
Vers une indexation sémantique d'images dans un fonds iconographique territorialisé

Marie-Noëlle BESSAGNET(*), Eric KERGOSIEN ()(***), Alain du Boisdulier (***)**,

** Laboratoire LIUPPA, FDEG et IAE, 64012 Pau Université Cedex, France*

*** Laboratoire LIUPPA, Faculté des Sciences et Techniques 64012 PAU Université Cedex, France.*

*(***) Société DIS (Document, Images, Solutions), Creaticité Bat C, Technopole Izarbel 64210 Bidart France*

RESUME

Dans cet article, nous abordons nos travaux pour indexer et faire des recherches sur une base d'images telle que celle mise à notre disposition par la médiathèque de Pau. Le but est de proposer une méthode, des outils et techniques et un modèle sémantique pour identifier automatiquement des objets ayant pour nous un intérêt visuel, par exemple un clocher, une statue. Afin de mener à bien cette indexation, nous avons besoin d'outils de pré-traitement d'images qui nous permettent d'extraire des signatures. A partir de ces dernières, nous souhaiterions pouvoir rechercher dans le fonds iconographique des images similaires, basées sur la signature d'un référent. Les outils de pré-traitement d'images travaillent au niveau symbolique alors que les outils de recherche d'images se font au niveau sémantique. Nous désirons améliorer la pertinence des résultats provenant de la phase de recherche en nous appuyant sur des informations permettant de géo localiser le contenu des documents, le but étant de proposer un affichage cartographique du résultat de la requête. Les images sur lesquelles nous travaillons sont des images N&B reflétant des scènes du Béarn aux XIXième et XXième siècles.

MOTS CLES : traitement d'images ; sémantique ; ontologie ; description textuelle automatique

1. Introduction

Grâce à la maturité des techniques de numérisation et de reconnaissance optique de caractères, des corpus documentaires conséquents deviennent de plus en plus facilement disponibles au format électronique, cependant leur accessibilité reste

problématique. Afin d'exploiter au mieux la valeur de ces documents dits hérités, il est important d'identifier et de formaliser leur potentiel informatif. Ce potentiel, une fois cerné, même partiellement, est alors utilisable comme clef d'accès pour dynamiser l'utilisation du corpus. Ainsi, la reconnaissance d'informations qui font «sens» dans le cadre de situations d'usage identifiées et modélisées constitue le challenge de recherche du projet de l'équipe. C'est par exemple l'idée qu'une navigation de type « touristique » ou « éducative » dans un ensemble de documents hétérogènes (récits, contes, cartes postales, lithographies, séquences filmées...) liées à un territoire conduit à des stratégies spécifiques d'appropriation, de réutilisation et de « construction du sens ».

Dans ce cadre, nous développons en partenariat avec la Médiathèque Intercommunale à Dimension Régionale (MIDR) de Pau le projet « Itinéraires virtuels dans le patrimoine Pyrénéen ». Ce projet correspond à un usage répondant à des besoins de type « tourisme culturel ». Il se présente sous la forme d'une visualisation cartographique d'un itinéraire (déterminé par l'utilisateur) permettant d'avoir un affichage structuré et géolocalisé du résultat de la recherche documentaire.

Les documents images (lithographies, cartes postales, photographies...) ont comme caractéristique principale l'absence de règles explicitées permettant d'en exprimer la sémantique, à l'inverse des textes et des graphiques. Dans notre projet, le but de l'indexation d'un document image est d'extraire et de représenter le contenu nécessaire et suffisant des cartes postales du fonds documentaire pour qu'il soit retrouvé par un utilisateur dans des applications à usage touristique ou pédagogique. Cette indexation se base donc sur une représentation (supportée par un modèle) et sur un processus d'extraction. Afin d'éviter d'extraire des informations non pertinentes, chacun des éléments précédents doit intégrer d'une manière ou d'une autre les besoins des utilisateurs.

Dans la plupart des applications actuelles, la recherche ou la classification d'images s'effectue principalement grâce à des données textuelles : nom de l'image, mots-clés ou texte descriptif en langage naturel. La plupart des outils que l'on trouve sur le web (Google Image, PicSearch, Agence Images) ayant à traiter des grandes bases d'images utilisent un système d'indexation et de recherche par mots-clés. Ces mots sont souvent hiérarchisés pour améliorer le recoupement entre les images.

Les systèmes de recherche d'images basés contenu (CBIR – Content-based Image Retrieval) traitent le plus souvent l'extraction de caractéristiques visuelles, généralement la couleur, la forme et la texture. (Iqbal et al, 2002) utilise ces caractéristiques dans le but de donner une description globale des images et apporte des résultats intéressants dans la quête du sens. Cependant, ceux-ci échouent lorsqu'on veut considérer la sémantique des objets décrits dans l'image. (Mezaris et al, 2003) et (Mojsilovic et al, 2002) traitent ces caractéristiques de façon locale dans l'image afin d'identifier les différents objets et apportent des résultats encourageants.

Nos travaux de recherche s'intéressent à un corpus particulier de documents iconographiques : des cartes postales en noir et blanc qui représentent des paysages

stéréotypés des Pyrénées. En accord avec (Schomaker, 2001), les objets à niveau de gris peuvent être extraits du contexte de l'image, grâce à des applications permettant de reconnaître et d'extraire les contours. Ces objets peuvent ensuite être utilisés à des fins d'indexation et/ou de recherche. Nous appelons ces objets des stéréotypes. Ainsi, un stéréotype est la représentation d'un sujet (objet, personnage, situation) dont les caractéristiques sont toujours les mêmes. L'objectif de nos travaux est de catégoriser un ensemble de documents images automatiquement grâce à la définition préalable d'un stéréotype.

