

D'un corpus d'images à une base d'images : une plateforme combinant syntaxe et sémantique et une méthodologie de prototypage

L. Besson, A. Da Costa, E. Leclercq, M.N. Terrasse
Laboratoire LE2I - Université de Bourgogne - France
E-mail :prenom.nom@u-bourgogne.fr

Résumé

Avec la multiplication des domaines qui utilisent l'image comme support d'information, le besoin de créer des bases d'images s'amplifie. Les techniques de recherche d'images par le contenu deviennent cruciales dans un tel contexte. Cependant, si les techniques proposées sont de plus en plus efficaces, elles sont aussi de plus en plus variées et spécialisées. Par conséquent, la tâche du concepteur de bases d'image –qui doit effectuer un choix difficile et lourd de conséquences– est ardue. L'objectif de cet article est de présenter une plateforme qui permet d'appliquer aux bases d'images les méthodes de développement des systèmes d'information.

Cette plateforme repose sur un modèle générique permettant de représenter les images de façon globale ou locale, en utilisant des attributs syntaxiques ou sémantiques, et avec des relations spatiales ou sémantiques entre objets de l'image. Le modèle générique doit être instancié pour une application donnée. L'instanciation du modèle est propagée vers le corpus d'images pour définir l'interface d'acquisition : choix des algorithmes d'extraction, stratégie de combinaison des extracteurs puis utilisation d'extracteurs pré-existants afin de produire les résumés des images. De même l'instanciation du modèle est propagée vers les utilisateurs pour définir la ou les interface(s) de recherche : indexation et classification des images. Une telle plateforme doit permettre d'une part de séparer les phases de conception et d'implémentation de la base d'image, d'autre part de construire rapidement un prototype qui pourra être validé avant la mise en exploitation.

Mots-clés : Ingénierie des bases d'image - Recherche par le contenu - Représentation de la syntaxe et de la sémantique.

Abstract

With the increasing number of domains that use images as an information media, the need for establishing image databases grows. Content-based image retrieval is becoming crucial in such a context. However, as the proposed search techniques are increasingly efficient, they are also increasingly diverse and specialized. Thus, an image database designer faces difficult choices. Our objective is to present a framework which first enables image

database designers to use information system engineering methodologies. Furthermore, our framework allows designers to combine syntactical and semantical descriptions of images.

Our framework is based on a generic model for representation of images. By using this model, it is possible to represent images either globally or locally (as a hierarchy of objects/zones), with syntactical and semantical attributes, with spatial and semantical relations between objects/zones. Such a generic model needs to be instantiated for a given application. All the choices made at the model instantiation are propagated to the lower level of our framework in order to generate an extraction interface. These choices are also propagated to the upper level of our framework in order to generate search interfaces (indexing and classification of images). Such a framework enables image database engineers to design, prototype and tests and then implement and optimize their databases.

Keywords : Image database engineering - Content-based image retrieval - Syntax and semantics representation.

1 Introduction

Qu'elle appréhende directement la réalité (comme une radiographie) ou qu'elle ait un rôle symbolique (comme les peintures rupestres), l'image est un important vecteur de transmission d'information. Avec le développement de nouvelles techniques de production, de diffusion et de traitement d'images, les techniques de recherche dans des bases d'images sont un des composants (et des problèmes) majeurs des technologies pour l'image.

La première difficulté (intrinsèque) des mécanismes de recherche d'images dans une base est liée à un double problème de volume : volume de la base en terme du nombre d'images qu'elle contient et volume de chaque image qui est une somme complexe d'informations. Les interfaces des bases de données images proposent donc deux approches pour traiter les problèmes de volume. La première approche consiste à diminuer virtuellement le volume de la base d'images en ne considérant plus comme unité d'accès une image mais un groupe d'images considérées comme similaires : c'est l'approche **classification**. La seconde approche consiste à diminuer virtuellement le volume d'une image en réduisant chaque image à un résumé : c'est l'approche **indexation**. Dans les deux cas il est essentiel de disposer d'un mécanisme permettant d'extraire de chaque image une partie significative "minimale et suffisante" afin de construire le résumé de l'image et, éventuellement, de décider si deux images appartiennent ou non à la même classe. Le cœur des deux approches est donc bien le même : extraire d'une image une sous-description représentative (le résumé de l'image) et être capable d'induire une similarité entre images à partir d'une similarité entre résumés. Les mécanismes permettant d'extraire un résumé à partir d'une image sont généralement classés en deux catégories : les mécanismes syntaxiques et les mécanismes sémantiques (dont les noms ont varié d'un auteur à l'autre : objets de l'image et objets du domaine [GWJ91], représentations numériques ou symboliques, notions d'image et image virtuelle [PSTT01], etc.). Dans la suite nous parlerons de zones dans le cadre d'une approche purement syntaxique

et d'objets le reste du temps. L'approche syntaxique consiste à réduire l'image (ou une zone de l'image) à un vecteur de paramètres physiques : les paramètres choisis ainsi que leur domaine peuvent varier considérablement d'une approche à l'autre. Certains systèmes utilisent plutôt la couleur [WJ97], d'autres plutôt les contours ou les textures [FFN⁺93], d'autres des caractéristiques propres au signal [LTMD01]. Cette approche syntaxique présente l'avantage d'être entièrement automatisable via des algorithmes de traitement d'images. Les approches sémantiques consistent à réduire l'image à un ensemble d'objets sémantiques (identifiés par leur signification dans le monde réel) reliés éventuellement par des relations sémantiques (qui correspondent à une interprétation de l'image comme reproduction d'une partie du monde réel). Cette approche nécessite l'intervention d'un expert et l'existence d'un modèle permettant de représenter la complexité du monde réel qui fait référence.

