

Vers une réduction du fossé sémantique dans le traitement des images de documents anciens à base d'ontologies : Application aux lettrines

Mickaël Coustaty*, Norbert Tsopze**, Alain Bouju*, Karell Bertet*, Georges Louis*

*Laboratoire L3i - Université de La Rochelle
Avenue Michel Crepeau, 17042 La Rochelle, France
{mcoustat, abouju, kbertet, glouis}@univ-lr.fr
<http://L3i.univ-larochelle.fr>

**Département d'Informatique - Université de Yaoundé I
BP 812 Yaoundé - Cameroun
tsopze@uy1.uninet.cm
<http://www.une-autre-page.html>

Résumé. Nous présentons dans cet article une approche de réduction du fossé sémantique dans le traitement des documents anciens, et plus particulièrement les images de lettrines. Cette approche modélise les connaissances de deux domaines (experts historiens et experts en traitement d'images) au sein d'ontologies. Cette démarche permet d'établir des liens entre ces domaines afin d'aider les historiens à interpréter ces images et à les situer dans le temps. Nous avons combiné trois ontologies (thésaurus défini par les historiens, ontologie du traitement d'image et ontologie spatiale) pour l'annotation de ces images; ainsi nous avons donc défini des règles permettant d'annoter certaines régions de l'image comme étant la lettre, une partie du corps de personnage ou encore de caractériser une lettrine comme figurative. Des expérimentations prouvent la pertinence du système dans l'annotation automatique d'images. Par ailleurs, la généralité de l'approche permet d'en envisager une exploitation dans d'autres contextes combinant analyse d'image, analyse spatiale et sémantique du domaine, comme par exemple des images de bandes dessinées Tsopze et al. (2012).

1 Introduction

De grandes campagnes de numérisation sont réalisées depuis près de dix ans à travers le monde entier, générant de grands volumes d'images de documents patrimoniaux. Ces documents représentent la mémoire de nos sociétés, et de nombreux historiens s'intéressent à la création et la vie de ces documents pour retracer leur histoire et celle de la France. Il devient donc nécessaire de proposer des outils adaptés pour rechercher leur contenu, et ces grands volumes imposent d'être capable de les organiser pour permettre une recherche rapide et dédiée.

Dans le cas des images de documents anciens, cette tâche devient difficile puisqu'il faut pouvoir mettre en concurrence des données textuelles avec le contenu des images (les pixels).

Afin de pallier ce problème, un domaine très actif en recherche (généralement connu sous le nom de Content-based Image Retrieval ou CBIR (Smeulders et al., 2000; Jeon et al., 2003; Liu et al., 2007)) vise à décrire et résumer le contenu des images en calculant des descripteurs sur l'image. Ces descripteurs extraient des informations bas niveau sur l'image (couleur moyenne, nombre de formes contenues, taille des formes, etc) et rechercher une image consiste alors à retrouver celles contenant des informations bas niveau similaires. Les descripteurs d'images présentent l'avantage de permettre une comparaison entre les images, mais ne permettent pas de comparaison directe avec une requête d'un utilisateur, c'est le problème de *fossé sémantique*. Il paraît donc nécessaire d'associer des mots du langage naturel aux images pour permettre une recherche naturelle pour l'utilisateur.

Plus spécifiquement, nous nous intéressons dans cet article aux images patrimoniales, et ces travaux visent à adresser le problème de fossé sémantique pour les images de lettrines issues de documents anciens, et à l'utilisation des ontologies qui semblent être une solution prometteuse pour définir une sémantique explicite entre concepts visuels et langage naturel. Ainsi, deux ontologies sont définies : une associée aux experts de la Renaissance ; une associée aux traitements des images de documents anciens et qui réutilise une ontologie spatiale. Ces ontologies sont peuplées à partir des connaissances des historiens et de traitements d'images appliqués sur les images, et sont enrichies à partir de règles d'inférences. Ces règles sont alors utilisées pour réduire le fossé sémantique, et proposer une annotation automatique de certaines zones de l'images par les mots-clés des historiens.

Ainsi, cet article commence par une présentation détaillée du problème de fossé sémantique. Puis, dans un second temps, nous présentons le cas d'étude illustratif : les images de lettrines issues documents anciens, pour proposer des outils de navigation et de recherche d'images de la Renaissance. Enfin dans une quatrième partie, les caractéristiques propres à ce cas d'étude sont identifiées dans le but de positionner notre approche dans un cadre générique plus large d'aide à la réduction du fossé sémantique.

Par ailleurs, la généralité de l'approche permet d'en envisager une exploitation dans d'autres contextes combinant analyse d'image, analyse spatiale et sémantique du domaine, comme par exemple des images de bandes dessinées Tsopze et al. (2012).

2 Fossé sémantique, ontologies et images

2.1 Fossé sémantique et connaissance du domaine

Afin de répondre à la croissance rapide des contenus multimédia disponibles, plusieurs tentatives ont abordé le problème très connu du fossé sémantique (Smeulders et al., 2000) pour proposer une analyse et une interprétation sémantique des images.

Clouard et al. (2010) proposent de réduire le fossé sensoriel et le fossé sémantique. Le *fossé sensoriel*, peut être vu comme l'écart existant entre la scène réelle et l'image acquise, tandis que le *fossé sémantique* correspond à l'écart entre la représentation de l'image par les utilisateurs et sa description basée sur des caractéristiques bas niveau.

Toujours à la recherche de modèles qui pourraient aider à mettre en corrélation les caractéristiques de bas niveau et la sémantique de haut niveau, des approches se sont basées sur les "connaissances contextualisées". En construisant des hiérarchies sémantiques ou en intégrant des connaissances à priori, on peut améliorer l'analyse des images et leur interprétation.

Ces approches partent du principe que les objets du monde réel sont toujours associés à leur contexte, et la représentation de ce contexte est essentielle pour l'analyse et la compréhension des images. Les connaissances contextuelles peuvent provenir de multiples sources d'information, et c'est la complexité de ces informations qui permet d'obtenir la description la plus fidèle possible. Dans un contexte particulier, l'introduction de ce type de connaissances aide à raisonner et à améliorer l'annotation d'images (Neumann et Möller, 2008; Bannour et Hudelot, 2011, 2012).

Sur la base de techniques en relation avec l'ingénierie des connaissances, une approche intéressante proposée par Newell (1982) définit, lors de la conception d'un système à base de connaissances, le niveau "connaissance" comme "*un niveau abstrait qui traite du contenu de la connaissance. A ce niveau, il s'agit de discuter du contenu de la connaissance qui doit être exprimée, et non de la manière dont la connaissance doit être codée dans un langage de représentation*". Le niveau connaissance de Newell peut être décliné en quatre catégories que sont : (i) les connaissances descriptives (relatives au domaine concerné), (ii) les connaissances opératoires (relatives aux opérateurs de traitement d'image), (iii) les connaissances stratégiques (relatives à la coordination des opérateurs) et enfin (iv) les connaissances système (relatives à l'implémentation "génie logiciel" d'une plate-forme d'interprétation).

Par conséquent, il est important d'utiliser des méthodes explicites et formelles pour représenter les connaissances. Cela permet de prendre en compte la connaissance associée au contexte général (connaissance du domaine), au contexte spécifique (connaissance issue des images), et d'améliorer l'interprétation de l'image, tout en s'affranchissant de l'implémentation utilisée.

De nombreux systèmes combinent plusieurs de ces techniques pour proposer des moteurs de recherche d'images à l'aide de mots-clés sémantiques. Parmi les représentations des connaissances du domaine existantes, les ontologies présentent de nombreux avantages. Elles fournissent un *cadre formel et explicite d'une conceptualisation partagée* (Gruber, 1995). Ainsi, les ontologies capturent la connaissance pertinente d'un domaine, fournissent une compréhension commune des connaissances de ce domaine, déterminent le vocabulaire reconnu du domaine, donnent la définition explicite de ce vocabulaire et des relations entre les termes du vocabulaire, et tout cela sous forme de modèles formels.

Le problème de recherche d'images de documents anciens par leur contenu impose deux contraintes majeures. Premièrement, il faut concevoir un système qui puisse modéliser et structurer les connaissances issues du domaine des historiens (concepts sémantiques), et celles issues du domaine du traitement des images (caractéristiques bas niveau). D'autre part, il faut que ce système puisse répondre au problème de fossé sémantique, et faire le pont entre les caractéristiques bas niveau et les concepts sémantiques utilisés pour l'interrogation par les historiens. Nous avons donc choisi d'utiliser le formalisme des ontologies pour notre approche, que nous présentons dans la suite de cet article.

2.2 Ontologies et annotation d'images

L'annotation automatique d'images, qui a été un des défis de cette dernière décennie, a été introduite au début des années 2000, et les premiers travaux se sont orientés vers les approches d'apprentissage statistiques et probabilistes. Ces approches fournissent des outils puissants et

efficaces pour créer des associations entre les caractéristiques visuelles et les concepts sémantiques (Jeon et al., 2003; Barnard et al., 2003; Maillot et al., 2004; Fan et al., 2008).

Cependant, même si les approches statistiques décrivent bien le contenu visuel des images, la sémantique est souvent limitée à sa manifestation perceptuelle au travers de l'apprentissage d'une fonction de correspondance associant les caractéristiques de bas niveau à des concepts visuels de plus haut niveau sémantique, sans permettre de modéliser le sens des images tel qu'il est perçu par les humains. Ces approches présentent également plusieurs limitations dans le cas où elles sont appliquées à de grandes bases d'images Liu et al. (2007) puisque les performances varient en fonction du nombre de classes considérées et des images traitées. Ces variations de performances peuvent s'expliquer par la grande variabilité qui réside en intra-classes dans le cas de grandes bases d'images (il peut être difficile de représenter toutes les formes de bâtiments à l'aide de caractéristiques visuelles), et des similitudes qui peuvent apparaître entre des classes. Ces disparités intra-classes, et ces similitudes inter-classes, peuvent conduire à des situations conflictuelles dans le processus d'annotation, et prouvent le manque de corrélation entre les caractéristiques sémantiques et bas niveau.

D'autres travaux de recherche proposent des études des techniques de réduction du fossé sémantique. Plusieurs catégories de techniques (Kompatsiaris, 2008) dans la résolution du fossé sémantique y sont identifiées, allant de l'utilisation des ontologies aux techniques d'apprentissage ou d'annotation automatique. L'annotation du contenu peut se faire de différentes manières : annotation libre où aucun vocabulaire n'est prédéfini à l'avance et l'utilisateur se sert de ses propres connaissances pour annoter ; annotation par les mots-clés où l'utilisateur se sert d'un vocabulaire (sans aucune relation entre les mots) prédéfini pour annoter ; l'annotation par une ontologie où il existe une hiérarchie entre les mots proposés à l'utilisateur.