Nous abordons dans une première partie la problématique d'extraction de connaissances dans une base d'images au cœur de systèmes tels que celui que nous souhaitons mettre en oeuvre pour notre fonds images. Nous proposons dans une deuxième partie les premiers traitements que nous avons mis en place à partir de fiches indexées manuellement par la MIDR. Nous proposons dans une troisième partie un modèle sémantique permettant de représenter les concepts pertinents de notre fonds iconographique. Nous détaillons les classes de sujets traitées dans notre fonds documentaire qui relèvent d'au moins 5 catégories (Etre, Lieu, Document, Epoque et Objet). Dans une quatrième partie, nous présentons, à partir d'expérimentations et développements d'outils, les traitements mis en oeuvre qui devront nous permettre d'interpréter le contenu d'une image sur les niveaux symbolique et sémantique..

2. Extraction de connaissances à partir de bases d'images

2.1. Travaux connexes

A l'instar du projet *imageVAL*¹, de nombreuses applications surtout dans le domaine médical s'intéressent à l'Extraction de Connaissances en s'appuyant sur l'Analyse d'Images. En effet, l'indexation et l'analyse d'images ont été particulièrement développées ces dernières années, devenant les axes de recherche parmi les plus importants dans le domaine du traitement d'images. En indexation et recherche d'images, le but est de trouver des caractéristiques permettant d'indexer l'image et de retrouver des images similaires lors d'une requête en fonction du contenu des images.

Dans tous ces travaux, on peut recenser 3 phases essentielles :

1. l'extraction, à partir des images, d'informations sur la forme des structures (des régions d'intérêts) et sur leurs positions relatives ;

¹ <http://www.imageval.org/presentation.html>

2. la mise en correspondance de ces informations avec les descriptions d'une base de connaissances sur le domaine (médical, patrimonial, historique, publicitaire, anatomique,...) ;
3. et la « traduction » des informations utiles pour l'utilisateur final, soit à des fins d'indexation, soit à des fins de recherche.

Comme l'a décrit (Smeulders et al., 2000), il existe deux fossés majeurs pour les applications tentant d'interpréter automatiquement le contenu d'une image:

- a. Le fossé sensoriel, défini comme le fossé existant entre le monde réel 3D et sa représentation en une image 2D,
- b. Le fossé sémantique, défini comme le fossé entre la représentation bas niveau d'une image et l'interprétation haut niveau que les humains en font.

Nous nous intéressons dans nos travaux au fossé sémantique. Ce fossé sémantique existant entre les descripteurs bas niveau et les concepts sémantiques rend ambigu la mise en correspondance des régions d'une image segmentée avec un objet sémantique. Par contre, nous pensons qu'en introduisant la prise en compte de connaissances du domaine, le gap entre sémantique et symbolique se réduit drastiquement (Zlatoff et al, 2004). En effet, il est nécessaire de posséder une description sémantique de l'image (basés sur des modèles conceptuels ou ontologiques) et de les faire correspondre à des signatures de bas niveau telles que la couleur, la texture, la forme et également des relations spatiales (Maillot et al, 2004). Dans nos travaux de recherche, nous avons adopté une telle démarche en définissant un modèle ontologique propre à nos cartes postales.

De plus, dans ce domaine, certains travaux de recherche tentent de mettre en place des systèmes et des techniques de recherche et d'extraction de contenus d'images basés sur les relations spatiales pouvant exister dans l'image. C'est le cas, par exemple, des travaux de (Di Sciascio et al, 2002). Ils tentent de réduire le gap entre les représentations visuelles des caractéristiques et le contenu des images segmentées. Nous intégrons une telle approche dans nos propres travaux car nous sommes persuadés que cette méthode correspond plus aux attentes des requêtes utilisateurs. Cela dit, nous définissons nos relations entre les éléments d'un document image à partir des coordonnées géographiques des entités spatiales identifiées dans le document (décrit dans le §4.1) et non pas à partir de leur position dans l'image. Nous cherchons également à identifier les éléments que nous estimons être principaux dans l'image. Des travaux de recherche tels que, (Schomaker, 2001) adoptent cette méthodologie. La méthode consiste à identifier des objets à partir de régions segmentées pour les interpréter ensuite grâce aux connaissances « domaine » (thèmes).

Le problème auquel nous sommes confrontés dans nos travaux est la définition de modèles et d'outils destinés à permettre d'indexer des cartes postales anciennes et de les retrouver en fonction d'usages définis par les utilisateurs finaux : des touristes, des pédagogues, des apprenants, des experts du domaine, etc. L'indexation à laquelle

nous nous attachons n'est pas triviale. En effet, le but est de marquer des zones d'intérêts dans des images référents permettant de classifier aisément notre fonds iconographique. Pour l'instant, l'hypothèse d'une indexation totalement automatique des cartes postales est une vue à long terme étant donné l'état de l'art en matière d'analyse d'images.

2.2. Protocole envisagé

L'approche la plus réaliste à court terme est de proposer un système d'indexation semi automatique, basé sur les fiches descriptives de la MIDR.

Nous avons défini un ensemble d'étapes permettant de mettre en œuvre un protocole que nous expérimentons actuellement.

Etape 1 : Traitement à partir des fiches indexées (niveau sémantique) :

- Définition et modélisation de la carte de concepts (détaillée dans le &4.1.) après un travail manuel consistant à identifier visuellement les concepts pertinents contenus dans notre fonds iconographique.
- Insertion de notre modèle (dictionnaire) dans le prototype PIV¹ actuel et traitement d'un groupe exemple d'images à partir des fiches indexées manuellement par la MIDR.

