à un critère d'arrêt puis propagation par exemple.

Dans une perspective de contrôle des résultats, l'ontologie domaine peut permettre de vérifier si la connaissance instanciée à l'issue de l'interprétation est en tous points conforme aux contraintes du modèle générique. Ainsi, si la fusion bas-niveau a, de manière erronée, fusionné deux objets sémantiques entre eux, et que l'un des deux est obligatoirement présent, le système est capable de préciser que la connaissance instanciée est incohérente. Dans cas là, il est ainsi possible de remettre en cause les traitements précédents.

Enfin, cette méthodologie vise à plus long terme à être étendue à tout domaine puisque seule l'ontologie est dépendante du domaine (les règles et la fusion bas-niveau pouvant s'interfacer à n'importe quelle ontologie). Les domaines concernés sont ceux qui peuvent être modélisés de manière *pertinente* par des connaissances de scène (relations d'adjacence, inclusion, de voisinage et critères bas-niveau). A ce titre,

nous envisageons de formaliser d'autres ontologies pour, par exemple, interpréter des photos aériennes et les confronter à des travaux précédents (Tellez, 1999).

8. Références

Crevier D. & Lepage R. (1990). Knowledge-Based Image Understanding Systems: A Survey. *Computer Vision and Image Understanding*, Vol. 67(2), pp. 161-185.

Draper B.A, Colins R, Brolio J, et al. (1989). The Schema System, *International Journal of Computer Vision*, Vol. 2(3), pp 209-250.

Egenhofer M. & Herring J. (1990). A Mathematical Framework for the Definition of Topological Relationships. In: *Proceedings of the 4th International Symposium on Spatial Data Handling* (pp. 803-813). Zurich.

Gersho A. & Gray R. (1991). *Vector Quantization and Signal Compression*, Boston, Kluwer Academic Publishers, 1992.

Goyal R. & Egenhoffer M. (1997). *The Direction-Relation Matrix: A representation for Direction Relations between Extended Spatial Objects*, University of Maine, 1997.
http://www.spatial.maine.edu/~rgoyal/ucgis/rkg_mje.html

Gruber T.R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, Vol. 5(2), pp. 199-220.

Hudelot C. & Thonnat M. (2003). An Architecture for Knowledge Based Image Interpretation. In: *Workshop on Computer Vision System Control Architectures*, Graz, Austria.

Hyvönen E, Styrman A, Saarela S. (2002). Ontology-Based Image Retrieval. In: *Proceedings of XML Finland 2002 Conference* (pp. 15-27), Helsinki.

Idrissi K, Lavoué G, Ricard J, et al. (2003). An Object of Interest based Visual Navigation, Retrieval and Semantic Content Identification System. *Computer Vision and Image Understanding: Special Issue on Colour for Image Indexing and Retrieval*, in press.

Kim W. Y. & Kim Y. S. (1999). A new region-based shape descriptor. *Technical Report, 15-01*.

Laws K. (1980). Textured Image Segmentation. Ph.D. Dissertation. University of Southern California, USA.

Marino Drews O. (1993). Raisonnement classificatoire dans une représentation à objets multi-points de vue. Thèse Informatique. Université J. Fourier, Grenoble.

Matsuyama T. & Hwang V. (1990). *SIGMA: A Knowledge-Based Aerial Image Understanding System*, New-York, Plenum.

Minsky M. (1975). A Framework for Representing Knowledge. *The Psychology of Computer Vision*, Edited by P. Winston. New-York: McGraw-Hill, pp. 211-281.

Niblack W, Barber R, Equitz W, et al. (1993). The QBIC Project: Querying Images by Content Using Color, Texture and Shape. In: *Proceedings of Storage and Retrieval for Image and Video Databases* (pp. 173-187), Bellingham, WA.

Noy N, Fergerson R, Musen M. (2000). The Knowledge Model of Protégé-2000: Combining Interoperability and Flexibility. In: *Proceedings of 12th European Workshop on Knowledge Acquisition, Modeling and Management* (pp. 17–32). Springer Verlag.

Noy N. & McGuinness D. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. *SMI Technical Report*.

Roussey C. (2001). Une méthode d'indexation sémantique adaptée au corpus multilingues. Thèse Informatique. Insa de Lyon, Lyon.

Schreiber A, Dubbeldam B, Wielemaker J, et al. (2001). Ontology-based Photo Annotation. *IEEE Intelligent Systems*, Vol. 16(3), pp. 66-74.

Schettini R. (1993). A Segmentation algorithm for color images. *Pattern Recognition Letters*, Vol. 14(6), pp. 499-506.

Smith J. & Chang S. (1996). Tools and Techniques for Color Image Retrieval. In: *Proceedings of The Symposium on Electronic Imaging: Science and Technology Storage and Retrieval for Image and Video Databases IV* (pp. 426-437).

Uschold M. & King M. (1995). Towards a Methodology for Building Ontologies. In: *Workshop on Basic Ontological Issues in Knowledge Sharing*, Montreal, Canada.

Tellez B. (1999). Méthodologie pour une structuration commune des photos aériennes et des plans cadastraux : application à la révision des bases de données urbaines. Thèse de Doctorat. Institut National des Sciences Appliquées de Lyon. Lyon.

Zlatoff N. (2003). Exploitation de connaissances domaine pour l'analyse et l'interprétation d'image. Application à l'indexation sémantique. Rapport de DEA Informatique. Université C. Bernard, Lyon.