Lavrenko et al. (2003) ont été les premiers à utiliser l'information issue du contexte (contexte perceptif) dans l'annotation d'image. Ils ont proposé un modèle statistique génératif qui estime la probabilité d'associer des mots avec les régions de l'image. Ainsi, une image est décrite par l'ensemble de ses régions et de ses mots-clés, ce qui lui confère une "sémantique". Cependant, cette méthode ne capture pas vraiment la sémantique des images, mais utilise plutôt des statistiques sur le contexte pour améliorer la description de l'image.

Zhang et al. (2012) dénombrent également un ensemble de méthodes pour annoter différentes régions de l'image en fonction de caractéristiques extraites. Ces méthodes basées sur l'apprentissage automatique (Support Vector Machines, arbres de décision, réseaux de neurones artificiels, réseaux bayésiens) permettent de classer les images (ou les régions) suivant les concepts préalablement appris mais se heurtent aux problèmes de l'apprentissage (temps élevé, nombre de classes limité et fixé à l'avance,...).

Dans (Tousch et al., 2012), les auteurs soulignent la place de l'utilisateur, son objectif ainsi que le contexte dans l'annotation des images. Les niveaux de généralité, de précision et le choix des mots du vocabulaire dépendent de l'application et des connaissances de l'utilisateur. Différents niveaux d'abstraction peuvent également être considérés dans l'annotation des images : contextuel, culturel, émotionnel, technique,...

Une méthode de construction automatique de la hiérarchie entre les concepts est proposée dans (Bannour et Hudelot, 2012) en combinant les informations visuelles, conceptuelles et contextuelles.

Dans (Neumann et Möller, 2008) les scènes (images) sont modélisées à l'aide de la logique de description, montrant ainsi son importance dans l'interprétation des scènes. Cette

étude note que la logique formelle permet d'éviter des erreurs qui peuvent survenir quand les connaissances et l'inférence sont construites de manière intuitive. Elle propose un modèle formel pour l'interprétation d'une scène et souligne l'importance des contextes spatial et temporel dans la tâche d'interprétation.

L'ontologie développée par Clouard et al. (2010) pour le traitement d'images compte 279 concepts, 42 rôles et 192 restrictions. Dans cette ontologie, les concepts sont organisés en différents niveaux : niveau physique pour traiter les informations allant de l'acquisition au stockage de l'image ; niveau perception pour traiter les primitives visuelles de l'image ; niveau sémantique qui établit les relations de généralisation et d'inclusion entre les concepts ; niveau tâche pour les procédures d'extraction d'objets et de détection des traits ; et niveau contrainte pour exprimer la robustesse et l'efficacité du système. Cependant il n'existe aucun lien permettant d'établir une correspondance entre les concepts décrits et les métadonnées de niveau sémantique. Le lien entre les connaissances utilisateur et les régions extraites de l'image n'est pas représenté car dépendant du contexte.

Les problèmes de vérité terrain (ou de base d'apprentissage pour les méthodes basées sur l'apprentissage automatique) et des mesures de qualité d'annotation se posent aux travaux cités précédemment. L'interprétation d'une image dépend des connaissances de l'utilisateur, du contexte et du degré de granularité de la connaissance. L'utilisation de l'ontologie a différents objectifs : description unifiée des caractéristiques de l'image, description visuelle des relations entre les caractéristiques (lignes, région,...), utilisation des informations contextuelles et enfin rapprochement du niveau visuel et du niveau sémantique. Plusieurs catégories d'ontologies (Hanbury, 2008) peuvent être distinguées dans l'annotation d'images :

Ontologie de haut niveau (créée à partir d'un thésaurus). Les images sont décrites par des métadonnées indépendantes du contenu de l'image (la date, le nom d'auteur, ...).

Ontologie de bas niveau. les images sont décrites par des métadonnées sur leurs caractéristiques de bas niveau (texture, couleur,...). L'ontologie représente alors les différentes méthodes d'analyse des régions de l'image.

Ontologie liant les éléments du bas niveau (texture, couleur,...) à ceux du haut niveau de l'image (voiture, bâtiment,...). Les régions de l'image sont décrites par leurs caractéristiques de bas niveau et l'ontologie permet de donner une signification à certaines régions. Cette catégorie permet de réduire le fossé sémantique.

Dans la première catégorie d'ontologies, on se contente d'associer à l'image un vocabulaire sans aucun traitement, les mots sont associés à l'image entière. Dans les deux dernières catégories les algorithmes de traitement d'images sont d'abord appliqués et l'ontologie est utilisée pour annoter les résultats de ces algorithmes, les mots sont alors associés au contenu ou à des régions de l'image.

Nos travaux se positionnent dans cet objectif de réduction du fossé sémantique. Nous avons pour cela choisi un contexte où la sémantique des images est clairement établie, celui des images de lettrines.

3 Cas d'étude : ontologie des lettrines

Nous avons défini une représentation ontologique des images de lettrines. Cette ontologie offre une représentation standardisée de données hétérogènes complexes décrivant ces images,

qui intègrent à la fois des caractéristiques numériques extraites de l'image et de certaines de ses régions d'intérêt, une information spatiale entre les régions, et la sémantique du domaine des historiens. Cette ontologie nous a également permis de réduire le *fossé sémantique* entre description pixellaire et description sémantique d'une image grâce à un mécanisme d'annotation automatique de certaines images ou de certaines régions de l'image à l'aide de propriétés déduites de l'ontologie. Un extrait de cette ontologie des lettrines est donné par la figure 2. Cette ontologie est accessible sur <http://www.navidomass.univ-lr.fr>. Dans la suite, nous utiliserons la légende présentée dans la figure 1 pour décrire les différentes ontologies. Ainsi, les classes seront représentées dans les figures par les symboles de la figure 1(a) (par exemple la classe "Image") et les relations entre elles par les flèches de la figure 1(b) : trait continu par la hiérarchie "is a" et trait discontinu pour les autres relations (par exemple les relations "hasLetter").

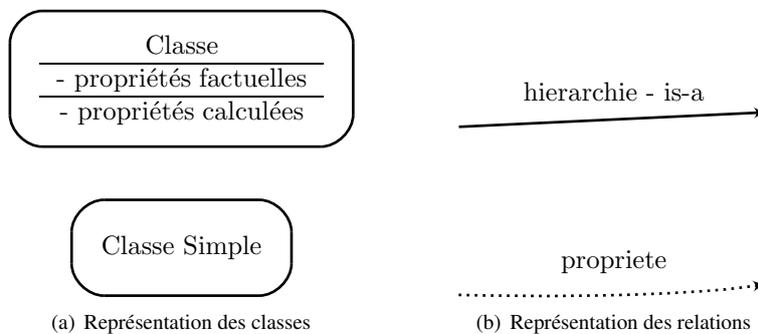


FIG. 1 – Description du formalisme utilisé pour décrire les ontologies.

Nous avons défini cette ontologie en plusieurs parties, illustrées par la figure Figure 3, chaque étape est décrite plus en détails dans les sections suivantes :

Ontologie des historiens (historian knowledge) (section 3.1) : la connaissance des historiens est décrite par un thesaurus (Jimenes, 2008) que nous avons transcrit en une ontologie possédant la classe *Lettrine* comme classe principale. Un fond de 4288 images de lettrines, annotées à la main par les historiens, est utilisé pour peupler cette ontologie.

Ontologie des régions d'intérêt (section 3.2) : cette ontologie, qui possède la classe *Image* comme classe principale, est composée de régions d'intérêt extraites à la fois de la couche forme et de la couche texture d'une image de lettrine, ainsi que de leurs caractéristiques bas niveau. Cette ontologie est peuplée par 909 images de lettrines décrites par 5588 régions de forme et par 451711 régions de traits. Nous avons enrichi cette ontologie des régions avec un mécanisme de partitionnement de l'image et des relations spatiales, dans le but de pouvoir localiser chacune des régions selon qu'elle se situe au centre de l'image, ou sur un bord.

Ontologie des lettrines (section 3.3) : l'ontologie des historiens fournit une représentation haut-niveau des lettrines, alors que l'ontologie des régions d'intérêt en est une représentation bas niveau. Nous les avons combinées en une seule ontologie des lettrines en reliant les classes principales des deux ontologies (classes *Lettrine* et *Image*), permettant

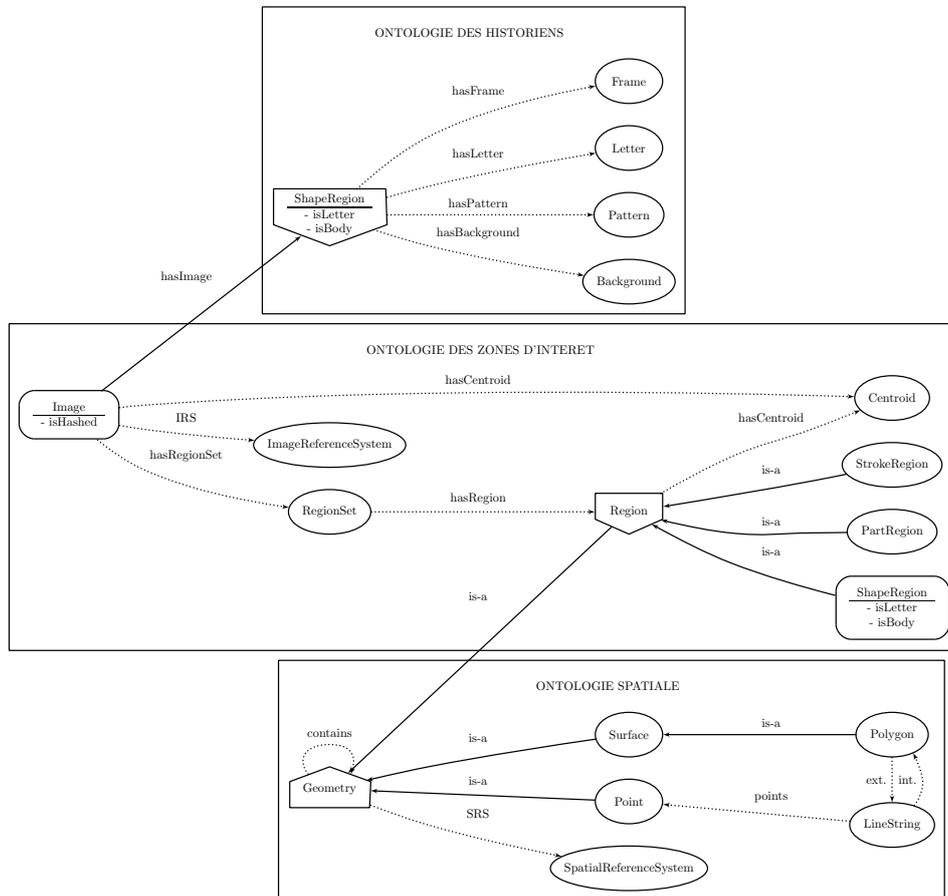


FIG. 2 – Extrait de l'ontologie des lettrines.

ainsi de représenter et d'interroger de façon homogène l'ensemble des données décrivant les lettrines. Nous avons ensuite cherché à réduire le fossé sémantique entre le niveau sémantique (haut-niveau) et le bas niveau par un enrichissement de l'ontologie reposant sur le mécanisme d'inférence.