Etape 2 : Classification supervisée :

- Obtenir les classes compactes d'images grâce à l'utilisation d'un outil tel que SIMILAR² (décrit dans le &5). On doit choisir la ou les image(s) référence(s) *a priori* puis lancer le traitement sur un ensemble d'images pour obtenir la classe pertinente après calcul de similarité ;
- Rechercher un stéréotype pouvant nommer les différentes classes obtenues.

Etape 3 : Trouver pour chaque référent de classe d'images une description plus atomique des éléments (objets ou concepts) qui pourraient faire sens. Extractions de descripteurs locaux identifiant les éléments principaux constituant le ou les référents (extraits de SIMILAR à l'étape 2).

Nous ajustons actuellement des outils de traitement d'images tels M.A.I.S.³ décrit dans le &5 afin d'obtenir automatiquement les descripteurs les plus discriminants pour chaque classe d'images.

Etape 4 : Après avoir appliqué le traitement étape 3 à chaque image d'une classe, prévision d'un traitement permettant de rechercher un élément (identifié par un

¹ Pyrénées Itinéraires Virtuels

² Produit logiciel proposé par l'entreprise Document, Images, Solutions

³ Motion Analyses for Image Sequences

vecteur caractéristiques) dans cette classe d'images afin d'identifier et de diminuer le bruit dans la classe (images non pertinentes au niveau sémantique car ne contenant pas les mêmes éléments principaux que le ou les référents).

Etape 5 : Re classification des images identifiées Etape 4.

Une fois ce protocole validé, nous souhaitons le généraliser à d'autres classes d'images.

3. Une première approche : la recherche textuelle

Il s'agit d'effectuer une extraction et une recherche directement à partir des informations saisies dans les notices réalisée par la MIDR. Ceci n'entraînera aucune modification des fiches index puisque l'on va se contenter des informations déjà existantes permettant ainsi une recherche variée sur plusieurs champs de méta données.

On interroge directement les fiches index sur lesquelles on cherche les mots clés de la requête pour pouvoir ensuite afficher les images correspondantes.

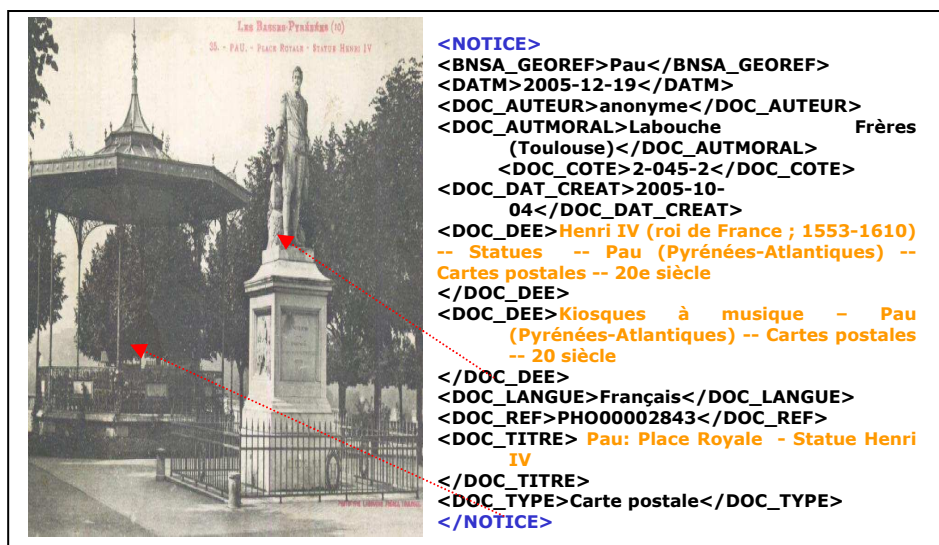


Figure 1: Une image et une partie de sa fiche index

Pour une telle recherche nous avons défini principalement 2 tags : *DOC_DEE* et *DOC_TITRE*. En effet ces deux champs suffisent pour effectuer une recherche textuelle simple puisqu'ils contiennent suffisamment d'information sur l'image pour une recherche par mot clé. Nous avons également utilisé le champ *DOC_TYPE* pour permettre aux utilisateurs de choisir le type de document.

Le prototype correspondant à cette recherche d'images textuelle nous a permis de tester les informations existantes dans les fiches et la faisabilité du système. Cette méthode reste simple mais nous sert de point de départ dans la phase de recherche et

nous permet dans un second temps d'évaluer en terme de pertinence des résultats de requêtes les systèmes de recherche que nous mettons en œuvre ensuite. Présentons la deuxième méthode de recherche par thème plus complexe et permettant plus d'interactivité via la carte de concepts.

Pour une base contenant un nombre d'images important telle que la notre, un classement par thèmes peut faciliter la recherche de documents. La classification peut se faire de façon supervisée (domaine connu à priori, permettant de définir des classes par la sémantique) ou non-supervisée (on définit les classes à partir de la structure des objets car la sémantique est plus difficile à déterminer). Nous avons opté pour la classification supervisée car notre corpus est particulier : des cartes postales en niveau de gris qui représentent des paysages stéréotypés des Pyrénées.

Nous définissons donc dans un premier temps manuellement un ensemble de classes représentatives de notre fonds iconographique. Les thèmes identifiant ces classes sont structurés sous la forme d'une carte de concepts, représentation correspondant à nos besoins. Présentons ce modèle.

4. Proposition d'un modèle sémantique pour l'analyse de documents image

4.1. Le modèle sémantique

Le modèle sémantique correspond à une description des notions stéréotypées qui apparaissent dans notre fonds iconographique.