La seconde difficulté – extrinsèque celle-là – relève de la prise en compte précise des besoins des utilisateurs. Ce problème a été étudié par Denos [DBM97]. Denos formalise la différence d'évaluation de la pertinence d'une réponse à une question par un humain (qui va travailler en fonction des données disponibles et du contexte de la recherche : c'est la notion de pertinence utilisateur) et par un système informatique (qui sera limité aux données disponibles : c'est la notion de pertinence système). Dans le cadre d'un système fermé (toutes les images font référence au même champ sémantique, tous les utilisateurs ont les mêmes objectifs et la même expertise par rapport aux images), la différence entre la pertinence utilisateur et la pertinence système peut-être contrôlée. Par contre, pour un système ouvert (en particulier dans le cadre d'applications réparties telles que le web) il est impossible de contrôler la différence entre la pertinence utilisateur et celle (figée) du système. Par exemple, en mode syntaxique, on peut espérer trouver des paramètres qui correspondent aux critères d'analyse des experts du domaine ; ces critères risquent fort d'être totalement indécryptables par des utilisateurs non experts. De même en mode sémantique, les annotations des experts d'un domaine peuvent ne pas être discriminantes pour les experts d'un autre domaine (ou pour des utilisateurs quelconques). Despres [DL00] en montre des exemples dans le cadre des systèmes bibliographiques, les ontologies en sont aussi des exemples flagrants [Gua95]. Santini & al. vont même plus loin en considérant qu'une partie de la sémantique d'une base de données provient de l'interaction des utilisateurs avec cette base [SGJ01].

La mise en œuvre d'une base d'images se réduit donc plus ou moins à un choix –difficile– entre un mécanisme stable mais de qualité moyenne (approche syntaxique) et un mécanisme instable avec de bons résultats ponctuels et un risque de totale inadéquation (approche sémantique). Quelques propositions ont cependant été faites pour tenter de diminuer l'écart entre ces deux extrêmes tels que le bouclage de pertinence ou l'introduction de thesaurus et de systèmes de déduction [Fou02]. En pratique, si on fait l'hypothèse que les techniques de recherche par le contenu sont assez efficaces pour satisfaire la plupart des besoins, le concepteur d'une base d'images reste néanmoins aux prises avec d'une part un corpus d'images et d'autre part les besoins des utilisateurs. Les caractéristiques du corpus d'images et du domaine d'application doivent être exploitées au mieux afin de satisfaire l'ensemble des utilisateurs. Par exemple, Chabot [OS95] utilise des

propriétés sémantiques (annotations) et des connaissances afin de piloter l'analyse des images ; KMeD [TCC+94] utilise une architecture multi-couches basée les contours, les formes et les objets, spécialisée pour le domaine médical et qui permet une identification quasi automatique des objets ; le système PICTION [SB94] combine les annotations textuelles et les techniques de traitement d'images pour proposer un module de traitement du langage naturel permettant l'évaluation des requêtes utilisateur.

Il est donc difficile de trouver le point d'équilibre entre toutes les contraintes pesant sur une base d'images tout en exploitant au mieux l'information potentiellement disponible. Nous pensons que les méthodologies mises en œuvre dans d'autres domaines des systèmes d'information peuvent être, avec grand profit, appliquées à la construction des bases d'images. À cet effet, nous proposons une plateforme de prototypage qui permet de passer –à moindre coût– d'un corpus d'images à une base d'images correctement profilée pour ses utilisateurs potentiels. Cette plateforme a pour cœur un **modèle générique** qui doit permettre de représenter les images de la plupart des applications. Ce modèle permet de décrire les images soit *globalement* (l'image est vue comme un tout, une unité de description) soit *localement* (l'image est constituée d'objets/zones individuellement identifiables). Les attributs associés aux images et aux objets/zones de l'image peuvent être soit *syntactiques* (e.g., couleur, intensité, forme) soit *sémantiques* (e.g., rivière, bâtiment, route). Dans le cas des descriptions locales, *plusieurs types de relations entre objets/zones* d'une même image peuvent être mis en place : composition (de l'image à partir des objets/zones), relations spatiales (e.g., au nord de, au sud-est de) et topologiques (e.g., voisinage, intersection), relations sémantiques (e.g., la porte fortifiée du camp romain). Dans le cadre d'une application donnée, ce modèle est instancié : on choisit les attributs (syntactiques, sémantiques ou mixtes) et les relations qui vont constituer les résumés des images. En utilisant la **couche basse de la plateforme**, l'instanciation du modèle est propagée pour définir l'interface d'acquisition des images. En utilisant la **couche haute de la plateforme**, l'instanciation du modèle est propagée pour définir l'interface de recherche.

2 Description de la plateforme et première approche de la méthodologie associée

L'architecture de notre plateforme est illustrée en Figure 1.a). Son noyau est constitué d'un couple **modèle générique** et **bibliothèque de mesures de similarité**. Le modèle générique permet –dans un formalisme unique– une description des images qui est adaptée à la fois à un traitement purement basé sur la syntaxe (comme par exemple les transformées en ondelettes) et à un traitement à forte composante sémantique. Notre modèle générique doit être instancié pour un corpus d'images donné et un jeu de fonctionnalités demandées par les utilisateurs potentiels, c'est-à-dire en fonction des informations disponibles, visibles dans les images d'une part et en fonction d'autre part des informations nécessaires aux traitements qui devront être effectués. La description d'une image dans ce modèle