Inférence sur les régions d'intérêt de la couche forme (section 3.4) : au niveau de la couche forme, nous avons ajouté des descriptions logiques de propriétés permettant d'identifier et d'annoter certaines régions forme comme la lettre de la lettrine (propriété *isLetter*), ou encore une partie d'un personnage du fond de la lettrine (propriété *isBody*).

Inférence sur les régions d'intérêt de la couche texture (section 3.5) : en ce qui concerne la couche texture, c'est l'ensemble des régions constitutives d'une image qui fournit une information pertinente. Nous avons utilisé les caractéristiques de l'ensemble des régions extraites pour déterminer si l'image est une lettrine à fond hachuré ou non (propriété

isHashed).

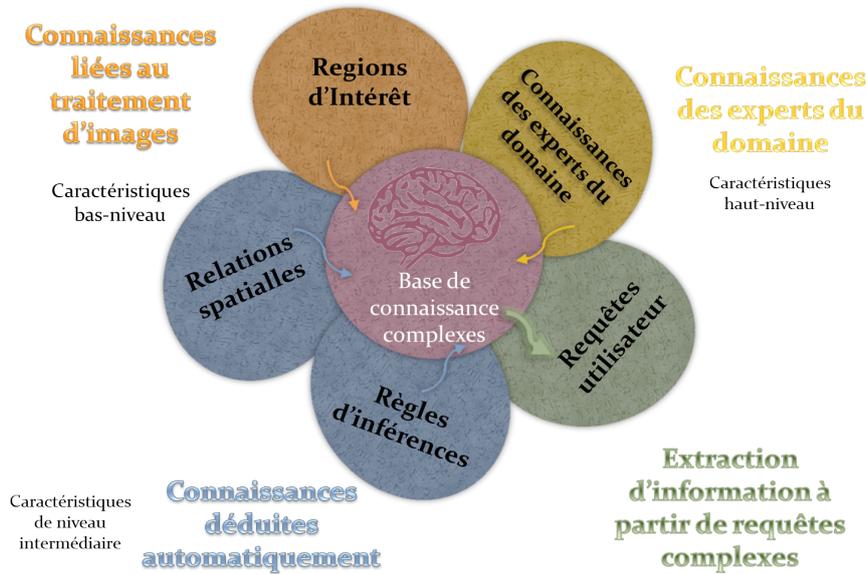


FIG. 3 – Représentation schématique de la base de connaissance complexe mise en place dans cet article sous forme de fleur ontologique des lettrines.

3.1 Ontologies des historiens

Les images de lettrines sont des images graphiques que l'on retrouve dans des documents anciens des *XV^e* et *XVI^e* siècles. La Figure 4 en donne un exemple. Il s'agit d'une lettre décorée, en début des paragraphes des livres de cette époque, obtenue à l'aide de tampon en bois, sculpté à la main. Composée principalement d'une lettre en son centre, une lettrine se caractérise également par un arrière-plan qui peut être décoratif, ou bien représenter des scènes sociales de l'époque (scènes figuratives). Les nuances et les ombres étaient obtenues à l'aide de traits parallèles (figure 4).

L'étude des lettrines fait l'objet de travaux au centre d'étude supérieur de la renaissance (CESR) de Tours. Les lettrines représentent une source d'information exploitée par les historiens pour situer les documents dans le temps, ou encore pour étudier les scènes sociales des fonds figuratifs. Afin d'imprimer ces lettrines, des tampons étaient utilisés, et l'usure associée à leur utilisation nombreuse permet de situer les documents les uns par rapport aux autres dans le temps. De plus, ils portent souvent des caractéristiques propres à celui qui les a sculptés.

La sémantique d'une image de lettrine décrite par un thésaurus issu de travaux historiens (Jimenes, 2008), est décomposée en quatre couches, chaque couche fournit une information spécifique (voir Figure 5) :

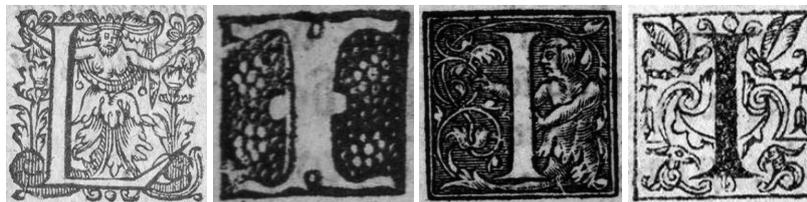


FIG. 4 – Exemples de lettrines.

Lettre (Letter) : positionnée au centre de l’image, la couche lettre caractérise en particulier la lettre qu’elle contient, sa couleur (noire ou blanche), l’alphabet (latin, grec, hébreu, ..) et la police utilisée (romain, gothique) ;

Motif (Pattern) : il est constitué des formes ornementales qui peuvent être décoratives ou figuratives (représentation d’une scène) ;

Fond (Background) : le fond peut être uniforme (noir ou blanc), hachuré ou criblé ;

Cadre (Frame) : enfin, le cadre correspond aux bords du tampon typographique. Il peut être composé de zéro, un ou deux traits.

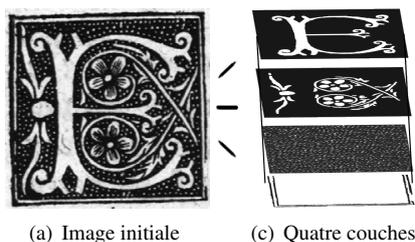


FIG. 5 – Décomposition en couches des historiens.

Nous avons retranscrit ce thesaurus en une ontologie des historiens. La Figure 6 décrit la T-Box (ensemble des concepts et de leurs propriétés dans l’ontologie) de cette ontologie. Elle est constituée d’une classe principale *Lettrine* reliée aux classes *Letter*, *Background*, *Pattern* et *Frame* représentant chaque couche sémantique par des propriétés (*hasLetter*, *hasBackground*, *hasPattern*, *hasFrame*). La classe *Letter* est quant à elle reliée aux classes *ColorLetter*, *TypeFont* et *Alphabet* par des propriétés. La propriété *IdentificationLetter* permet de préciser de quelle lettre il s’agit.

Nous avons ensuite peuplé la partie A-Box (ensemble des faits ou instances des concepts définis dans la TBox) de cette ontologie par 4288 images de lettrines annotées à la main par les historiens.

3.2 Ontologies des régions d’intérêt

Les images de lettrines sont des images particulièrement difficiles à traiter. En effet, ce sont des images qui se sont détériorées avec le temps - jaunissement du papier, pages abîmées,

Ontologies et Lettrines

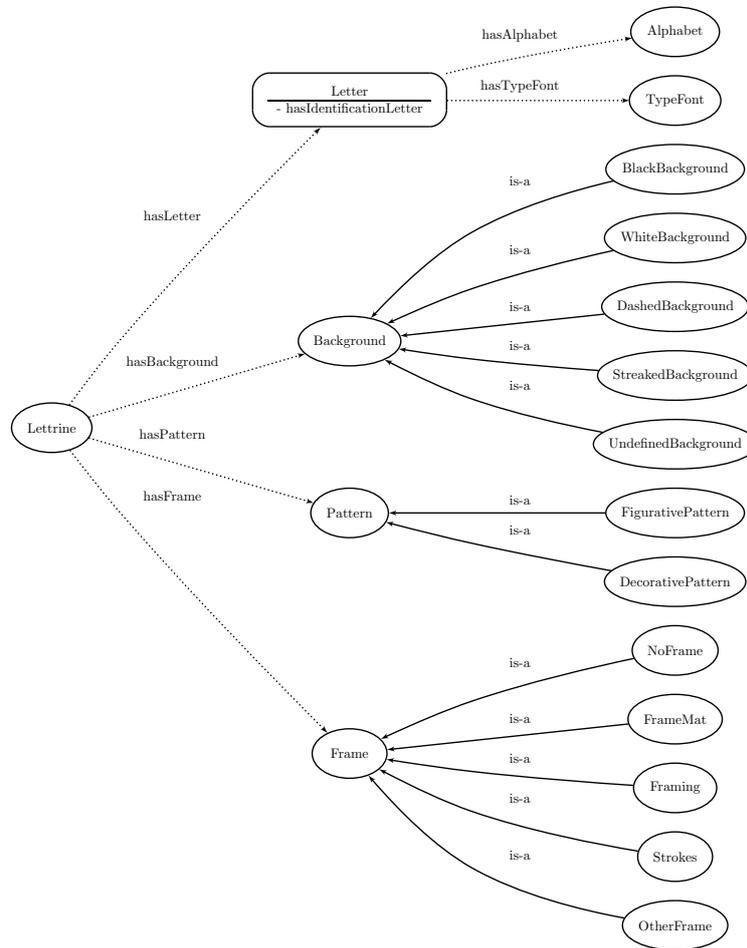


FIG. 6 – Extrait de l'ontologie des historiens (partie T-Box).

tâchées - et les techniques développées doivent être robustes à ces détériorations. Par ailleurs, il s'agit d'images binaires à base de traits sur lesquelles les approches texture classiques d'analyse d'images ne peuvent pas s'appliquer. C'est pourquoi une approche spécifique d'extraction d'information à partir d'images à base de traits a été proposée dans Coustaty (2011).