4.1.1. Les sujets traités

Après avoir analysé manuellement le fonds iconographique, nous avons regroupé les concepts traités dans notre fonds documentaire en classes relevant de cinq catégories que nous qualifions comme suit:

- ***des êtres*** : par exemple des personnages, des animaux ou encore de la flore
- ***des lieux*** : par exemple des montagnes, des forêts, des communes, des cours d'eau, des monuments, des rues, des places, ...
- ***des objets*** : par exemple des œuvres d'art type tapisserie ou meubles (dans le château), des transports tels le tramway, le funiculaire, les charrettes, les calèches, ...
- ***des époques*** : par exemple date sur les documents, période (XIXième), saisons. Cette catégorie nous permet de décrire les événements les plus pertinents pour le territoire à décrire.
- ***des documents*** : par exemple des images (cartes postales, lithographie, photographie,...), des textes (livres, journaux, poèmes,...), du son ou

encore de la vidéo.

La figure 2 représente une partie du modèle que nous proposons pour caractériser le fonds documentaire de la Médiathèque en développant pour chaque concept ses sous-concepts principaux. La liste pour chaque concept n'est pas complète. Nous avons choisi d'utiliser le standard Topics Map™ (Pepper & Moore, 2001), dans lequel chaque concept correspond à un nom de Topic. Ces derniers sont liés par des associations.

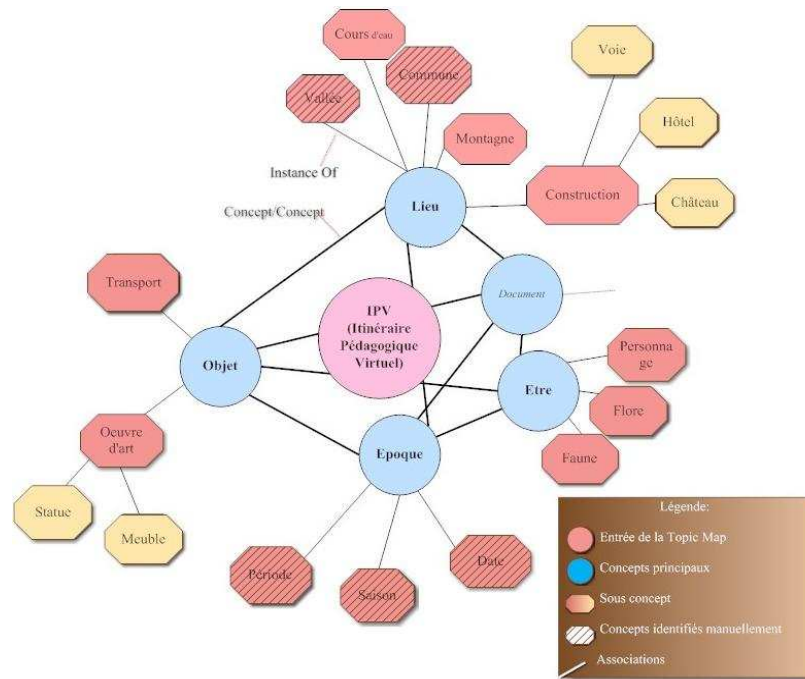


Figure 2 : Le détail des sujets abordés dans le fonds documentaire

4.1.2. Les associations

Pour définir les associations, nous avons tout d'abord dû regrouper chaque concept en concept plus générique afin de limiter le nombre d'associations dans notre carte de concepts et ainsi en faciliter la compréhension. Cette approche nous permet d'identifier les relations père/fils entre concepts et nous la définissons sous la forme « instance_of ». On remarque rapidement en étudiant notre base d'images que la plupart des images sont composées de plusieurs éléments définis dans notre carte de concepts que l'on peut relier entre eux par des relations diverses. Le sens global de l'image est donné par ces relations entre les différents éléments présents dans l'image. Nous définissons ce deuxième type d'association comme spatiale (Lesbegueries, 2006), basé sur le raisonnement spatial qualitatif (Cohn et al, 2001).

Une scène composée est en fait une relation concepts/concepts et elle peut être instanciée de différentes manières :

- **Distance** : Cette relation permet de mettre en avant une cible à l'aide d'une distance par rapport à un autre élément (par exemple, *à 1km de*)
- **Adjacence** : Cette relation permet d'évoquer la proximité d'un élément par rapport à un autre (par exemple, *à côté du* château Henri IV)
- **Orientation** : Relation géographique d'un lieu par rapport à un autre en s'orientant selon les quatre points cardinaux (par exemple, *au sud de...*)
- **Inclusion** : Relation d'inclusion d'un élément dans un autre élément (par exemple, *dans, à l'intérieur de...*)
- **Intersection** : Relation d'intersection entre deux éléments (par exemple, *entre* Château et Eglise, *entre* deux personnages)

Ces associations de type spatial entre concepts apportent un plus indéniable dans notre quête du sens car, en s'appuyant sur des systèmes d'informations géographiques puissants tels que Viamichelin¹, nous pouvons maintenant identifier une entité spatiale (ES) et la localiser en fonction d'autres entités spatiales présentes dans l'image. Nous identifions une ES à travers un processus découpé en 4 étapes. La première concerne la collecte d'images numérisées relatant du patrimoine culturel pyrénéen. La deuxième supporte une analyse linguistique puis sémantique (Abolhassani et al., 2003) afin d'extraire des ESs sous forme d'instance du modèle spatial de PIV. La troisième s'appuie sur des ressources géographiques (communes, lieux-dits,...) afin de valider les ESs détectées à l'étape précédente. La dernière propose le calcul de représentations spatiales géo-référencées.