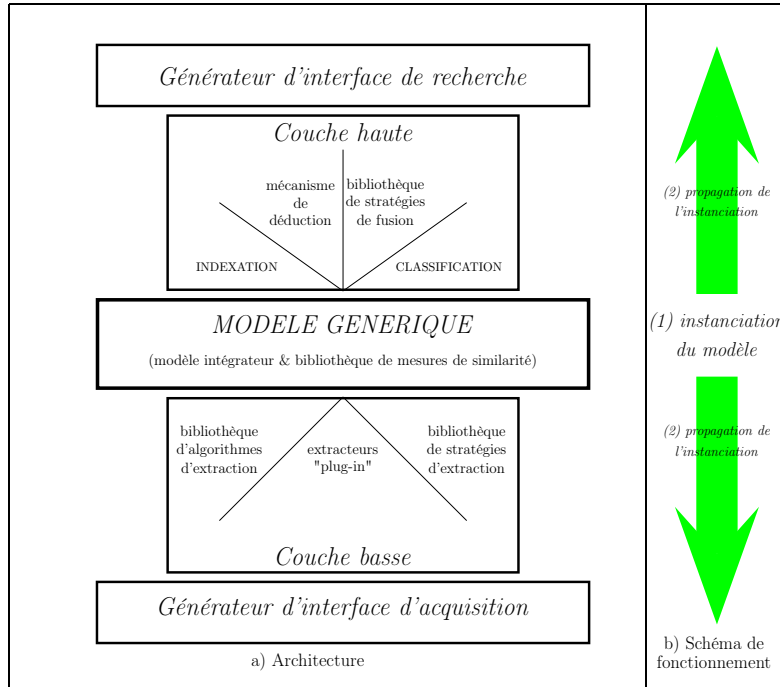


FIG. 1 – Présentation de notre plateforme

instancié constitue le résumé de l'image.

L'instanciation du modèle peut alors être propagée vers le corpus d'images (en utilisant les ressources de la **couche basse** de notre plateforme) pour mettre en place l'**interface d'acquisition** des images. Le concepteur de la base d'image utilise la bibliothèque d'extracteurs et de stratégies d'extraction– pour définir puis implémenter les outils d'extraction qui vont construire les résumés des images du corpus. Le choix des algorithmes d'extraction, la stratégie de combinaison des extracteurs sont déterminés dans une phase d'analyse. La phase d'implémentation "combine" des extracteurs pré-existants (fournis avec la plateforme) ou ad hoc (développés spécialement pour l'application concernée) pour mettre en place des outils d'extraction qui vont construire les résumés des images du corpus.

De même, l'instanciation du modèle est propagée vers les utilisateurs via la **couche haute** de notre plateforme –qui offre des mécanismes d'indexation et de classification, des mécanismes de déduction, une bibliothèque de stratégies de fusion syntaxique/sémantique– afin de mettre en place l'**interface de recherche** de la base d'images (ou éventuellement les interfaces de recherche appropriée aux différents types d'utilisateurs).

Dans la suite de cette section, nous donnons quelques détails sur la couche basse de la plateforme et sur la méthode de construction de l'interface d'acquisition.

La couche basse de notre plateforme Ainsi que nous l'avons brièvement décrit ci-dessus, le modèle instancié qui a été choisi pour une application donnée détermine le contenu des résumés des images (en termes d'attributs et de relations entre objets). Il est donc nécessaire de mettre en œuvre divers extracteurs d'information pour construire les résumés à partir des images. Nous distinguons trois types d'extracteurs qui peuvent être automatiques, semi-automatiques ou manuels (e.g., la plupart des annotateurs sémantiques). Le premier type, dit des *extracteurs syntaxiques*, calcule des paramètres physiques tels que couleur moyenne, énergie, etc. Le second type, dit des *extracteurs géométriques*, isole dans l'image des formes géométriques simples (cercles, triangles, rectangles, etc.) à partir d'un contour ou d'une texture. Le troisième type, dit des *annotateurs sémantiques*, permet –par intervention d'un expert du domaine sur lequel porte la base d'images– d'ajouter des méta-données à l'image. Notre bibliothèque d'extracteurs est organisée comme décrit en Figure 2. Le point d'accès privilégié est une bibliothèque d'algorithmes d'extraction (e.g., algorithme de Roth & Levine). Pour chacun de ces algorithmes sont données : la liste des paramètres qu'il extrait d'une image (e.g., couleur moyenne, histogramme d'intensité, relation de voisinage, datation, nature archéologique), la liste des paramètres dont il a besoin en entrée, la liste des modules de code implémentant cet algorithme. Les seuls paramètres considérés ici sont ceux qui doivent être extraits de l'image. Un dictionnaire de paramètres sert de référence afin de faciliter l'utilisation de cette bibliothèque. Cette bibliothèque est bien entendu évolutive : elle sera étendue au fur et à mesure des besoins et des expériences menées.

Construction de l'interface d'acquisition A partir de la liste des paramètres à extraire de l'image, le concepteur de la base d'image lance une construction incrémentale (automatique ou semi-automatique) de la liste des algorithmes d'extraction nécessaires. Ce processus consiste à supprimer à chaque étape au moins un paramètre de la liste, en ajoutant éventuellement de nouveaux paramètres à cette liste. Le processus est poursuivi jusqu'à une situation de point fixe pour laquelle il n'y a plus de paramètre à extraire qui ne soit pas couvert par au moins un des algorithmes retenus. À partir du moment où l'extraction de tous les paramètres est assurée, il reste à résoudre un ensemble de dépendances. Le concepteur de la base d'image lance alors un tri topologique [GT96] pour déterminer les ordres possibles pour l'enchaînement des algorithmes d'extraction. La dernière étape est le choix d'un des ordres possibles et la construction (par combinaison de composants) d'une interface d'acquisition à partir des modules d'implémentation des algorithmes sélectionnés et ordonnés. Au delà de la phase de prototypage, cette première version de l'interface sera optimisée. L'optimisation devra porter sur l'élimination des calculs redondants (à la charge de l'équipe d'informaticiens) ainsi que sur l'optimisation des modules d'extractions proprement dit (sous contrôle d'un expert en segmentation).