Tout d'abord, l'image est décomposée en plusieurs couches selon une décomposition décrite dans Dubois et al. (2008), résultat d'une série de projections (Hamidi et al., 2010). Cette décomposition est particulièrement pertinente pour l'analyse d'images de lettrines. En effet, on peut observer que la géométrie d'une image se retrouve principalement sur sa *couche forme*, alors que la *couche texture* permet d'isoler les traits qui la composent. Nous avons extrait des régions d'intérêt à partir de chacune de ces deux couches. Un exemple de décomposition est donné à la figure 7.

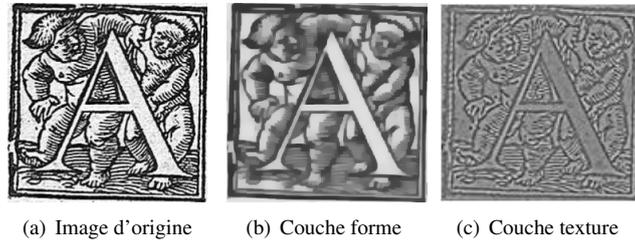


FIG. 7 – Décomposition en couches d'une lettrine : résultats obtenus par le traitement proposé par Coustaty (2011).

Régions d'intérêt de la couche forme. La lettre, ou encore des éléments de scènes naturelles des lettrines à fond figuratif, correspondent à des formes de l'image. C'est pourquoi nous avons mis en place une méthode d'extraction adaptée à la couche forme, décrite dans Coustaty et al. (2011) qui permet d'en extraire certaines régions forme constitutives de l'image. Ces régions d'intérêt sont obtenues en trois étapes :

1. Tout d'abord, l'utilisation d'une loi de Zipf (Pareti et Vincent, 2006) a été utilisée pour sa robustesse aux variations de niveaux de gris, et pour son indépendance à la couleur des composantes. Elle permet de segmenter la couche formes en composantes connexes.
2. Ensuite, les composantes connexes dont l'aire est supérieure à 1% de l'image sont retenues, les autres étant considérées moins pertinentes. Les composantes connexes ainsi retenues forment les régions d'intérêt de la couche forme.
3. Enfin, nous associons à ces régions d'intérêt des caractéristiques décrivant leur forme :
 - L'excentricité Ecc de la région (*Eccentricity*), donne une indication de forme. Elle est définie comme le rapport entre le rayon minimal r_m et le rayon maximal r_M de l'ellipse minimale englobant la région (Pratt, 2007) : $Ecc = \frac{r_M - r_m}{r_M + r_m}$
 - La moyenne en niveaux de gris (*GreyMean*), et leur déviation (*GreySTD*) permettent d'estimer la couleur de la région, ainsi que sa régularité.
 - Le nombre d'Euler E_n de la région (*Euler*) permet d'en estimer la compacité. Introduit dans Pratt (2007), il se calcule en fonction du nombre de trous H que contient une forme et se réduit à $E_n = 1 - H$ pour une seule région.

Ces régions d'intérêt ne sont pas obtenues par un traitement classique de segmentation d'image en régions, mais par une méthode adaptée décrite dans Coustaty (2011).

Régions d'intérêt de la couche texture. La couche texture se compose essentiellement des traits de l'image. Dans une image de lettrines, un ensemble des traits aux caractéristiques visuelles similaires (longueur, orientation, épaisseur, courbure) peuvent correspondre à un fond hachuré, ou encore à des zones d'ombres. Nous avons mis en place un traitement de la couche texture pour en extraire des zones constituées de traits similaires (Coustaty, 2011). Ces régions de traits sont obtenues en cinq étapes :

1. Nous appliquons tout d'abord une binarisation suivi d'un débruitage pour ne conserver que les traits dans l'image binarisée.

Ontologies et Lettrines

2. Une squelettisation nous permet ensuite d'extraire les traits qui peuvent être d'épaisseurs variables.
3. Nous avons ensuite défini des caractéristiques de traits prégnantes pour la vision humaine, à savoir la longueur, l'épaisseur, l'orientation, l'homogénéité et la courbure. Une classification non supervisée à partir de ces caractéristiques nous permet de créer des classes de traits similaires.
4. Enfin, des régions de traits sont obtenues en regroupant les traits aux caractéristiques similaires, et qui sont voisins dans l'image. Des caractéristiques sont ensuite recalculées pour l'ensemble des traits de chaque région : le nombre de traits d'une région (*StrokesNumber*); la longueur des traits (*StrokesLength*); l'épaisseur des traits (*StrokesWidth*); l'orientation des traits (*StrokesOrientation*); l'homogénéité (*StrokesHomogeneity*); la courbure des traits (*StrokesCurvature*).

Ontologie enrichie des régions d'intérêt. L'ontologie des régions d'intérêt est constituée, pour sa partie T-Box (cf. Figure 10), d'une classe principale *Image*. Des informations globales relatives à l'image sont introduites par les propriétés *ImageReferenceSystem*, *hasCentroid*, *hasLength* et *hasWidth*. Les couches forme et texture sont des images dérivées de l'image initiale, stockée dans la sous-classe *DerivatedImage*, alors que l'image initiale est stockée dans la sous-classe *SourceImage*

Les classes *ROI* (Region Of Interest) et *ROISet* permettent d'associer un ensemble de régions à une image. L'aire d'une région, ainsi que les coordonnées de son centre de gravité d'une région, permettant ainsi de la situer dans l'image, sont introduites par les propriétés *hasCentroid* et *hasArea*.

Pour distinguer les régions forme des régions traits, nous avons créé deux sous-classes de la classe *ROI*, à savoir les classes *ShapeROI* et *StrokesROI*. Chacune de ces classes est décrite par les caractéristiques des régions d'intérêt à l'aide de leurs propriétés.

1	2	3
4	5	6
7	8	9

FIG. 8 – *Partitionnement d'une image utilisée pour localiser les régions.*

Nous avons ensuite enrichi l'ontologie des régions d'intérêt par un mécanisme de partitionnement d'une image de lettrine : chaque image est partitionnée en neuf zones (voir figure 8), une zone correspond au centre de l'image, 4 partitions correspondent aux différents bords, et les 4 dernières aux coins de l'image. Les partitions sont des régions définies par la sous-classe *PartROI* de la classe *ROI*.

Les relations spatiales sont alors utilisées pour positionner une région dans un partitionnement. C'est sur les préconisations de la norme ISO sur la manipulation des informations spatiales (13249-3 :2002, 2002) que nous avons choisi d'utiliser l'algèbre RCC8 (Cohn et al., 1997)

qui propose huit types de relations spatiales entre deux objets spatiaux (cf. Figure 9), ainsi que le format SFS (Simple Feature Specification) pour représenter des objets spatiaux, standard dérivé des spécifications spatiales de SQL préconisées par le consortium OGC (OpenGIS Consortium) (Inc., 1999) (cf. Figure 11).

Ainsi, la classe *ROI* et ses sous-classes héritent de la classe abstraite *Geometry* du standard SFS. Les relations spatiales RCC8 sont des propriétés calculées permettant de comparer entre elles des régions d'intérêt. En particulier, elles permettent de situer des régions *ROIShape* et *ROIStrokes* dans une des neuf partitions *ROIPart*. Il est ici à noter que les relations spatiales entre une région d'intérêt et chacune des neuf partitions peuvent également se calculer lors de la phase d'extraction des régions, mais au détriment de l'aspect normatif.

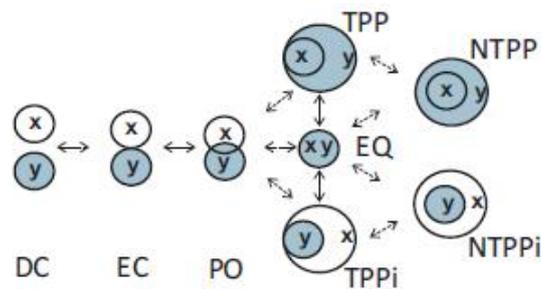


FIG. 9 – Relations spatiales de l’algèbre RCC-8 - schéma disponible sur le site wikipedia.org.

Nous avons sélectionné aléatoirement un jeu de test de 909 images de lettres à partir desquelles 5588 régions ont été extraites de la couche forme, et 451711 régions ont été extraites de la couche texture. La partie A-Box de l’ontologie des régions d’intérêt est peuplée par ces images et ces régions.

3.3 Ontologie des lettres

Les trois ontologies portent chacune une information pour une image de lettre : l’ontologie des historiens fournit une représentation sémantique des lettres, alors que l’ontologie des régions d’intérêt en est une représentation bas niveau. C’est pourquoi nous les avons combinées en une seule *ontologie des lettres* en liant la classe principale *Lettre* de l’ontologie des historiens à la classe principale *Image* de l’ontologie des régions d’intérêt pour la propriété *hasImage*. La T-box de cette ontologie regroupe ainsi l’ensemble des classes et propriétés des deux ontologies ainsi combinées. Par conséquent, chaque image de l’ontologie des régions d’intérêt est décrite à la fois par ses régions d’intérêt extraites (et leurs caractéristiques bas niveau), ainsi que par les annotations manuelles réalisées par les historiens. Notons cependant qu’une partie seulement des images annotées de l’ontologie métier sont présentes dans l’ontologie des régions d’intérêt. Par conséquent, certaines lettres de l’ontologie des historiens ne sont pas décrites par des régions d’intérêt.