4.1.3. Un modèle de navigation

Le travail consiste à récupérer les informations d'une fiche index décrivant une image. Lorsque l'on identifie un concept dans une fiche index, on ajoute à cette fiche l'information relative au topic comme indiqué figure 3 ; on ajoute le lien vers l'image au Topic présent dans la carte de concepts. L'image est alors classée dans notre carte de concepts.

```
<?xml version="1.0" encoding="UTF-8"?>
<doc doc_base="indexPau.xml" doc_original="Pau.jpg"
ls_doc="indexPau.xml" ls_sem="indexPau.xml">
<NOTICE>
<BNSA_GEOREF>Pau</BNSA_GEOREF>
<DATM>2005-12-19</DATM>
<DOC_AUTEUR>anonyme</DOC_AUTEUR>
<DOC_DEE>Henri IV (roi de France ; 1553-1610) -- Statues -- Pau
(Pyrénées-Atlantiques) -- Cartes postales -- 20e siècle</DOC_DEE>
<DOC_DEE>Kiosques à musique -- Pau (Pyrénées-Atlantiques) -- Cartes
postales -- 20 siècle</DOC_DEE>
.....
```

¹ <http://www.viamichelin.fr>

```

.....
<DOC_TITRE>Pau: Place Royale - Statue Henri IV</DOC_TITRE>
<PHO_LEGEND>Série &quot;Les Basses-Pyrénées&quot;</PHO_LEGEND>
<ES id="1">
  <libelle>Place Royale</libelle><type>Place</type>
</ ES >
  < ES id="2">
    <libelle>Statue Henri IV</libelle><type>Statue</type>
  </ ES >
  < ES id="3">
    <ES >
  <NOTICE>

```

Figure 3 : Proposition de fiche index après classification

Cette carte de concepts est mise à jour pour chaque ajout d'une nouvelle image ou groupe d'images dans notre corpus par le module de classification. Ce module ajoutera à chaque Topic les liens aux images correspondantes. L'utilisateur peut alors choisir un ou plusieurs Topics ou sous Topics parmi ceux présents dans la carte de concepts déjà chargée. Une fois le(s) topic(s) sélectionné(s), le module d'affichage montre à l'utilisateur toutes les images présentes dans la carte de concepts et relatives au(x) Topic(s) sélectionné(s).

Des images ne sont pas automatiquement classées lors de ce traitement. En effet, certaines fiches manquent d'informations pertinentes et les mots utilisés ne sont pas toujours ceux définis comme thèmes dans notre carte de concepts (problème d'orthographe ou utilisation de synonymes). Pour tenter de palier ce problème, nous définissons plusieurs noms identifiant chacun des concepts mais, bien que cela augmente la quantité des images ainsi classés, près de 5% de ces derniers restent indépendants du modèle. Prenons l'exemple du concept « chaise » (sous-concept de objet) qui est rarement défini dans les fiches index. La chaîne de traitement permettant d'extraire des informations directement à partir du contenu de l'image doit pouvoir apporter des solutions à ce problème.

Rappelons que l'indexation a été réalisée manuellement par la MIDR et on remarque que toutes les informations contenues dans les images que l'on juge nous, pertinentes, ne sont pas toujours représentées dans les fiches index (problème d'objectivité).

Après avoir défini notre modèle sémantique, nous pouvons proposer un modèle de navigation à travers cette Topics Map (TM)^o(Figure 4).

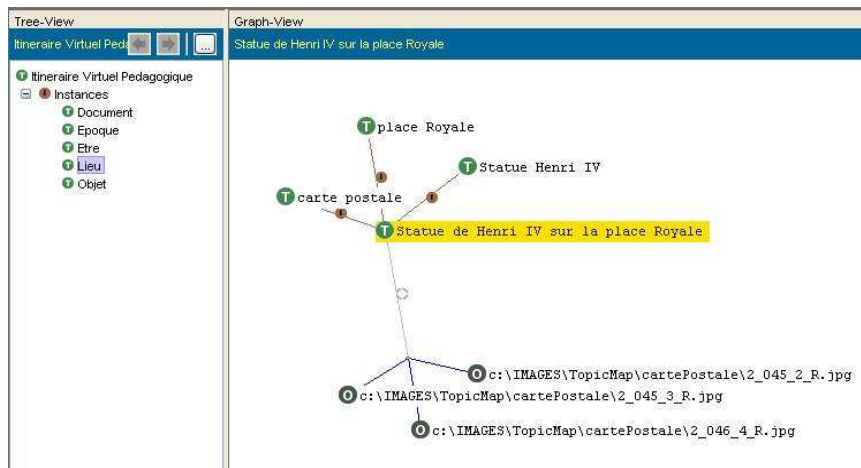


Figure 4 : Parcours de la Map, les occurrences de Statue de HenriIV sur la place royale (visualisée avec l'outil TMNav¹)

Le prototype du projet PIV dans lequel nous avons inséré notre dictionnaire de concepts nous permet d'ores et déjà d'obtenir des résultats encourageants dans notre quête du sens. Cependant, cette première approche a des limites :

- Le traitement se base sur l'extrait d'images indexées manuellement et notre but est à terme de rendre cette opération automatique ;
- Le contenu des images est décrit trop généralement dans les fiches index.

4.2. Le modèle sémantique avec géoréférencement

Des travaux réalisés dans l'équipe de recherche ont permis d'indexer les documents texte provenant du fonds documentaire de la Médiathèque (Lesbegueries et al, 2006), (Etcheverry et al, 2006). Ces travaux calculent des représentations spatiales d'entités géographiques trouvées dans les documents texte et dans les requêtes textuelles.

Ce travail permet à un lecteur-utilisateur de découvrir et de s'appropriier un document selon une approche « territoriale ». Cette finalité nécessite un marquage des contenus documentaires selon une facette spatiale. Nous appliquons ces concepts aux images du corpus afin d'afficher au lecteur-utilisateur les images relatives à une requête de type « au sud de Gan », « à côté du Gave », « à Pau »....