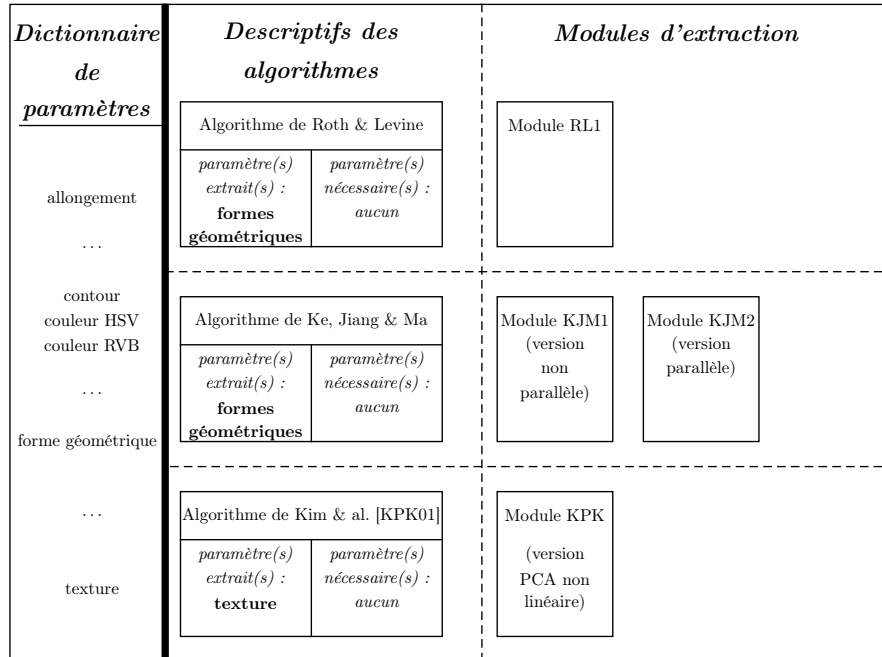


FIG. 2 – Structure de la bibliothèque d’algorithmes d’extraction

3 Le modèle de description des images

Ainsi que nous l’avons signalé, notre modèle décrit les images selon deux points de vue orthogonaux. Un premier point de vue traite de la granularité [AKDGB, ZAL01] de description des images : globale ou locale. Un second point de vue traite de la nature des attributs et relations utilisés pour la description : sémantiques ou syntaxiques. Le modèle que nous proposons est capable de répondre à toutes les combinaisons des niveaux de granularité avec la nature des informations représentées dans les attributs. Par exemple, une description **locale syntaxique** incluant la couleur est utilisée dans QBIC [GWJ91] ainsi que dans le projet IKONA [GB]. Nous avons testé un tel type de description avec des transformées en ondelettes appliquées aux images d’une base paléontologique. Chaque image produite par la transformée en ondelettes est considérée comme une "zone" de l’image. Les informations attachées à chaque zone sont calculées de façon automatique à partir des valeurs des paramètres de la transformée en ondelettes. De même, une description **globale syntaxique** peut être par exemple un histogramme de couleurs ou d’intensités comme dans ImageGrep [WJ97]. Un attribut syntaxique global est le plus souvent soit la valeur moyenne soit la valeur la plus significative d’un paramètre physique de l’image. Nous avons utilisé une description **locale sémantique** sur une base de vues aériennes de sites archéologiques. Dans

une telle description les zones identifiées dans l'image sont des objets (ayant une interprétation dans le monde réel). La composition de l'image à partir des objets est hiérarchique. Une description **globale sémantique** est encore souvent utilisée pour décrire par des mots-clés les images d'une base mise à disposition d'experts du domaine d'application (e.g., base d'images d'une agence photographique, base d'images médicales) : chaque image étant représentée par des attributs date, lieu, sujet, etc. Les descriptions sémantiques globales sont en général créées par des experts du domaine qui annotent les images.

Le modèle que nous présentons est générique et multi-granulaire. La généralité repose sur la possibilité d'instancier plusieurs éléments de la description d'une image en fonction du domaine d'application et des fonctionnalités prévues. Le choix des attributs décrivant une image, par exemple, est fortement lié au domaine d'application. Pour une application comme la base de vues archéologiques, la couleur n'est pas utilisable pour décrire les objets. En effet cette base étant composée de vues aériennes prises avec différentes techniques (photographie couleur, noir et blanc ou infrarouge, etc.) ainsi qu'à différentes saisons, un même objet peut avoir des couleurs différentes d'une image à l'autre. En revanche, pour cette application, la forme est un paramètre essentiel. Nous appelons *paramètre discriminant* un paramètre qui est "syntaxiquement" détectable dans les images du corpus à traiter et *paramètre significatif* un paramètre qui a un sens dans le contexte de l'application.

La multi-granularité repose sur une organisation hiérarchique de la représentation d'une image. L'image est (éventuellement) décomposée en objets (ou zones) qui peuvent eux-mêmes être décomposés. Dans la suite nous parlerons d'objets (pour objets/zones) afin de ne pas alourdir inutilement le texte. À chaque objet sont associés des attributs syntaxiques et sémantiques. Deux objets peuvent être associés entre eux par des relations syntaxiques, spatiales et sémantiques. Dans la suite de cette section, nous présentons la formalisation de notre modèle générique. Nous décrivons ensuite ce qu'est la phase d'instanciation du modèle que nous illustrons par un exemple.