Notre contribution à une réduction de ce fossé sémantique repose sur un enrichissement de l’ontologie par de nouvelles propriétés permettant d’annoter sémantiquement des images, ou

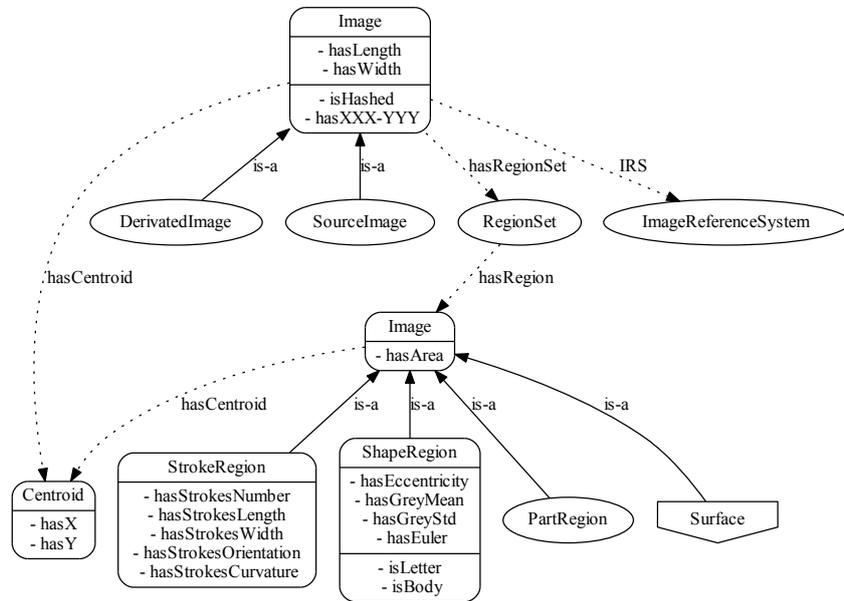


FIG. 10 – *Ontologie enrichie des régions d'intérêt (partie T-Box).*

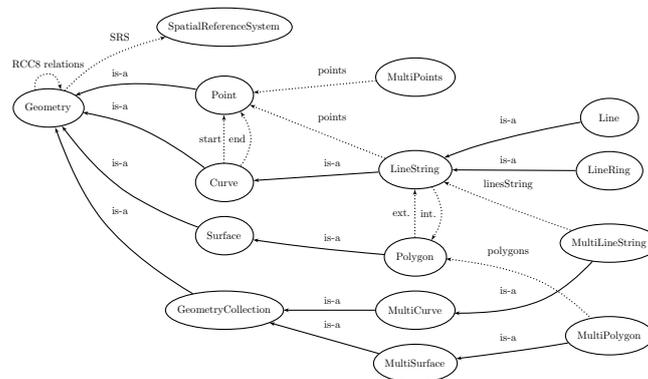


FIG. 11 – *Ontologie spatiale du standard SFS de l'OGC.*

des régions d'intérêt. Il s'agit de propriétés calculées, définies par une formule ou une règle logique, qui peuvent intégrer l'ensemble des données décrivant les images. Ces propriétés sont indiquées comme calculées sur le schéma de la figure 10.

3.4 Inférence sur les régions d'intérêt de la couche forme

La lettre, ou encore des éléments de scènes naturelles des lettrines à fond figuratif, correspondent à des formes de l'image. Plus précisément, la lettre correspond à une grande région d'intérêt de la couche forme située au centre de l'image et contenant peu de trous. C'est pourquoi nous avons cherché à identifier la région d'intérêt de la couche forme correspondant à la lettre à l'aide d'une propriété calculée *isLetter*, mécanisme d'annotation sémantique de certaines régions au sens image.

Par ailleurs, on observe que des parties de personnages des lettrines figuratives correspondent à de grandes régions d'intérêt de la couche forme que nous avons cherchées à identifier par la propriété calculée *isBody*. Il s'agit de régions suffisamment grandes qui ne touchent pas les bords de l'image et qui ne correspondent pas à la lettre.

Ces deux propriétés (cf. Figure 12) reposent sur les caractéristiques bas niveau des régions d'intérêt de la couche forme, ainsi que sur un partitionnement spatial de l'image. La propriété *isBody* s'exprime également en fonction de la propriété *isLetter* (une région lettre n'est pas une partie d'un personnage), et d'une information sémantique (seules des régions des lettrines figuratives sont à considérer).

Nous présentons ces deux propriétés et une évaluation de leur performance ci-dessous. Bien que ces travaux se situent dans un contexte de recherche d'information dans une base de documents, les deux propriétés présentées ont été évaluées en testant le taux d'erreurs générées par l'annotation automatique. Cette évaluation permet d'estimer la capacité du système à annoter automatiquement des images et à retrouver cette information.

Propriété *isLetter* : description et experimentation Cette propriété calculée, définie pour la classe *ShapeROI*, indique si une région est identifiée comme la lettre d'une lettrine. Elle s'exprime en fonction de quatre propriétés qui intègrent des caractéristiques des régions de la couche forme, mais également une information spatiale avec le partitionnement de l'image :

- (*isLetter*) : (1) zone d'aire maximale
 (2) située au centre de l'image (ie.contenu dans la partition au centre)
 (3) contenant peu de trous
 (4) contenant le centre de l'image de la lettrine
 dans son rectangle englobant.

1. *région d'aire maximale* : dans l'ontologie des régions, l'aire est indiquée pour chaque région (propriété *hasArea* de l'ontologie des régions). Parmi toutes les régions vérifiant les propriétés précédentes, celle d'aire maximale est retenue et annotée par la propriété *isLetter*.
2. *région située au centre de l'image* : il s'agit de vérifier que la région est contenue dans la partition centrale (propriété spatiale *contains*), et ne chevauche par les régions du bord. Comme l'image est partitionnée suivant la figure 8, cette propriété permet de la lettre se trouve dans la partition centrale (partition 5)
3. *région avec peu de trous* : le nombre d'Euler, associé à chaque région de la couche forme (propriété *hasEuler* de l'ontologie des régions), est ici utilisé. Après quelques expérimentations, nous avons positionné le nombre d'Euler entre -2 et +2 pour caractériser une région avec peu de trous.

4. *contenant le centre de l'image de la lettrine dans son rectangle englobant* : Cette propriété permet de préciser que le centre de l'image (propriété *hasCentroid* de l'ontologie des régions) doit appartenir au rectangle englobant de la région choisie, avec une marge de 15 pixels.

La propriété *isLetter* a été appliquée sur l'ensemble des 909 lettrines peuplant l'ontologie des régions. Après vérification manuelle, 816 régions s'avèrent être correctement annotées, et 103 ne le sont pas, d'où un taux d'erreur de 11%. Cependant, cette règle ne garantit pas l'existence d'une région annotée pour chaque lettrine.

Il serait ici possible de comparer la forme de la région avec celle attendue de la lettre, information présente dans l'ontologie des historiens (propriété *hasIdentificationLetter*). Une telle extension permettrait de détecter les cas où la lettre se décompose en plusieurs régions, et ainsi réitérer le calcul de la propriété *isLetter* jusqu'à ce que toutes les régions de la lettre soient détectées.

Propriété *isBody* : description et experimentation Cette propriété, définie pour la classe *ShapeROI*, indique qu'une région est identifiée comme la partie d'un personnage du fond d'une lettrine. Elle s'exprime en fonction de cinq propriétés qui intègrent des caractéristiques des régions forme, une information spatiale avec le partitionnement, ainsi qu'une information de l'ontologie du domaine :

(*isBody*) : (1) région d'une lettrine figurative
(2) située au centre de l'image,
(3) avec peu de trous,
(4) de couleur gris-clair,
et (5) qui n'est pas une lettre.

1. *région d'une lettrine figurative* : seules les régions des lettrines figuratives (propriété *hasPattern* de l'ontologie des historiens) sont considérées.
2. *région située au centre de l'image* : comme pour la propriété *isLetter*, cette propriété se teste à l'aide de la propriété spatiale *contains*.
3. *région avec peu de trous* : comme pour la propriété *isLetter*, il s'agit de tester le nombre d'Euler (propriété *hasEuler* de l'ontologie des régions) associé à chaque région dans l'ontologie des régions d'intérêt.
4. *région de couleur gris-clair* : la caractéristique colorimétrique d'une région (propriété *hasGreyMean* de l'ontologie des régions), doit être supérieure à 90 pour indiquer une couleur gris-clair. Le niveau de gris va de 0 noir à 255 blanc le plus intense.
5. *région qui n'est pas une lettre* : parmi toutes les régions vérifiant les propriétés précédentes, seules celles qui ne vérifient pas la propriété *isLetter* sont annotées.

La validation des régions annotées est un processus manuel. C'est pourquoi nous avons sélectionné aléatoirement pour la règle *isBody* un sous-ensemble de 45 images de lettrines. Cet ensemble est constitué de 27 lettrines à fond décoratif, 18 lettrines à fond figuratif, à partir desquelles 112 régions forme au total ont été extraites. La propriété *isBody* a permis d'annoter correctement des régions forme de 17 images sur 18 images figuratives, d'où un taux d'erreur de l'ordre de 2%.

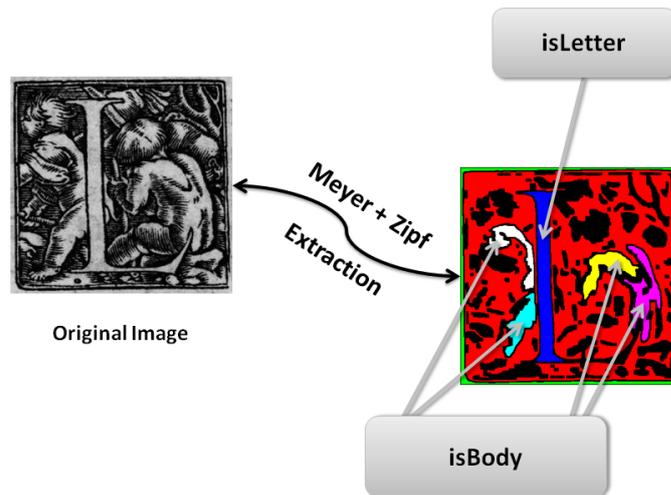


FIG. 12 – Extraction de formes d'une lettrine.

La figure 15 représente la lettrine figurative dont les régions ne sont pas correctement annotées. On peut y observer que la région jaune correspondant à une partie d'un personnage n'est pas annotée car son nombre d'Euler est trop grand. En revanche, la région blanche est annotée alors qu'elle ne correspond pas à une partie d'un personnage. L'utilisation d'un partitionnement plus complexe de l'image, ou encore d'autres descripteurs statistiques des régions, pourrait permettre d'annoter correctement les régions de ce type.

3.5 Inférence sur les régions d'intérêt de la couche texture

Un mécanisme d'inférence sur les régions d'intérêt de la couche texture a également été mis en place dans le but de réduire le fossé sémantique entre l'ontologie des régions d'intérêt et l'ontologie des historiens. On observe qu'une lettrine hachurée contient des régions de traits dont la superficie totale est suffisamment grande pour couvrir une bonne partie de l'image. Ce sont donc les caractéristiques de l'ensemble des régions de traits qui portent une information pertinente au niveau de l'image elle-même pour déterminer si elle est ou non hachurée. Nous avons cherché à identifier les images de lettrines hachurées à l'aide de la propriété calculée *isHashed* d'une image.