Ainsi, comme pour le texte, chaque image de notre corpus aura une ou plusieurs représentations spatiales présentée sous forme de boite englobante (Fig. 5). Ce terme provient du monde des SIG (traduit de l'anglais Minimum Bounding Rectangle). Une boite englobante est un rectangle orienté x, y dont les coordonnées sont (Xmax,Xmin,Ymax,Ymin). Il englobe de manière minimale une entité spatiale géo-

¹ <http://tm4j.org/tmnav.html>

référéncée. Le nombre de boites englobantes présentant une image dépendra du nombre d'entités spatiales figurant dans sa description.

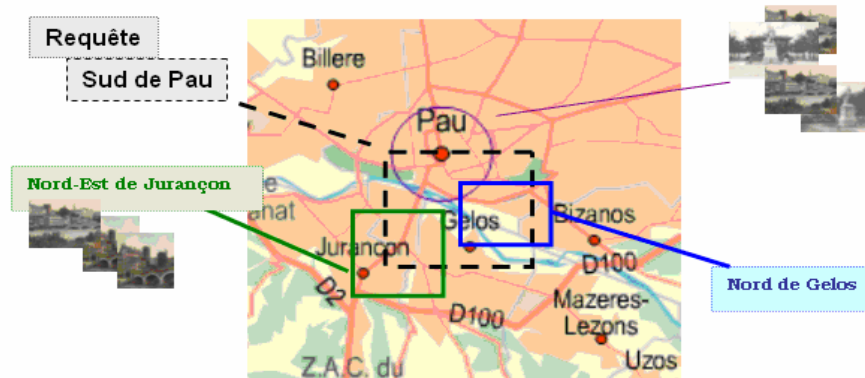


Figure 5 : *Le principe des boites englobantes*

En prenant en compte notre interprétation de l'image au niveau sémantique, nous ajoutons de nouvelles informations représentant les boites englobantes dans le fichier index de l'image décrite figure 6.

```

<?xml version="1.0" encoding="UTF-8"?>
.....
<ES id="1">
  <libelle>Place Royale</libelle>
  <type>Place</type>
  <representation>
    <bounding_box>
      <x_min>47331.65625</x_min>
      <y_min>1813025.375</y_min>
      <x_max>386240.46875</x_max>
      <y_max>1820997.375</y_max>
    </bounding_box>
  </representation>
</ES>
...

```

Figure 6 : *Représentation d'image en boite englobante*

Comparaison des résultats

Dans nos expériences, nous avons constaté ce que peut apporter une recherche par concept par rapport à une recherche d'image simple. Cette méthode offre bien à l'utilisateur une meilleure interactivité avec le système lui permettant d'obtenir plus aisément ce qu'il recherche. Avec une recherche géo référencée, pour une même requête, les images trouvées sont différentes de celles d'une recherche textuelle simple sans géo référencement.

En cherchant des images du « sud de Pau », on a obtenu des images déjà trouvées pour une recherche avec « Jurançon », qui est au sud de Pau. La prise en compte des

relations d'adjacence, de distance et d'orientation par la recherche ne sont pas pris en compte dans une recherche textuelle simple, d'où l'importance de ce travail de géo-référencement.

Cependant, ces premières expérimentations traitent la sémantique d'une image à un niveau que nous qualifierons de macro. Nous souhaitons caractériser plus finement les éléments des images. Afin de combler au mieux ces limites, nous expérimentons le protocole à partir de l'étape 2 dont le but est de valider et d'affiner les résultats obtenus précédemment. Les images sont alors traitées dans cette deuxième partie du protocole au niveau symbolique et le principe consiste à identifier des éléments de l'image caractérisés par un ou plusieurs vecteur(s) signature(s). Nous allons aborder les travaux de recherche en cours nous permettant de lier les résultats, obtenus à l'aide de ce traitement de l'image bas niveau, aux concepts définis dans notre carte de concepts (niveau sémantique).

5. Liaison entre les niveaux symbolique et sémantique

Nos expérimentations de boîtes à outils de traitement d'images nous laissent envisager plusieurs solutions ; elles dépendent grandement de résultats obtenus à partir du traitement symbolique de l'image. La première consiste en effet à ne prendre en compte que le vecteur caractéristique global de l'image pour ensuite le comparer à une base d'images traitées de la même façon afin d'y retrouver les images similaires. C'est la solution exploitée par l'outil SIMILAR qui propose D.I.S. La seconde consiste à extraire des vecteurs caractéristiques locaux de l'image pour ensuite les comparer à la base d'images afin d'y retrouver les images similaires.

5.1. Expérimentations

5.1.1. Extraction des caractéristiques globales

Cette première étape dans notre traitement de l'image bas niveau consiste à représenter sous forme d'un vecteur caractéristiques les données issues d'une chaîne de segmentation en régions d'images puis à calculer la distance entre ces images.

Dans la recherche d'images similaires par le contenu, diverses statistiques (basés sur la couleur, la texture et la forme) forment la description de l'image. Ces statistiques, que nous appelons ici descripteurs peuvent être mises sous la forme d'un vecteur. La distance entre deux images est alors la distance dans un espace à n dimensions, où n est le nombre de descripteurs choisis. Ce mode de représentation implique un choix définitif dans la nomenclature utilisée pour représenter les images : chaque image doit avoir le même nombre de descripteurs et chaque descripteur possède une place précise au sein du vecteur. La couleur, la forme et la texture sont les trois familles de descripteurs les plus couramment utilisés. Un vecteur caractéristique global caractérise l'image de façon générale.