Un aperçu du modèle formel La description des images est basée sur la description d'objets (*objets simples* détectables dans l'image, *objets complexes* qui sont composés à partir des objets simples, l'image elle-même qui est traitée comme un objet). Nous notons par O l'ensemble des objets apparaissant dans la description d'une image. Cet ensemble O est l'union de deux sous-ensembles O_s et O_c qui contiennent respectivement les objets simples et complexes. Dans le cas d'une description globale des images, l'ensemble O est réduit à un seul objet –considéré comme simple– qui est l'image elle-même.

Chaque objet de O est décrit par un tuple comprenant : un identificateur d'objet (noté *idf*), une géométrie (notée *geo*) qui doit permettre de retrouver l'ensemble des pixels de l'image associés à cet objet (cette géométrie est optionnelle), des tuples d'attributs physiques et sémantiques (respectivement notés *AttP* et *AttS*). Nous notons *Descr* l'ensemble des tuples décrivant les images et nous mettons ainsi en place une fonction de description d'objets notée d :

$$\begin{aligned}
d : \quad O &\rightarrow Descr \\
o &\rightarrow \langle idf(o), geo(o), AttP(o), AttS(o) \rangle
\end{aligned}$$

Entre les objets de l'ensemble O nous pouvons définir plusieurs relations : relations de composition qui déterminent une vision hiérarchique de l'image, relations spatiales (distance, direction, relation topologique), relations sémantiques. Il peut y avoir plusieurs relations de composition dans la description d'une image : ces relations correspondent alors à plusieurs points de vue sur la façon dont l'image est construite. Nous imposons qu'il existe pour chaque image une relation de décomposition privilégiée prenant en compte tous les objets simples et l'image elle-même. Les relations entre objets de O –quelque soit leur nature– sont représentées par des graphes étiquetés. Nous utilisons les définitions suivantes :

- Un graphe étiqueté R_i est un tuple $\langle ETIQ_i, A_i, \mathcal{E}_i \rangle$ tel que :
 - 1) L'ensemble des étiquettes est noté $ETIQ_i$. Dans certains cas, cet ensemble sera réduit à une constante : par exemple pour certaines relations de composition pour lesquelles il n'y a pas différents cas de mise en relation de deux objets.
 - 2) L'ensemble des arcs A_i est un sous-ensemble de $O \times O$ ($A_i \subset O \times O$). En fonction de la nature de la relation, diverses contraintes peuvent être mises en place sur cet ensemble d'arcs A_i .
 - 3) La fonction d'étiquetage \mathcal{E}_i associe à chaque arc une étiquette.
$$\begin{aligned}
\mathcal{E}_i : \quad A_i &\rightarrow ETIQ_i \\
a &\rightarrow \text{étiquette}
\end{aligned}$$
- L'ensemble R des graphes étiquetés est noté $\{\langle ETIQ_i, A_i, \mathcal{E}_i \rangle\}_{i=1..n}$. Pour faciliter les traitements ultérieurs nous imposons que la numérotation des graphes étiquetés constitutifs de R permette de regrouper les relations de même nature (composition, spatiale, sémantique).

Phase d'instanciation du modèle La phase d'instanciation du modèle a pour objectif de déterminer, pour une application donnée, la structure de description qui sera commune à toutes les images et à tous les objets des images ainsi que les relations entre objets qui vont être prises en compte. Pour cela nous proposons les étapes suivantes :

- Détermination (par un spécialiste de segmentation par exemple) des *paramètres discriminables* pour le corpus d'images. Il s'agit des paramètres qui sont calculables (si on doit extraire automatiquement les paramètres) ou visibles (si on travaille de façon manuelle ou semi-automatique).
- Détermination (par les spécialistes du domaine) des *paramètres significatifs* c'est à dire des paramètres qui sont sémantiquement utilisables. Il s'agit ici d'éliminer les paramètres qui n'ont pas de sens pour le corpus d'images (par exemple parce que leurs valeurs ne sont pas cohérentes d'une image à l'autre).
- Détermination (par les analystes et concepteurs de la base) des *paramètres nécessaires* à la réalisation des fonctionnalités demandées.
- Choix du jeu de paramètres du modèle à partir des paramètres discriminants, significatifs et nécessaires. Cette étape consiste à trouver un équilibre entre ce qui est nécessaire et ce qui est disponible. Trois cas principaux peuvent

se produire. Dans le cas le plus favorable, toutes les informations nécessaires sont disponibles. Dans le cas intermédiaire, certaines informations nécessaires sont couvertes par des paramètres discriminants mais pas significatifs. Il faut dans ce cas effectuer une analyse précise des méta-données associées aux images pour tenter de rendre significatifs les paramètres manquants. Dans le cas le plus défavorable, certaines informations nécessaires sont couvertes par des paramètres significatifs mais non discriminants. Dans ce cas, il faut soit améliorer la qualité du corpus d'images soit réduire le jeu de fonctionnalités qui sera proposé.

Les paramètres ainsi choisis sont ensuite répartis dans les catégories syntaxique ou sémantique et les tuples correspondants sont définis. Pour chaque paramètre il reste à définir son ensemble de valeurs possibles.

- Si le modèle est local, les relations entre objets (autres que la relation de composition privilégiée) sont définies dans les trois catégories : composition, spatiales et sémantiques. Pour chacune des relations on définit l'ensemble des étiquettes qui sera associé à son graphe.