Pour cela, nous avons tout d'abord enrichi la description d'une image avec des propriétés décrivant l'ensemble des régions de traits qui la composent. La propriété *isHashed* permet de déterminer si une lettrine possède un fond hachuré à partir de caractéristiques bas niveau, et se vérifie en la comparant avec la propriété *hasBackground* de l'ontologie des historiens. C'est pourquoi nous avons pu utiliser une approche classique de classification supervisée (à savoir un

arbre de décision C4.5) pour déterminer les seuils des différentes caractéristiques permettant de décider si la lettrine est hachurée ou non.

La figure 13 présente l'arbre de décision simplifié construite à cet effet à l'aide de l'algorithme C4.5, où

SV : sont des étiquettes des branches de l'arbre conduisant à la classe *isHashed* et sont décrits par : S est un opérateur de comparaison qui peut être $<$, \leq , $>$ ou \geq ; et V est une valeur seuil de l'attribut de comparaison.

YX : sont des nœuds de l'arbre et X correspond à chacune des 6 propriétés spécifiques des régions de traits stockées dans la classe *StrokesROI*, à savoir *hasStrokesNumber*, *hasStrokesLenght*, *hasStrokesWirdth*, *hasStrokesOrientation* et *hasStrokesHomogeneity* ; Y correspond à la valeur minimale, maximale, moyenne, ou à l'écart-type (*Min*, *Max*, *Avg* et *Std*) de chacune des 6 caractéristiques de l'ensemble des régions de traits constitutives de l'image.

(*isHashed*) : comparaison des valeurs minimale, maximale, moyenne, ou d'écart-type des caractéristiques des régions de traits par rapport à des seuil (7 feuilles) obtenus par apprentissage

Cette règle a été appliquée sur l'ensemble des 909 lettrines peuplant l'ontologie des régions parmi lesquelles 140 sont hachurées. 126 images sont annotées automatiquement. Une vérification automatique est alors possible en interrogeant l'ontologie métier (des historiens) peuplée manuellement : 123 sont effectivement hachurées. D'où un rappel de $\frac{123}{140} = 87,8\%$, et une précision de $\frac{123}{126} = 97,8\%$.

3.6 Outils utilisés

Afin de proposer un outil performant tant du point de vue qualité que vitesse de traitement, nous avons utilisé PROTÉGÉ 3.4.4 (Knublauch et al., 2004), éditeur d'ontologies, pour construire cette ontologie, et l'API JENA (Jena, 2011), pour la peupler. Nous avons également utilisé l'outil DLV¹, base de données déductive, qui permet de définir les règles d'inférence et de les tester en ligne de commandes. Basé sur la programmation logique déductive, DLV est simple à utiliser, et offre une facilité d'expression des requêtes, avec des temps de réponse relativement faibles. Nous utilisons également SEWELIS (SEmantic WEB Logical Information System) (Ferré, 2009) pour une recherche d'information dans l'ontologie construite à la fois par requête et par navigation, et qui laisse envisager une meilleure interaction avec l'utilisateur.

4 Discussion

Le problème de recherche d'images de documents anciens par leur contenu impose deux contraintes majeures. Premièrement, il faut concevoir un système qui puisse modéliser et structurer les connaissances issues du domaine des historiens (concepts sémantiques), et celles issues du domaine du traitement des images (caractéristiques bas niveau). D'autre part, il faut que ce système puisse répondre au problème de fossé sémantique, et faire le pont entre les

1. <http://www.dlvsystem.com/dlvsystem/index.php>

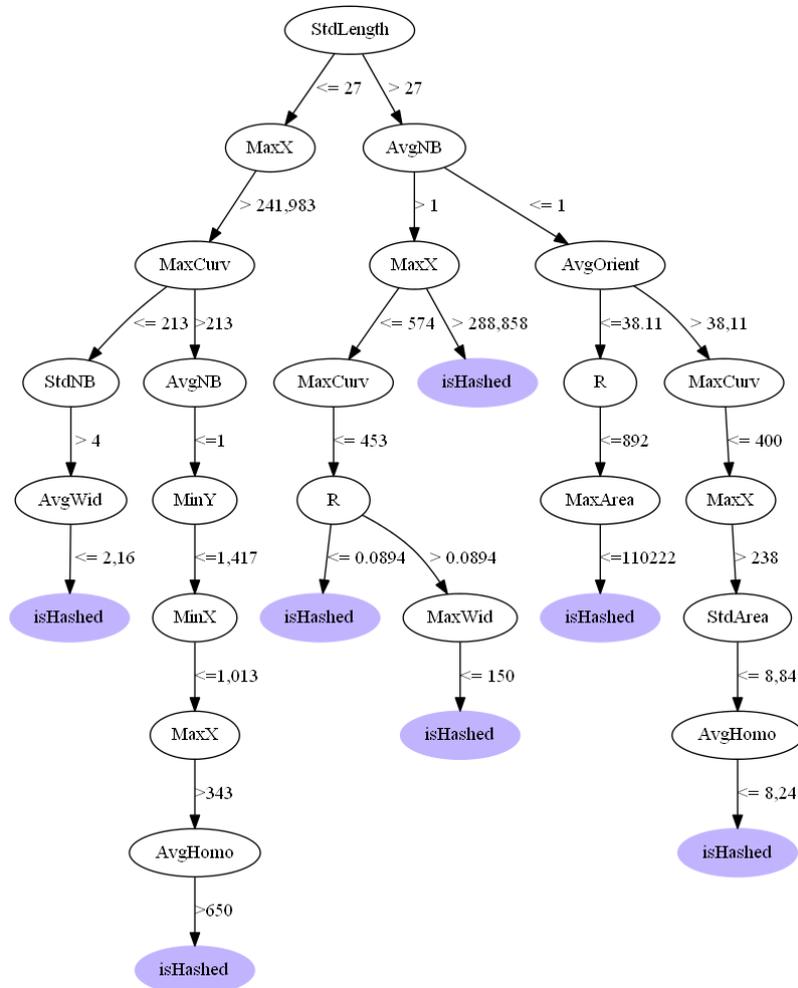


FIG. 13 – Arbre de décision simplifié permettant de trouver les règles.

caractéristiques bas niveau et les concepts sémantiques utilisés pour l'interrogation par les historiens. Nous avons donc choisi d'utiliser le formalisme des ontologies, dont nous décrivons les apports en terme de structuration (section 4.1). Les spécificités de notre approche pour les images de lettrines peuvent s'étendre à un cadre plus générique ainsi que nous le décrivons dans la section 4.2.

4.1 Structuration des données

Dans la construction d'une application qui, comme la nôtre, tend à combiner des traitements d'images (dits de bas niveau) et des traitements symboliques (dits de haut niveau), on

se trouve confronté à une grande masse de données hétérogènes. Les résultats des traitements d'images identifient des régions de l'image, avec leurs propriétés (aire, connexité, ...).

Au départ de ces données, des règles sont appliquées pour identifier certaines de ces régions comme des composants sémantiques du document de base (cette région est la lettre dans la lettrine, cette région est une partie d'un corps, ...). Toutes ces données doivent être rassemblées d'une manière cohérente pour permettre de nouvelles déductions, la navigation, la présentation de résultats sous forme d'images, etc...

Voici deux exemples (voir Figure 14) d'images produites pour présenter le résultat final d'un enchaînement de traitement d'images suivi de déductions sur les résultats. De telles images sont produites automatiquement à partir des données enregistrées. Elles sont utiles pour confronter des résultats à la réalité de terrain. Dans le premier exemple (cf. Figure 14), chaque image reprend la lettrine de départ, puis, en surimpression, la zone qui a été identifiée comme une lettre. La première image décrit une identification réussie, et la deuxième correspond à un échec partiel.

Le second exemple (cf. Figure 15) représente l'image de la lettrine figurative (au motif figuratif) dont les zones ne sont pas correctement annotées par la propriété *isBody*. Dans la seconde image apparaissent en surimpression l'ensemble des régions extraites de la couche forme. La zone rouge correspondant à la lettre sera correctement annotée par la propriété *isLetter*. En revanche, la zone jaune, qui correspond à une partie d'un personnage, ne sera pas annotée par la propriété *isBody* car son nombre d'Euler est trop grand. La zone blanche sera quant à elle annotée alors qu'elle ne correspond pas à une partie d'un personnage. L'utilisation d'un partitionnement plus complexe de l'image, ou encore d'autres descripteurs statistiques des zones, pourrait permettre d'annoter correctement les zones de ce type. De même une gestion plus souple de la classification (avec l'utilisation de rejet, de degré de confiance, classifieur flou,...) pourrait permettre de proposer à l'utilisateur des régions candidates, mais pour lesquelles le système ne peut prendre de décisions fermes.



FIG. 14 – Résultat final d'un enchaînement de traitement. d'images.

Parallèlement au problème de représentation, se pose la question du choix de modélisation des données, pour soutenir l'ensemble des traitements, déductions et calculs qui leur seront appliqués.

Le travail de modélisation pourrait être fait en UML. En général, on convertit le modèle UML dans un autre modèle de données comme le modèle relationnel. Mais un diagramme de classe UML ou un schéma de base de données est situé au niveau logique des méta-données, et il est essentiellement statique. Chaque nouveau traitement d'image, chaque nouvelle déduction demanderait la modification du modèle UML ou du schéma relationnel. Si un modèle

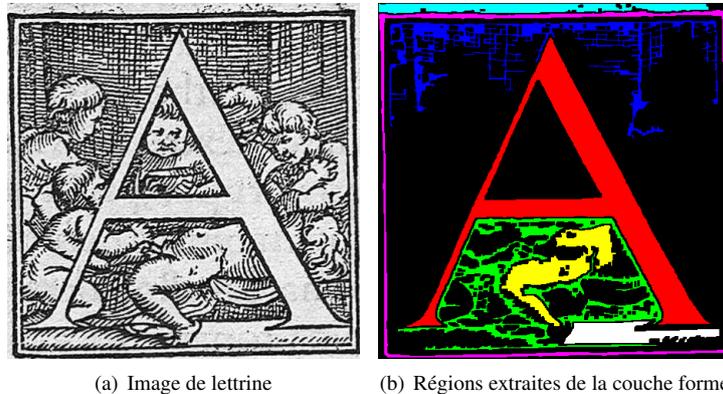


FIG. 15 – L'image où des zones ne sont pas correctement identifiées comme partie de personnage par *isBody*.