Dans le cas de l'outil SIMILAR, du fait de la normalisation et du filtrage passe-bas sur les images en amont de la chaîne de segmentation, la texture n'est plus pertinente, les petits motifs ayant disparu. L'outil caractérise donc l'arrière-plan et les deux plus grandes régions en terme de couleur, de forme et de localisation dans l'image.

La procédure retourne la liste des images considérées comme proches de l'image requête. Cette liste est ordonnée de façon croissante en fonction de l'indice de proximité de chaque image par rapport à l'image requête.

Le traitement est simplifié et nous permet d'obtenir le sens global d'une image. Par calcul de similarité des paquets d'images, nous pouvons maintenant associer manuellement les images à nos concepts. Cela dit cette approche nous confronte actuellement à deux problèmes distincts :

- si les résultats sont concluants sur des images couleurs vu le choix des descripteurs de l'outil SIMILAR développé initialement pour traiter des affiches publicitaires couleur, nous ne pouvons pour le moment valider cette approche sur le fonds iconographique mis à disposition par la MIDR car près de 80% de ces images sont en niveau de gris. La chaîne de traitement de l'outil SIMILAR est en cours de modification pour prendre en compte un segmenteur en niveau de gris.
- L'identification de plusieurs éléments (concepts) distincts pour composer une scène donnée est également problématique actuellement. Prenons l'exemple de l'image figure 7, où l'on ne pourrait choisir les deux concepts *Kiosque* et *Statue* pour obtenir l'image donnée.

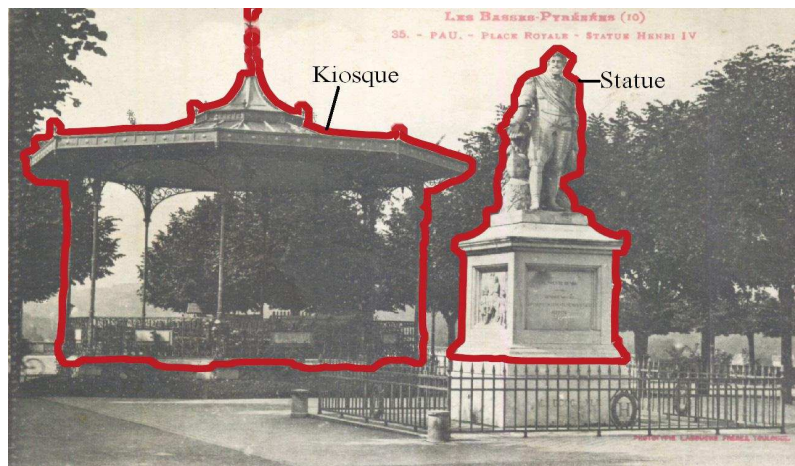


Figure 7 : Composition de concepts constituant une image

Cette première solution est donc intéressante pour les images constituées d'un ou deux élément(s) principaux mais elle devient limitée dans le cas où l'on effectue une recherche d'images à partir de plusieurs concepts. En effet, en reprenant l'image ci-

dessus composée de plusieurs objets ayant des caractéristiques, couleurs et textures, très différentes, le vecteur de caractéristiques extrait à partir de l'image entière perd les informations locales (les objets) et ne produit qu'une moyenne grossière du contenu de cette image.

5.1.2. Extraction des caractéristiques locales

La solution envisagée consiste à identifier ces éléments principaux constituant l'image et d'extraire de l'image les vecteurs caractéristiques de chacun plutôt que de recalculer une signature générale de l'image qui nous fera perdre de l'information pertinente. Un vecteur de caractéristiques local est alors un ensemble de caractéristiques d'une partie de l'image et leur poids dans l'image.

Ce traitement, dans notre projet, s'appuie entre autres sur l'environnement d'analyse d'image et de séquences vidéo M.A.I.S. développé dans notre laboratoire (Luthon & Clément, 2006). Grâce à une adaptation de cet environnement, nous devons être capable de repérer des formes significatives dans les images à partir de méthodes de détection de contours. Ainsi, la deuxième boîte à outils (M.A.I.S) devra nous permettre, à partir du résultat contour obtenu (Fig. 8), d'extraire la signature locale correspondant aux éléments (ici Statue ou Kiosque) sous forme d'un fichier .txt que nous associons ensuite aux concepts de notre Topics Map.

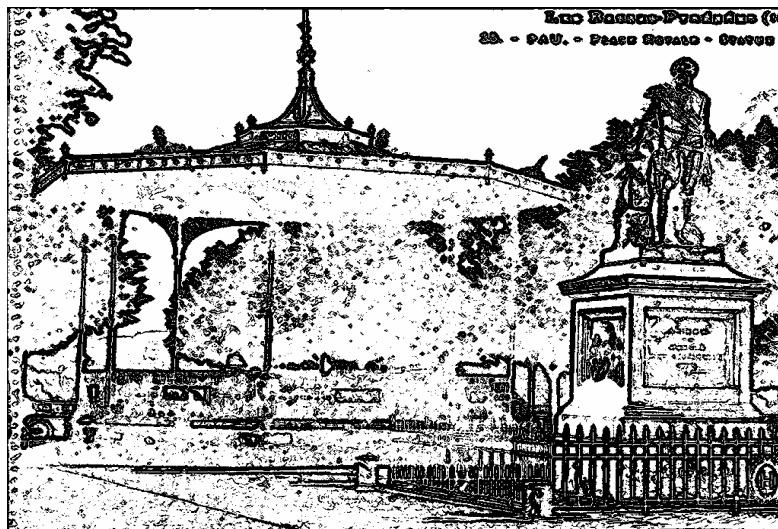


Figure 8 : Image segmentée dans M.A.I.S.