Exemple d'instanciation du modèle Dans cette partie nous allons présenter un exemple d'instanciation du modèle pour une application d'archéologie aérienne. Cet exemple d'instanciation est tiré d'un projet consistant à passer d'un corpus de vues aériennes de sites de la région Bourgogne (sous forme d'images et de fiches techniques sur papier) à une base d'images numérisées et annotées¹. Le corpus contient plus de 60000 images et a été produit sur une période de trente ans environ dans l'objectif d'inventorier des vestiges archéologiques. La base d'images est très hétérogène en terme de qualité des images car elle comprend des vues prises sous différents angles (ce qui provoque des déformations sur les images), avec différentes techniques (photographie traditionnelle, photographie infra-rouge etc.). À chaque image sont associées des informations sémantiques provenant de fiches techniques. Ces informations sont soit externes à l'image (telles que la date de prise de vue, le nom de l'opérateur, la localisation), soit liées à la nature archéologique du contenu ou du contexte de la photo (les mots clés utilisés dans cette partie font référence à un thésaurus spécialisé). Certaines des informations sémantiques sont elles aussi de qualité variable : lors de la prise de vue, toutes les images n'ont pas été localisées de manière exacte, pour certaines photos il n'y a aucune indication d'orientation, etc. L'exploitation de la base d'images passe par plusieurs types de requêtes tels que : la recherche d'images par des critères sémantiques globaux (on peut par exemple rechercher les images *“qui ont été prises pendant le mois d'octobre 1978”* ou bien rechercher les images *“qui ont été prises dans un rayon de 3 km autour du site romain des sources de la Seine”*), la recherche d'images par des critères sémantiques locaux (on peut par exemple rechercher les images *“qui contiennent une trace de motte féodale”*), la recherche d'images en fonction de leur contenu (on peut ainsi vouloir rechercher toutes les images *“qui contiennent une zone avec une texture de pavage en galets”*).

¹Cependant, l'image utilisée pour illustrer cette instanciation provient d'une autre base d'image d'archéologie aérienne [Ua].

Lors de la modélisation, l'image est décomposée en objets. Ces objets correspondent à des formes géométriques simples : bâtiments visibles en surface, infrastructures (telles que segments de routes, rivières, ponts) ou traces de bâtiments anciens. Pour ces derniers, les contours peuvent n'être que partiellement détectables sur la vue aérienne. Une fois les objets extraits, les relations spatiales (distance, direction, adjacence...) entre ces objets sont calculées puis modélisées. Dans le cadre de cette application, nous avons utilisé une interface d'extraction manuelle [Ler03] afin de simuler la détection des contours simples (géométriques ou courbes simples). Un expert (spécialiste d'archéologie) a indiqué quels sont les contours importants et quels sont leurs attributs sémantiques. Dans cette instantiation de notre modèle, les objets sont décrits par leur géométrie et des attributs sémantiques mais ils n'ont pas d'attributs physiques. Pour chaque objet, le tuple d'attributs sémantiques $AttS = \langle loc, tpo, tso, dat \rangle$ comprend :

- une localisation globale (e.g., Massala), notée *loc*, qui n'a de sens que sur l'image elle-même. Cette localisation est facultative ;
- un type principal d'objet, noté *tpo*, qui indique la nature –archéologique ou non– de l'objet (e.g., contrevallation, rivière). Ce type principal est obligatoire ;
- un type secondaire d'objet, noté *tso*, qui précise le type principal (e.g., fragment de contrevallation, segment amont de rivière). Ce type secondaire est facultatif ;
- une datation de l'objet (e.g., 73, contemporain) qui est optionnelle.

Nous détaillons ci-dessous l'instanciation du modèle pour cette application et nous donnons un exemple d'image (Figure 3.a) avec les contours conservés par l'expert (Figure 3.b) et la représentation de l'image dans le modèle instancié (Figure 3.c). Dans la Figure 3.c, les liens entre objets représentent la composition privilégiée, les cadres des objets simples sont en gras. Dans la Figure 4.a, nous donnons les directions de placement de l'un des objets (l'objet *AB*) par rapport à tous les autres objets. En Figure 4.b, nous illustrons les relations (relation topologiques en trait gras clair et relations de direction en traits gras sombres) entre cet objet *AB* et les autres objets. La seule étiquette utilisée pour la relation topologique est *D* pour *Disconnected*. Les étiquettes utilisées pour la relation de direction font référence à la table donnée en partie a) de la figure. La combinaison des deux structures (l'arbre qui décrit la composition des objets et les graphes qui décrivent les autres relations entre objets) modélise tous les aspects de l'image.

L'ensemble des graphes étiquetés utilisés dans cette instantiation est

$$\{ \langle ETIQ_i, A_i, \mathcal{E}_i \rangle \}_{i=1..3}$$

avec :

- Pour le graphe d'indice 1, qui est le graphe de la relation de composition privilégiée, un ensemble d'étiquettes réduit à une constante (cette étiquette constante n'est pas visualisée dans la Figure 4.a).
- Pour le graphe d'indice 2, qui est le graphe de la relation de direction, l'ensemble des étiquettes est l'ensemble des vecteurs de direction de la forme $\langle N, S, E, O, NE, NO, SE, SO \rangle$ (pour respectivement Nord, Sud, Est,