UML peut représenter directement la hiérarchie des concepts, un schéma relationnel ne la représente pas directement ni explicitement. Malgré la grande efficacité des systèmes de gestion des bases de données relationnelles, cette voie a donc été écartée. Mais nous avons utilisé un modèle UML pour définir la partie spatiale de l'ontologie que nous avons transformé en ontologie (Malki, 2009) en utilisant l'outil *eclipseuml2owl*².

Les ontologies formelles basées sur les logiques de description donnent une méthode très souple pour structurer des données. La T-Box, au niveau méta, décrit la structuration des données qui sont contenues dans la A-Box. L'approche est beaucoup moins statique que UML ou le schéma relationnel. Introduire un nouveau concept ou un nouveau rôle ne demande pas de précaution particulière (sauf qu'il faut vérifier la cohérence du modèle résultant).

Confrontés à de grands volumes de données leur temps de calcul peut devenir prohibitif, ce qui nous a amenés à en utiliser d'autres comme les règles SWRL³ ou DLV (Leone et al., 2005). DLV est une implémentation de Datalog (Gallaire et Minker, 1978)⁴, base de données déductive, qui permet de construire et peupler ces ontologies ; puis de les interroger en ligne de commandes. Basé sur la programmation logique déductive, DLV est simple à utiliser, et offre une facilité d'expression des requêtes, avec des temps de réponse relativement faibles. Nous utilisons également SEWELIS (Ferré, 2009) (SEmantic WEb Logical Information System) pour une recherche d'information dans l'ontologie construite à la fois par requête et par navigation, et qui laisse envisager une meilleure interaction avec l'utilisateur.

Toutefois, nous utilisons une démarche ontologique comme guide, même quand nous utilisons ces approches qui ne rentrent pas dans ce cadre.

2. <http://code.google.com/p/eclipseuml2owl>

3. SWRL : A Semantic Web Rule Language

4. <http://www.dlvsystem.com/dlvsystem/index.php>

4.2 Généricité de l'approche

Ces travaux autour des images de lettrines nous ont permis de dégager des caractéristiques propres à notre approche, à savoir l'intégration de données complexes au sein d'une même base ontologique de connaissances (connaissances propres au domaine ; connaissances propres aux traiteurs d'image ; information spatiale) accompagné d'un mécanisme d'inférence exploité dans un objectif de réduction du fossé sémantique (cf. Figure 16). Bien que ces traitements aient été exploités de manière spécifique, ils peuvent néanmoins être vus de manière générique pour être transposés à d'autres cas d'utilisation. Ainsi notre système global vise à décrire à partir d'un domaine d'expertise, de zones d'intérêts extraits automatiquement, des relations spatiales entre les zones pour proposer des requêtes complexes basées sur les mécanismes d'inférences. Nous présentons chacune de ces étapes ci-dessous.

Complex Knowledge Management

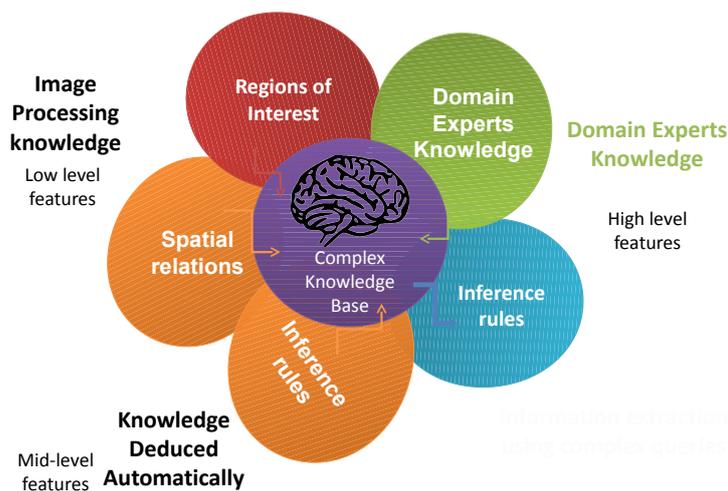


FIG. 16 – Fleur ontologique générique.

Ontologie du domaine. La description des images à l'aide d'un vocabulaire des experts repose sur une ontologie de domaine dans notre approche. Cette description est basée sur une annotation du contenu, qui peut se faire de différentes manières : annotation libre où aucun vocabulaire n'est prédéfini à l'avance et l'utilisateur se sert de ses propres connaissances pour annoter ; annotation par mots-clés où un vocabulaire (sans aucune hiérarchie) est proposé à l'utilisateur et ce dernier l'utilise pour annoter ; annotation par une ontologie où l'annotation est faite par l'utilisateur via un vocabulaire défini par une hiérarchie des mots. Dans le cas d'une annotation utilisant un vocabulaire hiérarchique permettant d'identifier les éléments sémantiques "cible", la problématique de réduction du fossé sémantique s'en trouve simplifiée (présence de deux vocabulaires, et d'experts

qui peuvent décrire les concepts et exprimer leurs contenus avec des formalismes différents). Nos travaux sur les images de lettrines, décrites sémantiquement par un unique thésaurus structuré et renseigné par les historiens, se positionnent dans ce contexte.

Ontologie des régions d'intérêts (analyse d'images). Afin de permettre une description des images par des zones d'intérêts dans notre modèle, nous avons défini une ontologie des régions d'intérêts. Ces régions sont issues d'un processus de segmentation d'image (initiale ou dérivée), décrites à partir de leur contenu et de leur position. Nous introduisons alors le concept d'image initiale et d'image dérivée obtenue par un traitement de l'image initiale. Dans le cas des lettrines, nous utilisons les images dérivées de la couche forme et la couche texture. Ce concept peut cependant être étendu à un cadre plus général. Notre approche est centrée sur la notion de région d'intérêt représentant une zone de l'image identifiée comme porteuse d'information, et qui peut se spécialiser (cf. Figure 17). Une telle représentation permet d'intégrer des caractéristiques soit au niveau de l'image initiale ou dérivée, soit au niveau des régions d'intérêt. On retrouve ainsi les notions de signatures globales et locales d'une image propres au domaine de l'analyse d'image.

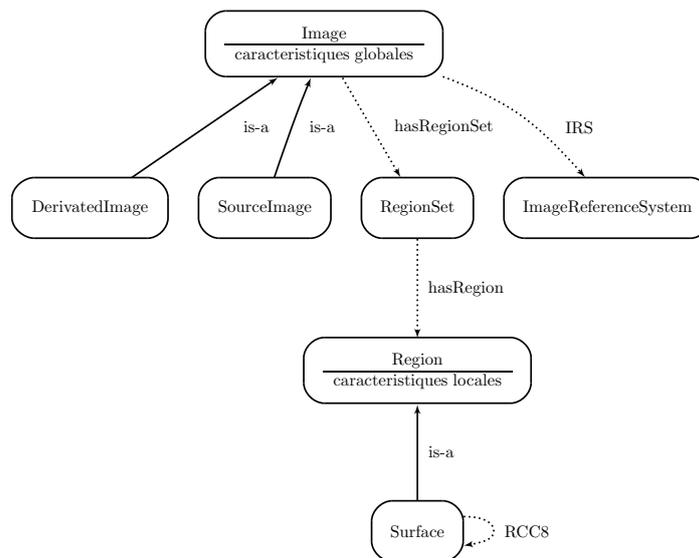


FIG. 17 – Ontologie générique des zones.

Ontologie spatiale. Toujours avec la vision globale de notre modèle, il paraît important et nécessaire de décrire le contenu des images à partir de ses régions d'intérêt et de leur position dans l'espace. Les relations spatiales regroupent des relations de topologie (RCC8), de distance (proche ou éloigné) ou d'orientation (N,S,E,W). Dans notre cas d'étude, nous nous limitons à l'exploitation des relations topologiques pour situer les régions d'intérêt de la couche forme par rapport à celle du partitionnement de l'image, dans le but d'identifier les régions qui se situent au centre de l'image.

De façon plus générale, il est possible de situer des régions de l'image les unes par rapport aux autres, qu'elles soient obtenues par une ou plusieurs méthodes d'extraction.

En particulier, une vérité terrain indiquant le positionnement spatial de certaines régions dans l'image peut être considérée comme un extracteur manuel. Un comparatif entre des zones extraites et une vérité terrain renseignerait alors sur la robustesse d'une méthode d'extraction, et permettrait d'en envisager des expérimentations à grande échelle. On retrouve cette notion d'extracteur sous le terme d'interprétation dans Lamiroy et Lopresti (2011) qui l'introduisent pour spécifier le résultat d'un algorithme de traitement d'images. Citons également Tsopze et al. (2012) où l'accent est mis sur un comparatif entre différents extracteurs pour évaluer les résultats d'un algorithme d'extraction par rapport à une vérité terrain pour des images de bandes dessinées.

Mécanisme d'inférence. La dernière brique élémentaire de notre système s'appuie sur des mécanismes d'inférences. Bien que ces mécanismes soient induits par la définition même des logiques de description, on parle également de règle d'inférence - terme qui sous-entend un mécanisme d'inférence en interactivité avec des demandes de l'utilisateur - ou encore d'extraction logique de connaissances. Le terme d'inférence en logique de description désigne plus précisément la possibilité d'exprimer des concepts ou propriétés non atomiques (et donc déductibles) à l'aide d'axiomes terminologiques (i.e. non atomiques) des logiques de description.

Une description logique de nouveaux concepts ou propriétés peut concerner certaines zones de l'image, ou bien l'image elle-même.

Dans l'ontologie des lettrines, la règle *isHashed* permet d'annoter des images, alors que les règles *isLetter* et *isBody* annotent des régions forme ou texture. L'annotation des images est facilitée lorsque la sémantique du domaine est clairement identifiée, les éléments sémantiques cibles étant alors connus. Elle peut alors s'envisager à grande échelle car son évaluation est automatique en comparant avec le résultat attendu de l'ontologie du domaine. L'annotation des régions informe quant à elle sur le positionnement d'information dans l'image, et s'apparente à une recherche par le contenu. Cependant, son évaluation reste limitée à un comparatif à la main, image par image, ou bien à l'existence d'une vérité terrain.

Enfin, deux stratégies peuvent être envisagées pour concevoir des règles afin de réduire le fossé sémantique. La première stratégie consiste à s'appuyer sur des mécanismes d'apprentissage automatique, dans le cas où seules les caractéristiques de bas niveau sont extraites. En effet, le mécanisme d'inférence s'apparente alors à une approche classique de fouille de données, qui peut être supervisée ou non. C'est le cas de la règle *isHashed* d'une image de lettrine pour laquelle des seuils ont été définis à l'aide d'un arbre de décision. Cet arbre, construit à partir des caractéristiques bas niveau de la couche texture, est supervisé par la propriété *hasBackground* de l'ontologie des historiens.