Il sera alors possible de lier chaque concept défini dans notre Topics Map à un vecteur caractéristiques référent choisi manuellement au préalable (dans une image que l'on a choisi).

De la même manière que dans l'outil SIMILAR, un calcul de similarité par intersection d'histogrammes normalisés permettrait d'identifier le vecteur

caractéristique considéré comme le plus proche dans la liste des vecteurs identifiant les référents définis dans notre carte de concepts. Cette procédure nous permettra de trouver dans notre base d'occurrences, des images constituées des mêmes concepts (éléments principaux de l'image ayant des vecteurs caractéristiques proches).

6. Conclusion et perspectives

Nous avons montré dans ce papier nos travaux actuels pour tenter d'apporter des éléments de solution afin de réduire le fossé entre le niveau sémantique et le niveau symbolique dans le traitement d'images.

Au niveau sémantique, nous avons proposé un modèle « ontologique » mettant en œuvre les divers concepts impliqués dans notre fonds iconographique de cartes postales. Ce modèle est au cœur du système et est utilisé à différentes fins : (i) Le modèle détaille la terminologie et les concepts avec lesquels nos méta données sont exprimées ; (ii) Le modèle offre différentes vues de base de recherche : les cinq sujets définis sont autant d'entrées pour formuler des requêtes ; (iii) Ce modèle offre un furetage sémantique permettant de faire des requêtes sur des relations spatiales à l'intérieur des images elles mêmes ainsi que sur des relations entre images du fonds. Ainsi nos images sont liées entre elles de plusieurs façons et offrent ainsi un intérêt de navigation amplifiée pour l'utilisateur final.

Au niveau symbolique, nous expérimentons actuellement la découverte automatique de descripteurs locaux afin d'obtenir une représentation des objets composant une image. Nous pourrons ainsi lier chaque concept défini dans notre modèle « ontologique » à un vecteur caractéristique référent.

A plus long terme, nous envisageons une interface qui permettra à l'utilisateur d'effectuer une recherche dans le fonds documentaire en saisissant une requête texte (« à Pau » par exemple) accompagnée d'une forme dessinée sous forme de contour avec la souris représentant un élément qu'il recherche (forme pointue représentant des clochers d'église par exemple). Le fonds documentaire sera ensuite interrogé à partir de la signature de l'objet dessiné et de la requête texte.

7. Remerciements

Nous tenons à remercier la MIDR pour l'accès à ses fonds documentaires. Nous remercions aussi D.I.S. pour l'accès à ses ressources logicielles.

8. Bibliographie

(Abolhassani et al., 2003) Abolhassani M, Fuhr N, Govert N, Information Extraction and Automatic Markup for XML documents, Intelligent Search on XML Data, LNCS Springer, p. 159-178, 2003.

- (Cohn et al, 2001) A. G. Cohn, S. M. Hazarika : Qualitative spatial representation and reasoning: an overview, *Fundamenta Informaticae*, pages: 1 – 29, 2001
- (Di Sciascio et al, 2002) Di Sciascio E, Donini, FM, Mongiello M., Spatial Layout representation for query-by-sketch content-based image retrieval, *Pattern Recognition Letters* 23, 1599-1612, 2002
- (Etcheverry et al, 2006) Etcheverry, P., Marquesuzaà, C., and Corbineau S. « Designing Suited Interactions for a Document Management System handling Localized Documents ». 24th ACM International Conference on Design of Communication, Myrtle Beach, USA, October 06, www.sigdoc2006.org, SIGDOC 2006.
- (Iqbal et al, 2002) Q. Iqbal and J. K. Aggarwal, CIRES: A System for Content-based Retrieval in Digital Image Libraries, Invited session on Content Based Image Retrieval: Techniques and Applications International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, pp. 205-210, December 2-5, 2002.
- (Lesbegueries et al, 2006), Lesbegueries, J., Sallaberry, C., and Gaio, M. « Associating spatial patterns to text-units for summarizing geographic information ». 29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval - GIR (Geographic Information Retrieval) Workshop, pp. 40-43, www.geo.unizh.ch/~rsp/gir06/accepted.html, ACM SIGIR 2006.
- (Luthon, F. Clement F., 2006). Luthon, F. Clement F., Macroscopic Quality Measurement of Plasma Polystyrene through Computer Vision, LIUPPA-LEGP. 2006.
- (Mezaris et al, 2003) Vasileios Mezaris, Ioannis Kompatsiaris, and Michael G. Strintzis, "Region-Based Image Retrieval Using an Object Ontology and Relevance Feedback," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 6, pp. 886-901, 2004.
- (Maillot et al, 2004) Nicolas Maillot, Monique Thonnat, Céline Hudelot: Ontology Based Object Learning and Recognition: Application to Image Retrieval. ICTAI 2004: 620-625
- (Mojsilovic et al, 2002) A. Mojsilovic and J. Gomes, Semantic Based Categorization, Browsing And Retrieval In Medical Image Databases, *ICIP (3) 2002*: 145-148
- (Pepper & Moore , 2001) S. Pepper, G. Moore "XML Topic Maps (XTM) 1.0 Specification", TopicMaps.Org, Aug. 2001. Available at <http://www.topicmaps.org/XTM/1.0>.
- (Schomaker, 2001) Lambert Schomaker, Image Search and annotation: from lab to web, Conférence invitée CIDE 01, 2001
- (Smeulders et al., 2000). Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. On Patt. Anal. And Machine Intell.*, vol.22, n.12, pp.1349-1380, Dec. 2000.
- (Zlatoff et al, 2004) Vision Gestalt et connaissances : une approche générique à l'interprétation d'images, Retrouvé sur le Web, <http://www-rech.enic.fr/coresa2004/articles/p029-zlatoff.pdf>. Dernier accès : 13/01/07