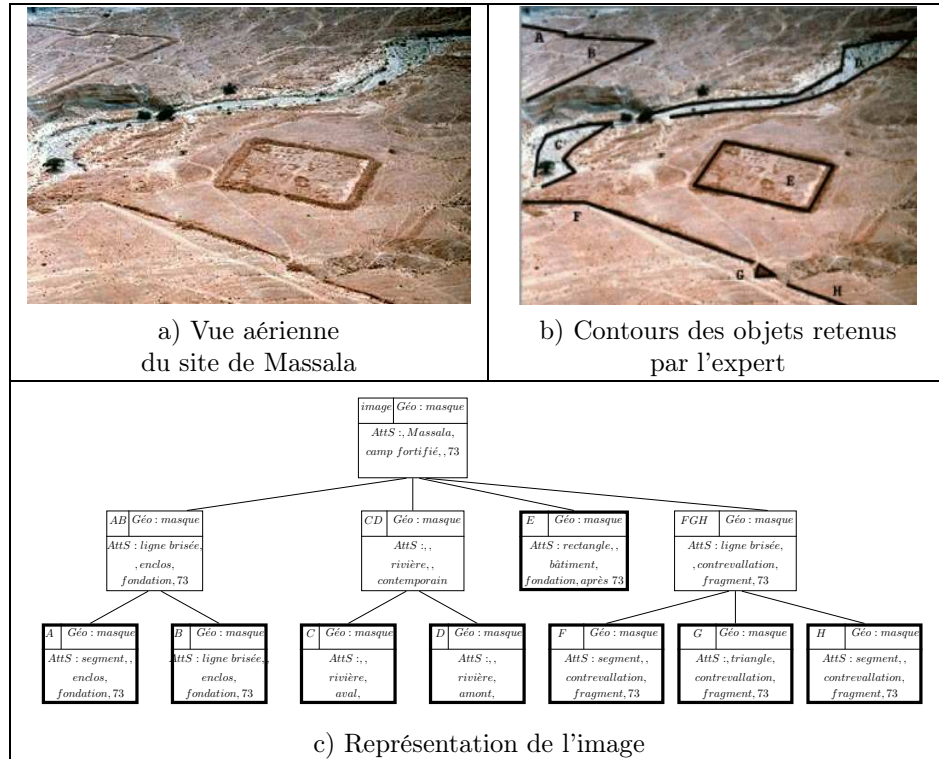


FIG. 3 – Le camp de Massala [Ua] : a) vue aérienne originale et b) avec les contours des objets retenus par l'expert (en utilisant une interface couplée avec une base de données pour le stockage des objets ainsi définis que nous avons développée [Ler03] sur la base d'un éditeur graphique en licence GNU [BLP]), c) représentation de l'image (objets et relation de composition privilégiée).

Ouest, Nord-Est, Nord-Ouest, Sud-Est et Sud-Ouest). La relation de direction que nous utilisons est une relation à base de logique floue que nous avons développée pour cette application archéologique [BTY01]. Les calculs des vecteurs de direction sont effectués dans la base de données liée à l'interface d'acquisition des objets.

- Pour le graphe d'indice 3, qui est le graphe de la relation topologique, nous utilisons certaines des relations topologiques issues de la méthode des neufs intersections d'Egenhofer [EF91].

4 Conclusion

Nous avons présenté les grandes lignes d'une plateforme pour la création de bases d'images. Cette plateforme permet –par combinaison de différents compo-

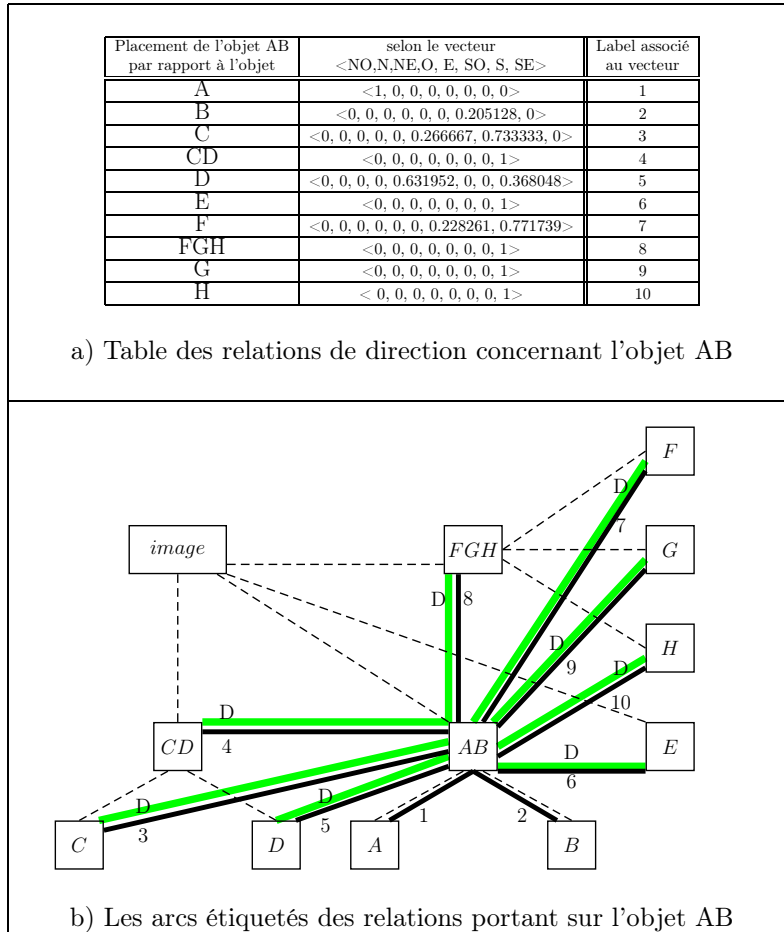


FIG. 4 – Le camp de Massala [Ua] : relations de l'objet AB avec les autres objets

sants autonomes mais intégrés via un modèle et une mesure de similarité génériques— d'utiliser dans la description des images des informations syntaxiques ou sémantiques qui peuvent être locales ou globales et d'accéder à la base en mode recherche (via un mécanisme d'indexation) ou en mode exploratoire (via un mécanisme de classification). La phase d'instanciation permet de cibler le système sur le type d'application visé par le choix des composants les plus adaptés au type d'images de la base (si la base n'est pas trop hétérogène). Le paramétrage de la mesure de similarité ainsi que certaines des relations spatiales proposées permettent d'adapter le système non seulement à l'application mais aussi aux mécanismes de perception humains.

Notre objectif à court terme est de définir les autres composants et particulièrement d'intégrer dans la bibliothèque d'extracteurs des modules qui doivent être

fournis par d'autres équipes de notre laboratoire.