La seconde stratégie consiste à s'appuyer sur une expertise pour définir les paramètres de la règle. Cette stratégie est illustrée dans notre cas par l'exploitation des relations spatiales, pour identifier qu'une lettre se situe au centre de l'image. En fouille de données, l'exploitation d'une information spatiale est classiquement réalisée à l'aide de l'outil graphe dans un objectif de classification. Cependant, ces approches sont souvent difficiles à paramétrer, et restent limitées à des données de faible dimension car il n'existe pas d'algorithme efficace garantissant une solution optimale (isomorphisme de graphes NP-complet). C'est pourquoi nous avons défini la propriété *isLetter* (propriété d'une région

forme qui intègre un positionnement au centre de l'image) non pas par apprentissage, mais par expertise.

Une combinaison de stratégies peut également être envisagée, afin de renforcer une connaissance peu fiable du domaine. Ainsi une *stratégie bottom-up*, où les règles intègrent des caractéristiques bas niveau et une information spatiale seulement (i.e. règles *isHashed*, et *isLetter*) peuvent être complétées par une *stratégie top-down* avec des règles définies par la seule sémantique. On peut ici noter que la règle *isBody*, est une combinaison de ces deux stratégies, puisqu'elle utilise les caractéristiques et la notion *figurative* issue du domaine.

Enfin, une règle reposant sur le résultat d'autres règles se positionne quant à elle dans une *stratégie itérative*, ou de bouclage, permettant une "montée" en sémantique. La règle *isBody*, qui annote une région (au centre) non précédemment annotée par la règle *isLetter*, en est un exemple.

5 Conclusion et perspectives

Cet article présente des travaux qui visent à réduire le fossé sémantique entre le vocabulaire utilisé par des historiens de la Renaissance et des résultats de traitement d'images. L'objectif de ces travaux est de proposer un outil de recherche d'image par le contenu intégrant des connaissances du domaine d'application et du domaine d'expertise : le traitement d'images de lettrine dans notre cas.

Nous avons donc utilisé des ontologies pour définir une sémantique explicite entre concepts visuels et langage naturel, et surmonter le problème de la représentation implicite d'une connaissance visuelle. De plus, ces ontologies sont utilisées comme une référence commune pour échanger des contenus sémantiques. Elles ont été peuplée par des données issues des experts historiens et par des résultats obtenus via des traitements d'images de lettrines. Partant de cette base de connaissances, des règles d'inférences sont utilisées pour créer un lien entre les mots-clés historiens et des régions des images, et ainsi réduire le fossé sémantique. Les résultats obtenus sur notre base d'image nous permettent de valider ce modèle et d'obtenir des résultats très corrects.

En perspectives des ces travaux, nous travaillons actuellement à la généralisation de ce modèle pour l'étendre à des images de bandes dessinées (Tsopze et al., 2012). De plus, il pourrait être intéressant de proposer une génération automatique de règles d'inférences à partir de bouclage de pertinence (l'utilisateur pourrait ainsi valider plusieurs images qui serviraient d'apprentissage à la règle). Enfin, bien que ce modèle propose des résultats très intéressants, il faudrait réfléchir à son optimisation pour l'étendre à de grandes bases de données qui ne sont actuellement pas directement intégrables via des ontologies.

Références

13249-3 :2002, I. (2002). *Information technology-Database languages ; SQL Multimedia and Application Packages ; Part 3 : Spatial*.

- Bannour, H. et C. Hudelot (2011). Towards ontologies for image interpretation and annotation. In *Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop on*, pp. 211–216.
- Bannour, H. et C. Hudelot (2012). Combinaison d'information visuelle, conceptuelle, et contextuelle pour la construction automatique de hiérarchies sémantiques adaptées à l'annotation d'images. In *Actes de la conférence RFIA 2012 - Reconnaissance des Formes et Intelligence Artificielle*, pp. 460–469.
- Barnard, K., P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, et M. I. Jordan (2003). Matching words and pictures. *J. Mach. Learn. Res.* 3, 1107–1135.
- Clouard, R., A. Renouf, et R. Marinette (2010). An ontology-based model for representing image processing application objectives. *International Journal of Pattern-Recognition and Artificial Intelligence* 24, 1181–1208.
- Cohn, A., B. Bennet, J. Gooday, et N. Gotts (1997). Representing and reasoning with qualitative spatial relations about regions. In *Spatial and temporal reasoning*, pp. 97–134. Kluwer.
- Coustaty, M. (2011). *Contribution à l'analyse complexe de documents anciens. Application aux lettrines*. Ph. D. thesis, Université de La Rochelle, La Rochelle, France.
- Coustaty, M., R. Pareti, N. Vincent, et J.-M. Ogier (2011). Towards historical document indexing : extraction of drop cap letters. *IJDAR* 14(3), 243–254.
- Dubois, S., M. Lugiez, R. Péteri, et M. Ménard (2008). Adding a noise component to a color decomposition model for improving color texture extraction. In *proc. of the 4th European Conference on Colour in Graphics, Imaging, and Vision (CGIV'08)*, pp. 394–398.
- Fan, J., Y. Gao, et H. Luo (2008). Integrating concept ontology and multitask learning to achieve more effective classifier training for multilevel image annotation. *Image Processing, IEEE Transactions on* 17(3), 407–426.
- Ferré, S. (2009). Navigating the semantic web with logical information systems. *Rapport Recherche Interne* (inria-00410959).
- Gallaire, H. et J. Minker (Eds.) (1978). *Logic and Data Bases, Symposium on Logic and Data Bases, Centre d'études et de recherches de Toulouse, 1977*, Advances in Data Base Theory. Plenum Press.
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.* 43, 907–928.
- Hamidi, A. E., M. Ménard, M. Lugiez, et C. Ghannam (2010). Weighted and extended total variation for image restoration and decomposition. *Pattern Recognition* 43(4), 1564–1576.
- Hanbury, A. (2008). A survey of methods for image annotation. *Journal of Visual Languages and Computing* 19, 617–627.
- Inc., O. G. C. (1999). OpenGIS simple features specification for sql. *OpenGIS Project Document* 99-049.
- Jena (2011). Jena - a semantic web framework for java. {<http://jena.sourceforge.net/>}, date visite : Mars 2011.
- Jeon, J., V. Lavrenko, et R. Manmatha (2003). Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the 26th annual international ACM SIGIR*

- conference on Research and development in informaion retrieval*, SIGIR '03, New York, NY, USA, pp. 119–126. ACM.
- Jimenes, R. (2008). Les bibliothèques virtuelles humanistes et l'étude du matériel typographique. Technical report, Centre d'Etude Supérieur de la Renaissance.
- Knublauch, H., R. Ferguson, N. Noy, et M. Musen (2004). The protégé owl plugin : An open development environment for semantic web applications. In *International Semantic Web Conference (ISWC)*, Volume 3298, pp. 229–243.
- Kompatsiaris, Yiannis ; Hobson, P. (Ed.) (2008). *Semantic Multimedia and Ontologies*, Volume 1 of *Theory and Applications*. Springer.
- Lamiroy, B. et D. Lopresti (2011). An Open Architecture for End-to-End Document Analysis Benchmarking. In *11th International Conference on Document Analysis and Recognition - ICDAR 2011*, Beijing, China, pp. 42–47. International Association for Pattern Recognition : IEEE Computer Society. ISBN : 978-1-4577-1350-7.
- Lavrenko, V., R. Manmatha, et J. Jeon (2003). A model for learning the semantics of pictures. In *NIPS'03*.
- Leone, N., G. Pfeifer, et W. Faber (2005). The DLV Project - A Disjunctive Datalog System (and more). <http://www.dbai.tuwien.ac.at/proj/dlv/>.
- Liu, Y., D. Zhang, G. Lu, et W.-Y. Ma (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition* 40(1), 262 – 282.
- Maillot, N., M. Thonnat, et A. Boucher (2004). Towards ontology-based cognitive vision. *Mach. Vis. Appl.* 16(1), 33–40.
- Malki, J., M. W. B. A. (2009). Une approche ontologique pour la modélisation et le raisonnement sur les trajectoires. prise en compte des règles métiers, spatiales et temporelles. In *JFO 2009 3ème édition des journées Francofones sur les Ontologies, Poitiers.*, pp. 157–168.
- Neumann, B. et R. Möller (2008). On scene interpretation with description logics. *Image Vision Comput.* 26, 82–101.
- Newell, A. (1982). The knowledge level. *Artif. Intell.* 18(1), 87–127.
- Pareti, R. et N. Vincent (2006). Ancient initial letters indexing. In *Proc. of the 18th International Conference on Pattern Recognition (ICPR'08)*, Hong Kong, China, pp. 756–759. IEEE Computer Society.
- Pratt, W. (2007). *Digital Image Processing : PIKS Scientific Inside* (4 ed.). Wiley-Interscience.
- Smeulders, A. W. M., M. Worring, S. Santini, A. Gupta, et R. Jain (2000). Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 1349–1380.
- Tousch, A.-M., S. Herbin, et J.-Y. Audibert (2012). Semantic hierarchies for image annotation : A survey. *Pattern Recognition* 45, 333–345.
- Tsopze, N., C. Guérin, K. Bertet, et A. Revel (2012). Ontologies et relations spatiales dans la lecture d'une bande dessinée. In *Actes des 23ème journées francophones d'ingénierie des connaissances*, Paris, France, pp. 175–182.
- Zhang, D., M. M. Islam, et G. Lu (2012). A review on automatic image annotation techniques. *Pattern Recognition* 45, 346–362.

Summary

This paper presents an ontology-based approach of the semantic gap problem resolution in the case of historical images, and more particularly to lettrines. This approach automatically creates some links between the regions of interest issuing from the computer vision algorithms in one hand, and their meaning on the other hand. These links will then be used to help historians. They need to identify the printed period of the documents and to find an automatic interpretation to the scene of the cap letter. Three ontologies are combined to annotate these images. In this way, we have defined some rules to annotate images (to characterize some lettrines as figuratives for example), or to annotate some regions (this region is a letter, this region correspond to a body-part, ...) as the letter, some body-parts, or to characterize. More particularly, the approach could be extended to other contexts mixing image processing, spatial relation and domain specific semantic, like in comics books Tsopze et al. (2012).