A moyen terme, nous prévoyons de mettre en place le mécanisme de déduction permettant de garantir la cohérence des informations sémantiques associées –à différents niveaux de détail– à une même image. Ce mécanisme de déduction doit aussi permettre d'utiliser alternativement des informations syntaxiques et sémantiques pour une même recherche. Ceci devrait nous permettre de passer à la phase d'utilisation de notre plateforme pour définir et tester des stratégies de fusion syntaxique-sémantique. Nous prévoyons de définir un protocole de validation des stratégies (et de construire une base d'images pouvant servir de "benchmark" pour comparer ces stratégies).

Références

- [AKDGB] W. Al-Khatib, Y. F. Day, A. Ghafoor, and P. B. Berra. Semantic Modeling and Knowledge Representation in Multimedia Databases. *Transactions on Knowledge and Data Engineering*, 11(1), 1999.
- [BLP] A. BASSET, C. LIGNIER, and P. PASCAL. Editeur graphique vectoriel. GNU General Public License, Available at URL <http://editeur-graphique.free.fr/>.
- [BTY01] L. Besson, M.N. Terrasse, and K. Yétongnon. A fuzzy approach to direction relations. In *CBMI 2001, Content-Base Multimedia Indexing*, pages 169–176, Brescia, Italy, September 2001.
- [DBM97] N. Denos, C. Berrut, and M. Mechkour. An image retrieval system based on the visualization of system relevance via documents. In *Proc. of the Int. Conf. on Database and Expert Systems Applications, DEXA '97*, pages 214–224, 1997.
- [DL00] M. Després-Lonnet. Thésaurus iconographiques et modèles culturels. *Document Numérique*, 4(1-2) :153–165, 2000.
- [EF91] M.J. Egenhofer and R. Franzosa. Point-Set Topological Spatial Relations. *International Journal of Geographical Information Systems*, 5(2) :161–174, 1991.
- [FFN⁺93] C. Faloutsos, M. Flickner, W. Niblack, D. Petrovic, W. Equitz, and R. Barber. Efficient and Effective Querying by Image Content. Technical report, Research Report, IBM Almaden Research Center, 1993.
- [Fou02] J. Fournier. *Indexation d'images par le contenu et recherche interactive dans les bases généralistes*. Phd-thesis, Université de Cergy-Pontoise, 2002.
- [GB] V. Gouet and N. Boujemaa. Object-based Queries Using Color Points of Interest. Imedia Project, Inria Rocquencourt, Available at URL citeseer.nj.nec.com/577185.html.
- [GT96] W. K. Grassmann and J.-P. Tremblay. *Logic and Discrete mathematics, A Computer Sciences Perspective*. Prentice Hall, 1996. ISBN 0-13-501206-6.

- [Gua95] N. Guarino. Formal Ontology, Conceptual Analysis and Knowledge Representation. *Int. Journal of Human and Computer Studies, Special Issue on Formal Ontology, Conceptual Analysis and Knowledge Representation*, 1995. guarino95formal.ps.
- [GWJ91] A. Gupta, T. E. Weymouth, and R. Jain. Semantic Queries with Pictures : The VIMSYS model. In G. M. Lohman, A. Sernadas, and R. Camps, editors, *17th International Conference on Very Large Data Bases, September 3-6, 1991, Barcelona, Catalonia, Spain, Proceedings*, pages 69–79. Morgan Kaufmann, 1991.
- [KPK01] K.I. Kim, S.H. Park, and H.J. KIM. Kernel Principal Component Analysis for Texture Classification. *Signal Processing Letters*, 8(2), 2001.
- [Ler03] Bernard Leroux. Combinaison des caractéristiques physiques et sémantiques pour la recherche dans les bases de données images. Technical Report, 2003.
- [LTMD01] J. Landré, F. Truchetet, S. Montuire, and B. David. Content-based Multiresolution Indexing and Retrieval of Paleontology Images. *SPIE Proc. of Storage and Retrieval for Media Databases - San Jose - CA - USA*, 4315 :482–489, 2001.
- [OS95] Virginia Ogle and Michael Stonebraker. Chabot : Retrieval from a Relational Database of Images. *IEEE Computer*, 28(9) :40–48, September 1995.
- [PSTT01] G. Petraglia, M. Sebillio, M. Tucci, and G. Tortora. Virtual Images for Similarity Retrieval in Image Databases. *Transactions on Knowledge and Data Engineering*, 13(6), November/December 2001.
- [SB94] R. K. Srihari and D. T. Burhans. Visual Semantics : Extracting Visual Information from Text Accompanying Pictures. In *Proc. of AAAI '94, Seattle, WA, USA*, 1994.
- [SGJ01] S. Santini, A. Gupta, and R. Jain . Emergent Semantics Trough Interaction in Image Databases. *Transactions on Knowledge and Data Engineering*, 13(3), 2001.
- [TCC⁺94] R. K. Taira, A. F. Cardenas, W. W. Chu, C. M. Breant, J. D. N. Dionisio, C. C. Hsu, and I. T. Jeong. An object-oriented data model for skeletal development. In *Proc. of the SPIE*, 1994.
- [Ua] URL=<http://www.archeologie-aerienne.culture.gouv.fr>. L'archéologie aérienne dans la France du nord.
- [WJ97] D.A. White and R. Jain. Imagegrep : Fast visual pattern matching in image databases. In *Proc. SPIE : Storage and Retrieval for Image and Video Databases*, volume 3022, pages 96–107, 1997.
- [ZAL01] X.M. Zhou, C.H. Ang, and T.W. Ling. Image Retrieval Based on Object's Orientation Spatial Relationship. *Pattern Recognition Letters*, 22, 2001.